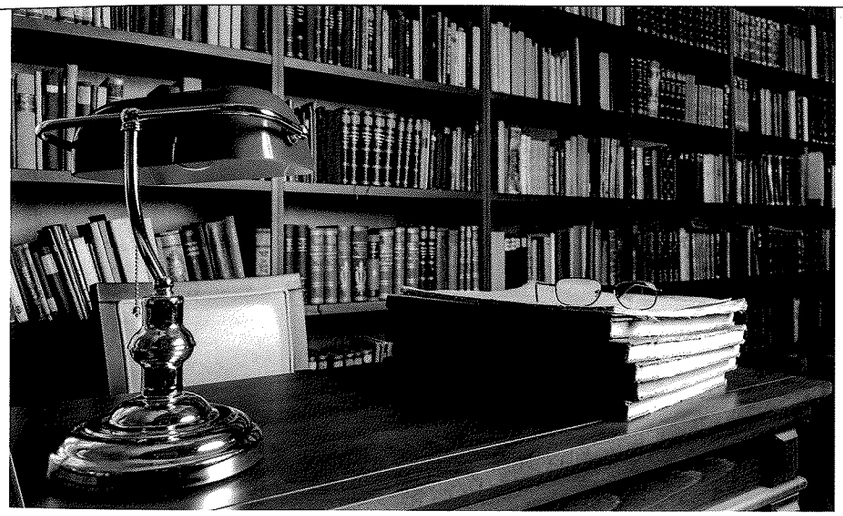
Gaetano Domenici • Pietro Lucisano • Valeria Biasi

Ricerca sperimentale e processi valutativi in educazione

Experimental Research and Evaluation Processes in Education



Con contributi di:
E. Botta, G. Castellana, C. Corsini, A.M. De Luca, G. Moretti, E. Rubat du Merac, R. Trinchero, I. Vannini, R.M. Viganò

McGraw Hill

Gaetano Domenici
Pietro Lucisano
Valeria Biasi
Ricerca sperimentale e processi valutativi in educazione
Experimental Research and Evaluation Processes in Education

5546-3
McGraw Hill

o Domenici • Pietro Lucisano • Valeria Biasi

erca sperimentale e processi
ta in educazione

ir Evaluation
s

LPN WE 42083 4286

e presen elli metodologico-proce-
li ricerca scientifica partire dalla nascita ed
e della pedagogia sperimentale con attenzione al contesto
lturale in cui è maturata la ricerca formativa nel panorama internazio-

no presentati sia gli orientamenti metodologici quantitativi (dalla
azione classica, alla "quasi" sperimentazione, ai modelli di micro-spe-
one, alle procedure della ricerca-azione e ricerca-formazione), sia gli
nti strettamente qualitativi (messi in luce in particolare nell'ambito
vazione sistematica, dell'analisi degli studi di caso, della conduzione
group), nonché una ricognizione delle principali procedure valutative
dio dei livelli di apprendimento. Si aggiungono prospezioni su metodi
ione di sistema, condotte con ricerche su grandi campioni, e sulla
della valutazione nei sistemi formativi on-line. Il volume consegna al
fine, indicazioni sintetiche per la stesura di un rapporto scientifico e si
per il ricercatore e anche per il docente in formazione ed in servizio,
mento di approfondimento delle metodologie empiriche di ricerca che
qualificare l'attività educativa

Domenici, professore straordinario alla Unicamillus - International University of
nces, è stato titolare della cattedra di Docimologia presso l'Università "Sapien-
a, e di Didattica presso l'Università "Roma Tre", dove ha ricoperto la carica di
irettore di facoltà e di dipartimenti di area educativa, e presieduto la "Fonda-
Tre-Education".

isano, professore ordinario di Pedagogia Sperimentale presso il Dipartimento
a dei Processi di Sviluppo e Socializzazione, "Sapienza" Università di Roma,
ieduto corsi di laurea in scienze dell'Educazione e della formazione e presiede
Laurea in Scienze della Formazione Primaria. È Presidente della Società
Ricerca Didattica e direttore del "Giornale italiano della ricerca educativa".

asi, professore ordinario di Pedagogia Sperimentale presso il Dipartimento di
ella Formazione" e il Dottorato di Ricerca in "Teoria e Ricerca Educativa e
ell'Ateneo "Roma Tre". Membro della Società Italiana di Ricerca Didattica
i interessi scientifici concernono lo studio dei processi cognitivi e motivazionali,
iche psico-pedagogiche dell'individualizzazione didattica e della valutazione.

€ 39,00 (i.i.)

ISBN 978-88-386-5546-3
9 788838 655463
www.mheducation.it

Gaetano Domenici • Pietro Lucisano • Valeria Biasi

Ricerca sperimentale e processi valutativi in educazione

Experimental Research and Evaluation Processes in Education

Con contributi di: E. Botta, G. Castellana, C. Corsini, A.M. De Luca, G. Moretti, E. Rubat du Merac, R. Trinchero, I. Vannini, R.M. Viganò

McGraw Hill

Comitato scientifico

Joao Barroso, *Universidade de Lisboa*; Richard Bates Deakin, *University of Melbourne*; Guido Benvenuto, *"Sapienza" Università di Roma*; Valeria Biasi, *Università "Roma Tre"*; Sara Bubb, *University of London, UK*; Jean-Émile Charlier, *Universitè Catholique de Louvain*; Cristiano Corsini, *Università "Roma Tre"*; Gaetano Domenici, *Unicamillus, International University of Health Sciences*; Peter EarLy, *University of London, UK*; James Levin, *City University of New York, USA*; Bruno Losito, *Università "Roma Tre"*; Fabio Lucidi, *"Sapienza" Università di Roma*; Pietro Lucisano, *"Sapienza" Università di Roma*; Massimo Marcuccio, *Università di Bologna*; Antonio Marzano, *Università di Salerno*; Romuald Normand, *École Normale Supérieure de Lyon, Institut Française de l'Éducation*; Paola Ricchiardi, *Università di Torino*; Vitaly V. Rubtzov, *Moscow State University for Psychology and Education*; Anna Salerni, *"Sapienza" Università di Roma*; Jaap Scheerens, *University of Twente, The Netherlands*; Noah W. Sobe, *Loyola University of Chicago*; Pat Thomson, *University of Nottingham, UK*; Roberto Trinchero, *Università di Torino*; Ira Vannini, *Università di Bologna*; Renata Viganò, *Università Cattolica di Milano*.

Mc Graw Hill

Copyright © 2021 McGraw-Hill Education (Italy), S.r.l.
Corso Vercelli, 40 - 20145 Milano (MI)
Tel. 02535718.1 - www.mheducation.it

I diritti di traduzione, di riproduzione, di memorizzazione elettronica e di adattamento totale e parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche) sono riservati per tutti i Paesi. Date le caratteristiche intrinseche di Internet, l'Editore non è responsabile per eventuali variazioni negli indirizzi e nei contenuti dei siti Internet riportati.

L'Editore ha fatto quanto possibile per contattare gli aventi diritto delle immagini che compaiono nel testo e resta a disposizione di chi non è stato possibile contattare.

Nomi e marchi citati nel testo sono generalmente depositati o registrati dalle rispettive case produttrici.

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.

Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi, Corso di Porta Romana n. 108, 20122 Milano, e-mail info@clearedi.org e sito web www.clearedi.org.

Portfolio Manager: Daniele Bonanno
Grafica di copertina: Feel Italia, Milano
Immagine di copertina: ©Milos Batinic/Shutterstock
Stampa: Logo S.p.A., Borgoricco (PD)

ISBN: 9788838655463

Printed in Italy
0 4 1 2 3 5 6 7 8 9 - 28 27 26 25 24

Indice

Introduzione XIII

Capitolo 1
Genesi e sviluppo della ricerca empirico-sperimentale in educazione 1
Renata Maria Viganò (Università Cattolica del Sacro Cuore)

1.1 Alcuni precursori 2
1.2 Fattori comuni e sviluppi differenziati nei diversi Paesi 3
1.3 Una possibile periodizzazione 5
1.3.1 Dalla fine del XVIII alla fine del XIX secolo 5
1.3.2 Dall'ultimo decennio del XIX secolo alla metà degli anni Trenta 5
1.3.3 Dalla metà degli anni Trenta agli inizi degli anni Cinquanta 9
1.3.4 Dagli inizi degli anni Cinquanta alla metà degli anni Sessanta 10
1.3.5 Dalla metà degli anni Sessanta ai giorni nostri 11
1.4 Guardando avanti 17
Riepilogo 21
Riferimenti bibliografici 22

Capitolo 2
Il paradigma scientifico: approcci e modalità di ricerca 25
Pietro Lucisano ("Sapienza" Università di Roma)

2.1 Il ricercatore e la sua ricerca 26
2.2 Il ricercatore e la scienza del suo tempo 27
2.3 Identificazione del problema e formulazione delle ipotesi di ricerca 31
2.3.1 Obiettivi della ricerca 31
2.3.2 Le ipotesi di ricerca 32

2.4 Atteggiamento e metodo scientifico 34
2.4.1 Metodo scientifico 35
2.4.2 Atteggiamento scientifico 35
2.5 Gli approcci di ricerca 38
2.5.1 L'approccio teorico 39
2.5.2 L'approccio storico 40
2.5.3 L'approccio comparativo 40
2.5.4 L'approccio clinico 41
2.5.5 La ricerca-azione 42
2.5.6 L'approccio empirico e sperimentale 43
2.5.7 L'approccio grounded 44
2.6. Le modalità prevalenti: descrivere, mettere in relazione, manipolare 45
2.6.1 Le modalità descrittive 45
2.6.2 Le modalità basate sulle relazioni 46
2.6.3 Le modalità sperimentali 46
2.7 Conclusioni 47
Riepilogo 48
Riferimenti bibliografici 48

Capitolo 3
Le fasi della ricerca e la scelta dei dati **51**
Pietro Lucisano ("Sapienza" Università di Roma), Giusi Castellana (Università "Roma Tre")

3.1 Le fasi della ricerca 52
3.2 La ripetibilità della ricerca 54
3.3 Il progresso scientifico e la serendipità 55
3.4 Le parole della ricerca 57
3.5 Le parole e i concetti 59
3.6 Le proposizioni 62
3.7 Leggi e teorie 63
3.8 I fatti del caso 65
3.9 Le variabili 66
3.10 Del togliere o accettare la complessità 67
3.11 Considerazioni conclusive 68
Riepilogo 68
Riferimenti bibliografici 69

Capitolo 4
La misura in educazione **71**
Gaetano Domenici (Unicamillus - International University of Health Sciences)

4.1 Osservazione e classificazione: i primi passi di ogni scienza 72
4.2 Qualitativo e quantitativo: un nesso assai stretto 73
4.3 Strumenti concettuali per comprendere e controllare gli eventi 73
4.4 Condizioni per misurare 74
4.5 Scale e tecniche di misurazione 75
4.5.1 La scala nominale 75
4.5.2 La scala ordinale 77
4.5.2.1 Principali operazioni consentite con la scala ordinale. Mediana e punteggi percentili 78
4.5.3 La scala ad intervalli 79
4.5.3.1 Principali operazioni consentite con la scala ad intervalli (media aritmetica, indici di variabilità o dispersione dei punteggi, correlazioni) 80
4.6 Misure di tendenza centrale e misure di variabilità dei punteggi 82
4.6.1 Media, mediana, moda 82
4.6.2 Deviazione standard 86
4.6.3 Distribuzione pentenaria 88
4.6.4 Punti Z e punti T 90
4.7 Campioni e statistiche. Cenni introduttivi 92
Riepilogo 96
Riferimenti bibliografici 96

Capitolo 5
Testing, ricerca e valutazione **99**
Valeria Biasi (Università "Roma Tre"), Gaetano Domenici (Unicamillus - International University of Health Sciences)

Parte prima
5.1 Teorie, costrutti, variabili, fattori, indicatori 100
5.2 Costruzione e validazione di scale di valutazione 102

5.2.1 Modelli e fasi per la costruzione di un Questionario e/o di un Test di profitto 105
5.2.2 Dai costrutti agli indicatori 106
5.2.3 Definizione degli item nei Test di atteggiamento 107
5.2.4 Standardizzazione, attendibilità e validità delle scale di valutazione 112
5.3 Testing e processi di valutazione degli apprendimenti 112
Parte seconda
5.4. Test oggettivi di verifica degli apprendimenti 114
5.4.1 Test oggettivi di profitto o prove strutturate 114
5.4.2 Tipologia e struttura dei quesiti 115
Vero/falso 116
Corrispondenze o item di confronto 117
Completamenti (o *cloze*) 118
Scelte multiple 118
5.4.3 Le Prove di Comprensione della Lettura (PCL) 121
5.4.4 Strutturazione dei quesiti, validità e attendibilità delle prove 123
5.4.5 Quando è necessario standardizzare le prove 124
5.4.6 Alcuni modi di classificare le prove oggettive di profitto 125
5.4.7 Come costruire test oggettivi di conoscenza 126
5.4.7.1 Determinazione degli obiettivi di verifica 127
5.4.7.2 Analisi degli obiettivi 128
5.4.7.3 Campionamento degli obiettivi, determinazione degli item e della lunghezza del test 129
5.4.8 Prove semistrutturate di verifica degli apprendimenti: caratteristiche e tipologie 132
5.4.9 Colloquio strutturato, semistrutturato e Riflessione parlata come strumenti di ricerca. 135
Riepilogo 139
Riferimenti bibliografici 140

Capitolo 6
Item analisi e misure di tratti latenti 143
Pietro Lucisano, Andrea Marco De Luca, Emanuela Botta ("Sapienza" Università di Roma)

6.1 Misurare con misura 144

6.2 Possiamo misurare le stesse cose in modo diverso 150
6.3 La facilità/difficoltà di una domanda 154
6.4 La discriminatività di una domanda 156
6.5 Le stesse cose misurate in modo diverso danno risultati diversi 163
6.6 Chi è senza errore scagli la prima pietra 164
6.7 Dalle prove standardizzate alle prove su misura 167
Riepilogo 169
Riferimenti bibliografici 169

Capitolo 7
Disegni sperimentali di ricerca, microsperimentazione didattica e osservazione sistematica 171
Valeria Biasi (Università "Roma Tre"), Gaetano Domenici (Unicamillus - International University of Health Sciences)

7.1 Disegni sperimentali: modello sperimentale classico e quasi-sperimentazioni 172
7.2 La sperimentazione a scuola 181
7.3 Microsperimentazione per lo studio dell'atteggiamento del docente e degli effetti sull'apprendimento degli allievi. Procedure e risultati di una ricerca 184
7.3.1 Auto-efficacia e atteggiamento motivante del docente 186
7.3.2 La ricaduta positiva sugli apprendimenti degli allievi 190
7.4 Auto-valutazione nella didattica laboratoriale universitaria on-line attraverso prove semistrutturate. Sintesi di un'apposita microsperimentazione. 191
7.4.1 Obiettivi dell'indagine 191
7.4.2 Struttura e caratteristiche delle attività del Laboratorio on-line 192
7.4.3 Ricadute sui livelli di apprendimento degli allievi 196
7.5 L'osservazione sistematica come strumento di ricerca e di qualificazione della didattica 196
7.6 Tecniche e strumenti dell'osservazione sistematica nei processi educativi 198
Riepilogo 203
Riferimenti bibliografici 203

Capitolo 8
La sperimentazione didattica. Un dialogo possibile con la ricerca-azione e la ricerca-formazione **205**
Ira Vannini (Università "Alma Mater" di Bologna)

8.1 La sperimentazione nella normativa scolastica italiana 206
8.2 Innovare e sperimentare nella didattica 211
8.2.1 La questione della validità interna nelle sperimentazioni 213
8.2.2 L'importanza di un approccio sperimentale nell'ambito della ricerca didattica 216
8.3 Nella scuola, l'insegnante... "sperimenta" 218
8.3.1 Le vie aperte dalla Ricerca-Azione 219
8.3.2 La sperimentazione: un dialogo difficile con le istanze della R-A 223
8.4 Fare ricerca e sperimentazione con la "ricerca-formazione" 224
Riepilogo 228
Riferimenti bibliografici 229

Capitolo 9
Studio di caso e focus group **233**
Emiliane Rubat du Mèrac ("Sapienza" Università di Roma)

9.1 Studio di caso 234
9.1.1 Cos'è lo studio di caso? 234
9.1.2 Quali sono le diverse categorie di Studio di caso? 236
9.1.3 Organizzazione del lavoro 237
9.1.4 La triangolazione 239
9.2 Focus group 240
9.2.1 Cos'è il Focus group? 240
9.2.2 Quando utilizzare il Focus group? 241
9.2.3 Quali sono le difficoltà e i limiti del Focus group? 242
9.2.4 Focus group: dimensione, composizione e durata 243
9.2.5 La scelta del luogo di svolgimento del Focus group e i Focus group virtuali 246
9.2.6 Preparazione della guida all'intervista 247
9.2.7 Prepararsi per l'interazione con i partecipanti 250
9.2.8 La raccolta dei dati 252
9.2.9 L'analisi dei dati 255

9.2.10 In conclusione, la restituzione 257
Riepilogo 257
Riferimenti bibliografici 258

Capitolo 10
Scelta degli strumenti per la ricerca educativa **265**
Gaetano Domenici (Unicamillus - International University of Health Sciences), Giovanni Moretti (Università "Roma Tre")

10.1 Coerenza e adeguatezza degli strumenti per la ricerca 266
10.1.1 Ipotesi, piano della ricerca e scelta degli strumenti 266
10.1.2 Strumenti di rilevazione dei dati come costrutti teorici. Il problema della loro coerenza ed adeguatezza agli scopi della ricerca. 268
10.2 Gli strumenti della ricerca. L'osservazione sistematica 270
10.2.1 L'osservazione sistematica tramite il sistema di categorie 271
10.2.2 Gli inventari o sistemi di segni 274
10.2.3 Le Scale di valutazione e i questionari di autovalutazione 276
10.2.4 I diari 279
10.2.5 Il diario di bordo 282
10.2.6 Il metodo degli episodi critici 285
10.2.7 Il portfolio 286
Riepilogo 288
Riferimenti bibliografici 288

Capitolo 11
Docimologia e valutazione educativa **291**
Gaetano Domenici (Unicamillus - International University of Health Sciences)

11.1 Nascita e sviluppo della docimologia 292
11.2 Produzione e diffusione dei test oggettivi di conoscenza e prime indagini sul profitto scolastico 294
11.3 Orientamenti della ricerca docimologica 295
11.4 Nuovi strumenti per la verifica dell'apprendimento 296
11.5 Modelli e strumenti valutativi come costrutti. Didattica e valutazione 299

11.6 Misurazione, verifica e giudizi valutativi 300
11.7 Attribuzione di significato ai dati valutativi. Criteri assoluti e criteri relativi 301
11.8 Restituzione e discussione dei dati valutativi 303
11.9 Strutturazione (documentazione) e comunicazione del giudizio valutativo 303
11.10 Autonomia scolastica, organizzazione didattica e valutazione 304
11.11 Promuovere la cultura della valutazione 306
11.12 La valutazione come risorsa 306
11.13 Informazioni, valutazione, decisioni in un'ottica sistemica 307
11.14 Quali informazioni 308
11.15 Determinazione, uniformità, disponibilità e disaggregazione delle informazioni 310
11.16 Valutazione e processi decisionali per l'apprendimento: come si struttura una decisione 311
Riepilogo 314
Riferimenti bibliografici 315

Capitolo 12
Pregi e limiti delle indagini internazionali su grandi campioni 319
Cristiano Corsini (Università "Roma Tre")

12.1 Le indagini internazionali in campo educativo 320
12.2 Le indagini in corso 324
12.3 La definizione dei costrutti: la "reading literacy" tra IEA e OCSE 332
12.4 Potenzialità e cautele: un bilancio 337
Riepilogo 340
Riferimenti bibliografici 340
Sitografia 342

Capitolo 13
Ricerca e processi valutativi nella formazione e-learning 343
Roberto Trinchero (Università degli Studi di Torino)

13.1 Tecnologie per apprendere, insegnare, valutare 344
13.2 Cosa ci insegna la ricerca sull'efficacia delle tecnologie per insegnare e apprendere 346

13.2.1 L'importanza di una buona guida istruttiva nell'uso delle tecnologie 349
13.2.2 L'importanza di un'elaborazione profonda e significativa dei contenuti da apprendere 350
13.2.3 Il ruolo del feedback e della metacognizione 351
13.2.4 Il ruolo dell'interazione con i propri pari 352
13.3 Integrare formazione e valutazione con l'ausilio delle tecnologie: un modello per la didattica in presenza e a distanza 353
13.4 Considerazioni conclusive 359
Riepilogo 360
Riferimenti bibliografici 360

Capitolo 14
Comunicazione dei risultati della ricerca scientifica e problematiche connesse 365
Valeria Biasi (Università "Roma Tre")

14.1 La comunicazione scritta dei risultati scientifici attraverso la stesura di un rapporto di ricerca 366
14.2 Tipologie di scritti scientifici 371
14.3 Questioni deontologiche nella conduzione della ricerca e nella rendicontazione dei risultati 374
14.3.1 Principi etici legati alla conduzione della ricerca 374
14.3.2 Principi etici legati al processo di valutazione e pubblicazione dell'articolo 374
14.3.3 Considerazioni etiche legate all'autore 375
14.4 La scelta delle Sedi Editoriali 375
14.4.1 Criteri per la proposta di un articolo a una rivista scientifica che adotta la procedura peer-review 375
14.5 La diffusione delle conoscenze scientifiche in ambito educativo: potenzialità e aspetti problematici 377
Riepilogo 379
Riferimenti bibliografici 380

Introduzione

Il presente volume è rivolto sia ai giovani ricercatori in formazione nei settori scientifico-disciplinari di area pedagogica e viciniori, sia ai docenti impegnati nei percorsi di prima formazione e di formazione in servizio: laddove, come sappiamo, l'elemento culturale e conoscitivo teorico-pratico più debole risulta essere in genere, di fatto, quello metodologico.

Se si considera poi che nella realtà educativa formale per cercare di risolvere nei migliori modi possibili i tanti problemi quotidiani - in *primis* quelli collegati alle difficoltà di apprendimento sul versante cognitivo come su quello affettivo e relazionale - i docenti oltre che fare appello alle loro specifiche conoscenze teorico-pratiche devono spesso, in qualche modo, operare essi stessi come ricercatori, si comprende allora che l'attenzione verso la metodologia della ricerca non è mai troppa.

Si può anzi dire, a nostro avviso, che nel panorama pedagogico italiano la preparazione metodologica sia ormai diventata *la questione delle questioni nella formazione dei nuovi ricercatori.*

Da queste esigenze ha preso avvio il percorso di stesura del presente volume che, a partire dalla nascita ed evoluzione della pedagogia sperimentale in Europa, con attenzione al contesto storico-culturale in cui è maturata la ricerca educativa empiricamente fondata nel panorama internazionale, affronta nei 14 capitoli allestiti, i principali temi di natura epistemologica e metodologica in ambito educativo.

L'obiettivo perseguito è quello di rendere comprensibili e quindi applicabili principii, metodi e strumenti utili per poter far compiere ai giovani ricercatori almeno i primi passi dell'impegno di ricerca in campo educativo e agli insegnanti le operazioni di problematizzazione degli ostacoli che spesso impediscono nella didattica quotidiana l'"apprendimento insegnato", e la prefigurazione delle più probabili linee risolutive, adattando ai vincoli di contesto le teorie, le procedure, gli strumenti più adeguati allo scopo.

Nei primi tre capitoli vengono prese in esame alcune questioni fondamentali della ricerca empirico-sperimentale in educazione d'apprima in chiave storica (Capitolo1), seguono la presentazione dei principali approcci di ricerca nell'ambito del paradigma scientifico (Capitolo 2) e delle fasi della ricerca (Capitolo 3)

Nei successivi quattro capitoli si approfondiscono le questioni relative alla misurazione in generale (Capitolo 4), alle procedure del Testing in educazione (Capitolo 5) e dell'elaborazione statistica compresa l'Item Analysis per la rilevazione di fattori latenti (Capitolo 6).

Vengono poi esaminati i principali disegni di ricerca: dal modello sperimentale classico alla quasi-sperimentazione, si offrono dettagliati riferimenti inoltre alle modalità della micro e della macro-sperimentazione (Capitolo 7), a quelle della ricerca-azione e della ricerca-formazione (Capitolo 8), agli studi di caso e alla metodologia del *focus group* nei contesti educativi (Capitolo 9).

A più riprese vengono presentati ed esaminati strumenti e procedure codificate per la raccolta dei dati: si affronta peraltro il tema centrale della coerenza e/o adeguatezza degli strumenti agli scopi della ricerca, ed il ruolo ricoperto dagli strumenti in qualità di veri e propri costrutti teorici (Capitolo 10).

Si passa quindi ad esaminare la nascita e lo sviluppo della Docimologia e dei principali aspetti formali dei processi di valutazione in campo educativo (Capitolo 11).

Segue l'esame, anche in chiave critica, delle principali ricerche condotte in ambito internazionale su grandi campioni (Capitolo 12) e la presentazione delle caratteristiche peculiari della ricerca sui processi valutativi nella formazione e-learning (Capitolo 13).

Il volume consegna infine al lettore indicazioni sintetiche per la comunicazione dei risultati di una ricerca, soprattutto attraverso la stesura di un rapporto scientifico, quale fase integrante dell'applicazione dello stesso metodo scientifico (Capitolo 14).

Per quanto estesa ed articolata sia risultata la trattazione dei problemi - peraltro trattati dagli autori in contesti diversi e qui ripresi e aggiornati nel quadro delle attuali tendenze di ricerca - non si poteva né si voleva che fosse esaustiva. E tuttavia, le sollecitazioni offerte per approfondimenti iniziali delle diverse questioni, vengono riprese ed ampliate, fuori del volume, ma come sua parte integrante fondamentale, su una apposita piattaforma e-learning.

Viene così messo a disposizione un ampio apparato didattico articolato in significative sintesi degli argomenti trattati, materiali di approfondimento anche per temi per cui non è stato possibile per motivi di spazio dare adeguato rilievo nei Capitoli allestiti, un apparato che potrà peraltro essere periodicamente aggiornato in relazione alle più cruciali tematiche emergenti. Un ricco dispositivo di esercizi di approfondimento ed estensione delle conoscenze via via acquisite dotato di prove di auto-valutazione conchiude l'opera *in progress* qui presentata.

Abbiamo scelto di utilizzare come *incipit* di tutti i capitoli del libro citazioni da un testo narrativo di un grande scrittore italiano: Italo Calvino. Il testo *Palomar* racconta per episodi le osservazioni e le riflessioni di un uomo che come dice l'autore "si mette in marcia per raggiungere, passo a passo, la saggezza. Non è ancora arrivato". L'esperienza di Palomar, il cui nome è ripreso da un grande telescopio, da uno strumento cioè che ha l'obiettivo di consentire di vedere lontano e di vedere meglio, rappresenta la difficoltà di cogliere la complessità del reale attraverso la sua modellizzazione. Il messaggio ai lettori consiste nel fatto che affrontando tematiche relative alla ricerca educativa, se è necessario fare ricorso al metodo scientifico e farlo nel modo più rigoroso possibile, è altrettanto necessario fare costantemente ricorso alla riflessione e all'intelligenza critica, per evitare di cadere nella trappola di una interpretazione solo normativa dei fenomeni con i quali ci confrontiamo. Ogni risultato della nostra ricerca, sia pure ottenuto con il massimo rispetto delle norme di un buon lavoro scientifico, rimane sempre un fondamentale elemento di partenza per una nuova riflessione critica.

GD, PL, VB

Roma, Luglio 2021

1

Genesi e sviluppo della ricerca empirico-sperimentale in educazione

Il capitolo ricostruisce la genesi e lo sviluppo della ricerca empirico-sperimentale in educazione, inquadrandola alla luce dell'intreccio di fattori culturali e scientifici, ma anche socioeconomici e politici, che ne spiegano le origini e le dinamiche evolutive, a livello internazionale e nei diversi Paesi.
Il primo punto descrive brevemente il contributo di alcuni studiosi, movimenti, orientamenti antecedenti la fine del XIX secolo i quali possono essere considerati, per ispirazione e intuizioni pedagogico-educative, precursori di tale indirizzo di ricerca anche se non propriamente corrispondenti ad essa.
Il secondo punto sintetizza i principali fattori che tra la fine del XIX e l'inizio del XX secolo concorrono all'emergere dell'approccio empirico-sperimentale nella ricerca educativa nei vari Paesi, per ciascuno dei quali indica le specificità più rilevanti e i contributi più autorevoli.
Il terzo punto si sofferma sullo sviluppo attraverso i decenni della ricerca in oggetto, proponendo una periodizzazione in fasi caratterizzate ciascuna dalla convergenza e prevalenza di aspetti culturali, scientifici, socioeconomici e politici i quali di volta in volta influiscono sul pensiero e sulla pratica della ricerca medesima.
Il quarto punto propone alcune riflessioni critico-prospettiche anche alla luce dei principali snodi problematici attuali della ricerca empirico-sperimentale in educazione.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Rapporti che storicamente sono intercorsi tra ricerca e fattori di contesto: culturali, scientifici sociali, economici, politici.
- Complessità del percorso epistemologico e metodologico della ricerca empirico-sperimentale in educazione.
- Centralità e problematicità del rapporto fra teoria, ricerca e pratica/che educative.
- Necessità di senso critico e ampia visione culturale anche per la ricerca empirico-sperimentale in educazione.

> Che sollievo se riuscisse ad annullare il suo io parziale e dubbioso nella certezza d'un principio da cui tutto deriva! Un principio unico e assoluto da cui prendono origine gli atti e le forme? Oppure un certo numero di principi distinti, linee di forza che s'intersecano dando una forma al mondo quale appare, unico, istante per istante?
>
> (Italo Calvino, *Palomar*)

1.1 Alcuni precursori

Una recensione e illustrazione sistematica della genesi e dello sviluppo della ricerca empirico-sperimentale in educazione implicherebbe un ampio cammino a ritroso nella storia dell'educazione. Gli obiettivi del presente testo impongono una sintesi e una semplificazione, soffermandosi su alcuni passaggi salienti dell'evoluzione dell'approccio di ricerca in parola. Di seguito sono perciò menzionati solo i più autorevoli precursori, sulla base delle informazioni documentate dagli studi storici (Depaepe, 2001; Mialaret, 2016).

All'inizio del XVIII secolo, Comenius (1592-1670) evoca implicitamente la necessità di una didattica sperimentale quando si rammarica che l'insegnamento scolastico o accademico non sia, come altre professioni, oggetto di studi scientifici. Due secoli dopo, Herbart (1776-1841) formula questo auspicio in maniera più precisa quando prevede di annettere una "scuola sperimentale" all'università. Quando nel 1808 è chiamato a succedere a Kant sulla cattedra di filosofia e pedagogia all'Università di Königsberg, presenta al curatore dell'Università un progetto che sarà poi ripreso da Dottrens.

Merita inoltre citare Cournot (1801-1877), il cui testo avrebbe potuto essere scritto all'inizio del XX secolo. Nel suo *Essai sur les fondements de nos connaissances* del 1851 apre una prospettiva nuova e non esita, come mostrerà poi Debesse, a elevare la pedagogia al rango delle discipline sperimentali. "In pedagogia come in medicina - scrive - la sperimentazione diretta non è impossibile anche se il rispetto per la natura umana e lo scopo dell'azione aggiungono alla sperimentazione difficoltà intrinseche e ne limitano la portata". Per acquisire informazioni sufficientemente valide Cournot sostiene inoltre l'uso delle statistiche per esaminare, ad esempio, i rapporti tra le diverse competenze o tra il livello di istruzione e il tasso di criminalità. Anticipando l'impiego di moderne tecniche di analisi fattoriale, propone anche di individuare le caratteristiche di personalità che risultano indipendenti da quelle fisiche, per confermare o smentire teorie deterministiche.

L'atteggiamento scientifico nella ricerca educativa, osserva Mialaret (2016), ha quindi origini più lontane nel tempo - seppur collegate a riflessioni non continuative e sistematiche - rispetto all'opinione diffusa che ancora sia alle battute iniziali.

1.2 Fattori comuni e sviluppi differenziati nei diversi Paesi

Nella genesi della pedagogia sperimentale occorre ritrovare nei diversi Paesi fattori e contributi che concorrono a diffondere il cosiddetto *ésprit scientifique* nella conoscenza dei processi educativi.

Nella sua dissertazione dottorale del 1948 Drese formula l'ipotesi che la genesi della pedagogia sperimentale sia avvenuta in ambito tedesco, sottolineando in particolare il ruolo primario di Lay e di Meumann. Su questa ipotesi concordano più studiosi (De Landsheere, 1986; Montealegre, 1963; De Corte et al., 1990).

Anche nel contesto inglese, nella prima metà del XX secolo, la psicologia applicata all'educazione e la costruzione di nuove tecniche di misura come i test hanno un rapido sviluppo; diversa però è la situazione della prospettiva sperimentale in chiave pedagogica. Secondo Bréhaut, studioso della realtà scientifica inglese, tale situazione è riconducibile a diversi fattori: l'insularità, la mancanza di fiducia verso l'approccio analitico allo studio dei fatti umani e, nella fase storica, l'avversione per la Germania e l'America (Bréhaut, 1973). In generale, nel contesto inglese lo sviluppo della sperimentazione pedagogica non è rapido né origina approfondimenti originali; i contributi più rilevanti attengono infatti all'ambito statistico e della misurazione, mentre la riflessione sulla struttura epistemologica della scienza educativa è molto meno presente. Fa eccezione in questo quadro Winch (1914) che, attento alla necessità di radicare lo spirito scientifico nella pratica e nell'innovazione educativa quotidiana, sviluppa contributi che toccano un orizzonte riflessivo e critico più ampio.

Gli studiosi concordano nell'indicare che le principali coordinate a cui collegare la genesi e lo sviluppo della ricerca empirico-sperimentale in educazione riguardano l'Europa occidentale alla fine del XIX secolo e nel volgere del XX; quanto accade in America va però considerato con particolare attenzione al contesto scolastico e accademico. La realtà culturale e scolastica americana ha caratteristiche peculiari che la differenziano dal continente europeo: Becchi (1969) pone in risalto il decentramento scolastico, il credo puritano nella produttività, il mito dell'efficienza e la fiducia nelle possibilità della scienza e della tecnica. In ambito pedagogico, l'impegno per la dimensione scolastica è prioritario rispetto allo sforzo di elaborazione e sistematizzazione dei concetti; la preoccupazione pragmatica prevale su quella teorica (Becchi, 1969). La volontà di innovare la scuola è animata da una duplice tendenza: da una parte l'istanza democratica e sociale, dall'altra quella volta al controllo scientifico dei fatti educativi; tali orientamenti hanno in Dewey e

in Thorndike due rappresentanti emblematici (De Landsheere, 1986). Le cautele speculative che in Europa avevano ostacolato l'applicazione della sperimentazione alla dimensione educativa sono assenti; la priorità è accordata al piano pratico. L'obiettivo che guida i ricercatori americani è il miglioramento della prassi scolastica; in forza di ciò, Thorndike traduce nella pratica lo scientismo radicale del Lay. In *An Introduction to the Theory of Mental and Social Measurement* afferma: "tutto quello che esiste, esiste in una certa quantità e dunque può essere misurato" (Thorndike, 1904, p. 58). In generale, ricercatori e psicologi statunitensi contribuiscono alla diffusione dell'orientamento sperimentale nella ricerca educativa impegnandosi nell'elaborazione di nuovi strumenti di misurazione. La loro riflessione non muove tuttavia dall'orizzonte pedagogico ma giunge ad esso in maniera indiretta; il loro contributo è essenziale sotto il profilo metodologico ma non tocca quello epistemologico.

Un discorso a parte merita Dewey la cui statura sicuramente giganteggia nel contesto statunitense. Con riferimento specifico alla pedagogia sperimentale va però situato nella posizione che ben delinea De Landsheere (1986, p. 15) quando precisa che Dewey "non accorda affatto importanza allo studio quantitativo dei comportamenti. Non solo questo tipo di studio è ai suoi primi passi in quel tempo, ma il suo carattere analitico, il suo riduzionismo appaiono a Dewey incompatibili con l'approccio globale ai fenomeni umani". Negli Stati Uniti il progressivismo di Dewey troverà concreta attuazione in Kilpatrick, in Rugg e in Brameld.

Approfondimento 1.1
JOHN DEWEY: PREFAZIONE A ESPERIENZA E EDUCAZIONE (1938)

Tutti i movimenti sociali comportano conflitti che si riflettono in dibattiti intellettuali. Non sarebbe segno di salute se un tema di interesse sociale del peso dell'educazione non fosse dunque un campo di scontri, teorici e pratici. Ma per la teoria, almeno per la teoria che costituisce una filosofia dell'educazione, i conflitti pratici e le controversie sollevate a partire da questi conflitti, pongono unicamente un problema. Compito di una intelligente teoria dell'educazione è quello di accertare le cause dei conflitti in essere e poi, invece di prendere parte per gli uni o per gli altri, indicare un piano di interventi che proceda da un livello più profondo e più comprensivo di quello rappresentato dalle pratiche e dalle in conflitto.
Assegnare questo compito alla filosofia dell'educazione non significa che essa debba tentare di trovare un compromesso fra opposte scuole di pensiero, una specie di via media, e neppure una combinazione di elementi presi qua e là da tutte le scuole. Significa che è necessario introdurre un nuovo ordine di idee che avvii nuovi modi di pratica. È questa appunto la ragione perché è così difficile svolgere una filosofia dell'educazione, quando ci si allontana dalla tradizione o dal costume. Per questa ragione dirigere una scuola, che si ispira a un nuovo ordine di idee, è molto più difficile che indirizzarla per il sentiero battuto. Ne consegue che ogni movimento verso un nuovo ordine di idee e di attività dirette da esse provoca, prima o poi, un ritorno a quelle che sembrano le idee e le pratiche più semplici e fondamentali del passato - come lo attesta ora nell'educazione il tentativo di far risorgere i principi della Grecia antica e del medioevo.
In questo senso alla fine di questo volumetto ho suggerito a coloro che guardano innanzi a un nuovo movimento nell'educazione, adatto alle esigenze di un nuovo ordine sociale, che si preoccupino dell'educazione in sé e per sé e non già di qualche "ismo" concernente l'educazione, come sarebbe per esempio il "progressismo". Poiché, suo malgrado, ogni movimento che pensa e opera in base a un "ismo" è talmente coinvolto nella reazione contro altri "ismi" che finisce con l'essere involontariamente controllato da essi. Il che lo induce a formulare i suoi principi per reazione ad essi invece di muovere da una comprensiva visione costruttiva dei bisogni, dei problemi e delle possibilità effettive. Quale esso sia, il valore di questo saggio consiste nel tentativo di richiamare l'attenzione su problemi dell'educazione di più largo respiro e più profondi, tali da suggerire il loro stesso inquadramento.

1.3 Una possibile periodizzazione

Ogni periodizzazione in segmenti temporali delimitati ha carattere artificioso; può tuttavia essere utile individuare, nella scia di De Landsheere (1986), cinque periodi generali in cui si dispiega lo sviluppo della ricerca empirico-sperimentale in educazione.

1.3.1 Dalla fine del XVIII secolo alla fine del XIX

Il primo periodo è detto prescientifico ed è collocabile dalla fine del XVIII secolo al termine del XIX. Già all'inizio del XIX secolo si avverte che la pedagogia non resterà estranea al fervore scientifico dell'epoca. Kant, Herbart, Spencer, Bay, Pestalozzi, Diesterweig, Barnard e altri provano a loro modo la necessità di sperimentare, ma restano al contempo profondamente influenzati dalle forti componenti storiche e filosofiche della scienza dello spirito tedesca, con le sue combinazioni logico-deduttive di aspetti ora empirici ora idealistici e con tutta la gamma intermedia fra i due estremi.

1.3.2 Dall'ultimo decennio del XIX secolo alla metà degli anni Trenta

Il secondo periodo, dall'ultimo decennio del XIX secolo alla metà degli anni Trenta, è contraddistinto dalla fioritura della ricerca quantitativa. Tra la fine del XIX secolo e l'inizio del XX sino alla guerra del 1914-1918 vi è uno sviluppo significativo di tale orientamento. L'approccio sperimentale è incentrato soprattutto sulla

misurazione quantitativa e si oppone con vigore al formalismo e al mentalismo, trascurando per lo più la dimensione teoretica del pensiero educativo.

Le trasformazioni culturali, economiche e sociali non sono senza impatto sulla sfera dell'educazione (Montalbetti, 2002). Nella seconda metà dell'Ottocento l'Europa è infatti sede di importanti cambiamenti economici e politici; di là dello scontro in Crimea del 1854 e della guerra franco-prussiana del 1870, l'epoca in oggetto è contraddistinta da una pace diffusa e dall'espansione europea in Africa e in Asia (Polanyi, 1974). Tali condizioni favoriscono la ripresa della rivoluzione industriale (Landes, 1978). L'impiego delle moderne scoperte scientifiche modifica il sistema di produzione e di organizzazione del lavoro; il processo di industrializzazione ingenera cambiamenti sul piano sociale: i lavoratori delle fabbriche si spostano dalle campagne alle città. Le necessità produttive imprimono spinta alla ricerca scientifica, applicata e tecnologica; in virtù dell'opera di studiosi come Bernard e Pasteur, il metodo sperimentale è esteso a nuovi campi: l'opera di Bernard *Introduction à la médicine expérimentale* (1865) non solo attesta il valore della sperimentazione in medicina ma rappresenta uno sforzo considerevole per stabilire la dottrina, il valore e gli effetti del nuovo metodo. In ambito filosofico, il positivismo interpreta e promuove l'entusiasmo per la scienza; con il *Cours de philosophie positive* (1877) Comte riduce la conoscenza umana allo studio dei dati del mondo sensibile eliminando ogni riferimento metafisico; la fiducia nelle possibilità della ragione e della scienza è illimitata e si diffonde l'idea di un progresso infinito volto al dominio dell'uomo sulla natura.

Secondo Depaepe l'insieme delle circostanze menzionate contribuisce, a diverso titolo, all'introduzione della sperimentazione in ambito pedagogico. Sarebbe però erroneo pensare a quest'ultima come a un mero precipitato di un preciso momento storico; in realtà le determinanti remote possono essere colte in alcuni contributi fondamentali della riflessione pedagogica precedente. Rousseau (1976) aveva insistito sulla necessità di osservare il fanciullo per attuare un intervento educativo efficace, auspicando che gli educatori modulassero la loro azione alla luce della conoscenza dell'allievo. In un altro contesto Kant (1959, trad.it.) aveva indicato come errore grossolano credere che nell'educazione non fossero necessari esperimenti rigorosi e si potesse decidere in linea esclusivamente teorico-razionale se qualcosa fosse buono o meno buono; Herbart (1997, trad.it.) aveva ulteriormente precisato tale orientamento. Dalla seconda metà dell'Ottocento tali considerazioni trovano un vero e proprio fattore di propulsione nel mutato quadro produttivo ed economico.

Tra gli effetti della società industriale vi è la necessità di costruire sistemi di istruzione su larga scala; essa necessita di un'organizzazione razionale del lavoro e di persone che sappiano utilizzare gli strumenti e che posseggano le conoscenze di base necessarie per il sistema produttivo. A fondamento della riorganizzazione

sociale ed educativa, lo spirito positivo teorizza l'estensione della sperimentazione dalle scienze naturali a quelle umane, quindi anche all'educazione: è compito della scienza determinare le leggi educative; in questa visione trova terreno fertile l'idea della sperimentazione (Montealegre, 1963).

Il naturalismo darwiniano in forte ascesa nel periodo in oggetto apre la strada all'idea che anche lo sviluppo umano sia determinato da fattori ereditari e dall'ambiente. Negli Stati Uniti, Hall è promotore dello studio scientifico del bambino; in Europa nasce il movimento della *pedologia*, intesa come scienza integrale e sperimentale del fanciullo, sotto l'impulso di Schuyten e Joteyko.

Lo sviluppo dell'approccio scientifico-sperimentale nelle scienze umane prende piede prima nella psicologia. Il riferimento è Wundt, che perviene alla costituzione di una scienza autonoma e sperimentale dei processi psichici e nel 1879 fonda a Lipsia il primo laboratorio di psicologia sperimentale. Progressivamente alcuni studiosi di area pedagogica iniziano a rifarsi agli studi psicologici e a mutuarne i metodi di ricerca; secondo Depaepe la psicologia sperimentale è "la fonte di ispirazione più concreta che contribuisce allo sviluppo della pedagogia sperimentale" (1998, p. 32).

Fra l'ultimo decennio del XIX secolo e l'inizio del XX il metodo sperimentale applicato allo studio dell'educazione si diffonde in diversi Paesi: in Germania con Lay e Meumann, in Inghilterra con Winch, in Francia con l'impulso di Binet, in Svizzera con il contributo indiretto di Claparède e Bovet. In Belgio in quel periodo domina il nome di Decroly: Il volumetto *Introduzione alla Pedagogia Quantitativa*, firmato con Buyse e pubblicato nel 1929, riprende e usa il lavoro di Mercante, studioso argentino che già a fine Ottocento applicava le statistiche allo studio dei problemi pedagogici. In Italia la personalità e le opere della Montessori influiranno sulla pedagogia di fine secolo.

Il sopraggiungere del conflitto mondiale fa però segnare in Europa una battuta d'arresto al movimento per la ricerca sperimentale in educazione; diversa è la situazione negli Stati Uniti dove invece il successo dei test ha rapida diffusione trovando impiego nei processi di selezione delle reclute militari e nell'organizzazione dell'industria, con una conseguente importante ricaduta nello sviluppo della ricerca sperimentale anche in campo educativo (Buyse, 1928). In Europa la situazione è più complessa poiché, nonostante l'attività sui test di Binet e Simon, sia l'impatto economico del conflitto sia la tradizione culturale in molte regioni europee costituiscono una remora all'accoglienza su larga scala di tale orientamento.

Nella genesi dell'approccio empirico-sperimentale alla ricerca educativa occupa un ruolo meritevole di attenzione anche la pedologia, termine con cui Crisman (1986) definisce lo studio del bambino secondo tutte le prospettive, per conoscere in modo profondo la sua natura. Con Blum (1899) la pedologia inizia a diffondersi proponendosi come scienza sperimentale e integrale del bambino, con l'ado-

zione dei metodi delle scienze naturali e la loro applicazione allo studio di tutte le dimensioni dello sviluppo infantile (Depaepe, 2001). Il movimento pedologico ha un rapido sviluppo in Europa occidentale, sino alla fondazione ad Anversa nel 1898 del primo laboratorio di Pedologia da parte dello Schuyten e l'ampia opera di sviluppo e diffusione scientifica negli anni seguenti a opera di Ioteyko, dell'*Université Libre de Bruxelles* (Ioteyko, 1912). L'entusiasmo pedologico subisce però un declino considerevole con l'arrivo della prima guerra mondiale e, al termine del conflitto, ritrova slancio ma progressivamente si esaurisce. In realtà, osserva Depaepe (2001), il rapido declino della pedologia va imputato non solo alle circostanze storiche, ma anche alla fragilità intrinseca del suo sistema concettuale legato a una visione riduttiva della scienza, alla fiducia cieca nelle potenzialità risolutrici del progresso scientifico, alla fede assoluta nell'oggettività del metodo positivo.

Nella progressiva estensione del metodo sperimentale applicato allo studio dell'uomo e in particolare dell'educazione, il modello di C. Bernard gioca un ruolo essenziale. Medico francese, studia le basi concettuali e metodologiche della ricerca fisiologica e nel 1865 espone le sue idee nell'*Introduction à l'étude de la médicine expérimentale*. Nel suo testo teorizza la possibilità di estendere il metodo sperimentale allo studio dell'uomo e riflette in particolare sullo statuto epistemologico della medicina nel passaggio di questa da sapere congetturale a sapere scientifico-sperimentale. Molti riconoscono a Bernard il merito di aver esteso il metodo sperimentale al campo delle scienze sociali. L'influsso in ambito pedagogico ottiene riconoscimento ufficiale con Buyse che, nella sua introduzione a *L'expérimentation en pédagogie* (1935), riprende il testo dello studioso francese sostituendo ai termini medici quelli pedagogici. Secondo Buyse, medicina e pedagogia condividono un modello epistemologico analogo; entrambe sono tripartite: l'una si fonda su fisiologia, patologia e terapia, l'altra su biologia, psicologia e didattica. In entrambi i casi, il ricercatore deve congiungere teoria e pratica, poiché "la testa e la mano non vanno separate" (Buyse, 1935, p. 11).

Tornando all'asse storico, il primo dopoguerra (1918-1940) presenta un panorama differenziato nello sviluppo della ricerca empirico-sperimentale in educazione (Mialaret, 2016). Lo shock di quattro anni di dura guerra, la rimessa in sesto dei servizi pubblici, la ripresa dell'attività economica e altro ancora fanno sì che i problemi della ricerca scientifica, in particolare nel campo educativo, divengano oggetto di attenzione variabile da parte delle autorità pubbliche e della società. I vari Paesi si muovono con velocità e iniziative differenziate; la situazione, almeno in Europa, appare a macchia di leopardo. Vi è una forte tendenza al rinnovamento in ambito pedagogico, favorita da un lato dai fautori di nuove scuole e dall'altro da quanti auspicano l'assunzione di un metodo sperimentale in campo educativo; in generale vi è l'orientamento a superare gli schemi pedagogici tradizionali (Mialaret e Vial, 1981). Anche l'Italia partecipa a queste dinamiche, seppure in maniera complessivamente più lenta. A titolo esemplificativo si può ricordare il ruolo svolto dalla "Rivista pedagogica" del Credaro che ospita una rubrica in cui sono pubblicati i risultati di alcune ricerche sperimentali condotte all'estero; nel 1912 viene tradotto in Italiano il volume del Claparède *Psicologia del fanciullo e pedagogia sperimentale*; nel 1909 esce il primo importante testo della Montessori *Il metodo della pedagogia scientifica applicato alle case dei bambini*, non esente tuttavia da critiche e rifiuti. Lombardo-Radice promuove la traduzione delle prime opere di Dewey accogliendo il desiderio di innovazione; la rivista "La nostra scuola", che nasce a Firenze nel 1913 con il contributo di maestri autorevoli e con l'appoggio di Prezzolini e di Lombardo-Radice, attesta il desiderio di svecchiare le pratiche didattiche. In generale nel nostro Paese prevale l'impegno a rinnovare la prassi educativa mentre è più debole la spinta a conferire scientificità alla pedagogia. Non mancano tuttavia iniziative che, seppur non possono essere nel loro insieme raccordate in una sorta di scuola di pensiero, segnalano attenzione all'esigenza di dare fondamenti scientifici alla ricerca educativa: dal lavoro di Pizzoli (1909) a quello di Pizzigoni (De Bartolomeis, 1953) sino alla Montessori, si diffondono iniziative che manifestano la conoscenza e l'accoglienza progressiva in Italia della pedagogia dell'attivismo; si guarda invece con diffidenza all'idea della sperimentazione che riceverà maggiore consenso solo con il tramonto dell'idealismo al termine del secondo conflitto mondiale (Bellerate, 1997).

1.3.3 Dalla metà degli anni Trenta agli inizi degli anni Cinquanta

Il terzo periodo, situabile dalla metà degli anni Trenta all'inizio degli anni Cinquanta, è caratterizzato da una fase prima di riflessione poi di stagnazione. La crisi economica che accompagna l'ascesa del fascismo comporta l'interruzione di molte ricerche e genera la contestazione sociale; la polemica e la lotta politica non lasciano sostanzialmente spazio alla ricerca in campo educativo. In questi anni hanno fortuna il movimento dell'*Educazione nuova* in Europa e il "progressivismo" negli Stati Uniti; la spinta ideale ha il sopravvento sulla ricerca, da cui attinge quanto è compatibile con la propria filosofia e può servire a suffragarla. La ricerca quantitativa sussiste ma cessa di progredire in maniera considerevole (Mialaret, 2016). Gli anni della guerra e della resistenza rallentano e quasi fermano la ricerca scientifica tranne nei settori strategici per l'attività bellica e nella riflessione impegnata a preparare la ricostruzione; l'indagine pedagogica segna il passo, soprattutto in Europa. Dal 1945 quest'ultima deve affrontare il compito di ricostruire, riavviare l'economia, riorganizzare le strutture amministrative, tessere la pace sociale. La riforma dell'istruzione entra nei programmi dei diversi Paesi, con l'obiettivo di rinnovare la scuola affinché sia leva per la nuova vita sociale e civile. L'impegno dell'immediato dopoguerra è anzitutto promuovere nuove strutture educative (il Piano Lan-

gevin-Wallon, le *comprehensive schools, Gesamtschule,* la scuola unica ecc.) frutto delle conquiste sociali e scientifiche degli anni precedenti il conflitto. Negli Stati Uniti si sviluppa il movimento dell'*operation research* che spinge la ricerca operativa anche in campo educativo, applicazione alla pratica pedagogica delle strategie di gestione e di modifica dei comportamenti messe a punto durante la guerra.

In questo quadro complessivo ritrova respiro anche un insieme di orientamenti di ricerca ricollegabili alla denominazione generale di "pedagogia sperimentale". È utile soffermarsi brevemente su questa espressione che ha generato più di un fraintendimento. Seppure con le interruzioni imposte dalle due guerre nella prima parte del Novecento, nei vari paesi riprendono iniziative di ricerca che si ispirano al movimento inaugurato da Binet a inizio secolo. Tale dinamica è particolarmente evidente nei paesi di lingua francese.

Spicca il Belgio che annovera una ricchezza di pionieri eccellenti: a Morlanwelz vi è l'azione efficace di Hotyat; a Bruxelles operano Jonckheere e poi Vandevelde; a Liegi Clausse seguito da Bayer, Crahay, de Landsheere; a Gand sono attivi Verbist e van Herrewegue; a Lovanio Buyse fonda il primo *Laboratoire de Pédagogie Expérimentale* attorno al quale si raccoglie un vero e proprio crogiuolo di sperimentalisti tra cui Frankard, Gilles, Bonboir. Nel laboratorio di Angleur è attivo Jadouille; a Forest-Bruxelles opera Goossens; Planchard insegna a Coimbra ma si ricollega alla scuola Belga. In Svizzera emerge Ginevra e, oltre a Piaget, hanno risonanza internazionale gli studi di Roller e Dottrens. In Francia si impongono Gal, Mialaret, Zazzo.

Le ricerche svolte nei vari centri perseguono criteri di scientificità e, sotto l'influsso di Dottrens, condividono l'emblema della pedagogia sperimentale. Tale espressione ha un significato che occorre comprendere collocandola nella sua fase storica: l'11ª edizione del libro di Claude Bernard, *Introduction à l'étude de la médecine expérimentale,* vede la luce nel 1939; la rivoluzione psicologica è in movimento e si parla molto di psicologia sperimentale; nel 1953 Dottrens, con la collaborazione dell'Università di Lione, organizza il primo incontro di una decina di ricercatori che affermano di fare ricerca in pedagogia sperimentale proprio per segnare la specificità scientifica del loro operato. Per cinque anni questo gruppo di ricercatori si incontra annualmente per aggiornarsi a vicenda, aggregando di volta in volta nuovi aderenti; nel 1954 è ufficialmente istituita l'Associazione Internazionale di Pedagogia Sperimentale di Lingua Francese (AIPELF), presieduta da Mialaret.

1.3.4 Dagli inizi degli anni Cinquanta alla metà degli anni Sessanta

Il quarto periodo è collocabile dalla metà degli anni Cinquanta alla metà del decennio successivo. In questi anni si mescolano diverse dinamiche socioeconomiche e politiche. Si fa tesoro dei progressi scientifici e tecnici spinti dalla guerra nonché

delle apparecchiature di produzione industriale sviluppate a fini militari; si è impegnati a ricostruire da zero le infrastrutture economiche; la domanda sociale è pressante dopo le distruzioni, le privazioni e le frustrazioni subite nel periodo bellico; sono immessi sul mercato molti nuovi prodotti che allettano il consumo e il commercio esplode. La contrapposizione Est-Ovest e le tensioni che l'accompagnano rilanciano l'industria dell'armamento e la ricerca scientifica nel campo delle scienze fisiche e umane a fini militari. Il denaro circola, i Paesi capitalisti raggiungono vertici di prosperità, si avvalgono dei contributi della scienza e in questa prospettiva vedono nell'educazione una chiave del progresso. Il lancio del primo Sputnik sovietico provoca negli Stati Uniti appassionati dibattiti sulla qualità dell'insegnamento specialmente in campo scientifico e sulla formazione dei maestri; ne consegue un considerevole afflusso di risorse per la ricerca educativa, provenienti sia da privati sia da autorità locali e dal governo federale. I Paesi anglosassoni in particolare guadagnano supremazia nella ricerca sperimentale applicata al settore dell'istruzione, seguiti da quelli scandinavi. Dagli anni Sessanta si osserva un riavvicinamento della ricerca alla pratica educativa, in particolare nei primi importanti lavori sulla definizione degli obiettivi dell'insegnamento e sulla costruzione dei curricoli.

1.3.5 Dalla metà degli anni Sessanta ai giorni nostri

Il quinto macro-periodo si estende dalla metà degli anni Sessanta ai giorni nostri ed è animato da una tensione di verifica epistemologica e di riconciliazione tra filosofia e scienze dell'educazione. Sul piano sociale raggiunge il suo punto più forte la contestazione originata dall'effetto congiunto di una molteplicità di fattori: il balzo in avanti scientifico appena compiuto ma anche la presa di coscienza dei limiti della conoscenza oggettiva, l'insoddisfazione generata dall'eccessivo consumo di beni materiali, l'analisi più lucida delle ideologie e dei meccanismi del potere economico-politico, la delusione di vedere l'acuirsi anziché l'appianamento delle sperequazioni economiche e sociali.

Proprio le scienze metodologicamente più avanzate sono scosse in tali anni da un profondo riesame epistemologico le cui ripercussioni incidono anche sulle scienze sociali e, negli anni Settanta, investono la ricerca educativa (Adorno e Popper, 1979; UNESCO, 1970). Con la sua *Epistémologie des sciences de l'homme,* Piaget (1972) ricostruisce un quadro generale in cui l'approccio di tipo storico ritrova un ruolo significativo.

Sono però i Paesi anglosassoni e scandinavi più avanzati nel campo della ricerca sperimentale a rimetterla direttamente in discussione. Dopo aver sviluppato ininterrottamente i metodi quantitativi, vedono i loro modelli e le loro metodologie contestate da due colleghi fra i più prestigiosi e autorevoli nel campo della valuta-

zione di programmi: Campbell e Cronbach.

Campbell ha un lungo passato di quantitativista; lo studio clinico di un caso occupa l'estremità inferiore della gerarchia scientifica nella sua nota distinzione tra lo studio del caso unico (*one-shot study*), la ricerca quasi sperimentale e la ricerca sperimentale. Nel 1974, nella sua comunicazione all'Assemblea dell'*American Psychological Association*, rimette tuttavia in discussione l'approccio esclusivamente quantitativo nella valutazione dei risultati di innovazioni sociali deliberatamente introdotte o, più in generale, nella valutazione di programmi. Dichiara così che l'obiettivo è giungere a una prospettiva che unifichi la conoscenza qualitativa e quantitativa, conforme ai fondamenti stessi della filosofia contemporanea della scienza. Campbell sostiene che la conoscenza quantitativa, pur se la oltrepassa, dipende dalla conoscenza qualitativa. E' deprecabile, aggiunge, che le scienze sociali abbiano depauperato il proprio modo di procedere, rifiutando di trovare, nel qualitativo, la controprova di una convalida fornita dal buon senso. Campbell apre perciò ampiamente la pista al concetto di *responsive evaluation*, di cui Stake si farà sostenitore, affermando: "La descrizione di un processo, ben preparata e coscienziosamente eseguita, appare come un componente auspicabile di ogni valutazione di programma; essa funge da controprova convalidante e permette la critica delle procedure di misurazione e dei dispositivi sperimentali" (Campbell, 1974, p. 18). Per la descrizione dei processi Campbell riconosce la qualità raggiunta dal metodo antropologico, al punto che il dibattito quantitativo/qualitativo è ormai da molti denominato nomotetico-antropologico.

Nella medesima Assemblea è relatore anche Cronbach e anch'egli ricolloca l'approccio storico in primo piano; l'esordio del suo intervento è eloquente "La separazione storica tra la psicologia sperimentale e lo studio delle differenze individuali ha creato ostacoli al progresso della ricerca psicologica" (Cronbach, 1975, p. 1). In realtà, già dal 1957 aveva sottolineato che, astraendo dalle differenze fra individui, numerose ricerche psicologiche e pedagogiche approdano a generalizzazioni di scarsa validità Per esempio, i risultati ottenuti con un determinato metodo d'insegnamento sono dovuti non solo a tale metodo ma anche alla sua interazione con le caratteristiche degli alunni (Cronbach, 1957). A distanza di anni asserisce che non basta fermarsi alle interazioni di primo livello ma occorre andare molto più a fondo per poter spiegare un risultato osservato, poiché fattori diversi intervengono agendo gli uni sugli altri e variando nel tempo.

Approfondimento 1.2

L. CRONBACH: DALL'INTERVENTO AL CONGRESSO DELL'AMERICAN PSYCHOLOGICAL ASSOCIATION (1975)

"Quando si considerano le interazioni, si entra in un palazzo degli specchi in cui l'immagine si riflette all'infinito. Anche se spingiamo la nostra analisi fino al terzo, fino al quinto grado, addirittura più oltre, possiamo sempre supporre di aver trascurato interazioni di un grado ancora superiore" (Cronbach, 1975, p. 7).

"Occorre frugare i dati per mettere in luce effetti locali dovuti a condizioni incontrollate o a risposte di livello intermedio" (Cronbach, 1975, p. 19).

"Anziché fare della generalizzazione la regola della nostra ricerca, propongo di capovolgere le nostre priorità. Un osservatore che raccolga dati in una situazione particolare si trova in grado di valutare le procedure praticate o proposte e di osservare gli effetti nel loro contesto. Cercando di descrivere e di spiegare quanto è accaduto, egli presterà attenzione non solo alle variabili controllate, ma anche alle condizioni incontrollate, alle caratteristiche personali e agli eventi che si producono durante il trattamento e la misurazione dei dati. Muovendo da situazione a situazione, il suo primo compito è quello di descrivere di nuovo e di interpretare l'effetto in ogni singolo caso, tenendo conto di fattori che agiscono unicamente in tale singolo caso o in una serie di eventi. Man mano che i risultati si accumulano, colui che cerca di comprendere farà del proprio meglio per scoprire come i fattori controllati potrebbero aver causato deviazioni locali dell'effetto modale. In altri termini, la generalizzazione interviene solo in una fase più tarda, e l'eccezione è considerata non meno seriamente della regola (...). Quando attribuiamo un peso adeguato alle condizioni locali, ogni generalizzazione è un'ipotesi di lavoro, e non una conclusione" (Cronbach, 1975, pp. 19-20).

In sintesi, la ricerca scientifica in campo educativo dovrebbe rinunciare a elaborare ogni teoria durevole; la sua ambizione si limiterà all'osservazione rigorosa e alla descrizione dell'uomo così come esiste, in un dato momento e in una data cultura

Nel processo di ripensamento della ricerca sperimentale in campo educativo un contributo fondamentale è segnato anche da Stake (1976) le cui idee maturano in sincronia con quelle di Campbell e di Cronbach. Come costoro, possiede una ricca esperienza in materia di valutazione quantitativa di programmi e giunge a formalizzare l'approccio qualitativo in un modello coerente (Hamilton e Jenkins et al., 1977). Per Stake, i dati analitici depurati della ricerca nomotetica si addicono alla ricerca scientifica ma non alla valutazione di programmi d'azione educativa. Per questi ultimi esige una rilevazione assai più articolata "che non esiti a far posto all'ambiguità" (Stake, 1976, p. 19). La descrizione è realizzata non solo dal valutatore ma anche, per esempio, da uno dei promotori del progetto, da un destinatario, da un soggetto con atteggiamento avverso ecc. Vanno registrati i momenti critici del programma, gli insuccessi in corso di realizzazione, le possibili opzioni di scel-

ta e le loro rispettive implicazioni. Al termine, gli obiettivi assegnati inizialmente al programma passano in seconda linea poiché il giudizio investe ciò che effettivamente si osserva e come esso si spiega in virtù dell'intera dinamica processuale.

La riflessione critica ingenerata da tale esperienza ha sollecitato un'evoluzione positiva della ricerca in campo educativo la quale ha progressivamente maturato un'identità epistemologica e metodologica più adeguata a cogliere la complessità del suo oggetto. Alcune acquisizioni hanno richiesto tempo ma sono di merito: ampia apertura alla dimensione qualitativa la quale fornisce una percezione più profonda delle realtà educative e permette elaborazioni quantitative più affinate; studio dei processi e dei cambiamenti che intervengono tra l'inizio e la fine dell'esperienza; consapevolezza del ruolo delle interazioni fra il trattamento sperimentale e le caratteristiche dei soggetti, il loro ambiente fisico e umano; riconoscimento dello status di "soggetto" a tutti gli attori coinvolti; superamento dei dispositivi sperimentali chiusi per tener conto dei risultati imprevisti; apertura alla dimensione locale, contingente; distinzione tra validità a breve e a lungo termine dei risultati di ricerca. Le nuove direzioni che così si delineano corrispondono a un progresso euristico di primaria importanza. Occorre tuttavia vigilare sulle possibili derive: la sterilizzazione da perfezionismo, indotta dalla tentazione irrealistica di tener conto di un alto numero di gradi d'interazione; l'alibi semplicistico che giustifica ogni ricerca approssimativa e mal impostata in nome della flessibilità e delle sfumature; l'assolutizzazione della sola ricerca-azione o in generale della ricerca partecipativa come necessariamente migliori rispetto a quelle condotte con altre impostazioni metodologiche.

Rimettere in discussione la ricerca educativa ha contribuito al suo arricchimento. Esempi autorevoli sono presenti già negli anni Ottanta in ricercatori per i quali studio qualitativo e rigore scientifico non si escludono tra loro, impegnati a definire una metodologia integrante le diverse sfaccettature della ricerca. Huberman e Miles (1984) propongono un insieme di tecniche di riduzione e di presentazione dei dati che consentano di ricavare un significato valido dai dati qualitativi. Come asserisce Huberman (1983, p. 502) "Tocca al ricercatore che si avvicina al modello di valutazione più fenomenologico o esistenziale di Stake essere ancora più metodico del ricercatore 'classico'".

La valutazione di programmi

Il ripensamento della ricerca empirico-sperimentale in campo educativo è innescato dalla valutazione di programmi. Questa, vòlta a definire o ridefinire una politica educativa e a preparare le decisioni che la traducono in atto, acquisisce valore di disciplina scientifica solo a partire dagli anni Settanta. Applicandosi alle programmazioni di lungo termine, essa tenta di stabilire in che misura i risultati conseguiti corrispondano alle intenzioni del programma – ossia in riferimento a criteri intrinseci – o possano indurre a riconsiderare tali intenzioni, alla luce di criteri estrinseci. Da un quarantennio circa essa è diventata un importante campo di ricerca, con la sua filosofia e i suoi metodi, ed è destinata a svolgere un ruolo crescente nel mondo dell'educazione. Resta però variabile la misura in cui i *policymakers* tengono conto dei suoi apporti per prendere le loro decisioni, come indicato da un'ormai consistente messe di studi e ricerche (Cfr. Pawson, 2006; Patton, 2010; Stame, 2007).

Avvicinandosi al XXI secolo, la consapevolezza critica della peculiarità metodologica della ricerca empirico-sperimentale in educazione trova ulteriori approfondimenti autorevoli. Il contributo di Van der Maren (2012), da molti anni studioso delle questioni metodologiche della ricerca educativa, incarna un problematicismo storicamente, culturalmente e socialmente documentato e delinea con lucidità gli snodi da percorrere. In accordo con la periodizzazione sopra delineata che si basa principalmente sugli studi storici di Depaepe (2001) e sulla sintesi elaborata da Mialaret (1981, 2016), definisce il processo di ripensamento compiuto dalla ricerca educativa dopo gli anni Sessanta come una critica postmoderna alla pedagogia scientifica. In particolare, si sofferma sull'effetto perverso che ha tradito l'obiettivo originario dei padri della ricerca educativa finendo per sottrarre l'educazione dalle mani degli insegnanti e affidandone il destino alle scienze sociali e alle politiche educative, come una mera loro applicazione. Analoga perversione è data dall'iper-frammentazione specialistica delle discipline scientifiche e relative sotto-discipline il cui obiettivo accademico è produrre conoscenza, senza interrogarsi rispetto ai sistemi di significato e ai modelli sociali e valoriali cui esse sono ricollegabili. La critica postmoderna vede quindi la ricerca passare dal dogma dell'oggettività a un'inclinazione ermeneutica con un orientamento fenomenologico quando si focalizza sul soggetto in quanto tale, e con un orientamento interazionista e simbolico quando si interessa alla cultura in cui si situano le azioni reciproche di insegnanti e studenti (Van der Maren, 2012).

L'autore belga-canadese illustra poi, come critica "post-postmoderna" alla ricerca postmoderna in educazione, un insieme di analisi sviluppate tra la fine del XX e l'inizio del XIX secolo.

Ceci n'est pas une pipe

M. Foucault così commenta il disegno di Magritte in Figura 1.1.: "...ci sono due pipe. Non bisogna dire piuttosto: due disegni di una stessa pipa? Oppure una pipa e il suo disegno, oppure due disegni che rappresentano ciascuno una pipa, oppure due disegni di cui uno rappresenta una pipa ma l'altro no; oppure due disegni che, né l'uno né l'altro, non sono e non rappresentano pipe, oppure un disegno che rappresenta non una pipa ma un altro disegno, il quale invece rappresenta una pipa, sicché sono costretto a chiedermi: a che cosa si riferisce la fase scritta sul quadro? Al disegno sotto il quale si trova immediatamente posta?" (Foucault, 1973, p. 26).

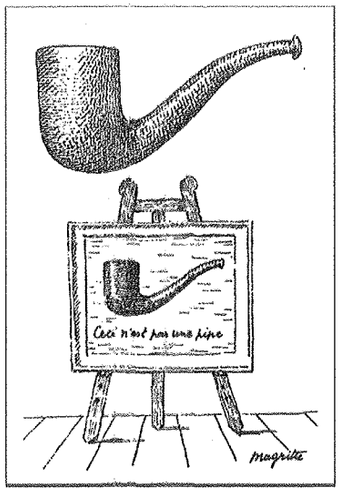


Figura 1.1 R. Magritte, Disegno.

In particolare, attinge alle ricerche sull'analisi del lavoro realizzate dagli studiosi di ergonomia in area francofona (Clot, 1999; Saujat, 2007; Theureau, 2004, 2006) e a studi di psicologia sociale (Blanchet, 2003); da tali lavori emerge una critica della ricerca postmoderna che utilizza principalmente l'intervista come strumento di rilevazione di dati. In generale, si osserva che il significato attribuito dagli attori alla loro pratica non corrisponde a quanto restituisce l'osservazione diretta delle loro azioni. L'analisi del lavoro mostra che il lavoro prescritto (scritto nelle guide, procedurale) non corrisponde a quello assunto (interpretato, interiorizzato, accettato, preso in conto) il quale a sua volta differisce dal lavoro svolto (direttamente osservato) che di per sé non corrisponde al lavoro raccontato (rilevato attraverso interviste e narrazioni di pratiche). In altre parole, capire le dinamiche di un intervento educativo richiede un'analisi dei rapporti e degli scarti tra prescrizione, internalizzazione, esecuzione e narrazione dell'azione svolta.

In risposta a questa critica, alcuni ricercatori hanno cercato di sviluppare metodi che permettano l'analisi incrociata dell'azione e del discorso sull'azione: per

esempio le interviste esplicitanti (Vermersch, 1994) e l'autoconfronto (semplice: un attore; incrociato: due attori) nel quale l'interpretazione dell'azione è costruita dall'attore stesso dinanzi alla materialità dell'azione osservata dopo essere stata videoregistrata (Van der Maren e Yvon, 2009; Yvon e Garon, 2006).

1.4 Guardando avanti

A conclusione dell'itinerario di sviluppo della ricerca empirico-sperimentale in educazione qui tratteggiato sono possibili alcune riflessioni prospettiche. L'obiettivo, forse un po' ambizioso, è tener desti l'inquietudine sistematica e lo sguardo a lungo termine che dovrebbero muovere ogni ricercatore spingendolo a interrogarsi sempre sul senso profondo della sua ricerca.

Una prima riflessione attiene al livellamento critico osservabile nella convinzione di molti ricercatori, per i quali la conoscenza importante si riduce alle pubblicazioni che considerano scientifiche degli ultimi dieci-vent'anni. Come se i protocolli e le attività di raccolta dei dati, talvolta una vera e propria mania del dato, potessero sostituire la cultura: il loro ruolo è invece quello di fornire spiegazioni, dimostrazioni, verifiche e, se possibile, previsioni dei vari fenomeni osservati.

Ciò che appare ancora più drammatico è quando i ricercatori ignorano ciò che è stato fatto prima che nascessero, non solo nel loro campo e nella riflessione teorica che pur adottano come riferimento chiamandola spesso impropriamente "la loro epistemologia", ma anche in altre discipline della conoscenza umana che possono far luce sui problemi di oggi e probabilmente di domani. Il teatro, il cinema, la letteratura spesso aiutano a cogliere le difficoltà della nostra esistenza più di qualsiasi ricerca scientifica; le migliori ricerche non sostituiranno mai una cultura aperta e una visione di lungo termine. La ricerca migliore, a condizione che le sue conclusioni resistano all'usura del tempo e al confronto con la realtà, resterà sempre un contributo alla costruzione di una cultura universale che è al di là di noi.

Approfondimento 1.3
G. BIESTA (2020). EDUCATIONAL RESEARCH. AN UNORTHODOX INTRODUCTION. LONDON: BLOOMSBURY, (P. 149)

I posed "the most difficult question" which was the question why we actually should be doing research at all. This is not the most difficult question in a general sense, but it is a pretty difficult question for researchers, as an honest answer to the question could mean that one would stop doing research or not even embark on it. Posing the question is important, in my view, because we cannot and shouldn't assume that for every problem we encounter, in education or elsewhere, research is the most approriate way for addressing or solving it. Rather than assuming that research start with the formulation of research questions - one of the ortodoxies one may find in introductions to educational research - it makes more sense to start with an identification of the problem in order then to ask whether research is the most appropriate way for addressing the problem. Sometimes the answer to that question will be "yes", in which case it is justified to begin developing an appropriate research design, but sometimes the answer will be "no", in which case researchers shoud withdraw.
While this sounds easy, it may actually be quite difficult to do. One reason for this has to do with the fact that we live in a time where there is (still, one might say) a strong believe in the positive power of research. Research, particularly if it is preceeded by the word "scientific", is often still perceived as a superior kind of knowledge with cognitive and normative power, a kind of knowledge that therefore overrule everyday knowledge.

Consideriamo però che la conoscenza quotidiana può avere aspetti illusori e che il cosiddetto "pensiero ingenuo" spesso prescinde da verifiche rigorose sulla natura e le origini dei fenomeni osservati: ciò rende ragione del ruolo sempre centrale delle metodologie scientifiche.

Una seconda riflessione attiene al modo in cui le politiche che inquadrano e governano la ricerca scientifica condizionano l'impatto di questa nelle professioni della relazione umana. Che i ricercatori in campo educativo lo accettino o meno, il loro lavoro non è mai neutrale. Il ricercatore non è necessariamente responsabile di ciò che scopre (Van der Maren, 1980) ma lo è di ciò che presenta come obiettivo di ricerca, di ciò che pubblica, di come e dove pubblica. Il modesto influsso esercitato fino ad oggi nei contesti di pratica si spiega anche con l'insufficiente comunicazione tra ricercatori e professionisti.

In realtà la questione dello scarso impatto della ricerca educativa va considerata secondo una duplice prospettiva: l'influsso generale esercitato sul sistema o su alcune delle sue componenti viste globalmente e il ricorso consapevole dei singoli educatori ai risultati della ricerca. Quest'ultimo resta assai raro, mentre l'impatto globale esiste ma non è facile documentarlo in modo oggettivo poiché è praticamente impossibile mostrare relazioni di causa-effetto tra ricerca e pratica scolastica. E' anche vero che la storia dell'educazione è scritta soprattutto da studiosi universitari e non da alunni e insegnanti o da educatori che descrivono ciò che accade, come essi lo vivono. A titolo di esempio, Van der Maren (1980) ricorda che nel 1960 erano già state pubblicate non meno di 4000 ricerche importanti sulla sociologia, la psicologia e la pedagogia dell'insegnamento della lettura, ma nel 1966 non vi era ancora alcuno studio sistematico sulla diffusione nella pratica delle idee fornite da queste ricerche.

Le tecniche di metanalisi possono assumere un ruolo importante in questa prospettiva. Nel 1921 Thorndike calcolava da 30 a 50 anni il tempo necessario affinché un risultato di ricerca penetrasse in maniera generalizzata nella pratica scolastica. Anche Mort nel 1964 ha indicato un arco temporale simile. Solo dopo il 1950 sono stati studiati e sviluppati i metodi e le tecniche di diffusione della ricerca e una percentuale dei finanziamenti ha iniziato a essere riservata anche a queste azioni.

Il problema tuttavia non si risolve con questo solo mezzo poiché il vero *vulnus* del sistema sta a monte. Ci si interroghi su quante opportunità reali abbiano gli insegnanti - dalla scuola primaria alla secondaria superiore - per un contatto diretto e una sufficiente frequentazione con la ricerca educativa nella loro formazione iniziale e in quella successiva; al più l'influsso si produce per presentazione indiretta: qualcuno presenta loro i risultati di ricerche. Tra i responsabili di sistemi scolastici molte decisioni continuano purtroppo a essere prese basandosi, nella migliore delle ipotesi, su sintesi di ricerche realizzate, spesso da collaboratori che a loro volta non lavorano su documenti di prima mano ma su presentazioni già semplificate, ristrutturate, sintetizzate da altri. A ben vedere, la posta in gioco va assai oltre il migliore adeguamento e la maggiore efficacia di particolari pratiche didattiche. Riprendendo l'espressione di Nisbet (Dockrell e Hamilton, 1980, p. 10), la ricerca educativa è più un modo di pensare e agire che una scorciatoia per giungere alle risposte; il vero influsso risiede nell'effetto che essa produce sugli atteggiamenti dei docenti. La qual cosa implicherebbe un vero e proprio ribaltamento nei percorsi di formazione degli insegnanti e degli educatori, vòlto a coltivare e consolidare un'identità professionale in cui la ricerca educativa non costituisca un semplice tassello in più di conoscenza che va ad accumularsi agli altri bensì rappresenti un *habitus* da interiorizzare e agire nella pratica professionale quotidiana (Viganò, 2016; 2020a; 2020b).

La terza riflessione introduce un'ipotesi che meriterebbe approfondimento, ossia che esista un "via italiana" alla ricerca educativa. Certamente il nostro Paese non è stato un pioniere nella ricerca empirico-sperimentale in educazione. Come la ricerca storica ha mostrato, tale disciplina ha faticato non poco ad affermarsi e a diffondersi e ancora fatica a essere riconosciuta come aspetto integrante necessario alla formazione di operatori e professionisti in campo educativo. Vi sono stati nuclei

accademici e studiosi eccellenti che dagli anni Cinquanta hanno espresso contributi autorevoli rappresentando veri e propri punti di riferimento, sino a essere considerati i capiscuola di tale ricerca in Italia (basti citare Visalberghi e Calonghi prima, Gattullo poi) ma, appunto, si è trattato molto più di singoli capaci di visione pedagogica e di strategia i quali hanno dato vita a scuole autorevoli che di un'uniforme movimento nazionale di apertura alla ricerca empirico-sperimentale in educazione. Le ragioni di tale dinamica sono state poste in luce e sono date dal convergere di diversi fattori ostacolanti: la forte tradizione filosofico-morale all'origine del pensiero pedagogico, la cautela di una parte del mondo cattolico dinanzi all'incedere della "modernità" di cui la cultura positiva e sperimentale era vista come espressione, l'avversione da parte dell'idealismo a riconoscere la pedagogia come scienza autonoma e l'influsso determinato dello stesso nella costruzione del sistema di istruzione e formazione. A oggi, nella comunità scientifica italiana di area pedagogica, gli studiosi attivi nel settore della ricerca empirico-sperimentale sono la minoranza rispetto a quelli degli altri settori. E' oggettivo altresì che la presenza italiana fra gli autori di contributi di ricerca educativa pubblicati nelle riviste scientifiche internazionali e negli organi decisionali e di *policy-making* resta percentualmente assai ridotta rispetto a quella degli omologhi, per esempio del centro-nord Europa e degli Stati Uniti.

Può essere tuttavia importante domandarsi se tale dinamica sia stata solo penalizzante o abbia generato anche aspetti di pregio; in questa sede ci si limita a ipotizzare come plausibile quest'ultima evenienza, rimandando ovviamente la verifica di tale considerazione generale a un vaglio critico assai più approfondito e analiticamente documentato. Parafrasando con una certa approssimazione quanto sopra menzionato in riferimento all'insieme di ostacoli al diffondersi della metodologia empirico-sperimentale nella nostra ricerca educativa, si può ravvisare una sorta di tessuto culturale profondo e remoto nelle sue origini in cui si intrecciano la cultura classica e quella cattolica. Tale terreno è stato rallentante ma non si può escludere che abbia anche contenuto eccessi di iper-razionalismo, iper-meccanicismo e iper-specialismi come quelli richiamati nelle pagine precedenti riguardo allo scenario internazionale.

Corre alla mente anche l'esempio dei cosiddetti cervelli in fuga: di là dello spreco di risorse umane che ciò rappresenta per l'Italia e della necessità di recuperare il saldo negativo che si trascina da anni senza miglioramenti, tra i motivi di successo dei talenti italiani all'estero ricorre spesso, accanto alla qualità della preparazione professionale specifica, la capacità di visione complessiva ossia una competenza culturale più ampia e profonda che integra e arricchisce l'azione dello specialista. Per usare una metafora automobilistica: non siamo stati e ancora non siamo in grado di dispiegare motori potenti e autostrade per correre veloci e percorrere in fretta grandi distanze; l'andar piano tuttavia, pur se penalizza e rischia di tenerci ai margini della gara, ci ha permesso di guardare meglio attorno, di cogliere aspetti del paesaggio che sfuggono a chi va veloce, di non trascurare sfumature che sono tali solo in apparenza ma che rappresentano la realtà quanto i colori primari.

Non si intende certamente, con questa ipotesi, trasformare i limiti oggettivi della ricerca empirico-sperimentale italiana in un elegante alibi intriso di provincialismo per non impegnarsi verso il miglioramento continuo; piuttosto, facendo tesoro di quanto la genesi e la storia di tale ricerca insegnano, occorre invece coltivare, rafforzare e corroborare a tutti i livelli - dal *policy-making* alla formazione alla ricerca - la visione ampia e il senso critico che sono la linfa di base della ricerca e, più in generale, della tensione per la conoscenza e la verità.

Riepilogo

La genesi e lo sviluppo della ricerca empirico-sperimentale in educazione possono essere compresi solo inquadrandoli alla luce dell'intreccio di fattori culturali, scientifici, socio-economici e politici che ne spiegano le origini e le dinamiche evolutive, a livello internazionale e dei singoli Paesi. Gli studiosi sono concordi nell'individuare le prime manifestazioni esplicite di tale orientamento di ricerca nei vari Paesi tra la fine del XIX secolo e gli inizi del XX; non mancano precursori illustri che hanno fornito intuizioni e ispirazioni rilevanti. E' possibile suddividere l'arco complessivo di sviluppo della ricerca empirico-sperimentale in educazione, dalle origini a oggi, nei cinque seguenti macro-periodi, contraddistinti ciascuno dalla convergenza e prevalenza di aspetti culturali, scientifici, socioeconomici e politici, i quali di volta in volta hanno influito sul pensiero e sulla pratica della ricerca: il *primo periodo*, detto prescientifico, collocabile dalla fine del XVIII secolo al termine del XIX; il *secondo periodo*, dall'ultimo decennio degli XIX secolo alla metà degli anni Trenta, caratterizzato dalla fioritura della ricerca quantitativa; il *terzo periodo*, dalla metà degli anni Trenta all'inizio degli anni Cinquanta, con una fase prima di riflessione e poi di stagnazione dovuta al conflitto mondiale; il *quarto periodo*, dalla metà degli anni Cinquanta alla metà degli anni Sessanta, con la spinta della ricostruzione e della ripresa economica che danno impulso alla ricerca educativa; il *quinto periodo*, dalla metà degli anni Sessanta ai giorni nostri, animato da una tensione di verifica epistemologica e di riconciliazione tra filosofia e scienze dell'educazione. Si impongono oggi alcune riflessioni critico-prospettiche anche alla luce dei principali snodi problematici attuali della ricerca empirico-sperimentale in educazione. Dall'analisi proposta dal capitolo è possibile ricavare, fra altre, alcune acquisizioni essenziali: il rapporto tra ricerca e fattori di contesto culturali, scientifici sociali, economici, politici; il complesso percorso epistemologico e metodologico della ricerca empirico-sperimentale in educazione; lo snodo centrale e problematico del rapporto fra teoria, ricerca e pratica/che educative; il senso critico e la vi-

sione culturale ampia come competenze essenziali anche per la ricerca empirico-sperimentale in educazione.

Riferimenti bibliografici

Adorno T., Popper K. (1979). *De Vienne à Francfort: la querelle allemande des sciences sociales*. Bruxelles: Editions Complexes.

Becchi E. (1969). *Problemi di sperimentalismo educativo*. Roma: Armando.

Bellerate B. (1997). Le origini della pedagogia sperimentale in Italia. In: Nanni C. (ed.), *La ricerca pedagogico-didattica. Studi in onore del Prof. Luigi Calonghi* (pp. 177-189). Roma: LAS.

Bernard C. (1865). *Introduction à l'étude de la médicine expérimentale*. Paris: Collège de France.

Binet A. (1914). *Les idées modernes sur les enfants*. Paris: Flammarion.

Binet A., Henry V. (1898). *La fatigue intellectuelle*. Paris: Schleicher.

Blanchet A. (2003). *Dire et faire dire, l'entretien*. Paris: Armand Colin.

Blum E. (1899). La pédologie: l'idée, le mot, la chose. *Année Psychologique*, 5, 300-310.

Bovet P. (1932). *Vingt ans de vie de l'Institut J.J. Rousseau de 1912 à 1932*. Genève-Paris: Delachaux et Niestlé.

Bréhaut W. (1973). British research in education: some aspects of its development. In: H.-J. Butcher, H.B. Pont (eds.), *Educational Research in Britain* (pp. 1-18). London: University of London Press.

Buyse R. (1928). *La pratique des tests mentaux*. Bruxelles: Lamertin.

Buyse R. (1935). *L'expérimentation en pédagogie*. Bruxelles: Lamertin.

Buyse R., Decroly O. (1929). *Introduction à la pédagogie quantitative*. Louvain: Librairie Félix Alcan.

Campbell D.T. (1974). *Qualitative knowing in action research*. Paper presented at the meeting of the American Psychological Association. New Orleans: Louisiana (Usa).

Clot Y. (1999). *La fonction psychologique du travail*. Paris: Presses Universitaires de France.

Comte A. (1877). *Cours de philosophie positive*. Paris: Librairie J.B. Baillière et Fils.

Crisman O. (1986). *Paidologie. Entwurf zu einer Wissenshaft des Kindes*. Jena: Tesi di dottorato, Facoltà di Filosofia.

Cronbach L.J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671-684.

Cronbach L.J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.

De Bartolomeis F. (1953). *G. Pizzigoni e la Rinnovata*. Firenze: La Nuova Italia.

De Corte E., Geerligs T. et al. (1990). *Les fondements de l'action didactique*. Bruxelles: De Boeck Université.

De Landsheere G. (1986). *La recherche en éducation dans le monde*. Paris: PUF.

Depaepe M. (2001), La recherche expérimentale en éducation de 1890 à 1940. Les processus historiques sous-jacents au développement d'une discipline en Europe de l'Ouest et aux États-Unis. In: R. Hofstetter, B. Schneuwly (eds.), *Le pari des sciences de l'éducation* (pp. 29-56). Bruxelles: De Boeck Université.

Dockrell W.B.,Hamilton D. (Eds.) (1980). *Rethinking Educational Research*. London: Hodder and Stoughton.

Drese O. (1948). *Les origines allemandes de la pédagogie expérimentale*. Louvain: Thèse de doctorat, Université Catholique de Louvain.

Foucault M. (1973). *Ceci n'est pas une pipe*. Montpellier: Fata Morgana.

Hamilton D., Jenkins D. et al. (1977). *Beyond the Numbers Game. A reader in Educational Evaluation*. London: Macmillan.

Herbart J.F. (1997). *Pedagogia generale dedotta dal fine dell'educazione*. Firenze: La Nuova Italia.

Huberman M. (1983). Recipes for busy kitchens: a situational analysis of routine knowledge use in schools. *Knowledge*, 4, 478-510.

Ioteyko J. (ed.) (1912). *Premier Congrés international de pédologie*. Bruxelles: Librairie Misch et Thron.

Kant E. (1859). *La pedagogia*. Firenze: La Nuova Italia.

Landes D.S. (1978). *Prometeo liberato. Trasformazioni tecnologiche e sviluppo industriale nell'Europa dal 1750 ai giorni nostri*. Torino: Einaudi.

Lay A.W. (1909). *Experimentelle didaktik*. Leipzig: Nemnich.

Meumann E. (1913). *Vorlesugen zur Einfurhung in di experimentelle Pädagogik*. Leipzig: Engelman.

Mialaret G. (2016). Les origines et l'évolution des sciences de l'éducation en pays francophones. *Les Sciences de l'éducation pour l'Ère nouvelle*, 49, 53-69.

Mialaret G., Vial J. (eds.) (1981). *Histoire mondiale de l'éducation*, IV. Paris: PUF.

Miles M.B., Huberman A.M. (1984). *Qualitative data analysis: a sourcebook of new methods*. Beverly Hills: Sage Publications.

Montalbetti K. (2002). *La pedagogia sperimentale di Raymond Buyse*. Milano: Vita e Pensiero.

Montealegre A. (1963). La sperimentazione in pedagogia. In: AA.VV. *Questioni di storia della pedagogia* (pp. 861-88). Brescia: La Scuola.

Montessori M. (1913). *Il metodo della pedagogia scientifica applicato alle case dei bambini*. Roma: Loescher.

Mort P.R. (1964). Studies in educational innovation from the Institute of Administrative Research: an overview. In: Miles M.B. (ed.) *Innovation in education* (pp. 78-91). New York: Teachers College Press.

PATTON M.Q. (2010). *Developmental evaluation: applying complexity concepts to enhance innovation and use.* New York: Guilford Press.
PAWSON R. (2006). *Evidence-based policy: a realist perspective.* London: Sage.
PIAGET J. (1972). *Epistémologie des sciences de l'homme.* Paris: Gallimard.
PIZZOLI U. (1909). *Pedagogia scientifica.* Milano: Vallardi.
POLANYI K. (1974). *La grande trasformazione.* Torino: Einaudi.
ROUSSEAU J.J. (1876). *Emile.* Paris: Baudouin.
SAUJAT F. (2007). Enseigner: un travail. In: Dupriez V, Chapelle G (eds.) *Enseigner* (pp. 179-88). Paris: Presses Universitaires de France.
STAKE R. (1976). A theoretical statement of responsive evaluation. *Studies in Educational Evaluation,* 2, 19-22.
STAME N. (ed.) 2007. *Classici della valutazione.* Milano: Franco Angeli.
THEUREAU J. (2004). *Le cours d'action. Méthode élémentaire.* Toulouse: Octarès.
THEUREAU J. (2006). *Le cours d'action. Méthode développée.* Toulouse: Octarès.
THORNDIKE E.L. (1904). *An Introduction to the Theory of mental and social measurement.* New York: Teachers College Press.
THORNDIKE E.L. (1921). *The Teacher's Word Book.* New York: Teachers College Press.
UNESCO (1970). *Tendences principales de la recherche dans les sciences sociales et humaines. I: Sciences sociales.* Parigi-l'Aia: Mouton-Unesco.
VAN DER MAREN J.-M. (1980). De la recherche universitaire en éducation. *Revue des sciences de l'éducation,* 6, 3-19.
VAN DER MAREN J.-M. (2012). Historique des perspectives de recherche: de la vérité aux énoncés hypothétiques. *Recherches Qualitatives,* 12 (Hors Série), 9-28.
VAN DER MAREN J.M., YVON F. (2009). L'analyse du travail: entre parole et action. *Recherches qualitatives,* 7 (Hors Série), 42-63.
VERMERSCH P. (1994). *L'entretien d'explicitation en formation initiale et en formation continue.* Paris: ESF.
VIGANÒ R. (2016). Ricerca educativa fra pratiche e politiche istituzionali. *Giornale Italiano della ricerca educativa,* 16, 71-84.
VIGANÒ R. (2020a). La validità della ricerca educativa tra criteri scientifici, contesti di pratica, responsabilità politica. *Pedagogia Oggi,* 1, 323-34.
VIGANÒ R. (2020b). Parlare di metodi di ricerca pedagogica non è questione di tecniche. *Pedagogia Oggi,* 1, 27-35.
WINCH W.H. (1914). The necessity of experimental pedagogy. *Forum of Education,* 5, 1-15.
WINCH W.H. (1914). *Inductive versus deductive method. An experimental research.* Baltimore: Warwick and York.
YVON F.,GARON R. (2006). Une forme d'analyse du travail pour développer et connaître le travail enseignant: l'autoconfrontation croisée. *Recherches qualitatives,* 26, 51-80.

2

Il paradigma scientifico: approcci e modalità di ricerca

In questo capitolo viene approfondito il modo in cui un ricercatore si avvicina alla sua ricerca a partire dalla considerazione di quali sono i condizionamenti di cui deve tenere conto, quelli che attengono al suo essere parte della cultura e della comunità scientifica del suo tempo, e quelli del contesto storico culturale in cui si trova a operare. Si cercherà di verificare come il ricercatore costruisce le sue ipotesi di ricerca in relazione ai problemi che pensa possano essere risolti attraverso il suo lavoro. Al ricercatore sono necessari atteggiamento scientifico, capacità di pensiero critico e metodo e di essere aperto al nuovo, curioso.
Il metodo rappresenta la dimensione del sapere consolidato e del rigore nel gestire l'indagine. Nel capitolo sono pertanto approfonditi i diversi approcci per fare ricerca educativa e le principali modalità di svolgimento della ricerca.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- I condizionamenti che il ricercatore porta con sé nell'identificare e nel condurre la ricerca.
- L'identificazione del problema e la costruzione delle ipotesi di ricerca.
- Atteggiamento scientifico e metodo scientifico.
- I diversi approcci alla ricerca educativa.
- Le modalità prevalenti di conduzione della ricerca: descrivere, mettere in relazione, manipolare.

Per tutto questo uno prima ancora di mettersi a osservare gli altri dovrebbe sapere bene chi è lui. La conoscenza del prossimo ha questo di speciale: passa necessariamente attraverso la conoscenza di se stesso; ed è proprio questa che manca a Palomar. Non solo conoscenza ci vuole, ma comprensione, accordo con i propri mezzi e fini e pulsioni, il che vuol dire possibilità d'esercitare una padronanza sulle proprie inclinazioni e azioni, che le controlli e diriga ma non le coarti e non le soffochi. Le persone di cui egli ammira la giustezza e naturalezza d'ogni parola e d'ogni gesto sono, prima ancora che in pace con l'universo, in pace con se stessi.

Italo Calvino, *Palomar*

2.1 Il ricercatore e la sua ricerca

Ogni attività di ricerca nasce dalla consapevolezza di una situazione problematica e dall'idea che a questa possa essere trovata una soluzione. La ricerca educativa muove dunque dalla percezione delle difficoltà e delle contraddizioni che incontriamo nel nostro lavoro.

Se non incontrassimo problemi non avremmo ragione di fare ricerca, e ancora, se fossimo convinti che ai problemi incontrati non ci siano soluzioni non avremmo parimenti alcuna ragione di fare ricerca. Merita pertanto soffermarsi a riflettere sulla situazione che dà origine alla ricerca e su come il ricercatore vive e percepisce questa situazione.

La situazione è percepita e situata, si colloca in un contesto storico culturale, e la stessa percezione della situazione come problematica ha origine in un percorso che ne determina le caratteristiche. Essa non è mai solo *qui e ora*, perché

"in questo qui e ora che percepisco come critico di fatto interviene tutta la mia storia, ovvero la cultura in cui sono cresciuto e sono immerso, le mie esperienze personali, le mie relazioni, il mio status e il mio ruolo che determinano la mia capacità di cogliere, ritagliare, sentire la situazione come problema nonché la mia capacità di immaginare e proporre soluzioni alternative a quello che sta accadendo".

Dunque, prima ancora di osservare a quali problemi educativi possiamo cercare soluzioni attraverso la ricerca, dobbiamo soffermarci a riflettere su come il ricercatore e il suo problema si determinino a vicenda, superando l'idea che problema e ricerca esistano separatamente, e che il problema sia "oggettivo" e la ricerca consista in una serie di procedure atte a trovare soluzioni "oggettive" a problemi "oggettivi".

Pertanto, prima di procedere nel nostro lavoro volto ad approfondire le caratteristiche degli strumenti di osservazione e dei "fatti del caso" (Dewey, 1938) che consideriamo osservabili, ci soffermeremo a considerare l'osservatore, la sua storia, e le ragioni per cui ciascuno dei lettori che stanno costruendo i significati di questo testo si trova "qui e ora" a cooperare per interpretarne i significati in funzione dei suoi obiettivi.

2.2 Il ricercatore e la scienza del suo tempo

Ciascuno di noi vive in un contesto storico culturale. Questo contesto sviluppa il nostro modo di leggere la realtà, sviluppa e condiziona la nostra attenzione su alcuni aspetti piuttosto che su altri, i nostri schemi mentali, il nostro modo di percepire le cose. Ci muoveremo ora per analizzare quali sono gli antecedenti che incidono sulla percezione del problema da parte del ricercatore, perché riteniamo importante che di questa dimensione il ricercatore debba essere consapevole per evitare di assumere posizioni dogmatiche scambiando quello che è il suo punto di vista con l'idea di disporre di una visione assoluta della realtà con la quale si trova a interagire.

Per chi si occupa di educazione è abbastanza facile comprendere come la ricerca sia in realtà un percorso di apprendimento in cui il soggetto assimila gli elementi con cui si trova a interagire e al tempo stesso li accomoda sulla base delle risposte che riceve dall'ambiente (Piaget).

Kuhn considera il paradigma come una prospettiva teorica condivisa e riconosciuta dalla comunità di scienziati di una data disciplina, basata sulle precedenti acquisizioni della disciplina stessa che indirizza la ricerca sia nella formulazione di ipotesi entro cui collocare la spiegazione del fenomeno osservato, sia nella scelta delle tecniche di ricerca empirica. Ogni ricercatore si muove all'interno di un discorso che integra assieme teorie, metodi e criteri, che determinano la legittimità sia dei problemi che delle soluzioni proposte. (1962, trad.it. p. 138).

Possiamo cercare di esaminare i condizionamenti che il ricercatore porta con sé nel momento in cui sceglie di intraprendere una attività di ricerca, distinguendo tra condizionamenti interni e condizionamenti esterni. Consideriamo condizionamenti interni l'appartenenza culturale del ricercatore e le impostazioni prevalenti della comunità scientifica a cui appartiene, i suoi valori e le modalità con le quali è uso affrontare i problemi.

L'appartenenza culturale, ossia la dimensione storico-culturale nella quale siamo immersi, determina il rilievo che attribuiamo alle situazioni (per esempio, percepire come critico un certo comportamento o ritenere normale un altro) e può indirizzare dunque la nostra attenzione su un aspetto piuttosto che su un altro della realtà con cui ci troviamo a interagire. Si pensi a quanto possa essere diverso il concetto di "educato" declinato in epoche diverse della nostra storia o in culture diverse. All'interno del suo contesto storico culturale, il nostro ricercatore appartiene a una comunità scientifica e condivide quello che Kuhn ha definito come il *paradigma scientifico*.

Ogni ricercatore poi, pur muovendosi all'interno di un paradigma, ha dei valori propri che in questo caso gli derivano dal modo in cui ha vissuto le esperienze nel suo contesto più ristretto, quello della sua famiglia, della sua comunità locale, del suo percorso personale. Senza addentrarci troppo nell'analisi di questi valori, possiamo però affermare che essi costituiscono degli organizzatori remoti del suo comportamento.

Infine, ogni ricercatore ha appreso a utilizzare certe procedure e certi strumenti per realizzare la sua attività. In qualche misura queste procedure e questi strumenti costituiscono a un tempo il suo punto di forza e i suoi limiti perché lo portano a osservare la realtà cercando di individuare in essa quegli elementi sui quali agire con gli strumenti che gli sono più familiari. Un ricercatore che ha dimestichezza con le tecniche del testing cercherà di trovare soluzioni ai problemi pediante l'utilizzo di test, di tabelle, di dati. Un ricercatore che ha sviluppato una maggiore dimestichezza con tecniche di ricerca di tipo qualitativo tenderà a ritenere più importante avere colloqui con i suoi allievi cercando in ciascuno le specificità, piuttosto che osservarne i comportamenti comuni. Le ipotesi, cioè le proposte di soluzione che il ricercatore cercherà di verificare, saranno indirizzate dalle strategie di soluzione dei problemi che gli risultano più familiari.

Approfondimento 2.1
VALORI DEL RICERCATORE

Ciascuno di noi tende ad avere una visione complessiva del senso della vita che può essere fatta risalire a concezioni di ciò che è bene e di ciò che è male, di ciò che è giusto e di ciò che è sbagliato. Pensare che tutti gli esseri umani siano uguali e abbiano diritto allo stesso benessere, o che non sia lecito uccidere nemmeno in nome dello Stato, o pensare che gli studenti e i lavoratori siano "capitale umano" piuttosto che "persone", è legato a giudizi di valore che hanno ricadute sul nostro modo di comprendere la realtà, vivere e, dunque, fare ricerca.

È utile a questo proposito riportare le considerazioni che Travers propone nel suo Manuale di Psicologia dell'Educazione: «Molti tentativi su base scientifica di pianificare l'educazione esprimono giudizi nascosti riguardo ai valori da realizzare nella vita. [...] Molti dei piani per l'educazione basati sulla psicologia sperimentale ritengono implicitamente che il fine principale dell'educazione sia plasmare il comportamento dei giovani in modo che essi possano inserirsi nel tipo di società industriale che già esiste. Tale idea implica che l'educazione abbia il ruolo statico di mantenere la società così com'è, e che i comportamenti debbano essere plasmati in modo da conformarsi a quello che la società attualmente prescrive. Altri vedono la scuola come luogo che prepara i ragazzi a operare cambiamenti fondamentali nella natura della società» (1973, trad. it., p. 39).

Di fronte al dato "il 10% degli studenti non ha superato l'esame finale di un corso di formazione" consideriamo come questa affermazione possa essere percepita da ricercatori con visioni diverse. Per un ricercatore ispirato da valori del tipo "non uno di meno", il dato si rileva come segnale di un fallimento del modello formativo e/o degli insegnanti, per un ricercatore ispirato da principi meritocratici il dato è indice di una bassa selettività degli esami. La stessa realtà non solo assume significati diversi, ma questi significati guideranno poi la ricerca di dati ulteriori che accompagneranno l'interpretazione del fenomeno. Il primo ricercatore analizzerà i condizionamenti sociali e gli effetti negativi sugli studenti che non hanno superato il corso; il secondo cercherà di confermare i bassi livelli di competenza in uscita dal corso e i loro effetti negativi sul futuro dei ragazzi e sulla società. E con buona pace della avalutatività della scienza, entrambi potranno portare argomenti solidi per avvalorare la propria ipotesi.

Il tempo e il contesto in cui il ricercatore si muove incidono poi sulle sue scelte, anche in modo più diretto. In questo caso parliamo di condizionamenti esterni, cioè delle leggi, dei regolamenti e delle regole formali che determinano ciò che è lecito, o degli indirizzi politici che vengono espressi dalle istituzioni e le risorse disponibili per la ricerca stessa.

Chiunque si accinga a fare ricerca deve tenere conto delle leggi che la comunità internazionale e la comunità nazionale a cui appartiene si sono date in relazione al terreno su cui si svolge l'indagine. In relazione all'organizzazione di ciascun Paese si dovrà fare riferimento a norme nazionali, regionali e locali.

Chi si occupa di ricerca in educazione non può fare a meno di conoscere a fondo le caratteristiche e l'evoluzione del sistema normativo in cui opera per avere chiaro quali relazioni abbiano le soluzioni ai problemi che intende affrontare con le regole della comunità a cui appartiene. Alcune proposte di cambiamento risultano coerenti con le norme vigenti, altre possono essere in contrasto. L'introduzione recente di norme a tutela della privacy comporta, per esempio, non pochi problemi ai ricercatori abituati a raggiungere con relativa difficoltà i dati personali degli allievi e, tuttavia, rappresenta un importante punto di arrivo di un percorso a cui la ricerca stessa ha dato un rilevante contributo.

Chi si occupa della scuola, per esempio, deve conoscere le leggi sulla scuola e l'assetto istituzionale che definisce le competenze sulla scuola, quali sono le competenze dell'amministrazione centrale dello Stato, delle Regioni, degli Enti Locali, e chi sono gli interlocutori per un determinato progetto.

Allo stesso modo, chi si occupa di formazione sul lavoro deve conoscere le norme relative ai rapporti di lavoro, alla formazione continua, alla sicurezza sul lavoro, alle assunzioni, e ancora deve conoscere, ad esempio, i contratti nazionali di riferimento del settore in cui intende operare.

In relazione ad alcuni settori di ricerca, il dibattito sulla opportunità di consentire sperimentazioni ha coinvolto l'intera opinione pubblica. Si pensi, ad esempio, agli esperimenti di manipolazione genetica, per i quali è in discussione il fatto stesso che lo Stato possa intervenire per limitare la ricerca o se invece debba essere la stessa comunità scientifica a porre dei limiti alla propria attività.

Gli indirizzi politici, ovvero gli obiettivi che le singole istituzioni ai diversi livelli si pongono, costituiscono un condizionamento esterno per la ricerca, perché ne influenzano attraverso i finanziamenti gli obiettivi e talvolta anche le stesse modalità.

L'Unione Europea, ad esempio, ha definito degli obiettivi per il funzionamento dei sistemi educativi dei paesi membri e ha indirizzato risorse al raggiungimento di questi obiettivi; di conseguenza l'attenzione dei ricercatori è spinta a spostarsi su tematiche che godono di maggiori finanziamenti, trascurando aspetti di ricerca di base altrettanto rilevanti.

Risorse e tempo sono vincoli fondamentali per l'impostazione di un piano di ricerca. Quale che sia l'obiettivo, si possono fare solo ricerche per le quali siano disponibili mezzi sufficienti.

La ricerca richiede comunque una pianificazione dei mezzi e delle risorse necessarie a raggiungere l'obiettivo desiderato. Il piano economico di una indagine deve pertanto essere disegnato in modo altrettanto serio del piano sperimentale e prevedere il costo di ciascuna delle attività previste. Le risorse ovviamente incidono sulle dimensioni e sull'accuratezza delle indagini, sulle dimensioni del gruppo di ricerca e del campione raggiungibile.

Tra i vincoli normativi vanno inoltre considerate le regole che le stesse comunità scientifiche si danno, sia come codici di comportamento sia come norme da rispettare per la conduzione rigorosa delle ricerche. Ad esempio l'AERA (American Educational Research Association), l'APA (American Psychological Association) e l'NCME (National Council on Measurement in Education) hanno una commissione congiunta che definisce e pubblica indicazioni sugli standard a cui attenersi per il testing educativo e psicologico (Standards for Educational and Psychological testing). Si tratta di indicazioni di grande rilievo e di grande utilità a patto di essere usate con intelligenza e non come protocolli da seguire in modo pedissequo.

I condizionamenti esterni rappresentano dunque elementi di opportunità e di vincolo con i quali il ricercatore deve rapportarsi. Il fatto che esistano non comporta certo un'adesione totale e acritica da parte del ricercatore, il quale, anzi, deve considerarsi costantemente in una posizione dialettica con il contesto in cui opera. Pensare che si possa non essere condizionati da questi indirizzi o non tenerne conto rappresenta invece una posizione ingenua che prelude dei rischi per la qualità del lavoro di ricerca.

2.3 Identificazione del problema e formulazione delle ipotesi di ricerca

2.3.1 Obiettivi della ricerca

Spesso la preoccupazione di chi si accinge a fare ricerca si concentra sul come, sul metodo, sugli strumenti, sulla raccolta delle informazioni e dei dati, mentre l'aspetto più importante da tenere presente è lo *scopo*. Per prima cosa bisogna chiedersi che cosa si intende fare con le informazioni raccolte, ovvero quale uso dei dati sarà fatto una volta raccolti. Questa domanda impatta in modo determinante il problema della validità che, come vedremo più avanti, non riguarda solo gli aspetti operativi dell'attività di ricerca, ma anche i suoi obiettivi.

Nel formulare gli obiettivi, si dovrà perciò cercare di esplicitare cosa proponiamo di fare con i risultati della ricerca, operazione questa strettamente collegata alla formulazione delle ipotesi e alla definizione delle altre fasi della ricerca.

Poiché gli obiettivi definiscono anche la qualità delle interazioni che si andranno a realizzare con il contesto, le istituzioni, coloro che dovranno collaborare con noi, questi finiscono per influenzare in modo significativo il contesto complessivo dell'indagine.

Ad esempio, quando in Italia è stato istituito un sistema di valutazione nazionale che aveva tra i suoi obiettivi la verifica dell'efficacia delle scuole, i media hanno veicolato l'informazione che alla fine della valutazione si sarebbero premiate le scuole più efficienti, e sono iniziati fenomeni di *cheating* che mai erano stati pre-

senti in indagini i cui obiettivi venivano definiti solo come diagnostici

Il tipo di collaborazione richiesto da ricerche di tipo educativo reclama di definire in modo chiaro quali siano i punti di arrivo del nostro lavoro, quali ricadute pensiamo possano avere sugli utenti finali, considerando tra questi non solo la comunità scientifica, ma anche gli insegnanti, gli studenti, le loro famiglie, le istituzioni e i decisori politici.

2.3.2 Ipotesi di ricerca

L'identificazione del problema, la definizione degli obiettivi che ci proponiamo e la formulazione delle ipotesi di ricerca sono tre operazioni che in qualche modo avvengono contestualmente. Tuttavia ciascuna di esse richiede attenzioni specifiche. Dal punto di vista metodologico, la formulazione delle ipotesi è la fase più importante della ricerca, ma anche la più difficile da chiarire e ancor più difficile da codificare.

Abbiamo visto che la formulazione dell'ipotesi segue la percezione di una situazione come problematica e le suggestioni di soluzione che provengono dall'esperienza pregressa e/o dagli stimoli del contesto.

Possiamo definire l'ipotesi di ricerca come "un'asserzione ritenuta valida al fine di vagliarne la verità" (McBurney, 1983). L'ipotesi non deve essere dunque formulata come una domanda, bensì come un'affermazione in modo da poter essere verificata o confutata.

L'ipotesi deve contenere una risposta adeguata al problema posto, e questa risposta deve essere verosimile. Proporre una risposta adeguata non vuol dire che l'ipotesi debba necessariamente esaurire il problema in questione, talvolta può limitarsi a considerarne solo alcuni aspetti: «Una ipotesi spiegherà, generalmente soltanto una parte dei fatti e nella scienza non bisogna temere la lunghezza del cammino» (Fraisse-Piaget, 1963, trad. it. pp. 149-150). Formulare un'ipotesi verosimile vuol dire, invece, che essa dev'essere coerente con il complesso delle conoscenze di cui disponiamo in relazione al problema in questione.

Infine, la nostra ipotesi deve essere verificabile, cioè deve essere costruita in modo tale da consentirne un controllo sul piano logico o sul piano empirico.

Infatti, mentre a una domanda possono essere proposte molteplici risposte, una affermazione dichiarativa può essere definita alla fine del processo di ricerca solo "vera" o solo "falsa". Più specificatamente, mentre sul dichiarare una proposizione alla fine di un lavoro di ricerca "vera", in realtà non possiamo mai raggiungere una certezza assoluta, paradossalmente non abbiamo lo stesso limite quando verifichiamo che questa risulta "falsa".

Perché l'ipotesi sia verificabile è necessario che i termini, i concetti con cui viene enunciata, siano definiti utilizzando un linguaggio univoco. Questo richiede

un'attenzione particolare poiché siamo abituati a utilizzare, in particolare nelle nostre discipline, concetti del linguaggio naturale che per loro natura sono polisemici, si prestano cioè a una molteplicità di interpretazioni.

È bene tenere conto della distinzione tra linguaggio comune e linguaggio tecnico o speciale. Il linguaggio tecnico o settoriale, che riguarda specificamente i singoli settori di ricerca, si basa il più possibile sull'uso di termini univoci, ossia che hanno un solo significato. Tale univocità dei termini è stabilita dagli stessi specialisti che danno le definizioni e stanno attenti a non discostarsene (si pensi, ad esempio, ai termini raffreddore/rinite) (De Mauro, 1980).

L'ipotesi contenuta nell'affermazione "Le ragazze di 13 anni sono più mature dei ragazzi dello stesso livello di età" è un esempio di questo tipo di errore, essendo il concetto di maturità nel linguaggio comune talmente ricco di contenuti da non significare quasi nulla. Non è chiaro se si tratta di crescita fisica, di sviluppo psicologico, di equilibrio nei comportamenti, la confusione del concetto rischia di confondere e vanificare qualsiasi tentativo di verifica. Potrebbe avere natura di ipotesi l'affermazione: "Le ragazze maturano sessualmente (hanno il menarca) prima che i ragazzi dello stesso livello di età abbiano le prime polluzioni". Si tratta in questo caso di un'ipotesi ampiamente esaminata in letteratura.

Il problema della verificabilità rimanda, dunque, alla natura dei concetti che utilizziamo nelle nostre affermazioni. Che vuol dire "maturità"? Il fisico Percy Bridgman ci chiederebbe di procedere a una definizione operativa di questo concetto, e cioè di legare questo concetto a operazioni osservabili, a prestazioni che tutti possono vedere o eseguire. Secondo questa impostazione se un concetto non è traducibile nei termini che abbiamo detto, non è un concetto scientifico (Bridgman, 1927). Sulla base di questa impostazione spesso i concetti scientifici vengono definiti in relazione alle operazioni con cui essi sono misurati. Ciò, in alcuni casi, è strettamente necessario. Infatti, se dovessimo svolgere un lavoro sugli scolari "studiosi" avremmo la necessità di definire, per esempio, la soglia di ore di studio oltre la quale si ha il diritto di fregiarsi dell'aggettivo "studiosi".

La difficoltà nel definire in modo operativo alcuni concetti ha portato spesso a legare i concetti agli strumenti utilizzati per la loro misurazione. Se è formalmente corretto dire che la capacità di saper scrivere è la capacità di ottenere buoni punteggi nelle nostre prove di scrittura, o che l'intelligenza è quella capacità che è misurata dai test di intelligenza, questo non ci esime dal giustificare teoricamente le scelte sulla base delle quali sono state costruite le prove in questione, o dal giustificare i motivi della scelta di uno strumento piuttosto che di un altro in relazione ai nostri obiettivi, poiché la validità non può mai risiedere nello strumento, ma è un elemento che caratterizza l'intero processo di ricerca.

La definizione dei concetti deve fare i conti con i modelli teorici di cui dispo-

niamo e con la possibilità di declinarli sul piano operativo.

La *verifica delle ipotesi* si dice *diretta* quando i termini del problema sono tutti osservabili. In alcuni casi, tuttavia, l'ipotesi è più complessa e richiede la considerazione di variabili che non è possibile osservare direttamente, ma dalle quali possiamo ricavare il comportamento sulla base del loro effetto su altre variabili del nostro esperimento. In questi casi abbiamo dunque una *verifica indiretta*.

Fraisse, per rilevare la complessità dei fenomeni che sono oggetto delle indagini sperimentali nelle scienze sociali, sottolinea il fatto che "la verifica è praticamente sempre parziale" (Fraisse-Piaget, 1963). I criteri che si utilizzano per la verifica delle ipotesi sono prevalentemente di tipo probabilistico.

L'ipotesi di una ricerca è formulata sotto forma di un'affermazione espressa attraverso un costrutto tipico: «se in particolari condizioni, si fanno determinate affermazioni, e la previsione è esatta, allora si dovranno ottenere i seguenti risultati». Può essere utile a tal proposito distinguere tra ipotesi che scaturiscono dall'osservazione empirica di situazioni educative che chiameremo *ipotesi indotte*, o ipotesi che nascono dallo studio e dalla riflessione sui fenomeni educativi che chiameremo *ipotesi dedotte*.

2.4 Atteggiamento e metodo scientifico

"Atteggiamento" e "metodo scientifico" concorrono assieme alla costruzione della conoscenza. Per esaminarli tuttavia merita fermarsi a riflettere sul significato che attribuiamo al termine "scienza". Dewey dedica a questo tema il saggio "L'unità della scienza come problema sociale", riportato in apertura del primo volume dell'International Encyclopedia of Unified Science (curata da Otto Neurath, Rudolph Carnap e Charles Morris). Si intende per *scienza* l'insieme delle conoscenze che l'umanità ha costruito nel corso della storia e assieme si intende per scienza il modo in cui queste conoscenze sono state acquisite e consolidate. Ed è evidente come atteggiamento e metodo precedano le conoscenze consolidate che si trovano riportate nei libri e nei trattati scientifici. Se atteggiamento e metodo si formano nella soluzione dei problemi quotidiani, nell'esperienza concreta in cui l'individuo interagisce con il contesto materiale e definisce scopi e piani di azione, ne consegue che queste dimensioni sono comuni all'esperienza di tutti gli esseri umani e non solo degli scienziati. "L'insieme di conoscenze e di idee che è il prodotto del lavoro, è il risultato di un metodo che è stato seguito da un numero molto più grande di persone, che hanno interagito in modo intelligente e con apertura mentale con gli oggetti e gli eventi dell'ambiente comune" (Dewey, 1938).

Dewey, tuttavia, tende a utilizzare "i termini "atteggiamento scientifico" e "metodo scientifico" come se coincidessero. È invece abbastanza evidente come questi due termini siano tra loro in un rapporto dialettico.

2.4.1 Metodo scientifico

Metodo è un termine che deriva dal greco, composto dal prefisso μετά che indica il seguire qualcosa e dalla parola ὁδός che significa strada, percorso. E dovendo scegliere un percorso evidentemente non possiamo che convenire con la definizione che propone Cartesio «Per metodo [...] intendo delle regole certe e facili, osservando le quali esattamente, nessuno dirà mai per vero ciò che sia falso, e senza consumare inutilmente alcuno sforzo della mente, ma gradatamente aumentando sempre il sapere, perverrà alla vera cognizione di tutte quelle cose di cui sarà capace».[1] In questa prospettiva il percorso più consolidato può essere considerato il metodo scientifico, quel processo codificato da Galilei che possiamo sintetizzare come "integrazione tra una dimensione teorica che elabora teorie e una componente sperimentale che ne verifica la consistenza". Queste due componenti costituiscono un continuo senza che si possa definire un ordine necessario: la teoria può nascere da osservazioni empiriche e può anticipare l'esito di future osservazioni.

Il metodo finisce, dunque, per costituire la componente regolativa della ricerca scientifica, ne garantisce la certezza, ma nel contempo contiene tutti i limiti di conservatorismo che affliggono ogni dimensione normativa. Se il metodo è la strada sicura, il rischio è che confini il percorso alla strada sperimentata e resista alla ricerca di nuovi itinerari "Chi lascia la strada vecchia per la nuova sa quel che lascia e non sa quel che trova" dice un vecchio proverbio, ma è altrettanto evidente che se si vogliono trovare cose nuove si devono percorrere anche strade diverse.

2.4.2 Atteggiamento scientifico

Consideriamo ora con attenzione la definizione di *atteggiamento scientifico* che Dewey propone: «In breve, l'atteggiamento scientifico, com'è qui concepito, è una qualità che si manifesta in ogni passo della vita. E allora che cos'è? Se lo definiamo per negazione, è libertà dalla schiavitù, dall'abitudine, dal pregiudizio, dal dogma, dalla tradizione accettata in modo acritico, dal puro egoismo. In termini positivi è il desiderio di ricercare, esaminare, discriminare, tracciare conclusioni solo sulla base dell'evidenza, dopo essersi presi la pena di raccogliere tutti i dati possibili. È l'intenzione di raggiungere credenze, e di provare quelle che risultano accettabili, sulla base dei fatti osservati, riconoscendo al tempo stesso che i fatti sono privi di senso a meno che non indichino idee. È, d'altra parte, l'atteggiamento sperimenta-

[1] Descartes R. (1628), *Regulae ad directionem ingenii*, in *Opuscola postuma, physica et matematica*, Blaeu, Amsterdam, 1701 (traduzione italiana in *Opere*, Laterza, Bari, 1967, vol. I, regola IV, p. 26).

le che riconosce come, mentre le idee sono necessarie per l'organizzazione dei fatti, esse sono al tempo stesso ipotesi di lavoro da verificare sulla base delle conseguenze che producono.».

Se l'atteggiamento scientifico corrisponde al desiderio del ricercatore di cercare una sintonia, una comprensione profonda della realtà e di considerare l'esperienza come uno stimolo a cercare nuove soluzioni, è proprio questo atteggiamento che spinge a esplorare strade nuove e a mettere in discussione anche le metodologie più consolidate. L'atteggiamento scientifico pone il ricercatore in una posizione critica rispetto al metodo e lo spinge a mettere in discussione le regole di conformità, anche quelle della stessa comunità scientifica a cui appartiene.

Approfondimento 2.2

I RISCHI DELL'ASSUNZIONE DEL METODO IN MODO DOGMATICO E L'EDUCAZIONE

Le regole sono connaturate a qualsiasi forma di interazione con la realtà. Anche l'attività più semplice comporta, se si vuole ottenere il risultato atteso, il rispetto di certe regole, e l'esperienza è il luogo in cui fin da bambini impariamo ad ascoltare le risposte del mondo reale e a rispettarne le indicazioni. Col tempo queste indicazioni apprese rischiano di diventare da parte degli adulti regole imposte ai giovani senza dar loro il tempo di farne esperienza. In modo analogo ciò avviene nella costruzione e nella difesa del metodo scientifico.

Supponiamo che un certo numero di studiosi si sia per molti anni impegnato a elaborare una teoria sull'educazione e che abbia accumulato una notevole quantità di evidenze in grado di sostenerla e, ancora, che negli anni questa teoria sia stata utile a spiegare un alto numero di esperienze educative e dunque sia sostenuta dalla raccolta di numerose osservazioni empiriche. Questi studiosi matureranno un atteggiamento ostile ai tentativi di metterla in discussione, e questo atteggiamento li porterà a sottostimare gli indizi che mettono in discussione l'incoerenza della teoria, o a dare loro una spiegazione che li riporti nell'alveo della teoria stessa.

Quando una teoria diventa egemone, coloro che la sostengono ricevono finanziamenti, mentre coloro che cercano di metterla in discussione non ricevono nulla.

Quando una teoria si impone la sua assunzione acritica assume i tratti di ideologia. Questo comporta che il sistema mediatico tende a confermarla e a deleggittimare coloro che la mettono in discussione.

I giovani per loro natura sono portati a mettere in discussione i modelli consolidati delle generazioni precedenti, e il tentativo di "addomesticarli" rischia di distruggere la loro grandissima capacità di immaginazione, la loro spontaneità e anche la loro capacità di vedere che "il re è nudo". In questo senso la ricerca educativa più di ogni altro tipo di ricerca deve essere attenta a non assumere tratti di rigidità.

In questa prospettiva il conflitto generazionale rappresenta una delle difese contro l'immobilismo. I giovani hanno meno esperienza e vogliono fare le loro esperienze, gli adulti hanno vissuto, hanno combattuto, hanno vinto o si sono arresi. Quelli che hanno vinto difendono i loro nuovi confini, quelli che si sono arresi sono stanchi di rischiare. I giovani, se non educati alla servitù volontaria, vogliono invece tentare strade nuove. Molti pagheranno a caro prezzo questi tentativi.

Volare alto è rischioso. Ma se oggi voliamo dobbiamo qualcosa anche a Icaro.

In un tempo quale il nostro in cui sembrano dominare i protocolli e i modelli, l'atteggiamento scientifico chiede al ricercatore una continua messa in discussione del metodo e al tempo stesso un grande rigore nell'applicarne i criteri una volta che si è deciso di utilizzarli. Attenendosi alla norma si ha il vantaggio di avere esperienze i cui esiti sono già stati verificati da altri, ma è evidente che camminando su strade già battute è difficile fare scoperte o nuovi incontri. La crescita della conoscenza non può che venire da nuovi tentativi, dall'assunzione di rischi e di incertezze, dunque solo in condizioni di massima libertà: "Più specificamente si può dimostrare quanto segue: data una norma qualsiasi, per quanto "fondamentale" o "necessaria" essa sia per la scienza, ci sono sempre circostanze nelle quali è opportuno non solo ignorare la norma, ma adottare il suo opposto."(Feyerabend, 2010, p. 21).

Del resto assumere il metodo e le regole che ci siamo dati nel tempo in modo acritico contiene numerosi rischi, e la stessa "demarcazione" che Popper propone tra discorso scientifico e altre forme di conoscenza non è così facile da applicare quando ci si confronta con una dimensione complessa come quella dei fenomeni legati all'educazione, in cui la condizione della ripetibilità dell'esperienza è sostanzialmente inapplicabile e in cui dunque - se è possibile indicare percorsi che hanno maggiore possibilità di essere efficaci nella soluzione di problemi - è impossibile proporre ricette o soluzioni certe ed è consigliabile avere sempre uno sguardo alle possibili strade secondarie.

Nell'uso comune, poi, ciascuna scuola o corrente di pensiero utilizza il termine "metodo" per descrivere il proprio modo di affrontare la ricerca. Questo uso rischia di mettere in discussione la componente del rigore che invece è necessaria se vogliamo distinguere i processi conoscitivi che muovono nell'alveo del discorso scientifico, da quelli che si fondano su basi diverse.

In questo capitolo abbiamo scelto di usare questo termine con parsimonia, limitandone l'uso solo a quel percorso che definisce l'impianto complessivo dell'attività di ricerca e ne è costitutivo della sua scientificità. Può essere utile, invece, distinguere tra approcci e modalità diversi per procedere all'impostazione di una ricerca e alla verifica delle ipotesi.

Per ragioni di chiarezza espositiva, presentiamo i principali percorsi di ricerca in ambito educativo come approcci , seguiti da una ulteriore classificazione basata sulla modalità prevalente utilizzata per esaminare e mettere in relazione le informazioni.

2.5 Gli approcci di ricerca

Possiamo fare ricerca educativa facendo riferimento a una molteplicità di approcci, cioè a prospettive dalle quali osservare le esperienze educative.

Approfondimento 2.3
LE ESPERIENZE EDUCATIVE E LA LORO NATURA DI PRAXIS

Francis Imbert è, senza dubbio, lo studioso che meglio ha formalizzato la contrapposizione tra *praxis* e *poiesis* nell'educazione. Riprendendo questa tematica da Aristotele e chiarendola attraverso i lavori di Hannah Arendt (1993) e di Comélius Castoriadis (1975), ha mostrato bene come qualunque impresa educativa sia profondamente influenzata da questa opposizione (Imbert, 1985, 1987, 1992).

La *poiesis* si caratterizza per il fatto di essere una fabbricazione che termina una volta raggiunto il suo scopo. L'oggetto che si pone come fine rende necessaria la messa in opera di mezzi tecnici, sapere e saper fare, di capacità e competenze che producono un risultato, oggettivabile e definitivo, che si stacca dal suo autore e che non riguarda più quest'ultimo. La poiesis è, per essere esatti, un'attività; nel senso aristotelico, non è un "atto".

La *praxis*, al contrario, si caratterizza per il fatto di essere un'azione senza altro fine che se stessa: qui non ci sono più oggetti da fabbricare, oggetti per i quali si disporrebbe in anticipo di una rappresentazione che ne permetterebbe la produzione e che li racchiuderebbe, in qualche modo, nel "risultato", ma un atto da compiere nella sua continuità, un atto mai veramente concluso, perché non comporta alcun fine esteriore a se stesso e stabilito a priori.

Per questo motivo, l'educazione non può essere mai davvero poiesis, anche se comporta inevitabilmente aspetti di "costruzione" che rinviano a un'immagine di conformità sociale stabilita in anticipo. Ridurre l'educazione a una poiesis significherebbe trattare il soggetto educato come una "cosa" di cui si potrebbe dire, prima di intraprendere il processo educativo, cosa debba essere e in base a cosa esattamente si potrà verificare che corrisponda proprio al nostro progetto. Significherebbe, in realtà, negare l'educazione asserragliandosi nella contraddizione che abbiamo visto all'opera in numerose occasioni: per essere "riuscito", l'educato deve assomigliare all'educatore, ma questa somiglianza implica che, come lui, deve diporre di una libertà che gli consenta proprio di essere diverso da quello che si ha in progetto per lui.

Philippe Mierieu, *Frankenstein Educatore*

Ciascun approccio presenta vantaggi e svantaggi e spesso è necessario integrare più approcci per meglio comprendere le esperienze educative. È tuttavia utile una descrizione dei principali approcci presenti nella ricerca educativa. I più praticati possono essere classificati in sette grandi filoni:

a. approccio teorico;
b. approccio storico;

c. approccio clinico;
d. approccio comparativo;
e. approccio di ricerca-azione;
f. approccio sperimentale;
g. approccio grounded;

2.5.1 Approccio teorico

Per approccio teorico intendiamo quella ricerca che tende ad approfondire con strumenti conoscitivi di tipo logico ed epistemologico gli apparati concettuali e i costrutti teorici che stanno alla base della ricerca educativa. Abbiamo visto nella definizione stessa di metodo che non esiste ricerca scientifica che non si confronti costantemente con il problema di una forte dimensione teorica. Anche in ambiti scientifici più consolidati, come per esempio la fisica, esiste una dimensione di ricerca teorica che anticipa, assiste e accompagna la ricerca sul campo e anzi, come afferma Popper, ne rappresenta in qualche modo il criterio di verità. «L'osservazione e l'esperimento non possono stabilire nulla in modo definitivo, perché c'è sempre la possibilità di un errore sistematico, dovuto all'interpretazione sistematicamente errata di qualche fatto. Ma l'osservazione e l'esperimento costituiscono certamente una parte importante della discussione critica delle teorie scientifiche. Essenzialmente ci aiutano a eliminare le teorie più deboli, e così offrono un sostegno, anche se solo temporaneo, alla teoria che sopravvive, cioè, alla teoria che è stata severamente controllata, ma non confutata» (Popper, 1969, p. 133).

Nelle scienze dell'educazione, la ricerca teorica deve comunque fare i conti con il distacco relativamente recente della pedagogia dalla filosofia.

Ricerca teorica non può voler dire ricerca estranea alla dimensione empirica dell'esperienza. Il lavoro di controllo critico della validità delle teorie, di confronto tra teorie, di elaborazione logica di modelli, di costruzione di ipotesi sul piano concettuale, nasce anch'esso dall'esperienza diretta o indiretta del ricercatore e deve tornare all'esperienza per trovare una conferma.

Anche quando si tratta di valutare, prima ancora di mettere alla prova, una teoria, accanto agli aspetti di coerenza formale non possiamo che considerare «come preferibile la teoria che ci dice di più, ossia la teoria che contiene la maggiore quantità di informazioni empiriche, o che ha il maggior contenuto empirico, che è logicamente più forte, che ha il maggior potere esplicativo e predittivo, e che perciò può essere sottoposta a controlli più severi, confrontando i fatti previsti con le osservazioni. In breve, preferiamo una teoria interessante, audace e altamente informativa, a una teoria banale» (ibidem, p. 164).

Diverso è invece pretendere che la discussione teorica debba necessariamente produrre qui e ora un esito empirico. Come afferma Popper, «dovremmo mettere

bene in chiaro, qui, che nella storia della scienza ci sono molti esempi di teorie che in qualche stadio del loro sviluppo non erano controllabili, ma che diventarono controllabili in uno stadio successivo» (ibidem, p. 131)

Se la ricerca teorica è ben impostata, non importa molto che si chiami filosofia dell'educazione, pedagogia generale, o filosofia tout court.

2.5.2 Approccio storico

Si tratta di un approccio che ha dato vita a un settore fortemente sviluppato degli studi pedagogici che ha dedicato i suoi sforzi all'esame delle dottrine pedagogiche, delle pratiche educative, della scuola e delle altre istituzioni formative.

L'impianto metodologico è quello dell'indagine storica rigorosa legata all'esame delle fonti e ai diversi paradigmi storiografici che regolano la ricostruzione, la narrazione e l'interpretazione delle modalità in cui le esperienze educative sono state condotte nel corso della storia.

Qualsiasi ricerca educativa comporta una collocazione delle esperienze con cui si confronta in un contesto che richiede indagine storica per essere compreso.

2.5.3 Approccio comparativo

Gli studi comparativi hanno come obiettivo il confronto sistematico tra esperienze educative e sistemi formativi di differenti contesti regionali o nazionali. L'educazione comparata può essere, infatti, definita come «la scienza che scopre, analizza e spiega le somiglianza e le differenze fra i fatti educativi nelle loro relazioni con il contesto politico, economico, sociale e culturale» (Lệ Thận Khôi, in Mialaret 1985, traduzione italiana, p. 77).[2]

Si tratta di lavori estremamente complessi perché richiedono di realizzare tipologie e classificazioni riferendo a contesti normativi, istituzionali e culturali sostanzialmente diversi. Le indagini comparative tendono a costruire indicatori sui quali basare il confronto,[3] e per questo richiedono studi di sfondo a carattere comparativo senza i quali non sarebbe possibile procedere in modo valido ad alcuna rilevazione empirica. Come ha scritto Benedetto Vertecchi, presentando le Ricerche Valutative Internazionali del 2000 condotte dall'Istituto Nazionale per la Valutazione del Sistema dell'Istruzione (Vertecchi, 2001c, p. 11): «Esaminare i fenomeni da un punto di vista comparativo consente, infatti, quel tanto di decentramento senza il quale le analisi e le interpretazioni che ne derivano rischiano di essere condizionate dal senso comune prevalente entro un determinato quadro culturale. In altre parole, ampliare il quadro dei riferimenti è necessario se si vuole sfuggire a un limite sempre in agguato quando si ha a che fare con l'educazione, quello della naturalizzazione degli intenti perseguiti, nonché delle soluzioni o degli stili d'intervento».

[2] Per approfondire il tema dell'oggetto dell'educazione comparata si veda l'intero capitolo di Lệ Thận Khôi riportato in Mialaret, *Introduzione alle scienze dell'educazione* (1985). Si vedano anche Mitter (1997), Todeschini-Ziglio (1995), e *The Encyclopedia of Comparative Education and National System of Education* (Pergamon, Oxford, 1988) diretta da Neville Postlethwaite.

[3] Gli indicatori educativi sono indici statistici che danno informazioni sullo stato del sistema educativo e possono essere costituiti sia da indici singoli (numero medio di alunni per classe), sia da indici composti che misurano combinazioni di eventi. Per un approfondimento su questo argomento segnaliamo, tra le tante pubblicazioni, oltre al rapporto annuale OCSE "Education at a glance", il programma PISA e le pubblicazioni CERI-OCDE, tradotti in lingua italiana dalla casa editrice Armando, o i rapporti dell'Unione Europea (per citarne uno, il rapporto Eurydice "Key Data on Education" che dal 1994 offre informazioni quantitative e qualitative sui diversi livelli di istruzione).

2.5.4 Approccio clinico

L'approccio clinico nella ricerca educativa nasce dall'esigenza studiare i processi formativi più in profondità. Il termine clinico (dal greco *klinein* e *klinè*, che significano rispettivamente "chinarsi" e "letto") rinvia alla pratica del medico che si china sul letto dell'ammalato per poterlo osservare da vicino per coglierne gli aspetti non immediatamente visibili. L'approccio clinico dunque cerca di ampliare per il possibile la dimensione empirica della ricerca e il contatto diretto con l'esperienza che si intende esaminare.

In campo educativo è a Jean Piaget che si può far risalire l'approccio clinico. Lo studioso ginevrino nell'introduzione a "La rappresentazione del mondo nel fanciullo" (1926), presentando i vantaggi di questo metodo rispetto al metodo dei reattivi e dell'osservazione pura, scrive:

«[...] il clinico può (1) parlare con il malato seguendolo anche nelle risposte, così da non perdere nessuna eventuale idea delirante; (2) condurlo dolcemente verso le zone critiche [...] senza sapere dove affiorerà l'idea delirante, ma mantenendo costantemente la conversazione su un terreno fecondo. L'esame clinico partecipa così dell'esperimento, nel senso che il clinico si pone problemi, formula ipotesi, varia le condizioni, e infine controlla ogni ipotesi in base alle reazioni provocate dalla conversazione. Ma l'esame clinico partecipa anche dell'osservazione diretta, nel senso che il buon clinico, pur dirigendo, si lascia dirigere, e tiene conto di tutto il contesto mentale, invece di cader vittima di "errori sistematici" come spesso accade allo sperimentatore puro».[4]

L'approccio clinico è segnato dunque dall'attenzione a comprendere e spiegare la singolarità e la materialità presenti nell'esperienza educativa. Questa impostazio-

[4] Piaget illustra il metodo clinico anche nei seguenti scritti: *Il linguaggio e il pensiero del fanciullo*, Ginevra, 1923 (Boringhieri, Torino, 1966) e nella *Prefazione* a *La genesi del numero nel bambino* scritto in collaborazione con Szeminiska, Neuchatel-Paris nel 1941 (traduzione italiana pubblicata da La Nuova Italia, Firenze, 1968).

ne teorizzata in Italia da Riccardo Massa che ne ha sviluppato attenta disamina epistemologica ha dato luogo anche a modalità specifiche di intervento educativo di aiuto.

2.5.5 La ricerca-azione

La "ricerca-azione" è un tipo di indagine orientata all'azione e al cambiamento, spesso condotta da ricercatori che al tempo stesso stanno intervenendo per cambiare la situazione che stanno studiando per apportare miglioramenti. La ricerca-azione infrange dunque l'idea della terzietà del ricercatore rispetto alle esperienze sulle quali sta svolgendo la sua indagine. La fusione dei due termini - ricerca e azione - ne indica la loro indissolubilità e complementarietà. Il ricercatore non osserva, ma partecipa all'esperienza e al tempo stesso la studia giocando insieme i ruoli di osservatore e di osservato. «Una pratica con scarsa fondazione teorica, quindi, anche se dotata di forte carica ideologica e politica; un territorio di azione conquistato con fatica e ancora esposto a incursioni e occupazioni», così scriveva Egle Becchi nel 1992 a commento della sua raccolta di definizioni di ricerca-azione.

Il fatto che il ricercatore sia coinvolto e condivida pienamente gli obiettivi di cambiamento dell'esperienza educativa alla quale partecipa, spesso anche con un ruolo di indirizzo, fa sì che spesso ciò che è necessario rafforzare siano le cautele metodologiche che consentono poi di assumere il contributo delle ricerche svolte con questa modalità e di generalizzarne le indicazioni.

Le ricerche-azione seguono percorsi fortemente individualizzati, utilizzano tecniche differenti e hanno in comune quasi sempre la partecipazione collaborativa, l'orientamento all'azione e finalità di miglioramento o di modifica di condizioni educative o sociali. Caratteristiche queste che finiscono per corrispondere agli interventi di insegnanti all'interno della scuola, o di educatori che si confrontano con situazioni di marginalità, di povertà educativa e di rischio. Per queste caratteristiche la ricerca-azione sembra sfuggire a canoni tradizionali e pertanto al rigore metodologico che ne legittimi gli esiti. Non segue parametri ritenuti indispensabili nella ricerca e viene, per tale motivo, guardata con un certo scetticismo da parte del mondo accademico.

Per Kemmis (1985) «la ricerca-azione è una forma di indagine di auto-riflessione, intrapresa da coloro che partecipano in situazioni sociali ed educative, allo scopo di migliorare la razionalità e la fondatezza (a) delle loro pratiche sociali o educative, (b) della loro comprensione di tali pratiche, (c) delle situazioni in cui tali pratiche sono messe in atto». Per Rapoport (1970) si tratta di «una ricerca sociale applicata, caratterizzata dal coinvolgimento immediato del ricercatore nel processo di azione. Il suo obiettivo è di fornire un contributo, nello stesso tempo, alle preoccupazioni pratiche delle persone che si trovano in situazione problematica e allo sviluppo delle scienze sociali».

È forse questo nuovo ruolo del ricercatore e dell'operatore, ovvero questa maggiore condivisione e sovrapposizione di ruoli e di campi che determina la forte novità della ricerca-azione. A volte, ricercatore e operatore, in quanto persone diverse, hanno un rapporto di collaborazione e di forte condivisione di presupposti, approcci e finalità; altre volte, è lo stesso operatore a essere ricercatore.

Esistono due livelli di ricerca-azione, uno è quello che coinvolge direttamente l'operatore sul campo, l'altro è quello in cui delle persone esperte aiutano gli operatori a imparare a usare questo strumento di ricerca. Le esperienze più diffuse sono state quelle di ricerca-azione di secondo livello nel campo della formazione in servizio degli insegnanti, in particolare in Gran Bretagna. L'esperienza condotta a questo livello da John Elliott ha evidenziato quanto questo tipo di azione abbia una valenza di crescita professionale anche per gli esperti.

Nei vari casi di ricerca-azione portata avanti allo scopo di aiutare lo sviluppo professionale degli insegnanti l'elemento comune è sempre stato la collaborazione all'interno dei gruppi, in particolare nei progetti di ricerca sviluppatisi in Gran Bretagna all'interno di istituzioni scolastiche o universitarie

2.5.6 Approccio empirico e sperimentale

In senso lato tutti gli approcci scientifici alla ricerca hanno una dimensione empirica, tuttavia nell'uso comune possiamo ritagliare un ambito di approccio alla ricerca che è caratterizzato dalla raccolta di dati e dalla analisi e o manipolazione delle variabili. L'uso dei due termini - empirico e sperimentale - è necessitato dal fatto che in ambito educativo è difficile un approccio sperimentale in senso stretto. Tuttavia, in questo approccio è dominante l'attenzione al metodo, la cura nella rilevazione dei dati e nelle operazioni di misura, la ricerca di definire il campo di osservazione in termini campionari. Possiamo considerare come approcci empirici dunque quei lavori di raccolta di informazioni che tendono poi ad analizzarle con strumenti statistici lavorando su campioni di diverse dimensioni, dalle grandi indagini internazionali dell'IEA (International Association for the Evaluation of Educational Achievements), all'OCSE PISA, alle indagini condotte su realtà di minori dimensioni ma sempre con un'attenzione alla metodologia e alla giustificazione degli strumenti e della procedura per garantire la validità e la generabilità dei risultati.

L'esperimento in senso stretto nasce da un disegno specifico che prevede l'intervento attivo dello sperimentatore che opera introducendo una o più variabili indipendenti per verificare gli effetti delle stesse sulla variabile dipendente, per cercare di stabilire un rapporto di causa-effetto tra l'intervento effettuato e il risultato osservato. Il ricercatore produce così un'alterazione della situazione "naturale"

introducendovi deliberatamente elementi nuovi con l'obiettivo di produrre cambiamenti (alterare la situazione) e interpretare il processo di mutamento che ne risulta. La difficoltà pratica ed etica di realizzare veri e propri esperimenti su esperienze educative ha portato alla teorizzazione di disegni quasi sperimentali, introducendo delle attenuazioni alle richieste che demarcano il modello sperimentale in senso stretto.

2.5.7 Approccio grounded

La Grounded Theory è un approccio che nasce dai lavori di Barney Glaser e Anselm Strauss. Si tratta di una metodologia induttiva per la raccolta e l'analisi di dati e per la produzione di teorie "radicate nei dati". Il termine "grounded" richiama proprio il forte radicamento di questo approccio nell'esperienza e nei dati che scaturiscono dall'esperienza stessa, attraverso l'immersione del ricercatore nei contesti reali oggetto di studio. Glaser e Strauss contrappongono il concetto di grounded theory a quello di grounded theory (teoria di ampio raggio). Pur riconoscendo infatti ai grandi esponenti delle scienze sociali il merito di aver prodotto modelli e linee guida per la generazione di teorie, Glaser e Strauss (1967) ritengono necessaria un'innovazione di approccio per la costruzione di "dense teorie locali", in grado di facilitare la comprensione profonda di fenomeni reali in tutta loro ricchezza e complessità.

L'uso dell'approccio grounded si è diffuso in diversi ambiti tra cui l'antropologia, la psicologia, la pedagogia, la medicina, utilizzando il termine sia in riferimento al processo della ricerca, sia al prodotto di tale processo, cioè la teoria stessa che emerge dai dati.

Lavorare con questo approccio comporta interpretare la ricerca non come un percorso predefinito, da svolgersi seguendo un disegno strutturato, progettato per verificare un'ipotesi, ma come un viaggio esplorativo, finalizzato alla comprensione profonda di fenomeni complessi, le cui tappe e i cui obiettivi si delineano a poco a poco durante il percorso. Quindi, non si inizia una ricerca con una teoria da provare, ma si individua un'area di indagine per poi approntare «un percorso euristico che consenta che ciò di quell'area è rilevante possa emergere nella forma di una teoria che fedelmente interpreti il fenomeno» (Mortari, 2007, p. 150). La raccolta, l'analisi, l'interpretazione dei dati e la costruzione della teoria non sono più fasi sequenziali, ma vi è tra loro una relazione circolare, ricorsiva: si raccolgono i dati, li si analizzano, si formula una teoria provvisoria che, eventualmente, porta alla ridefinizione delle modalità di raccolta dei dati, quindi si raccolgono nuovi dati, e si riformula la teoria, per sua natura concettualmente densa, fortemente radicata nell'esperienza e, allo stesso tempo, sempre soggetta a continue revisioni (Cohen, Manion e Morrison, 2011).

Quando si adotta l'approccio grounded si raccolgono attraverso interviste, integrate con l'osservazione, documenti, statistiche e dati di diverso tipo, le cui fonti sono *in primis* le esperienze osservate. Ogni tipologia di dati getta uno sguardo particolare sul fenomeno da indagare, accrescendo la "sensibilità teorico-concettuale" del ricercatore e la sua capacità di "far parlare i dati" (Mortari e Ghirotto, 2019, p. 84).

2.6. Le modalità prevalenti: descrivere, mettere in relazione, manipolare

2.6.1 Modalità descrittive

Rientrano in questa modalità tutte quelle ricerche che si propongono di analizzare le caratteristiche di esperienze educative per sistematizzarle e modellizzarle a partire dalla rilevazione accurata delle informazioni ritenute significative. Qualsiasi raccolta di informazioni muove da un modello e punta a costruirne uno più aggiornato. La ricerca descrittiva, infatti, comporta una selezione di informazioni e non si ferma mai alla semplice raccolta e classificazione dei dati, ma cerca di ricostruire attraverso questi un quadro unitario dell'esperienza presa in considerazione. Cerca inoltre di fornire un'interpretazione di eventi, situazioni, atteggiamenti, opinioni, tendenze, sviluppi attraverso un pertinente e controllato rilievo di aspetti direttamente presenti al ricercatore o ricavabili mediante l'analisi di documenti o di altre informazioni strutturate. Il ricercatore non interviene, cioè non manipola o predispone gli elementi da osservare, ma lavora sulle informazioni che trova e a cui è interessato in relazione a quanto si aspetta possa emergere da una conoscenza più approfondita di ciò che osserva.

La ricerca descrittiva dispone di molteplici tecniche di raccolta dei dati: osservazioni, interviste, questionari, test, e strumenti di raccolta multimediali quali registrazioni, fotografie, filmati. Come per ogni tipo di ricerca, sarà compito del ricercatore individuare quale o quali strumenti utilizzare in considerazione della natura dell'esperienza che osserva e del tipo di ipotesi da verificare o di domande di ricerca dalle quali fare emergere ipotesi. Quale che sia l'approccio prevalente del ricercatore, anche la ricerca descrittiva muove dalla scelta di una situazione da osservare (e dunque considerata problematica o da approfondire) rispetto alla quale aspetti quali ipotesi o domande di ricerca guideranno la scelta delle modalità e degli strumenti per la raccolta delle informazioni e le successive modalità di integrazione delle stesse in un quadro complessivo che avrà necessariamente un carattere interpretativo e tenderà alla possibile riproposizione del modello descrittivo utilizzato in esperienze con caratteristiche simili.

2.6.2 Modalità basate sulle relazioni

La ricerca è basata sulla rilevazione di informazioni rilevanti dell'esperienza educativa e l'analisi delle reciproche relazioni, facendo riferimento ad analisi statistiche fondate su modelli di correlazione. Anche in questo caso il ricercatore non interviene nell'esperienza se non per il modo in cui rileva le informazioni e per il modo in cui la rilevazione viene percepita da coloro che sono coinvolti nell'esperienza educativa osservata.

In questa modalità l'obiettivo di sviluppare relazioni e modelli interpretativi è più esplicito e richiede una maggiore formalizzazione delle informazioni che devono essere raccolte, classificate e organizzate in modo da prestarsi poi ad analisi statistiche.

Questo tipo di ricerca può essere "relazionale" o "predittiva". Nella ricerca relazionale si analizzano i rapporti tra le variabili per vedere se vi sono relazioni statisticamente significative, mentre nella ricerca predittiva si parte dalle relazioni tra le informazioni raccolte e si usano modelli statistici per individuare la probabilità che in condizioni analoghe possano riprodursi i fenomeni osservati.

2.6.3 Modalità sperimentali

La modalità sperimentale prevede che il ricercatore, sulla base di una ipotesi robusta, intervenga direttamente sull'esperienza di cui intende verificare una caratteristica, da un lato cercando di isolarla o controllarla, dall'altro introducendovi degli elementi per verificarne gli effetti su aspetti specifici opportunamente definiti.

L'intervento del ricercatore consiste nella manipolazione di una o più variabili indipendenti (variabili stimolo) e generalmente viene definito trattamento.

In altri termini, dato che l'obiettivo che la ricerca sperimentale si propone, nella sua forma essenziale, è di verificare la natura del rapporto tra le due variabili, l'intervento si basa sul far variare una delle due variabili e di osservare gli effetti sull'altra, cercando di isolare per il possibile le due variabili per evitare l'influenza di altri fattori. Mira dunque a verificare una ipotesi del tipo: "Il trattamento, cioè l'introduzione della variabile *X* (variabile stimolo) provocherà (causerà) una modifica nella variabile *Y*, che consentirà di affermare che *X* è causa del cambiamento di *Y*".

La modalità sperimentale classica nasce in laboratorio (dove è possibile ottenere il massimo isolamento delle variabili da esaminare) e su fenomeni facilmente isolabili e riproducibili.

La modalità sperimentale non sempre è facile da realizzare quando si tratta di problemi educativi, sia per la difficoltà di isolare e controllare il contesto, sia per la complessità del trattamento che non può ovviamente essere ridotto a uno stimolo puntuale, sia per motivi pratici o etici che condizionano il trattamento stesso.

In questa prospettiva è stata introdotta la distinzione proposta da Claude Bernard tra esperimento provocato ed esperimento occasionale. Nell'esperimento occasionale la manipolazione della variabile dipendente avviene senza l'intervento del ricercatore. Per fare un esempio, se riteniamo che la continuità didattica, in termini di presenza degli stessi insegnanti per un ciclo di studi, sia un fattore di miglioramento degli apprendimenti, non potremmo (per motivi etici) esporre una classe al ricambio continuo di insegnanti per verificarne gli esiti, ma dato che questo purtroppo avviene, possiamo scegliere classi equivalenti di cui alcune hanno subìto un costante ricambio di insegnanti per verificare gli effetti della discontinuità didattica. Fortunatamente o sfortunatamente è largamente possibile accedere a queste situazioni senza provocarle.

L'esperimento di laboratorio nasce in settori di indagine diversi da quello delle scienze sociali, tuttavia è possibile costruire disegni di ricerca sperimentali o quasi-sperimentali che consentono la formulazione di spiegazioni causali di quanto avviene nelle esperienze educative.

2.7 Conclusioni

Abbiamo provato a distinguere tra approcci di ricerca e modalità (descrivere, cercare relazioni o sperimentazioni) per comprendere le esperienze educative e trovare soluzioni che le migliorino. È tuttavia utile considerare che quale che sia l'esperienza problematica che vogliamo affrontare è necessario fare riferimento a tutti gli approcci e a tutte le modalità, anche se nella pratica seguiremo un approccio e una modalità prevalenti.

Le esperienze e i problemi che queste propongono non sono "disciplinari" e "settoriali" e non sono organizzati sulla base delle definizioni esposte in questo o in altri testi. Le esperienze problematiche hanno un retroterra storico, una dimensione teorica, richiedono comparazioni e confronti, e anche se non siamo nella condizione di ricerca-azione, il fatto stesso di intraprendere una ricerca comporta motivazioni e impegno che hanno analogie con questo approccio. Così come l'osservare un fenomeno per descriverlo - pur cercando di farlo nella massima neutralità - comporta l'osservazione delle relazioni tra le cose che osserviamo e anche, in modo più o meno esplicito, nella nostra mente, la ricerca di rapporti causali. Le definizioni che abbiamo proposto sono quindi da utilizzare solo come riferimenti e con mente aperta.

La tendenza a specializzarsi in un approccio non deve infatti andare a scapito dell'attenzione alla complessità delle esperienze concrete che stiamo studiando e della possibilità di integrare approcci (*mixed methods*) (Burke e Onwuegbuzie, 2004) e modalità di raccolta delle informazioni (qualitative e quantitative) che aiutino la comprensione di fenomeni complessi quali quelli legati ai processi educativi.

Riepilogo

In questo capitolo è stato esposto il modo in cui un ricercatore si avvicina al problema, e quali sono i condizionamenti che porta con sé e che deve conoscere e di cui deve acquisire consapevolezza.

Il ricercatore si muove all'interno di un paradigma scientifico ed è condizionato dai suoi valori, dalla sua esperienza e metodologia di lavoro, e subisce subisce condizionamenti esterni legati alle norme, agli indirizzi politici, ai finanziamenti disponibili per le attività di ricerca. Abbiamo poi visto in che modo il ricercatore costruisce le proprie ipotesi di ricerca come asserzioni su possibili soluzioni di problemi di cui cerca la verifica o la falsificazione.

Sono state illustrate le caratteristiche dell'atteggiamento scientifico e del metodo scientifico che debbono guidare il lavoro di ricerca pur essendo tra loro in rapporto dialettico. Sono stati inoltre esaminati i diversi approcci con i quali si può condurre la ricerca in campo educativo e le principali modalità per affrontare la ricerca, descrivere, cercare relazioni o sperimentare.

Riferimenti bibliografici

AERA, APA, NCME (2014). *Standards for educational and psychological testing.* Washington: AERA

Bernard C. (1952). *Introduction à l'étude de la médecine expérimental.* Paris: Flammarion.

Becchi E. (1992). *Ricerca-azione: riflessioni su voci di dizionari, manuali, enciclopedie,* in *Scuola e Città,* 4. 145.

Bridgman P. (1927). *The logic of modern physics.* New York: Macmillan. (trad. it. *Logica della fisica moderna.* Torino: Einaudi, 1952.)

Cohen L, Manion L, Morrison K. (2007). *Research methods in education.* 7th ed. NY: Routledge.

De mauro T. (1980). *Guida all'uso delle parole.* Roma: Editori Riuniti.

Dewey J. (1916). *Democracy and education.* New York: Macmillan (trad. it. *Democrazia e educazione.* Firenze: La Nuova Italia, 1988.)

Dewey J. (1929). *The sources of a science of education.* New York: Livering Publishing Corporation. (trad. it. *Le fonti di una scienza dell'educazione.* Firenze: La Nuova Italia, 1951.)

Dewey J. (1933). *How we think.* Boston: Heath. (trad. it. *Come pensiamo.* Firenze: La Nuova Italia, 1973.)

Dewey J. (1938). *Logic, the theory of inquiry.* New York: Henry Holt (trad. it. *Logica, teoria dell'indagine.* Torino: Einaudi, 1974.)

Dewey J. (1939a). *Theory of valuation.* Chicago: University of Chicago Press (trad. it. *Teoria della valutazione.* Firenze: La Nuova Italia, 1967.)

Dewey J. (1939b). Unity of science as social problem. In: Neurath O, Carnap R.C. Morris C (eds.) Fondations of the Unity of Science. Toward an International Encyclopedia of Unified Science. Chicago-London: University of Chicago Press, pp. 29-38. (trad. it. in P. Lucisano, *L'unità della scienza come problema sociale di John Dewey.* Cadmo, 22, 2000, pp. 33-37.)

Dewey J., Bentley A.F. (1949). *Knowing and the known.* Boston: The Beacon Press (trad. it. *Conoscenza e transazione.* Firenze: La Nuova Italia, 1974.)

Feyerabend P. (2010). *Against method.* 4th ed. New York: Verso Books. Trad. it. *Contro il metodo,* Milano: Feltrinelli.

Fraisse P., Piaget J. (1963) (eds.). *Traité de psychologie expérimentale.* Paris: Presses Universitaires de France (trad. it. *Trattato di psicologia sperimentale,* vol. I, *Storia e metodo.* Torino: Einaudi, 1972, pp. 123-200.)

Glaser B., Strauss A. (1967). *The discovery of Grounded Theory: strategies for qualitative research.* Hawthorne: Aldine de Gruyter.

Kemmis S., McTaggart R. (1982). *The action research planner, deakin university press: geelong.*

Kemmis S. (1985). *Action research.* In: Husén T., Postlethwaite T.N. (eds.) *The International Encyclopedia of Education,* vol. 1, p. 35.

Johnson R.B., Onwuegbuzie A.J. (2004). *Mixed methods research: a research paradigm whose time has come,* 33, Issue 7

Kuhn T. (1962). *The structure of scientific revolutions.* Chicago: University of Chicago Press (trad. it. *La struttura delle rivoluzioni scientifiche.* Torino: Einaudi, 1978).

Massa R. (2004). *La clinica della formazione. Un'esperienza di ricerca.* Milano: Franco Angeli.

Mcburney D.H. (1986). *Experimental Psychology.* Belmont: Wadsworth (trad. it. *Metodologia della ricerca in psicologia.* Bologna: Il Mulino).

Mialaret G. (éd.) (1985). *Introduction aux sciences de l'education.* Bureau International de l'Education, Genéve: UNESCO. (trad. it. *Introduzione alle scienze dell'educazione.* Roma-Bari: Laterza, 1992.)

Mierieu P. (2006). *Frankenstein educatore.* Bergamo: Junior 2007.

Mortari L. (2007). *Cultura della ricerca e pedagogia. Prospettive epistemologiche.* Roma: Carocci.

Mortari L., Ghirotto L. (2019). *Metodi per la ricerca educativa.* Roma: Carocci.

Piaget J. (1926). *La représentation du monde chez l'enfant.* Paris: Presses Universitaires de France (trad. it. *La rappresentazione del mondo nel fanciullo.* Torino: Einaudi, 1955.)

Popper K.R. (1934). *The logic of scientific discovery.* London: Hutchinson (trad. it. *La logica della scoperta scientifica.* Torino: Einaudi, 1969).

Popper K.R. (1969). *Scienza e filosofia. Problemi e scopi della scienza.* Torino: Einaudi.

RAPOPORT R.N. (1970). *Three dilemmas in Action Research.* Human Relation, n. 5, p. 499.

TRAVERS R.M.W. (1973). *Educational psychology.* New York: Macmillan (trad. it. *Psicologia dell'educazione.* Torino: Loescher, 1986).

VERTECCHI B. (2001). *Sull'incerto confine tra comparazione e conformità.* In: Istituto Nazionale per la Valutazione del Sistema d'Istruzione (CEDE), *Ricerche valutative internazionali 2000,* Frascati: Franco Angeli-CEDE).

3

Le fasi della ricerca e la scelta dei dati

La ricerca si sviluppa secondo una sequenza logica di fasi che seguono il modo in cui si sviluppa il nostro pensiero riflessivo. In questo capitolo vengono esaminate le caratteristiche e i problemi della ricerca scientifica con attenzione alle caratteristiche dei problemi educativi. Alla base del metodo scientifico c'è ad esempio l'idea che ogni ricerca possa essere replicata, cosa che è realmente molto difficile nei nostri campi di indagine. Sarà poi preso in considerazione il rapporto che c'è tra le scoperte casuali e la loro assunzione in termini di conoscenze scientifiche.
Nella seconda parte del capitolo sono approfonditi alcuni elementi base della ricerca quali l'uso delle parole e dei concetti, la costruzione di proposizioni, la formulazione di leggi e teorie. Segue un'attenta considerazione sui rischi che si corrono nel trasformare la realtà in cui stiamo facendo ricerca in dati. Anche se la ricerca richiede operazioni di riduzione dei fenomeni osservati, il ricercatore deve essere capace di cooperare con la complessità.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Fasi della ricerca.
- Principio della ripetibilità della ricerca e la sua applicabilità alla ricerca educativa.
- Ruolo della serendipità nel progresso scientifico.
- Parole, concetti, proposizioni, leggi e teorie.
- Dati e variabili.

Nella vita del signor Palomar c'è stata un'epoca in cui la sua regola era questa: primo, costruire nella sua mente un modello, il più perfetto, logico, geometrico possibile; secondo, verificare se il modello s'adatta ai casi pratici osservabili nell'esperienza; terzo, apportare le correzioni necessarie perché modello e realtà coincidano. Questo procedimento, elaborato dai fisici e dagli astronomi che indagano sulla struttura della materia e dell'universo, pareva a Palomar il solo che gli permettesse d'affrontare i più aggrovigliati problemi umani, e in primo luogo quelli della società e del miglior modo di governare.

Italo Calvino, *Palomar*

3.1 Le fasi della ricerca

Secondo il modello sviluppato da John Dewey la ricerca, in modo analogo al pensiero riflessivo, si sviluppa secondo un modello circolare che prende le mosse dalla percezione dell'esistenza di un problema e si sviluppa attraverso le fasi della intellettualizzazione del problema, della scelta di una strategia per affrontarlo, della scelta degli strumenti, della raccolta dei dati necessari, della verifica empirica della soluzione e della valutazione.[1]

L'indagine nasce dunque da un primo giudizio (che sarebbe più corretto chiamare pre-giudizio) e si conclude con un secondo giudizio che ne rappresenta il risultato compiuto e che, tuttavia, sarà a sua volta origine o elemento per indagini successive. Il primo giudizio valuta la situazione in esame come problematica, incerta, il secondo la valuta come definita, risolta.[2] Tra questi due estremi, spiega

[1] «L'esigenza di risolvere una difficoltà è il fattore permanente che guida l'intero processo della riflessione [...]. La natura del problema fissa il fine del pensiero, ed il fine controlla il processo del pensiero» (Dewey, 1933, trad. it. p. 75).

[2] «L'intero processo del pensare consiste in una serie di giudizi posti in relazione reciproca in modo da condurre a un giudizio finale - la conclusione. [...] Giudicare è tanto l'atto di scegliere e pesare le conseguenze dei fatti e delle suggestioni così come si presentano, quanto l'atto di decidere se i fatti allegati siano realmente fatti o se l'idea usata è un'idea fondata o una pura fantasticheria» (Dewey, 1933, trad. it. pp. 194-195).

Dewey, il nostro modo di procedere attraversa alcune fasi che non si succedono necessariamente in un ordine prestabilito.

Cinque fasi, o aspetti, del pensiero riflesso

Tra questi due estremi, come fasi del pensiero, vi sono:

- le suggestioni, in cui la mente si spinge innanzi verso una possibile soluzione,
- una intellettualizzazione della difficoltà o della perplessità avvertita (direttamente esperita) in un problema da risolvere, in una questione per cui deve essere cercata una risposta;
- l'uso di una suggestione dopo l'altra come idee conduttrici, o ipotesi, per iniziare e guidare l'osservazione e altre operazioni nella raccolta del materiale di fatto;
- l'elaborazione mentale dell'idea o della supposizione proprio come idea o supposizione (ragionamento) nel senso in cui un ragionamento è una parte, e non il tutto dell'inferenza;
- il processo di controllo dell'ipotesi tramite l'azione diretta o nell'immaginazione.

(Dewey, 1933, trad. it. p. 180).

L'intero processo di ricerca è dunque un processo di valutazione: è una valutazione la percezione del problema come tale, e noi abbiamo aggiunto che questa prefigura un'idea della possibile soluzione del problema stesso; è un insieme di giudizi l'intellettualizzazione del problema e la formulazione di una ipotesi, così come ogni assunzione di dati fatta per verificare l'ipotesi contiene le dimensioni della misurazione e della valutazione.

Dunque, abbiamo visto che il metodo scientifico «pur potendo sembrare complicato nelle sue forme più sofisticate è in fondo molto semplice» (Russell, 1931, trad. it. p. 39).

Prima di procedere a un esame più approfondito delle singole fasi della ricerca educativa è utile riassumerne gli aspetti principali dell'itinerario, ossia il suo percorso ciclico (Figura 3.1). Ciò significa, come abbiamo detto e ripeteremo più volte in queste pagine, «che dobbiamo sempre accettare tutti i risultati in modo provvisorio, consapevoli del fatto che indagini successive possono rilevarli sbagliati» (Bailey, 1982, trad. it. p. 30).

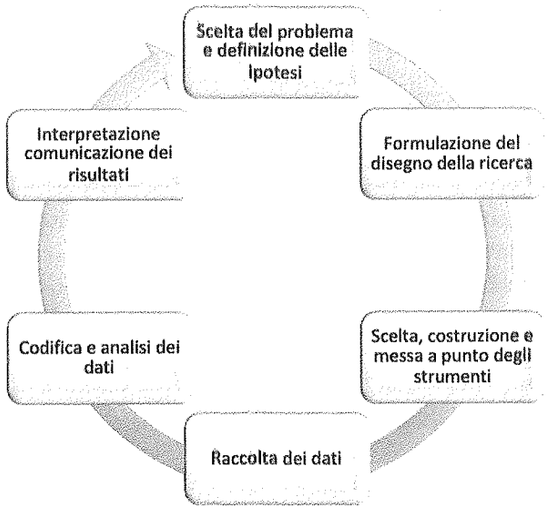


Figura 3.1 Le fasi della ricerca

3.2 La ripetibilità della ricerca

L'ipotesi di ricerca deve essere costruita in modo tale da poter essere sottoposta a verifica, e il fatto di avere verificato una ipotesi consente di ritenerla vera fino a prova contraria. Ogni nostra conoscenza deve poter essere sottoposta a ulteriori verifiche, e per questo motivo ogni verifica deve essere impostata in modo da poter essere ripetuta.

Affinché uno studio sia replicabile è necessario che il ricercatore dichiari in modo esplicito le ipotesi, i metodi di verifica, gli strumenti, il campione utilizzato e renda disponibili alla comunità scientifica i suoi dati e il dettaglio delle procedure seguite.

Illustrare chiaramente le procedure seguite nelle diverse fasi non è però un risultato facile anche per i vincoli di spazio che sono posti dall'esterno nella pubblicazione dei lavori.

La ripetizione delle indagini inoltre non è molto comune dato che i ricercatori preferiscono provare a cercare nuove soluzioni piuttosto che verificare lavori condotti da altri.

Ricordiamo inoltre che le esperienze educative hanno caratteristiche che rendono spesso difficile la ripetizione della ricerca, si svolgono nel tempo, con soggetti in crescita che apprendono dalle esperienze precedenti e che vivono in contesti diversi, ciascuno in grado di influenzare i loro stati d'animo, apprendimenti e comportamenti.

Tra i regali più recenti e noti della serendipità, merita citare la pillola sindenfil, più nota come Viagra. Ian Osterloh ha raccontato che la scoperta è avvenuta del tutto per caso. Si stava sperimentando un farmaco per l'angina pectoris, il dolore al petto che precorre gli attacchi cardiaci, quando la maggior parte dei partecipanti ai test raccontarono di avere osservato un effetto non previsto: un deciso e visibile miglioramento delle erezioni.

Recentemente studiosi dell'Istituto di ricerca cardiaca "Victor Chang" di Sydney, guidati dal dott. Mingong Khou, hanno compiuto un decisivo passo avanti nella ricerca sul cancro, con la scoperta, avvenuta "per caso", di un gene capace di fermare la crescita incontrollata delle cellule, caratteristica dei tumori. I ricercatori stavano cercando un gene responsabile dell'ingrossamento dei muscoli del cuore - e quindi dell'arresto cardiaco - e hanno scoperto un gene, il *TD14*, che ferma la crescita delle cellule. Questo gene opera come un interruttore dei segnali che determinano la crescita anormale delle cellule.

Tuttavia a ogni scoperta deve seguire una fase di verifica sperimentale che ha la funzione di validare i risultati ottenuti.

3.3 Il progresso scientifico e la serendipità

«È difficile spiegare in un libro che cosa sia esattamente la ricerca perché anche chi fa ricerca spesso la descrive in maniera diversa da come la fa realmente» (McBurney, 1983; trad. it. p. 43). Nella vita quotidiana come nella ricerca molti apprendimenti avvengono a prescindere da una intenzionalità esplicita e da procedimenti formalizzati.

A dispetto del rigore metodologico richiesto dai manuali, la conoscenza si sviluppa anche in ragione della creatività, degli errori e del caso. Quando la conoscenza nasce da eventi non previsti si parla di *serendipità*. Il temine si ispira a un antico racconto persiano ripreso da Horace Walpole (1717-1797) nel quale assume il significato di "capacità di trovare ciò che non si sta cercando". Nel racconto i tre figli di Jafer, re-filosofo di Serendip (Ceylon) scoprono continuamente per caso e per sagacia cose che non vanno cercando. Questo fenomeno è stato analizzato dal famoso sociologo Merton, in un libro tradotto recentemente in italiano, ma che risale al 1958. Merton sostiene che la serendipità ha giocato un ruolo importante nella storia della scienza (Merton e Barber, 2002). Anche se in qualche modo per cogliere elementi non attesi dalle esperienze che facciamo non basta avere solo una mente aperta, occorre una mente preparata.

Approfondimento 3.1

Popper sintetizza per punti nella conferenza *Problemi, scopi e responsabilità della scienza*[3] (1969, trad. it. pp. 136-138) alcune delle tematiche che abbiamo esposto.

«Mi sia ora concesso di riassumere la prima parte di questa chiacchierata riformulando tutte le cose controverse che sono venuto dicendo, in un certo numero di tesi, che tenterò di mettere nella forma più stimolante che mi sia possibile.

1. Tutta la conoscenza scientifica è ipotetica o congetturale.
2. L'accrescimento della conoscenza, e specialmente della conoscenza scientifica, consiste nell'imparare dagli errori che abbiamo commesso.
3. Quello che possiamo chiamare il metodo della scienza consiste nell'imparare sistematicamente dai nostri errori; in primo luogo, osando commetterli - cioè, proponendo arditamente teorie nuove, e, in secondo luogo, andando sistematicamente alla ricerca degli errori che abbiamo commesso: andandone alla ricerca, cioè, mediante la discussione critica e l'esame critico delle nostre teorie.
4. Tra gli argomenti più importanti usati in questa discussione critica ci sono gli argomenti tratti dai controlli sperimentali.
5. Gli esperimenti sono costantemente guidati dalla teoria, da mezze idee teoriche, di cui spesso lo sperimentatore non è consapevole, da ipotesi sulle fonti possibili di certi errori sperimentali, da speranze o congetture intorno a quello che sarà un esperimento fruttuoso: cioè a dire, da mezze idee teoriche che l'esperimento di un certo tipo sarà teoricamente fruttifero.
6. Quella che è chiamata obiettività scientifica consiste soltanto nell'approccio critico; nel fatto che se abbiamo pregiudizi nei confronti della nostra teoria favorita, qualcuno dei nostri amici o dei nostri colleghi (o, in mancanza di questi, qualcuno degli scienziati della generazione successiva) sarà ansioso di criticarci, vale a dire, di confutare, se può, la nostra teoria favorita.
7. Questo fatto può incoraggiarvi a tentare di confutare da voi le vostre teorie; cioè può imporre su di voi una certa disciplina.
8. Nonostante ciò, sarebbe un errore il pensare che gli scienziati sono più «obiettivi» dell'altra gente. A farci tendere all'oggettività non è l'obiettività, o il distacco, del singolo scienziato, ma la scienza stessa, o quella che possiamo chiamare la cooperazione allo stesso tempo amichevole e ostile fra gli scienziati, cioè la loro prontezza a criticarsi reciprocamente.
9. Per il dogmatismo e i pregiudizi dei singoli scienziati c'è addirittura una giustificazione metodologica. Poiché il metodo della scienza è il metodo della discussione critica, è estremamente importante che le teorie criticate siano difese tenacemente. Infatti, solo in questo modo possiamo imparare quale sia il loro potere reale; e solo se le critiche incontrano resistenza, impariamo la piena forza di un'argomentazione critica.
10. La parte fondamentale che nella scienza hanno le teorie, o ipotesi, o congetture, fa sì che sia importante distinguere tra teorie controllabili, o falsificabili, e teorie non controllabili o non falsificabili.
11. Solo una teoria che asserisca o implichi che certi eventi concepibili non accadranno nel fatto, è controllabile. Il controllo consiste nel tentare di far accadere, con tutti i mezzi di cui possiamo disporre, precisamente quegli eventi di cui la teoria ci dice che non possono accadere.
12. Così si può dire che ogni teoria che possa essere sottoposta a controlli vieta che certi eventi accadano. Una teoria parla della realtà empirica solo nella misura in cui le pone limiti.
13. Dunque, ogni teoria può essere messa nella forma «questa cosa così e così non accade». Ad esempio, la seconda legge della termodinamica può essere messa nella forma: non esiste una macchina del moto perpetuo del secondo tipo.
14. Nessuna teoria può dirci qualcosa intorno al mondo empirico, a meno che non sia in linea di principio in grado di entrare in collisione con il mondo empirico; e ciò significa, esattamente, che deve essere confutabile.
15. La controllabilità ha gradi: una teoria che asserisca di più, e dunque assuma rischi più grandi, è controllabile meglio di una teoria che asserisca molto poco.
16. Analogamente, i controlli possono essere graduati secondo che siano più o meno severi. Ad esempio, i controlli qualitativi sono, in linea generale, meno severi dei controlli quantitativi; e i controlli delle predizioni quantitative più precise sono più severi dei controlli delle predizioni meno precise.
17. L'autoritarismo nella scienza era collegato con l'idea di fondare, cioè a dire, di provare o di verificare, le sue teorie. L'approccio critico è collegato con l'idea di sottoporre a controlli, ossia di tentare di confutare, o di falsificare, le sue congetture».

[3] Conferenza tenuta al XLVII raduno annuale della Federazione delle Società americane per la Biologia Sperimentale, Atlantic City (New York), 17 aprile 1963.

3.4. Le parole della ricerca

3.4.1 Dare un nome alle persone, agli animali, alla natura e ai nostri sentimenti

Nel paragrafo precedente abbiamo sostenuto che la ricerca nasce dalla prefigurazione di una soluzione a un problema, quindi dalla ipotesi di un punto di arrivo. Il metodo scientifico di cui disponiamo è il nostro punto di partenza, e al tempo stesso è il punto di arrivo delle generazioni che ci hanno preceduto.

Ora è necessario approfondire alcuni aspetti del modo in cui elaboriamo e conserviamo la conoscenza. Gli strumenti con i quali costruiamo, comunichiamo a noi stessi e agli altri e conserviamo la conoscenza sono soprattutto strumenti linguistici. «Negli stadi intermedi di transizione e trasformazione della situazione indeter-

minata viene impiegato come mezzo il discorso, mediante l'uso di simboli. Per usare la terminologia logica generalmente accettata, vengono necessariamente portate in campo le proposizioni, ossia i termini e le relazioni fra di essi» (Dewey, 1938, trad. it., p. 136).

Per fare un esempio, la frase "gli studenti non amano la scuola" acquista significato sulla base del significato che noi attribuiamo al termine "studenti", al verbo "amare", e a quel che ritorna nelle nostre menti quando pensiamo alla "scuola". Non c'è dubbio che ciascuno interpreta le tre parole in relazione alla propria esperienza. Tuttavia, nonostante i molti significati di ciascuno dei termini utilizzati per costruire questa frase, è possibile trovare un certo consenso nella sua interpretazione.

Lo sforzo della scienza è di costruire consenso intorno ai concetti e alle proposizioni che legano insieme i concetti. La costruzione del sapere scientifico è dunque un processo sociale, collettivo, la cui prima regola è la collaborazione tra persone e il cui risultato, "sempre provvisorio", è un accordo tra persone. In prima approssimazione, questo accordo può riguardare soprattutto coloro che si occupano a tempo pieno di un settore di conoscenza, la comunità scientifica, gli esperti. Le conquiste degli esperti devono in qualche modo potersi estendere all'intera collettività sociale. Una conoscenza che non viene comunicata o che risulta intangibile dalla comunità sociale, infatti, non ha senso e non può definirsi scientifica.

Secondo Dewey «Un singolo ricercatore in un dato campo specifico si richiama alle esperienze della comunità dei suoi colleghi per la conferma e la correzione dei suoi risultati. Finché non sia raggiunto l'accordo da parte di coloro che devono integrare le condizioni prospettate, le conclusioni annunciate da un singolo ricercatore hanno il carattere di un'ipotesi, specialmente se non s'accordano con la tendenza generale dei risultati già accettati. E anche l'accordo fra le attività e i loro risultati, quali si realizzano rispetto al pubblico più ampio (non tecnicamente scientifico), pur rimanendo su di un piano differente, costituisce nondimeno parte integrante di un controllo completo di conclusioni fisiche, ogniqualvolta la loro pubblica ripercussione sia degna di rilievo» (Dewey, 1938, trad. it., pp. 607-608).

Talvolta ci vogliono anni e a volte secoli prima che la comunità sociale tutta comprenda, o addirittura accetti, il lavoro di uno scienziato. Tuttavia, che questo avvenga o sia avvenuto nella storia dipende dal fatto che lo scienziato ha lasciato tracce del suo lavoro, e che queste tracce siano state disponibili a chi, anche molto tempo dopo, ha deciso di ripercorrere il suo cammino.

Il nostro punto di partenza, dunque, richiede che vengano presi in esame alcuni aspetti che rappresentano la sintesi provvisoria di cui disponiamo e che saranno gli strumenti che utilizzeremo per metterci al lavoro.

3.5 Le parole e i concetti

In alcuni casi le parole che usiamo si riferiscono a cose immediatamente osservabili come quando diciamo tenendo in mano una penna "Questa è una penna". In altri casi i concetti si riferiscono a fenomeni più complessi e non immediatamente osservabili come quando ci riferiamo a stati d'animo o sentimenti o a cose non immediatamente osservabili, come parole del tipo "interesse", "amore" o simili. Infine, la nostra mente è in grado di produrre un terzo tipo di concetti che non hanno alcun riscontro reale, come quando parliamo di "angeli" o "demoni". Harari sostiene che la capacità di trasmettere informazioni su cose che non esistono affatto sia il tratto distintivo della rivoluzione cognitiva dei Sapiens: "Per quanto ne sappiamo, solo i Sapiens sono in grado di parlare di intere categorie di cose che non hanno mai visto, toccato od odorato" (p. 36).

I concetti del primo e del secondo tipo hanno piena cittadinanza nel discorso scientifico, quelli del terzo tipo vanno invece maneggiati con maggiore prudenza.

Nel linguaggio naturale, ragioni di economia ci hanno spinto a utilizzare le stesse parole con più significati, rimandando la disambiguazione ai rapporti tra coloro che parlano e al contesto in cui vengono inserite. Per esempio, se diciamo "penna" e chiediamo al di fuori di un contesto specifico "cos'è una penna?", le risposte possibili sono molteplici e vanno dalle piume degli uccelli al piccolo pezzo di plastica usato per far vibrare le corde di una chitarra. Nel linguaggio scientifico si cerca di limitare il più possibile la polisemia delle parole e di utilizzarle con molta cautela, sostituendole dove è possibile con simboli meno equivoci, e in altre situazioni cercando di affiancare al concetto in uso una sua definizione.

Il ricorrere a una definizione condivisa è poi ancor più necessario quando usiamo concetti che fanno riferimento a dimensioni non immediatamente osservabili, in questo caso la definizione condivisa fa riferimento a quegli aspetti osservabili che assieme caratterizzano il concetto non osservabile a cui si fa riferimento (definizioni operative). Nel formulare una teoria dovremmo cercare di formalizzare le nostre proposizioni in modo da poter ottenere su di esse il massimo consenso possibile. Per raggiungere questo obiettivo facciamo ricorso a linguaggi specializzati, a linguaggi nei quali i concetti vengono utilizzati in modo univoco, e in particolare alla logica formale, o ancor più spesso a linguaggi di tipo matematico o statistico. A proposito della funzione della terminologia e del suo rapporto con il consenso, Dewey in *Conoscenza e transazione* si esprime così:

«La scienza adopera con piena efficienza le proprie denominazioni tecniche, che servono a caratterizzare come provvisoriamente accettabili certe regioni del campo dell'indagine scientifica, permettendo così di prendere più attentamente in considerazione altre regioni che rimangono problematiche. Codesta efficienza si basa sulla possibilità che il ricercatore ha di mantenere stabili tali denominazioni

– di sapere cioè quel che con esse propriamente designa – sia attraverso i successivi stadi del suo procedimento di indagine che nel dialogo con i suoi colleghi» (Dewey e Bentley, 1949, trad. it., p. 58).

Approfondimento 3.2

Nel framework del PISA (Programme for International Student Assessment), dopo avere richiamato i presupposti teorici della ricerca si procede a una definizione operativa di ciò che si intende misurare. Il dettaglio con cui viene spiegata ciascuna delle parole utilizzate è un esempio del rilievo che ha la definizione dei termini nell'attività di ricerca (OECD (2019), Pisa 2018 *Assessment and Analytical Framework*, Paris: OECD Publishing)

La *literacy* in lettura è la capacità degli studenti di comprendere, utilizzare, valutare, riflettere e interagire con i testi al fine di raggiungere i propri obiettivi, sviluppare la proprie conoscenze e potenzialità e partecipare alla società.

Si procede poi a una spiegazione di questa definizione parola per parola in modo da chiarire alla comunità scientifica e agli altri utenti della ricerca il significato di ciò che si vuole prendere in esame.

Literacy in lettura ...

Il termine *literacy in lettura* viene utilizzato al posto del termine "lettura" perché trasmette in modo più preciso a un pubblico non esperto ciò che l'indagine sta misurando. La "lettura" è spesso intesa come semplice decodifica (ad esempio, conversione del testo scritto in suoni), o anche come lettura ad alta voce, mentre l'intenzione di questa ricerca è di misurare costrutti molto più ampi e comprensivi. La literacy in lettura include una vasta gamma di competenze cognitive e linguistiche, dalla decodifica di base alla conoscenza delle parole, della grammatica e delle strutture linguistiche e testuali più ampie necessarie per la comprensione, nonché l'integrazione del significato con la propria conoscenza del mondo. Comprende anche le competenze metacognitive: la consapevolezza e la capacità di utilizzare una varietà di strategie appropriate durante l'elaborazione dei testi. Le competenze metacognitive vengono attivate quando i lettori pensano, monitorano e regolano la loro attività di lettura per un particolare obiettivo. Il termine "literacy" si riferisce tipicamente alla conoscenza di una persona di una materia o di un campo, sebbene sia stato più strettamente associato alla capacità di un individuo di apprendere, utilizzare e comunicare informazioni scritte e stampate. Questa definizione sembra vicina al concetto che il termine literacy in lettura intende esprimere in questo quadro l'applicazione attiva, mirata e funzionale della lettura in una serie di situazioni e per vari scopi. [...]

... la capacità degli studenti di comprendere, utilizzare, valutare, riflettere...

Il termine *comprensione* è facilmente collegato al concetto ampiamente accettato di "comprensione della lettura", secondo il quale tutta la lettura implica un certo livello di integrazione delle informazioni del testo con le conoscenze pregresse del lettore. Anche nelle prime fasi della lettura, i lettori devono attingere alla loro conoscenza dei simboli (per esempio, le lettere) per decodificare i testi e devono usare la loro conoscenza lessicale per generare significati. Tuttavia, questo processo di integrazione può anche essere molto più ampio: include, per esempio, lo sviluppo di modelli mentali di come i testi si relazionano al mondo.

La parola *utilizzo* si riferisce alle nozioni di applicazione e funzione: fare qualcosa con ciò che leggiamo.

Valutare è stato aggiunto per PISA 2018 per incorporare l'idea che la lettura è spesso diretta a un obiettivo, e di conseguenza il lettore deve soppesare fattori come la veridicità degli argomenti nel testo, il punto di vista dell'autore e la rilevanza di un testo agli obiettivi del lettore.

Riflettere è stato aggiunto a "comprensione", "utilizzo" e "valutazione" per enfatizzare l'idea che la lettura è interattiva: i lettori attingono ai propri pensieri ed esperienze quando interagiscono con il testo. Ogni atto di lettura richiede una riflessione, in cui i lettori rivedono e mettono in relazione le informazioni all'interno del testo con le informazioni esterne al testo. Man mano che i lettori sviluppano le loro riserve di informazioni, esperienze e convinzioni, testano costantemente ciò che leggono rispetto alla conoscenza esterna, rivedendo e rivedendo continuamente il loro senso del testo. Riflettere sui testi può includere la valutazione delle affermazioni dell'autore, il loro uso della retorica e di altri mezzi di discorso, nonché inferire la prospettiva dell'autore. Allo stesso tempo, in modo incrementale e forse impercettibile, le riflessioni dei lettori sui testi possono alterare il loro senso del mondo. La riflessione potrebbe anche richiedere ai lettori di considerare il contenuto del testo, applicare la loro conoscenza o comprensione precedente o pensare alla struttura o alla forma del testo. Ciascuna di queste abilità comprese nella definizione – "comprensione", "utilizzo", "valutazione" e "riflessione" – sono necessarie, ma nessuna è sufficiente per una buona capacità di lettura.

... interagire con ...

Una persona che è istruita nella lettura non solo ha le capacità e le conoscenze per leggere bene, ma apprezza e usa la lettura per una varietà di scopi. È quindi un obiettivo dell'educazione coltivare non solo la competenza ma anche l'impegno nella lettura. L'impegno in questo contesto implica la motivazione a leggere e comprende un insieme di caratteristiche affettive e comportamentali che includono l'interesse e il godimento della lettura, un senso di controllo su ciò che si legge, il coinvolgimento nella dimensione sociale della lettura e pratiche di lettura diverse e frequenti.

... i testi ...

La parola "testi" intende includere tutta la lingua utilizzata nella sua forma grafica: scritta a mano, stampata o su schermo. In questa definizione, escludiamo come testi artefatti del linguaggio puramente uditivo come registrazioni vocali, film, TV, immagini animate e immagini senza parole. I testi includono visualizzazioni come diagrammi, immagini, mappe, tabelle, grafici e fumetti, che includono un po' linguaggio scritto (per esempio, didascalie). [...]

3.6 Le proposizioni

La nostra conoscenza procede[4] con il collegare i concetti in proposizioni. Si usa distinguere le proposizioni in descrittive, esplicative e predittive.

Le *proposizioni descrittive* sono quelle che descrivono il "come" o il "quando" o il "modo" di un fenomeno. Se qualcuno ci chiedesse di spiegare meglio la frase «gli studenti non amano la scuola», dovremmo anzitutto descrivere con più precisione che cosa intendiamo per studenti (età, genere, background socioeconomico ecc.), poi dovremmo cercare di descrivere cosa intendiamo per amore e da che cosa induciamo questo "non amore", e ancora dovremmo chiarire che cosa intendiamo per scuola e descrivere i momenti e le manifestazioni di questa mancanza di passione.

Le *proposizioni esplicative*, invece, sono quelle che spiegano il "perché" del fenomeno in esame. Un ricercatore che si occupa di educazione, assunto che gli studenti non amino la scuola, dovrà chiedersi perché questo avviene. Dovrà allora costruire proposizioni ipotetiche del tipo "se... allora" e cercare di stabilire legami tra due fenomeni. Il più forte di questi legami è quello che stabilisce che un fenomeno è causa di un altro.

Le *proposizioni predittive*, infine, non si limitano a spiegare a posteriori il fenomeno osservato, ma hanno la forza di prevedere che a un dato fenomeno corrisponde con molta probabilità, o meglio con probabilità nota, un altro fenomeno.

La nostra conoscenza comprende molte proposizioni predittive. Siamo in grado, per esempio, di fare previsioni meteorologiche molto accurate grazie all'impiego di una serie di strumenti che ci consentono di prevedere, sulla base della misura di alcuni fenomeni, le condizioni atmosferiche dei giorni successivi. Siamo però meno in grado di dire quali siano le cose da cambiare per evitare che "gli studenti non amino la scuola". Abbiamo già molte informazioni su questo problema, ma non ancora sufficienti da permetterci di costruire proposizioni predittive che abbiano la forza di essere scientificamente e socialmente convincenti.

La classificazione delle proposizioni in descrittive, esplicative e predittive va considerata in modo pragmatico. Molti sostengono ad esempio che una buona descrizione contenga già la spiegazione di un fenomeno, e Bailey ci ricorda che «la precisa natura del rapporto tra previsione e spiegazione è stata argomento di dibattito tra i filosofi della scienza. Alcuni sostengono che la spiegazione e la previsione sostanzialmente coincidano: differiscono soltanto perché la previsione precede l'evento, mentre la spiegazione ha luogo dopo che l'evento è accaduto».

Bailey si limita a rinviare il dibattito ad altra sede scrivendo: «ci limiteremo ad affermare che la nostra soddisfazione sarebbe maggiore se riuscissimo non soltanto a spiegare l'evento dopo che si è verificato, ma anche a prevederlo per poterlo analizzare nello stesso momento in cui si manifesta» (1982, trad. it. p. 58).

[4] In realtà questa frase è un artificio espositivo dato che il rapporto tra proposizioni e concetti non è tale da stabilire un ordine di priorità.

In che modo, infatti, potremmo scegliere tra una spiegazione a posteriori e un'altra spiegazione a posteriori di un fenomeno, se non mettendo alla prova queste spiegazioni come previsioni?

3.7 Leggi e teorie

Quando abbiamo raggiunto una conoscenza consolidata della relazione tra due fenomeni e siamo in grado di formulare previsioni attendibili, utilizziamo per la descrizione della regolarità di questa relazione il termine legge. La legge è l'affermazione di un'associazione regolare tra fenomeni. Nel formulare una legge non è necessario avere raggiunto una conoscenza che ci consenta di stabilire nessi assoluti tra due fenomeni in esame, è sufficiente stabilire che tra essi esiste una regolarità di associazione. Per tornare al ragionamento fatto poco sopra sui concetti, usiamo il termine legge nella scienza e nel diritto con significati abbastanza simili, il senso è di affermare che in linea di massima le cose dovrebbero andare in questo modo. In entrambi i casi le eccezioni confermano la regola, e quando le eccezioni si fanno numerose si rende necessario rivedere la regola.

Per avere una legge non è necessario che tra i fenomeni osservati vi sia una associazione perfetta, nelle scienze sociali è sufficiente osservare che nella maggior parte dei casi a un certo tipo di comportamento segue un certo tipo di risposta.[5] Quando è possibile cerchiamo di esprimere il livello di approssimazione delle nostre leggi in termini di probabilità statistica.

Non vi è dubbio che la proposizione «gli studenti non amano la scuola», sia pure precisata linguisticamente, possa essere assunta al rango di legge, dato che la constatazione di questo fenomeno è abbastanza consolidata sia dal punto di vista diacronico sia dal punto di vista comparatistico, e che le eccezioni non sono sufficienti a sconfessare la regola.

La scoperta e la definizione di leggi è un fatto importante dal lato pratico, ma non è che un primo passo nel cammino della conoscenza. Il passo successivo (o precedente) è la costruzione di teorie.

Il termine teoria deriva dal greco (teoria), e il suo uso iniziale era per definire assieme uno spettacolo (*theà*) e la partecipazione allo spettacolo stesso (*horân*). Nel

[5] Su questo aspetto cfr. Weber.*Il metodo delle scienze storico-sociali*, pp. 207-237. Pietro Rossi, nell'Introduzione al volume, afferma che «Weber realizza l'abbandono del modello classico di spiegazione causale e il passaggio a uno schema di spiegazione che non è più causale, ma è invece condizionale. Nel porre in luce una serie finita di fenomeni - diversi a seconda del punto di vista dell'indagine - da cui dipende un certo fenomeno considerato nella sua individualità, le scienze storico-sociali non ne stabiliscono i fattori determinanti, ma ne individuano un certo gruppo di condizioni che, accanto ad altre lo rendono possibile. Alla relazione di causa-effetto, interpretata come relazione necessaria, subentra un rapporto di condizionamento» (pp. 27-28). Per una posizione simile cfr. il capitolo sulla Logica di Dewey su "L'indagine sociale", in particolare pp. 633-634.

tempo ha assunto poi il significato di contemplazione intellettuale e quindi di pensiero speculativo in contrapposizione alla pratica.

In senso tecnico usiamo il termine teoria per riferirci a una costruzione complessa che raccoglie un insieme di leggi e proposizioni. Dunque, nel nostro campo di indagine, una teoria è descrivibile come una serie di leggi e proposizioni che collegate fra loro mirano a spiegare un determinato fenomeno sociale.[6]

La teoria rappresenta il consolidamento di esperienze ripetute e il punto di arrivo del nostro sforzo di concettualizzare, ordinare e collegare tra loro i fatti.

Questo collegamento di eventi singoli in leggi e di leggi in teorie costituisce in senso lato il processo di accumulazione e organizzazione delle conoscenze, e in senso stretto quella che chiamiamo spiegazione scientifica. La fecondità di una teoria è legata al numero di fenomeni o di leggi che riesce a collegare. È necessario precisare tuttavia che leggi e teorie hanno la stessa natura che hanno le ipotesi, esse cioè sono da verificare. La loro validità infatti può essere sottoposta a un numero inesauribile di conferme e può essere comunque messa in discussione, "falsificata", da un tentativo non riuscito di conferma.[7]

Nella costruzione delle nostre conoscenze le teorie svolgono la funzione di coordinare e organizzare le conoscenze e le leggi e hanno anche forza di fare intravedere nuove prospettive di lettura e interpretazione dei fenomeni e nuove leggi.

La storia ha visto alternarsi, in momenti diversi, il prevalere dello sforzo di sintesi legato alla costruzione di teorie generali in grado di spiegare la maggior parte dei fenomeni osservabili e il prevalere dello sforzo di analisi legato alla precisazione di teorie più puntuali, meno ampie negli obiettivi ma capaci di spiegare in modo più approfondito i fenomeni considerati.

Non è questa la sede per approfondire tematiche di filosofia della scienza e di epistemologia, tuttavia possiamo concludere questo paragrafo cercando di sintetizzare quelle che sono le caratteristiche della teoria che fa da sfondo a tutte le nostre attività di ricerca. I fenomeni sociali in generale, e in particolare il sottoinsieme di fenomeni sociali che noi studiamo, quelli cioè legati ai processi educativi, presentano delle regolarità che sono comprensibili dal nostro pensiero. La possibilità di cogliere questa regolarità ci consente di descrivere i fenomeni educativi e di classificarli, di riportarli a concetti e di costruire tipologie e tassonomie. Lo studio di queste regolarità ci consente inoltre di formulare spiegazioni e leggi sia associando regolarità sia stabilendo rapporti di causa ed effetto tra fenomeni.

[6] Mc Burney (1983, trad. it. p. 22) propone due diverse definizioni di teoria: (1) *teoria in senso lato*: una asserzione o una serie di asserzioni riguardanti le relazioni tra variabili; (2) *teoria in senso stretto*: una asserzione o una serie di asserzioni sulle relazioni tra variabili che includa almeno un concetto non direttamente osservabile, ma necessario a spiegare le relazioni tra variabili.

[7] Popper in *La logica della scoperta scientifica* (1934) scrive: «La scienza non posa su un solido strato di roccia. L'ardita struttura delle sue teorie si eleva, per così dire, sopra una palude. È come un edificio costruito su palafitte. Le palafitte vengono conficcate dall'alto, giù nella palude: ma non in una base naturale o "data"; e il fatto che desistiamo dai nostri tentativi di conficcare più a fondo le palafitte non significa che abbiamo trovato un terreno solido. Semplicemente, ci fermiamo quando siamo soddisfatti e riteniamo che almeno per il momento i sostegni siano abbastanza stabili da sorreggere la struttura» (trad. it. p. 107).

La possibilità di queste operazioni rende possibile che lo sforzo di conoscere sia collegabile alla possibilità di decidere meglio. Il fatto che la solidità delle conoscenze si possa associare a decisioni che tengano in conto queste conoscenze rimane, invece, un problema lontano dall'essere risolto.

3.8 I fatti del caso

Nella considerazione di ciò che percepiamo in una esperienza educativa e che riteniamo di poter migliorare con la nostra ricerca c'è una fase di selezione di quegli elementi che possono esserci utili a comprendere, definire e affrontare il problema: i dati del problema che Dewey preferisce chiamare "fatti del caso" legando fortemente il problema della raccolta delle informazioni alla situazione in cui queste si presentano. «L'intendere o l'interpretare sta nell'ordinare quei materiali che sono stati accertati come fatti; cioè, nella determinazione delle loro relazioni. In ogni data materia esistono molte relazioni di molte specie. Occorre determinare quel particolare gruppo di relazioni che è applicabile al problema trattato. Concetti teorici adatti entrano in gioco soltanto se il problema trattato è chiaro e definito; da sola, cioè, la teoria non può decidere quale gruppo di relazioni si debba istituire, o come si debba intendere un dato complesso di fatti. Un meccanico, per esempio, comprende le varie parti di una macchina, di un'automobile o altro, solo se conosce in che modo le varie parti lavorano assieme; infatti, è il modo in cui lavorano assieme che fornisce il principio d'ordine in base al quale esse vengono collegate fra loro. Il concetto del "lavorare assieme" involge il concetto di conseguenze: la significanza delle cose risiede nelle conseguenze che producono interagendo con le altre cose ben specificate. Nucleo centrale del metodo sperimentale è la determinazione della significanza delle cose osservate mediante l'istituzione deliberata di modi d'interazione» (Dewey, 1938, trad. it. pag. 633).

I dati sono dunque importanti, tuttavia una cattiva pratica scientifica ha portato molti ricercatori a fare assumere ai dati una natura indipendente dal contesto e dal momento in cui vengono rilevati. È bene tenere presente che i dati per essere tali devono essere prima presi, e che qualunque operazione di presa di dati li estrae dal continuum spazio-temporale in cui essi esistono per trasformarli in qualcosa di diverso.

I dati dipendono dunque dal modo in cui vengono presi, dalle nostre intenzioni, dagli strumenti che utilizziamo per prenderli. Il fatto di trasformarli in numeri (informazioni in una matrice, percentuali) e poi applicare a questi numeri criteri di calcolo fa parte di un processo di semplificazione che ha effetti molto potenti sul

nostro modo di operare con le informazioni, ma che presenta pure svantaggi notevoli.

In generale, quale che sia il caso problematico che stiamo osservando nella scelta dei fatti, siamo costretti a operare semplificazioni. La realtà è così complessa che il tentativo di osservarne tutte le componenti si trasforma in un "rumore assordante", non possiamo che scegliere ed essere consapevoli che le informazioni che abbiamo rilevato sono frutto di una scelta.

Di volta in volta l'ampliarsi della potenza dei nostri strumenti di osservazione ci fa cadere nell'illusione di poter controllare l'intero della realtà che vogliamo studiare. In realtà si tratta solo di ritagli. I fatti del caso sono ritagli di realtà più complesse che esistono in un insieme di relazioni. A questi ritagli attribuiamo delle proprietà che ricaviamo dall'osservazione delle loro relazioni con noi, con i nostri strumenti e con altri elementi della realtà stessa.

3.9 Le variabili

Nel linguaggio della ricerca utilizziamo il termine "variabile" per riferirci a elementi di realtà osservabili che possiamo classificare in base alle loro relazioni con altri elementi simili e misurare in base a nostre convenzioni di misura. Ad esempio, possiamo distinguere gli esseri umani in base alla conformazione fisica in "maschi" e "femmine" (il che non ha niente a che vedere su come percepisce la propria identità di genere la persona nel momento in cui l'abbiamo classificata e se manterrà la stessa identità in un momento successivo) o classificare gli studenti in base al titolo di studio conseguito, o al punteggio riportato in una prova.

Dunque, possiamo applicare il concetto di variabile alla maggior parte degli elementi che entrano in relazione con noi.

In altri termini, per studiare empiricamente un fenomeno dobbiamo eliminare una parte della sua complessità e ridefinirlo scegliendo alcune variabili che riteniamo abbiano un ruolo importante nella caratterizzazione del fenomeno in esame. L'identificazione delle variabili richiede un forte collegamento con le teorie che sono sottese alla ricerca. Le teorie sono costrutti ipotetici, le variabili sono il riferimento materiale di questi costrutti.

In genere distinguiamo le variabili in relazione alla loro natura e al tipo di operazioni che consentono di realizzare con le seguenti definizioni:

Variabili fisiche e non fisiche: consideriamo fisiche tutte quelle variabili di cui è possibile una osservazione diretta. Ad esempio, altezza, peso, titolo di studio, e così via; mentre consideriamo non fisiche tutte quelle variabili che si riferiscono a concetti non direttamente osservabili, quali la motivazione ad apprendere, l'intelligenza, gli atteggiamenti.

Variabili quantitative e qualitative: le variabili quantitative variano in grandezza, mentre quelle qualitative variano in genere. Si tratta di una distinzione che è facile in teoria, ma che presenta difficoltà nell'applicazione a casi particolari. Un esempio di variabile quantitativa è la velocità di esecuzione di una prova, il numero di libri di una biblioteca, il punteggio ottenuto in un test. Sono esempi di variabili qualitative il girare a destra o a sinistra in un labirinto, l'appartenenza a una associazione giovanile o a un sindacato, il tipo di corso di studi seguito o di lavoro svolto, il genere.

Variabili continue e discrete: è definita continua qualsiasi variabile la cui misurazione vari senza soluzione di continuità. Esempi di variabili continue sono l'età, il numero di anni di scolarità, il tempo, il reddito. In una scala continua il numero degli intervalli possibili è infinito. Una variabile è definita, invece, discreta quando le sue misurazioni procedono a salti. Per esempio, il numero degli studenti di una classe è una variabile discreta o discontinua, il punteggio ottenuto in una prova rappresenta una variabile discreta. I dati qualitativi sono sempre discreti, mentre i dati quantitativi possono essere continui o discreti.

Variabili indipendenti e dipendenti. Si usa poi distinguere le variabili in base al modo in cui definiamo i loro rapporti nella nostra ricerca, attribuendo il ruolo di variabile dipendente alla variabile che vogliamo spiegare e di variabile indipendente alla spiegazione ipotizzata. In una relazione asimmetrica la variabile che utilizziamo per indurre un mutamento nell'altra è chiamata variabile indipendente. La variabile il cui valore dipende dall'altra, e che assumiamo non possa a sua volta influenzarla è denominata variabile dipendente. In una relazione causale, attribuiamo il ruolo di causa alla variabile indipendente e quello di effetto alla variabile dipendente.

3.10 Del togliere o accettare la complessità

Nel ragionare sui dati è necessario tenere sempre presente che la relazione tra chi fa ricerca e le cose o le persone su cui fa ricerca è una interazione, non un rapporto unidirezionale. Anche se il ricercatore si limita a osservare, la sua osservazione interagisce con ciò che osserva. In questo ci è di lezione quanto abbiamo appreso dalla fisica quantistica. Come afferma Rovelli (2020) «dobbiamo partire dalla semplice constatazione che lo scienziato, come il suo strumento di misura, sono anch'essi parte della natura. Quello che la teoria dei quanti descrive è il modo in cui la natura si manifesta a un'altra parte della natura». Secondo questa interpretazione la *teoria dei quanti* descrive come qualunque oggetto fisico si manifesti a un qualunque altro oggetto fisico. I vecchi modelli della conoscenza ci hanno abituato a pensare una realtà formata da oggetti, ciascuno con proprie caratteristiche permanenti, ma in realtà questi "oggetti" non fanno che agire continuamente uno con l'altro. Per comprendere dobbiamo concentrarci su queste interazioni, sulla espe-

rienza, e non sulla natura degli oggetti isolati. Ne consegue l'impossibilità di separare le caratteristiche di un osservabile (oggetto) dalle interazioni dove le sue proprietà si manifestano, e dagli oggetti (osservatori e/o strumenti) a cui si manifestano. In questa prospettiva l'attribuzione di una proprietà a uno studente, a un elaborato, al di fuori della interazione in cui la manifesta, può essere fuorviante. Così un punteggio a un test o un voto acquistano significato solo all'interno delle interazioni che li hanno determinati. Evocare la teoria dei quanti potrebbe sembrare eccessivo per spiegare dimensioni che da educatori sperimentiamo continuamente verificando che lo stesso studente non è lo stesso studente nell'interazione con due insegnanti diversi, e che lo stesso compito scritto può provocare giudizi diversi, anche opposti, in due diversi correttori. Per semplificarci la vita spesso preferiamo non tenere conto di questa complessità, operiamo delle riduzioni, ma la ricerca di semplificazioni non sempre porta a una conoscenza corretta delle cose.

Così dobbiamo sempre ricordare che le variabili che rileviamo non descrivono le cose: descrivono il modo in cui le cose si manifestano le une alle altre e che queste manifestazioni non sono permanenti, sono eventi istantanei. «A volte questi eventi formano catene che suscitano l'illusione di essere permanenti, ma solo in particolari circostanze e solo per un periodo di tempo estremamente breve in ciascun caso singolo» (Rovelli 2020). Queste considerazioni debbono dunque spingerci a una grande prudenza, sia nella fase in cui cerchiamo di rilevare informazioni sia nel loro uso per formulare spiegazioni e previsioni.

3.10 Considerazioni conclusive

Sia che le tecnologie vengano utilizzate come supporto trasmissivo per le informazioni, sia che vengano utilizzate come strumenti di telepresenza, sia che vengano utilizzate come strumenti di costruzione collaborativa, l'efficacia della didattica messa in atto dipende da come l'insegnante è in grado di organizzare un buon ambiente di apprendimento, in grado di combinare sinergicamente didattica e valutazione.

Riepilogo

In questo capitolo muovendo dalle fasi che sviluppano una ricerca abbiamo esaminato i punti di forza e i limiti dell'approccio scientifico tradizionale e i problemi che si propongono nell'applicarlo alla ricerca educativa. In particolare, abbiamo esaminato alcune caratteristiche del metodo scientifico e il modo in cui anche le scoperte che avvengono in modo casuale debbano poi essere rigorosamente verificate per poterne poi generalizzare la portata. Ci siamo soffermati poi a esaminare

il modo in cui costruiamo la nostra conoscenza della realtà attraverso il linguaggio, e i punti di forza e i limiti dei nostri strumenti, le parole, i concetti, le proposizioni, e come da tutto ciò procediamo alla costruzione di leggi e teorie.

Infine, ci siamo occupati della complessità che attiene al nostro rapporto con la realtà che a un tempo cerchiamo di osservare e di conoscere ma di cui allo stesso tempo facciamo parte. In questa prospettiva abbiamo esaminato come i fatti relativi ai problemi educativi che affrontiamo subiscano delle riduzioni al momento in cui vengono trasformati in dati per le nostre elaborazioni.

In sintesi, abbiamo visto come la ricerca richieda delle riduzioni e delle semplificazioni, ma come il ricercatore non debba mai rinunciare ad affrontare la complessità delle esperienze educative.

Riferimenti bibliografici

Bailey K.D. (1982). *Methods of social research.* New York: The Free Press (trad. it. *Metodi della ricerca sociale.* Bologna: Il Mulino, 1995.)

Dewey J. (1933). *How we think.* Boston: Heath (trad. it. *Come pensiamo.* Firenze: La Nuova Italia, 1973).

Dewey J. (1938). *Logic, the theory of inquiry.* New York: Henry Holt (trad. it. *Logica, teoria dell'indagine.* Torino: Einaudi, 1974).

Harari N.Y. (2012). *From animals into Gods: a brief history of humankind.* Scott Valley (Ca, US): CreateSpace Independent Publishing Platform (trad. it. *Sapiens: da animali a dei.* Milano: Bompiani, 2017.)

McBurney D.H. (1983). *Experimental psychology.* Belmont Wadsworth Publishing Company (trad. it. *Metodologia della ricerca in psicologia.* Bologna: Il Mulino, 1986.)

Merton R.K., Barber E. (2002). *Viaggi e avventure della serendipity. Saggio di semantica sociologica e sociologia della scienza.* Bologna: Mulino.

OECD (2019). *Pisa 2018. Assessment and analytical framework.* Paris: OECD Publishing.

Rovelli C. (2020). *Helgoland.* Milano: Adelphi.

Russell B. (1931). *The scientific outlook.* London: Allen & Unwin (trad. it. *La visione scientifica del mondo.* Roma-Bari: Laterza, 1988).

Weber M. (1958). *Il metodo delle scienze storico-sociali.* Torino: Giulio Einaudi.

4

La misura in educazione

In questo capitolo viene presentato il concetto di misura partendo dalla necessità, via via sempre più avvertita anche nel campo delle scienze dell'uomo, di porre in relazione aspetti qualitativi e quantitativi della propria realtà di studio. Con l'accrescersi delle conoscenze diverrà utile, anche a tale scopo, introdurre misure quanto più efficaci ed adeguate alle nuove necessità: la teoria risulterà sempre più indispensabile per la costruzione di modelli d'indagine e di misurazione; la misurazione si rivelerà particolarmente necessaria per la verifica della teoria di riferimento. Vengono approfondite, perciò, nel testo, le conoscenze necessarie all'avvio delle procedure di misurazione in ambito educativo; precisati i concetti di osservazione e classificazione dei dati qualitativi e quantitativi, dimensioni misurabili attraverso scale nominali, ordinali o ad intervalli. Ci si sofferma, inoltre, sulle operazioni consentite rispetto alla natura delle variabili allo studio, sul calcolo delle principali misure di tendenza centrale e di dispersione dei fenomeni, sulle ragioni che rendono necessario il campionamento e l'impiego delle principali statistiche descrittive e inferenziali, parametriche e non parametriche, compresa quella di rilevare la specifica distribuzione dei dati relativi ai fenomeni indagati, e le loro caratteristiche peculiari.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Classificazione e misurazione di aspetti significativi di un dato fenomeno.
- Rapporti e relazioni tra qualità, quantità, dato, informazione, interpretazione.
- Caratteristiche distintive delle scale di misura nominale, ordinale e a intervalli, e operazioni formali con esse consentite.
- Significato quali-quantitativo delle principali misure di tendenza centrale e di variabilità dei punteggi e/o dei fenomeni.
- Calcolo e significato dei principali punti standard e di alcune statistiche descrittive e inferenziali, parametriche e non-parametriche.

Il signor Palomar ora cerca di limitare il suo campo d'osservazione; se egli tiene presente un quadrato diciamo di dieci metri di riva per dieci metri di mare, può completare un inventario di tutti i movimenti d'onde che vi si ripetono con varia frequenza entro un dato intervallo di tempo. La difficoltà è fissare i confini di questo quadrato, perché se per esempio lui considera come lato più distante da sé la linea rilevata d'un'onda che avanza, questa linea avvicinandosi a lui e innalzandosi nasconde ai suoi occhi tutto ciò che sta dietro; ed ecco che lo spazio preso in esame si ribalta e nello stesso tempo si schiaccia.
Comunque il signor Palomar non si perde d'animo e a ogni momento crede d'esser riuscito a vedere tutto quel che poteva vedere dal suo punto d'osservazione, ma poi salta fuori sempre qualcosa di cui non aveva tenuto conto..."

Italo Calvino, *Palomar*

4.1 Osservazione e classificazione: i primi passi di ogni scienza

Tra i primi passi che ogni scienza compie nel suo costituirsi come complesso di conoscenze ordinate, coerenti e affidabili, dotato di proprie procedure e strumenti di indagine, assumono particolare rilievo l'*osservazione* e la *classificazione* delle caratteristiche dei fenomeni studiati. In una prima fase evolutiva l'osservazione ha lo scopo di cogliere le peculiarità con cui si manifesta il fenomeno indagato per descriverlo verbalmente nel modo più esatto possibile, e per classificarlo principalmente secondo criteri di omogeneità. Questa classificazione assume, seppur a un primo livello di complessità, i requisiti della misurazione.

La *classificazione*, in senso stretto, consiste infatti nel raggruppare diversi elementi in funzione di una determinata caratteristica in modo da formare *classi* che siano il più omogenee possibile rispetto alla proprietà in rapporto alla quale gli elementi sono stati inclusi in ciascuna classe. Via via che le conoscenze si estendono e le procedure metodologiche si affinano, si accresce anche la necessità di introdurre misure sempre più efficaci e adeguate alle nuove necessità, e viceversa: la teoria è indispensabile per la costruzione di modelli d'indagine e di misurazione; la misurazione è necessaria anche per la verifica della teoria.

4.2 Qualitativo e quantitativo: un nesso assai stretto

Nel campo delle cosiddette scienze umane, meglio definibili scienze dell'uomo, per diverse ragioni le relazioni tra *qualitativo e quantitativo* sono state per lungo tempo interpretate come espressione di una insanabile opposizione e contraddizione. Tale interpretazione, sebbene sia stata superata nell'ambito della ricerca, permane ancora purtroppo nelle convinzioni e negli atteggiamenti di molti che di quelle stesse scienze si occupano professionalmente, mentre è facile comprendere invece che tra i due termini sussiste un nesso assai stretto persino a livello di linguaggio comune: in molti casi le qualità espresse in forme di aggettivi possono essere ordinate per gradi attraverso l'impiego di taluni avverbi, così contribuendo alla costruzione di uno spazio per alcuni versi definibile già come quantitativo. Ma anche nelle pur complesse scienze sociali ed educative, dove la misurazione è stata introdotta tardivamente, essa ha raggiunto, come vedremo, avanzati livelli di sofisticatezza che hanno dato un contributo non secondario all'accrescimento dei quadri concettuali e teorici di quelle discipline.

4.3 Strumenti concettuali per comprendere e controllare gli eventi

Chiunque abbia esperienza d'insegnamento sa bene come nella pratica educativa, ma anche a maggior ragione nello svolgimento di attività di ricerca, si manifesti di frequente la necessità di dover *rilevare, trattare, analizzare e interpretare* dati e informazioni su diversi fattori che si suppone abbiano un qualche peso nell'ostacolarne o nel facilitarne la riuscita. Si tratta spesso di dover operare per meglio comprendere e controllare certi eventi, per verificare ipotesi già formulate o per formularne di esplicative dei risultati comunque registrati - positivi o negativi che siano - per compiere valutazioni sulla cui base si assumeranno decisioni, talvolta teoriche, più frequentemente operative. Le informazioni che però si possono trarre dallo svolgimento di queste attività sono assai numerose.

Tuttavia la ricchezza delle informazioni che ogni situazione educativa (pur determinata e circoscritta) produce, può non essere colta se non si posseggono gli strumenti concettuali e operativi necessari per poter rilevare e dare organizzazione agli elementi con cui la varietà dei fatti si manifesta, o se, in altri termini, le informazioni potenziali non vengono fatte affiorare a livello di consapevolezza. A tale scopo occorre disporre, tra l'altro, di un modello interpretativo del fenomeno (anche se aperto a elementi imprevisti) e di strumenti e procedure di rilevazione, misurazione e analisi dei dati più significativi. Anche quando ci si trovi di fronte a un fatto relativamente semplice da studiare e si impieghi il più elementare livello procedurale d'indagine, vi è la necessità di dover risolvere almeno due problemi:

- assumere tra gli ipotetici infiniti dati relativi al fenomeno, quelli congruenti con il modello interpretativo adottato (salvo poi a scoprirne, magari, la scarsa o nulla connessione e significatività: ma anche questa evidenza costituisce un arricchimento del quadro conoscitivo di partenza);
- rilevare i dati seguendo procedure formali in grado di renderli espressivi di grandezze tra loro commensurabili.

4.4 Condizioni per misurare

Ponendo in secondo piano le delicate questioni dei modelli interpretativi, è necessario innanzitutto garantire l'omogeneità in termini formali dei diversi elementi informativi sulla cui base dovranno compiersi le operazioni di analisi, confronto e interpretazione dei fenomeni. Ci sono delle condizioni affinché si determini tale situazione:

- una *prima condizione* è data dalla capacità di saper scegliere e impiegare criteri, scale e strumenti di misura adeguati sia agli scopi dell'indagine, sia alle circostanze, al campo e all'oggetto specifici dell'intervento, nonché alle caratteristiche con cui si manifesta una determinata variabile;
- una *seconda condizione* è la conoscenza delle operazioni che ogni particolare tipo di misurazione consente formalmente di compiere senza però che con il loro svolgimento si alterino la rappresentatività e la significatività delle informazioni e dei dati trattati.

In sintesi estrema, infatti, misurare significa usare una grandezza fissata arbitrariamente come misura campione unitaria, per poter determinare in quale rapporto un'altra grandezza, da misurare, la contiene o è da essa contenuta.

Saper compiere *misurazioni valide* e *attendibili* e con il più alto grado di precisione reso volta a volta disponibile/possibile, saper allestire strumenti di misurazione, saperli impiegare e sapere quali operazioni sono consentite e quali non sono lecite con i dati rilevati, sono condizioni basilari che costituiscono il fondamento non solo di ogni procedura di ricerca, ma anche di una consapevole conduzione delle attività formative, in particolare delle procedure valutative.

Approfondimento 4.1
COSA SIGNIFICA MISURARE

Misurare è un'operazione convenzionale che consiste nell'associare un numero o un simbolo a un oggetto qualitativamente definito, o a una proprietà dell'oggetto, secondo regole determinate, in modo che a quella stessa proprietà si possano attribuire alcune caratteristiche dei numeri che così vengono a rappresentarla. Attraverso un'operazione di confronto, da condurre ovviamente secondo le regole codificate, dell'oggetto (o di una sua proprietà) da misurare con altro analogo scelto come riferimento (o come misura campione) diviene possibile associare al primo un numero (o un simbolo) che ne esprime, appunto, la misura.

Come si può notare, si tratta di una definizione operativa, che indica perciò con esattezza:

- la necessità di definire preliminarmente, e precisamente, le proprietà o le caratteristiche di un oggetto;
- le regole di associazione e di relazione di queste con un numero, nonché i procedimenti da impiegare formalmente per classificare certi eventi, oggetti e/o loro proprietà, in modo non ambiguo.

4.5 Scale e tecniche di misurazione

Una delle più importanti funzioni della misurazione è quella di rappresentare in modo tendenzialmente univoco una determinata realtà o un certo concetto di essa - qualunque sia il suo grado di astrazione - per consentire, su aspetti che la riguardano, una comunicazione intersoggettiva non ambigua (che faciliti, deweyanamente, un "accordo per l'azione") o comunque tale da permettere il confronto delle interpretazioni e la verifica di particolari proposizioni e asserzioni. Ne deriva che è di grande importanza impiegare correttamente le *procedure di trasformazione delle differenze d'ordine qualitativo in differenze quantitative*, ovvero quelle regole di misurazione rigorosamente codificate attraverso cui si compie l'associazione di un numero a un oggetto o, più precisamente, a certe sue determinate qualità e caratteristiche.

A seconda delle regole seguite, che derivano dalle proprietà dei numeri che si associano alle qualità dell'oggetto, si potranno avere quattro livelli di misurazione a cui corrispondono altrettante *scale di misura*: *nominale, ordinale, ad intervallo* e *di rapporto* (quest'ultima generalmente non impiegabile in ambito educativo in quanto in tale scala lo 0 non ha come nella scala a intervalli un valore convenzionale, ma indica assenza di misura), che posseggono caratteristiche formali cumulative. Ognuna di esse, a partire da quella ordinale, possiede infatti le proprietà di cui godono le scale precedenti, più altre qualità che la caratterizzano peculiarmente.

4.5.1 La scala nominale

Poniamo che un docente dopo aver trattato una certa questione concettuale desideri verificare se gli allievi la padroneggino o meno. Sulla base degli esiti delle sollecitazioni offerte (prove tradizionali, oggettive, semistrutturate ecc.) e distinguendo la qualità delle prestazioni in due categorie - prestazioni accettabili e prestazioni non accettabili - può collocare ogni studente in ciascuna delle due

"classi" così nominate, pervenendo a un quadro informativo utilizzabile per eventuali interventi di feed-back tempestivo. La scala di misura dell'apprendimento in tal modo impiegata è la scala nominale.

Una scala, dunque, che permette di collocare le prestazioni in due distinte classi omogenee, semplicemente in funzione dell'apparire o non apparire negli allievi della specifica qualità che si è voluta "misurare".

In campo educativo molte delle misurazioni che si compiono sono di tipo nominale, sia quando interessino qualità o variabili di fatto discontinue (del tipo: maschio-femmina), sia - soprattutto - quando per comodità interpretativa si rendono volutamente discontinue o discrete (attraverso opportune graduazioni) alcune variabili, e continue o indiscrete altre (quali l'apprendimento per esempio, ma anche caratteristiche che rimandano a categorie rilevabili attraverso l'osservazione sistematica o l'impiego di interviste, questionari e strumenti analoghi).

L'operazione che si compie è di vera e propria classificazione delle proprietà rilevate: si raggruppano in classi omogenee diversi elementi o variabili di un evento o di un processo, per il loro solo apparire, in funzione di una determinata caratteristica. Ognuna delle classi - che possono ovviamente essere più di due - conterrà elementi dello stesso tipo od omogenei rispetto alla categoria impiegata come criterio di classificazione.

Le operazioni consentite a questo livello di scala sono essenzialmente il calcolo delle frequenze, cioè il numero dei dati (nel nostro caso: prestazioni e/o allievi) presenti in ciascuna della classi in cui abbiamo diviso l'intero universo.

Possono inoltre calcolarsi percentuali e proporzioni (almeno quando i dati di ognuna delle classi siano superiori a venti-trenta unità). Questi ultimi valori sono particolarmente utili quando, sempre usando la scala nominale, si vogliono fare confronti tra classificazioni riferite a popolazioni di diversa consistenza numerica.

A livello di scala nominale i più diffusi calcoli di relazioni (che in questa sede non è necessario illustrare) tra le variabili di un fenomeno sono la distribuzione binomiale e la distribuzione del *chi quadro* (χ^2) che consente di verificare se siano o no significative le differenze delle frequenze riscontrate in due o più campioni.

Approfondimento 4.2

COME AUMENTARE IL GRADO DI PRECISIONE DELLA SCALA NOMINALE

Anche se per alcuni versi la scala nominale viene ritenuta grossolana, non quantitativa in senso stretto, poiché non consentirebbe la registrazione delle possibili gradazioni del possesso di una specifica qualità degli elementi inclusi in una stessa classe, è anche vero che restringendo l'ampiezza delle qualità rilevate, soprattutto nelle variabili continue, o determinando più classi, si può aumentare il grado di precisione delle operazioni di misura.

Impiegando questa scala, l'apprezzamento di molti comportamenti e atteggiamenti degli allievi (che viene spesso compiuto dagli insegnanti con i voti in decimi o con il linguaggio verbale scritto) può infatti diventare più preciso, persino univocamente determinabile e interpretabile. Facciamo un esempio. Per ciascuna delle pur estese qualità (conoscenze, abilità, capacità, competenze) indicate nella prima colonna di un apposito elenco, associamo due o più classi di comportamento, la prima che potremmo indicare con 'A' (comportamento cognitivo che mostra il possesso di quella abilità o competenza) e la seconda con 'B' (comportamento cognitivo che mostra l'assenza, ovvero il non possesso di quella stessa abilità o competenza). Nelle due classi (A e B) collocheremo rispettivamente gli allievi che mostrano di possedere o non possedere la qualità indicata. Avremo compiuto in tal modo una misurazione così precisa da suggerire, essa stessa, a noi e agli allievi, le azioni da intraprendere per consolidare l'apprendimento oppure, se non si è strutturato, per promuoverlo attraverso altre vie (funzione proattiva della valutazione; Domenici, 2001).

La stessa comunicazione intersoggettiva relativa ai risultati registrati tra docenti e allievi, tra docenti e docenti, tra docenti e genitori diviene meno ambigua. Un prezioso impiego della scala nominale si ha con l'uso delle liste di *descrittori dell'apprendimento* (Domenici, 1981).

4.5.2 La scala ordinale

Con le scale ordinali la proprietà dei numeri che empiricamente si impiegano è quella delle relazioni ordinali (non cardinali come erroneamente ma assai diffusamente si crede!) che denotano vere e proprie posizioni di graduatoria. Con esse si può quindi stabilire una relazione di maggioranza, uguaglianza o minoranza nel possesso delle qualità oggetto della misurazione. La distanza tra i diversi punti della graduatoria che così si ricava non è però né fissa né quantitativamente determinabile.

In campo educativo un tipico esempio di scala ordinale sono i voti che si impiegano nella scuola e negli esami universitari. I voti in decimi o in trentesimi, e simili, indicano infatti valori di posizione la cui distanza non è univocamente determinabile, così che essa può derivare, tra l'altro, dalle caratteristiche personali di chi li usa: la distanza che intercorre tra un 6 e un 5, considerata generalmente assai più ampia che tra un 9 e un 8 o tra un 3 e un 2, varia col variare del concetto di sufficienza, 6, posseduto dal singolo docente o strutturatosi in un determinato contesto.

L'unico dato certo rilevabile con la lettura dei punteggi assegnati con una scala ordinale è, infatti, la posizione nell'ordine di graduatoria assunta da un determinato punteggio rispetto a tutti gli altri. Solo con uno specifico riferimento al contesto in cui i punteggi sono stati assegnati (p.e. una data classe scolastica), si può asserire che chi ad esempio abbia conseguito il voto 8, sappia di più e forse meglio di chi invece abbia riportato come voto 7.

Al di fuori della classe scolastica in cui quei voti sono stati assegnati, la loro capacità di offrire informazioni precise, di per sé già bassa, si riduce fortemente: così come per il podista arrivato terzo si può dire che è stato più veloce di chi è arrivato quinto, ma appunto solo in riferimento a quella determinata gara. In un'altra gara (cioè in un contesto differente), con altri concorrenti potrebbe, pur con la stessa velocità, piazzarsi anche ultimo. Un voto 7 in matematica conseguito nel mese di maggio in una terza liceo scientifico potrebbe corrispondere a un 4 assegnato a un altro allievo, di una stessa terza classe, ma di un altro liceo, seppure nello stesso periodo, in riferimento agli stessi argomenti, temi, problemi sottoposti a controllo, e con identica prestazione. Evidentemente per fare confronti decontestualizzati bisognerebbe impiegare una scala diversa da quella ordinale.

Fatta questa precisazione, torniamo a esaminare le principali caratteristiche di questa scala.

Le classi che determinano un certo grado di intensità della qualità presa in considerazione con la scala ordinale espresse attraverso i numeri, vengono definite *ranghi*, e la scala *rating scale*.

La posizione di graduatoria può essere determinata oltre che con i numeri ordinali anche con termini, soprattutto aggettivi o avverbi, che indichino il diverso grado di possesso di una certa qualità. Si possono, così, porre in ordine di graduatoria i generi letterari preferiti da un certo campione di allievi, la disponibilità a compiere certe attività di studio ecc., impiegando una scala ordinale i cui punti siano definiti appunto da aggettivi e avverbi: 'ottimo', 'buono', 'sufficiente', 'moltissimo', 'abbastanza', 'poco', 'per niente', eccetera.

4.5.2.1 Principali operazioni consentite con la scala ordinale. Mediana e punteggi percentili

Quando si impiega la scala ordinale, oltre alle operazioni proprie della scala nominale, se ne possono compiere altre, come il calcolo della *mediana* e degli indici del valore di posizione dei punteggi quali i *percentili* (non è però consentito il calcolo della media aritmetica tra due o più punteggi: in un certo senso sarebbe come pretendere di calcolare la media tra 'poco' e 'pochissimo', e simili).

La mediana è quel valore che in una distribuzione di dati sistemati in ordine crescente o decrescente occupa la posizione centrale: essa rappresenta infatti un valore di posizione. Con un numero pari di dati, si situerà tra i due valori centrali, anche se ovviamente potrebbe corrispondere a essi.

Dato che la mediana esprime il valore centrale di una distribuzione, confrontando i valori assunti dalle mediane di due differenti distribuzioni, magari riferite ai voti assegnati da uno stesso docente o una stessa prova in tempi diversi o a compiti differenti, oppure da docenti diversi, a uno stesso compito in classe, si ricava con immediatezza e pregnanza una indicazione preziosissima sulle più macroscopiche differenze tra i due 'campioni' o dello stesso campione in tempi differenti. Dal confronto delle mediane si comprenderà immediatamente se la metà più uno degli allievi di ciascuno dei campioni si situa sotto o sopra quella che è stata stabilita come sufficienza.

Se si vuole poi rendere esplicito il risultato del confronto tra diverse prestazioni di uno stesso allievo o di allievi differenti (appartenenti allo stesso gruppo o a gruppi distinti), i punteggi assegnati con una scala ordinale possono essere trasformati in *punteggi percentili*.

Il significato logico dei percentili si basa sul concetto di percentuale, tant'è che permettono di cogliere qual è la posizione di una misura o di una prestazione rispetto all'intera distribuzione cui essa appartiene, fatta uguale a 100. Esprimono infatti la posizione di un punteggio mettendo in evidenza proprio quale percentuale delle prestazioni (comprendendo anche quella considerata) sta al di sotto di esso e quale percentuale sta invece al di sopra. Al 73° percentile, o centile (che convenzionalmente viene indicato con P73), corrisponderà quindi quel punteggio della graduatoria che sta al di sopra del 73% di tutti i punteggi ed è superato (comprendendo se stesso) dal 27%; analogamente il 40° percentile (P40) sarà dato invece dal punteggio che supera (incluso se stesso) il 40% delle prestazioni poste in graduatoria, e sta al di sotto del rimanente 60%, e così via.

Ovviamente il 50° centile dividendo in due la distribuzione, rappresenta la mediana.

4.5.3 La scala a intervalli

Nella scala a intervalli la distanza tra due punti consecutivi è costante. L'intervallo viene così a rappresentare un preciso indicatore quantitativo, non solo qualitativo, della differenza o distanza tra i suoi due estremi. Il punto zero della scala assume però valore convenzionale, come nella misura delle temperature con la scala centigrada, non già, come si sarebbe portati a pensare intuitivamente, valore nullo cioè assenza totale, assoluta, della qualità e/o quantità misurata.

Ma vediamo alcuni esempi di impiego della scala a intervalli in campo educativo e alcune forme di rappresentazione grafica dei dati con essa rilevati.

Nelle prove oggettive di profitto il peso (o valore in termini di punteggio) assegnato a ciascun quesito consente di determinare a priori il punteggio massimo conseguibile qualora tutte le risposte risultino esatte. Tale punteggio massimo teorico costituisce l'ampiezza della scala che viene impiegata per misurare le prestazioni rilevate con quello strumento. Dato che per ogni risposta esatta viene assegnato a priori un punteggio ben definito, e che quindi le differenze delle diverse

prestazioni sono determinate univocamente dagli scarti di punteggio, la scala a intervalli può essere impiegata a tal fine in modo pertinente. Con essa, oltre a poter considerare allo stesso modo gli elementi (prestazioni e/o allievi o altre variabili) che cadono nello stesso intervallo (come per la scala nominale) e la differenza d'intensità del possesso della caratteristica misurata tra due diversi elementi o punteggi, ovvero la posizione d'ordine di diverse prestazioni (come è possibile con la scala ordinale), diventa anche determinabile, sempre con una certa approssimazione, l'ampiezza del rapporto di tale intensità.

Come si può notare, nel *sistema empirico* messo a punto, una volta fissato arbitrariamente il valore numerico da associare come misura a una data caratteristica (che anche se *continua* possiamo in questo caso considerare *discontinua*), e utilizzandolo come unità di misura, diviene agevole individuare il valore di tutti gli elementi presi in esame.

4.5.3.1 Principali operazioni consentite con la scala a intervalli (media aritmetica, indici di variabilità o dispersione dei punteggi, correlazioni)

Con misure compiute a livello di scala a intervalli si possono fare tutte le operazioni indicate (e non indicate, ma formalmente previste) per le scale precedenti, più il calcolo della media aritmetica, di alcuni indici di variabilità o dispersione dei punteggi, di molte correlazioni. Si possono infine impiegare numerosi altri test statistici di analisi e interpretazione dei dati.

Com'è facile immaginare, è altissima la quantità delle operazioni, dei test di analisi, descrizione e inferenza statistica impiegabili a livello di scala ad intervalli.

Dovendosi così trattare delle operazioni che possono compiersi con le scale di misura a intervalli, si privilegerà la dimensione "apprendimento degli allievi", e quindi degli strumenti (in particolare delle prove oggettive di profitto e di quelle semistrutturate) nonché delle procedure utili per poter misurare, conoscere, analizzare e interpretare quella dimensione.

Da un punto di vista metrico-formale le questioni non sono dissimili se le rilevazioni si riferiscono a variabili diverse dall'apprendimento. Il vincolo formale è dato solo dall'impiego della scala a intervalli che può ben essere usata per misurazioni e valutazioni relative ad atteggiamenti, a variabili sociali, culturali, economiche e altro ancora, o riguardare elementi costitutivi o aspetti organizzativi del sistema formativo e delle singole unità scolastiche.

Per compiere l'analisi e l'interpretazione dei risultati della somministrazione di una prova di verifica dell'apprendimento si possono seguire, come vedremo meglio nel Capitolo 10, due vie diverse:

- quella che fa valutare le prestazioni in rapporto a uno standard predefinito, a una soglia di riferimento (*criterio assoluto*);
- quella che viceversa fa apprezzare gli esiti complessivi e quelli individuali in rapporto all'andamento generale della prova (*criterio relativo* o *normativo* dal riferimento alla curva normale di distribuzione dei punteggi).

Nel primo caso, si ritengono accettabili le prestazioni (gli apprendimenti e simili) il cui punteggio cade sopra la soglia; non accettabili le altre.

Nel secondo caso la lettura può spingersi ben oltre quella dicotomica appena accennata. Ciò richiede una trattazione statistica, per quanto semplicissima, dei dati assunti con gli strumenti valutativi impiegati (prove oggettive o strutturate e alcune prove semistrutturate) oppure con altri strumenti di rilevazione dei dati relativi a qualsiasi altra variabile, che permettano, ovviamente, l'uso della scala a intervalli.

In questa sede si può solo dire che tra le informazioni, i dati, le misure che interessano generalmente gli attori principali dei processi formativi formali (docenti e allievi, ma anche gli altri decisori interessati ai fatti formativi, genitori, capi d'istituto, politici, opinione pubblica), vi sono quelli che danno conto del fenomeno indagato, in termini sia complessivi sia analitici.

In una scuola, o in una classe, interessa sapere, per esempio, come procede l'apprendimento complessivo di tutta la popolazione studentesca (misura di tendenza centrale), ma anche, parallelamente e contestualmente, quello dei singoli, nonché delle relazioni reciproche, ovvero di come si pone l'apprendimento di ognuno rispetto a quello di tutti (misure di variabilità o dispersione di punteggi).

Le sole misure di tendenza centrale e di variabilità del fenomeno indagato, per esempio dell'apprendimento, calcolabili quando si impieghino strumenti per rilevare dati valutativi con scale a intervalli, aiutano, come vedremo, più di mille altri mezzi a conoscere la situazione e a strutturare decisioni davvero pertinenti ed efficaci.

Possiamo ora dire che, come sappiamo, la "media" corrisponde al rapporto tra la somma dei dati numerici e il numero dei dati stessi, la "mediana" corrisponde al valore centrale tra i dati numerici considerati e ordinati in modo crescente o decrescente, e la moda o le mode al punteggio o ai punteggi più ricorrenti. La misura di variabilità o dispersione dei punteggi è invece data da un valore che esprime quanto i punteggi si discostano mediamente dalla media aritmetica degli stessi.

4.6 Misure di tendenza centrale e misure di variabilità dei punteggi

In campo educativo, la *scala di misura quantitativa* per eccellenza è, come già abbiamo potuto notare, quella *a intervalli regolari*. Una volta adottato un qualunque strumento di rilevazione dei dati relativi alle variabili di una ricerca che impieghi una scala a intervalli si pensi a una prova oggettiva di profitto, o a taluni questionari o, ancora, a griglie per l'osservazione sistematica e simili, con i dati assunti si possono compiere tutte le operazioni di trattamento e analisi dei dati consentiti da questa scala quantitativa. Operazioni che possono essere assai sofisticate e perciò capaci di farci cogliere molte informazioni sia sugli aspetti qualitativi sia su quelli quantitativi del fenomeno indagato e delle variabili studiate. Dal momento che la ricerca educativa, sperimentale e non, quantitativa e qualitativa, direttamente o indirettamente finisce con l'avere come obiettivo esplicito o implicito il miglioramento soprattutto della capacità di apprendimento di ogni soggetto o allievo preso in considerazione e di gruppi o universi di soggetti omologhi, nel dar conto delle operazioni fondamentali da compiersi quando si usino tali scale si è scelto come riferimento esemplificativo l'uso di una prova oggettiva di profitto (si veda il Capitolo 5). Per analogia le stesse operazioni possono/devono compiersi quando gli strumenti adottati nella ricerca siano strutturati a livello di scala a intervalli.

Così, dopo aver calcolato i punteggi, se si dispongono gli stessi in ordine crescente (o decrescente), attraverso il semplice calcolo della *media aritmetica*, della *mediana* e della *moda* la capacità di "leggere" i risultati della somministrazione di una prova aumenta: con questa ulteriore operazione si è in grado infatti di sapere anche qual è la tipicità complessiva dei risultati conseguiti dal gruppo di allievi considerato nel suo insieme.

Inoltre, se si vuole conoscere in modo sintetico la variabilità delle prestazioni, ovvero la loro dispersione o il loro addensamento attorno al valore medio (evitando di fare improbabili confronti punteggio per punteggio con i valori centrali della distribuzione), allora è necessario calcolare la *deviazione standard*, che tra l'altro consente la trasformazione dei punteggi grezzi in punteggi standardizzati, i soli che permettano di compiere confronti precisi tra prestazioni di gruppi diversi di allievi relativi a una stessa prova o a prove diverse.

4.6.1 Media, mediana, moda

Quando si dispone di molti dati relativi a un evento o a un fenomeno (nel nostro esempio i dati sono costituiti dai punteggi grezzi riportati dagli allievi a una prova oggettiva di profitto), si può fare ricorso ad alcune misure - prima accennate - di tipicità del fenomeno per conoscere sinteticamente il suo andamento complessivo.

Anche quando si considerino solo punteggi conseguiti da un ristretto numero di allievi quali quello di una classe, se non si ricorre a tali misure vi sono reali difficoltà nella interpretazione complessiva del fenomeno apprendimento, nel senso che sarà difficile esprimere con poche misure la situazione generale della classe per trarne informazioni certo globali o generali, ma non per questo poco significative, circa l'accettabilità o meno dei processi attivati e dei risultati formativi raggiunti.

Quando poi il numero dei soggetti aumenta, le complicazioni crescono: i semplici dati (o punteggi) grezzi potranno farci vedere, per così dire, l'albero non già la foresta. Ebbene, per conoscere le caratteristiche generali del fenomeno indagato bisogna ricorrere alle misure di tipicità. Ve ne sono diverse, ciascuna con caratteristiche peculiari. La loro scelta deriverà primariamente dalla qualità delle informazioni che sull'evento si vogliono assumere.

La più conosciuta e diffusa di queste misure è la *media aritmetica*. Essa è data dal quoziente della somma dei valori di ciascun dato e il numero totale dei dati, così come indicato nella seguente formula:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

che si può scrivere anche:

$$\bar{X} = \frac{\sum X_i}{n}$$

dove: $\overline{X}$ = media aritmetica; $x_1 \ldots x_n$ = serie dei dati; n = numero totale dei dati.

La media si calcola quindi compiendo la sommatoria (Σ) di tutti i valori dei dati (x_i), per i che va da *1* a n, e dividendo il risultato per il numero totale dei dati (n). Considerando i seguenti punteggi elencati in ordine crescente:

30; 31; 31; 35; 37; 38; 40; 40; 41; 41; 41; 41; 42;
46; 46; 47; 49; 50; 51; 52; 52; 53; 55; 55; 56

la media risulta uguale a 44, ossia alla somma aritmetica dei singoli punteggi (che è 1.100) divisa per il numero complessivo dei punteggi (n, che è 25).

Altra misura di tipicità è la *mediana* (Md). Essa è costituita dal valore del punteggio che occupa la posizione centrale di una distribuzione sistemata sempre in ordine crescente o decrescente.

Nella distribuzione sopra riportata, ordinata in modo crescente, essendo dispari il numero dei dati (n = 25), la mediana, cioè il punteggio che occupa la posizione di mezzo e che ha ovviamente alla sua sinistra e alla sua destra un ugual numero di dati, è 42.

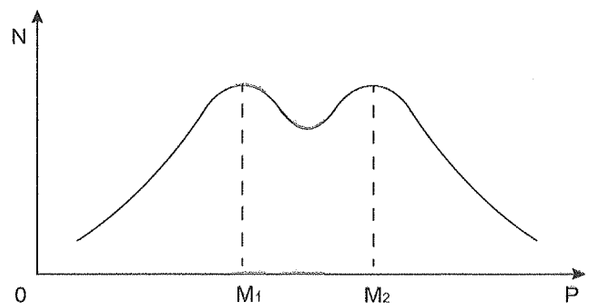
Quando il numero dei dati di una distribuzione è pari, la mediana può non corrispondere a un valore reale della distribuzione bensì alla media aritmetica dei due valori centrali, così come si indica nel seguente esempio:

2; 4; 6; 7 $n = 4$ (numero pari);
la mediana è data dalla media dei due valori centrali, 4 e 6, sarà perciò = 5.

L'ultima delle misure di tipicità qui considerate è la *moda* (M), che esprime il valore che in una distribuzione ricorre con maggiore frequenza, cioè il maggior numero di volte. Nella serie di 25 punteggi prima riportata, la moda è 41. È questo infatti il punteggio che ricorre (quattro volte) più di ogni altro. Naturalmente in una distribuzione si possono riscontrare più valori modali. Nella seguente serie di dati:

3; 4; 4; 4; 6; 7; 7; 9; 9; 9; 10

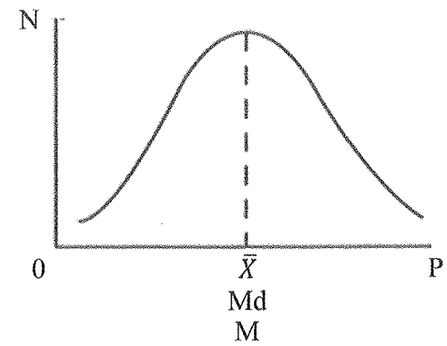
si hanno due mode: 4 e 9. In questo caso la distribuzione è detta *bimodale*:



La moda è di grande utilità soprattutto quando i dati della distribuzione sono molto numerosi. In tali casi però più che la moda è opportuno individuare la classe di frequenza in cui i dati compaiono in maggior numero: ciò presuppone, ovviamente, il raggruppamento dei dati in classi aventi intervalli omogenei.

Il *confronto tra media, mediana e moda*, in particolare quello tra le prime due misure, assume grande importanza per il rilievo delle informazioni sul fenomeno indagato che da tale confronto si può ricavare.

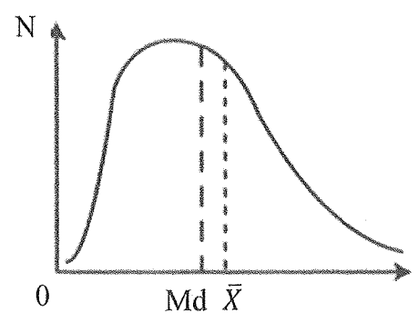
Soprattutto quando i dati sono abbastanza numerosi, se le tre misure coincidono (o tendono allo stesso valore), la distribuzione dei dati è detta *normale* (o tendenzialmente normale) oppure *casuale*, ed è rappresentata dalla curva qui disegnata:

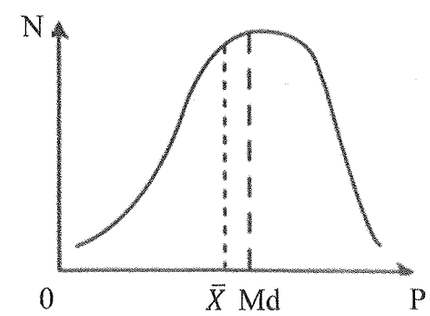


Semplificando, si può dire che quando si verifica questo caso il numero e la distanza dei punteggi situati alla sinistra (punteggi bassi) e di quelli situati alla destra (punteggi alti) dai valori centrali, sono tendenzialmente simmetrici. Se i dati della distribuzione sono costituiti dai punteggi riportati in un test di profitto, si può affermare con estrema sintesi, che gli studenti sono divisi in tre gruppi: uno abbastanza consistente situato attorno ai valori centrali; gli altri due più ristretti attestati rispettivamente sui valori bassi e sui valori alti dei punteggi.

Per l'interpretazione dei dati e del fenomeno che con questi semplici calcoli si vogliono analizzare, riveste particolare significato il *confronto tra mediana e media*.

Può accadere, oltre al caso precedentemente esaminato, che la mediana assuma un valore inferiore a quello della media, oppure un valore superiore. Le curve riprodotte di seguito rappresentano rispettivamente questi due casi.





Se le distribuzioni sono il risultato di due prove di verifica dell'apprendimento, nel primo caso (mediana < media) si può dire che oltre la metà degli allievi ha conseguito punteggi bassi o comunque inferiori al valore medio complessivo. Se infatti la mediana (che divide in due il numero dei soggetti) ha un valore inferiore a quello medio, ciò significa che oltre la metà dei soggetti ha conseguito risultati più bassi della media, il cui valore è stato evidentemente innalzato solo dalle buone pre-

stazioni di pochi allievi.

Come si può ben notare la sola media darebbe in questa circostanza una informazione "tendenziosa" sulla situazione indagata: un suo valore, che di per sé potrebbe anche essere accettabile, nasconderebbe i risultati della maggioranza dei soggetti di fatto inferiori alla media.

Didatticamente ciò potrebbe significare accettazione dei risultati complessivi dell'apprendimento e di conseguenza della decisione di inopportunità di interventi di modificazione significativa del processo formativo.

Ma la stessa curva può esprimere non già ciò che si è appena detto, bensì un alto livello di difficoltà della prova. Un'attenta analisi degli item e dell'intera prova in rapporto agli obiettivi di verifica dell'apprendimento, darebbe conto della giustezza o meno di questa interpretazione.

Nel secondo caso (mediana > media) le indicazioni da trarre sarebbero esattamente opposte a quelle indicate per la prima delle due distribuzioni. È evidente che quando gli strumenti di verifica sono davvero rispondenti alla situazione in cui si opera, la seconda curva rappresenta un risultato del processo di insegnamento-apprendimento certamente più auspicabile ed accettabile di quello espresso dalla prima delle due curve e da quella in cui coincidono media, mediana e moda. La seconda distribuzione configura infatti un momento di transizione verso la cosiddetta curva a «J», che indica il raggiungimento da parte della totalità o quasi degli allievi, della padronanza delle abilità-obiettivo fissate dalla programmazione didattica e quindi il pieno successo dell'attività di insegnamento.

4.6.2 Deviazione standard

La *media*, la *mediana*, la *moda* e altre misure simili *indicano*, come si è detto, la tipicità di un fenomeno, o ancor più propriamente *la tendenza centrale dei punteggi* che lo definiscono (le tre misure indicate tendono a collocarsi, infatti, verso il centro della distribuzione). Ma esse da sole non bastano a esprimere seppur sinteticamente alcuni altri caratteri emblematici oltre che significativi del fenomeno. A una stessa media e/o mediana e/o moda possono corrispondere infatti distribuzioni tra loro molto differenti i cui elementi caratteristici non sono però indicati o espressi da queste misure.

Se consideriamo ad esempio le seguenti distribuzioni, ciascuna ordinata secondo valori crescenti - 3, 5, 6, 6, 10 e 5, 6, 6, 6, 7 - ci si accorge che media, mediana e moda coincidono per entrambe essendo 6 il loro valore. Ma quest'ultimo da solo non indica la variabilità dei punteggi che vi è in ciascuna serie: nella prima il campo di variazione, cioè l'intera estensione dei punteggi (quindi delle abilità o di qualsiasi altra qualità ad essi corrispondenti) è: 10 - 3 = 7; più precisamente la misura di questa estensione, definita *gamma* (γ) si calcola aggiungendo 1 alla differenza tra massimo e minimo: $\gamma = (P_{max} - P_{minimo}) + 1$.

Nel primo dei casi considerati si avrebbe $\gamma = (10 - 3) + 1 = 8$; nel secondo invece 3. Ciò significa che a una stessa misura di tendenza centrale, 6, corrispondono andamenti delle distribuzioni molto diversi tra loro e quindi una diversa omogeneità (o variabilità) delle prestazioni (o altre specifiche qualità) degli allievi cui le serie dei punteggi si riferiscono.

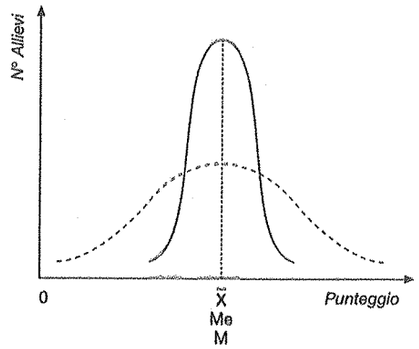
Se ci si accontentasse della semplice media come indicatore dei risultati, in entrambi i casi si potrebbe, a ragione, ritenere opportuno di non intervenire per modificare il processo di formazione apparendo i risultati formativi come abbastanza soddisfacenti, ma basterebbe calcolare la misura di variabilità che si è qui introdotta per rendersi conto dell'errore di una simile valutazione. Quando poi il numero dei dati è abbastanza elevato, l'utilità di compiere oltre a quelle di tendenza centrale altre misurazioni che informino con pertinenza della variabilità della distribuzione si accresce enormemente. Ma se il campo di variazione che è una di tali misure dà una idea approssimativa e talvolta erronea della variabilità dei punteggi (per il suo calcolo si tiene conto solo dei due valori estremi della distribuzione, cioè di quello più piccolo e di quello più grande che ovviamente potrebbero essere i valori meno probabili, cioè meno diffusi e perciò più lontani dalla maggior parte di quelli che compongono la distribuzione), vi è un'altra misura di variabilità che, considerando tutti i punteggi, e non solo quelli estremi, non induce in errori interpretativi. Questa misura è lo *scarto quadratico medio* o *deviazione standard* ed è espressa con il simbolo σ (sigma). Essa è data dalla radice quadrata della media dei quadrati degli scarti dei singoli valori dalla loro media:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}}$$

dove: σ = deviazione standard; $\sum$ = simbolo dell'operazione di sommatoria; X_i = singoli valori della distribuzione; $\bar{X}$ = media dei valori; n = numero dei valori.

Per meglio comprendere il significato di questa misura di variabilità che a primo acchito può apparire astrusa, e per vedere quindi quali importanti informazioni da essa si possono trarre, è utile dire che: in una ipotetica distribuzione avente valori tutti uguali, sigma risulterebbe uguale a zero (in quanto tutti gli scostamenti dalla media darebbero zero); l'elevazione al quadrato degli scarti dalla media fa sì che i valori più distanti della media della distribuzione abbiano un peso maggiore nel determinare il valore della deviazione standard. Ciò significa che quanto più alto è il valore di sigma (curva tratteggiata del grafico appresso riportato, chiamata platicurtica, "piatta" rispetto ovviamente alla curva normale) tanto più vi è dispersione dei punteggi rispetto alla media. In altri termini, a bassi valori di sigma corrisponderanno distribuzioni molto aggregate e compatte (curva continua, dello stesso

grafico detta invece *leptocurtica*, "appuntita"), cioè un risultato formativo dell'intero gruppo (o di una qualsiasi altra variabile rilevata) abbastanza omogeneo; viceversa, con alti valori di sigma ci troveremo dinanzi a distribuzioni molto disaggregate, cioè a un esito formativo (o di possesso di una qualunque altra qualità rilevata) riferito all'intero gruppo, molto eterogeneo e differenziato.



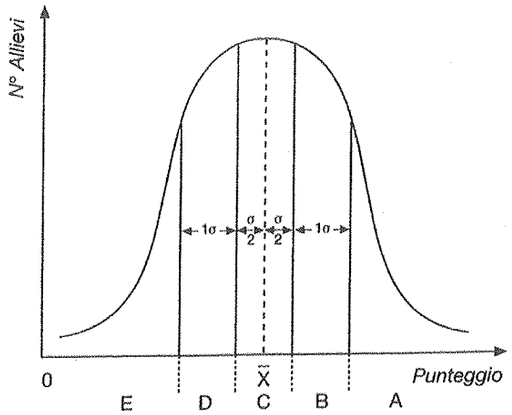
La deviazione standard indica, in estrema sintesi, *quanto mediamente ciascun punteggio di una distribuzione si discosta dalla media.*

È dunque evidente l'importanza che, nella ricerca in generale, nei processi di valutazione dei risultati delle prestazioni degli allievi appartenenti soprattutto a gruppi estesi, in particolare, e quindi nelle assunzioni di decisioni di intervento operativo, riveste la conoscenza della deviazione standard. Essa inoltre per le sue caratteristiche formali (indica tra l'altro i punti di flesso della curva di distribuzione dei punteggi e ben precise percentuali di eventi o valori interni a una certa area o ampiezza della distribuzione) è spesso utilizzata per standardizzare i punteggi grezzi delle rilevazioni, divenendo così una vera e propria *unità di misura degli scostamenti dalla media.*

4.6.3 Distribuzione pentenaria

Consiste nel distinguere le posizioni relative delle diverse prestazioni (misure, punteggi) rispetto alla loro media aritmetica in cinque fasce (A, B, C, D, E) che esprimono posizioni d'ordine decrescenti, ciascuna dell'ampiezza di una deviazione standard.

La fascia centrale C viene a essere costituita da tutti i punteggi compresi tra il valore individuato sottraendo alla media aritmetica mezza deviazione standard, e il valore che si ha aggiungendo alla media mezza deviazione standard. Aggiungendo al punto di destra, o massimo di C, una deviazione standard, si determinerà la fascia B cui seguirà, sempre con l'estensione di una deviazione standard, la fascia A. Operando simmetricamente a sinistra, si determineranno la fascia D ed E.



Con tale procedura sarà quindi semplice paragonare gli esiti delle misurazioni di una certa variabile appartenenti a due diversi campioni.

Ancora, conoscendo o confrontando le fasce in cui cadono per esempio due punteggi grezzi conseguiti da uno stesso allievo in prove dissimili, relative a uno stesso o a differenti ambiti disciplinari, diviene agevole individuare le analogie o le differenze di posizione delle due diverse prestazioni. Altrettanto può dirsi se il confronto prende in esame prestazioni di allievi differenti. In altri termini con tale stratagemma si facilita il confronto sincronico e quello diacronico tra esiti di diverse prestazioni di uno stesso soggetto e di allievi o di diversi gruppi di allievi.

Ma consideriamo ora alcuni altri aspetti positivi della tipologia delle analisi che la distribuzione pentenaria consente di fare.

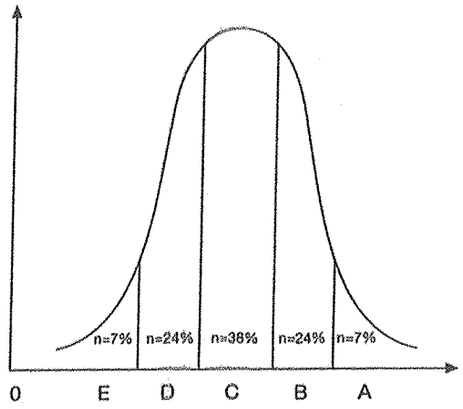
Come si può notare la curva sopra riportata è una curva gaussiana. Dato che la distribuzione dei dati che essa rappresenta è casuale, per rilevazioni numericamente elevate, all'interno della fascia C cadrà circa il 38% dei casi (misurazioni, punteggi ecc.); in ciascuna delle fasce B e D si registrerà ii 24% circa dei punteggi, mentre in ognuna delle fasce estreme, A ed E, il 7% circa. Quando si registrino distribuzioni di dati assai prossime a quelle indicate, evidentemente la popolazione presa in esame è di tipo casuale.

Quelle frequenze sono quindi un utile punto di riferimento nella interpretazione dell'andamento di molte rilevazioni compiute in particolari situazioni: pur rappresentando una distribuzione teorica hanno una forte valenza pratica, giacché assai spesso certi requisiti delle popolazioni esaminate si distribuiscono in modo analogo.

Lo scostamento della frequenza delle misure di certe caratteristiche degli allievi, magari rilevate all'inizio di una procedura sperimentale e al suo termine, dalle frequenze teoriche della curva normale, può immediatamente dar conto, infatti, di quanto quelle caratteristiche si discostino da quelle di una popolazione casuale.

È auspicabile che certi fenomeni positivi e intenzionali, come è appunto quello formativo scolastico, producano un cambiamento così significativo nella popolazione da distribuirsi in modo ben diverso da quello casuale.

La distribuzione pentenaria permette di compiere, anche da questo punto di vista, una prima, immediata, interpretazione del fenomeno indagato.



4.6.4 Punti Z e punti T

Vanno considerati punti standard in senso stretto, i punti Z e i punti T che derivano dai primi. Permettono di compiere con estrema precisione il confronto fra punteggi riportati in tempi diversi, in prove diverse da uno stesso allievo o da più allievi. A differenza della distribuzione pentenaria in cui ogni singola fascia include un ventaglio relativamente ampio di punteggi grezzi di diverso peso, tanto da far classificare in una stessa fascia punteggi tra loro più distanti di quelli che cadono in fasce differenti, i punti standard denotano e connotano ciascuno dei punteggi grezzi: ad ogni punto standard corrisponde un equivalente punto grezzo. Esprimono con immediatezza il significato che i punteggi assumono in rapporto alla media e all'ampiezza della deviazione standard della distribuzione cui essi appartengono.

Così, se un allievo riporta 32 punti in un test (max teorico 40) e 45 in un altro test (max teorico 50), è vero che con una banale proporzione si può sapere in quale delle due prove si è avvicinato di più al massimo teorico, ma per conoscere la migliore o la peggiore posizione relativa, cioè rispetto all'andamento dell'intero gruppo cui la prova è stata somministrata, occorre calcolare i corrispondenti punti standard.

Operando solo con quella proporzione si prenderebbero dei pericolosi abbagli. Con una più accorta analisi si potrebbe infatti scoprire che nella prova per la quale ha riportato il punteggio più vicino al massimo conseguibile l'allievo risulta in

ultima posizione, e che si situa invece nella prima proprio per la prestazione con il punteggio più distante da quello massimo teorico. Se solo sulla base di quel calcolo o della diversa ampiezza dei punteggi grezzi, che pure intuitivamente sembrano assai utili, si dovesse consigliare al soggetto in esame il corso di studi più congruente con le sue migliori attitudini, si compirebbe un errore tanto grossolano quanto grave.

Per sapere quindi se, nonostante le apparenze, il primo dei due punteggi non valga più del secondo, che pure ha valore assoluto maggiore, è necessario considerare le medie e le deviazioni standard delle distribuzioni cui appartengono, poiché dalla loro ampiezza derivano i significati dei singoli punteggi. Quanto più alta sarà la media e più bassa la deviazione standard, tanto maggiore sarà il valore dei punteggi che superano quello medio.

Sinteticamente si può dire che i punti Z indicano a che distanza dalla media (distanza misurata in termini di deviazioni standard) si situa un certo punteggio grezzo o, per converso, di quante deviazioni standard, sopra o sotto la media aritmetica, si trovi il particolare punteggio grezzo considerato.

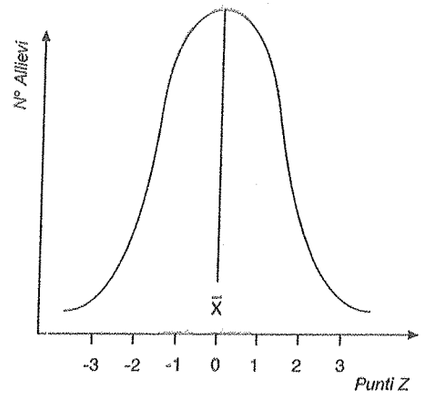
I punti Z sono infatti dati dal rapporto tra la differenza del punteggio X dalla media dei punteggi e il valore della deviazione standard:

$$Z = \frac{X_i - \bar{X}}{\sigma}$$

Nel caso in questione, il confronto tra i due punti Z indicherà quale sarà da considerarsi come prestazione migliore. Alla migliore prestazione corrisponderà il più alto punteggio Z.

Dalla formula si comprende che Z è neutro se $\overline{X}$ è uguale a zero; che Z è positivo quando $Z > \overline{X}$ (cioè quando il punteggio è più alto della media); e che infine Z è negativo se $Z < \overline{X}$ (quando cioè il punteggio è più basso della media).

Nella curva di distribuzione normale, 6 punti Z includono la quasi totalità dei dati in essa racchiusi. Più precisamente, tra 0 e 1 punto Z cadrà circa il 34% dei casi; tra 1 e 2 punti Z il 14% circa; infine tra 2 e 3 punti Z si avrà solo il 2% circa degli eventi rilevati. Simmetricamente altrettanto avverrà tra 0 e −3 punti Z. È possibile, ma non certo probabile, che un punteggio grezzo assuma un valore la cui distanza dal valore medio risulti superiore a 3Z o inferiore a −3Z:



I punti Z consentono di compiere con immediatezza confronti tra punteggi conseguiti in diverse prove, o da soggetti di differenti gruppi, proprio perché rendendoli omogenei permettono di inserirli tutti in un'unica distribuzione d'ordine: dal confronto tra due o più punteggi Z, risulterà migliore quello con maggiore valore relativo.

Hanno tuttavia il limite di assumere anche valori negativi che, benché rispecchino situazioni non soddisfacenti, certo non giovano sul piano della motivazione degli allievi verso il compito di apprendimento. Per ovviare a questo, ma anche ad altri inconvenienti di calcolo procedurale con i numeri negativi, è possibile compiere le stesse operazioni di confronto utilizzando altri punteggi standard quali i punti T. Con essi si pone, convenzionalmente, la media aritmetica delle distribuzioni dei punteggi uguale a 50 e il sigma uguale a 10, così da garantire punteggi tutti col segno positivo, ancorché tra essi nettamente differenziati:

$$T = 50 + \frac{10\,(X_i - \bar{X})}{\sigma}$$

Lo scostamento sopra o sotto 50 indicherà il valore e il significato dei punteggi riportati da uno stesso allievo in test differenti o da allievi appartenenti a gruppi o campioni diversi, ovvero la posizione di un gruppo rispetto a un altro di confronto.

Ovviamente, se si vuole trasformare direttamente un punto Z in punto T basterà sostituire nella formula i valori della parentesi e del denominatore con il punto Z: T = 50+10 (Z).

4.7 Campioni e statistiche. Cenni introduttivi

Una volta che in ogni particolare indagine siano stati raccolti i dati seguendo tempi, modalità e strumenti previsti dal piano della ricerca - ma anche altri dati o elementi conoscitivi magari connessi all'influenza di eventuali variabili intervenienti o parassitarie - potrebbe rendersi necessario andare oltre i calcoli fin qui messi in evidenza, essenzialmente misure di tendenza centrale e di variabilità del fenomeno, per riuscire ad analizzarlo, comprenderlo e interpretarlo meglio. Anche se gran parte dei metodi e delle tecniche di raccolta, analisi e interpretazione dei dati servono, fondamentalmente, a "descrivere" le caratteristiche essenziali dei fenomeni indagati, può tuttavia presentarsi la necessità di andare oltre la descrizione pur attenta e rigorosa del fenomeno, per cercare di cogliere dai fatti studiati conclusioni o interpretazioni che, seppur interlocutorie, abbiano valore generale, estensibile all'intera popolazione o universo cui appartiene il campione sul quale si è operato. Per poter sapere se ed entro quali limiti sia legittimo cogliere dai fatti osservati nel campione in esame - cioè in uno dei "campioni" possibili, "inferenze" o conclusioni generali che possano estendersi e risultare valide per l'intero "universo" cui quel campione appartiene - occorre misurare il livello di approssimazione del campione all'*universo*, anzi, sarebbe più razionale prevedere a priori quali caratteristiche deve possedere il "campione" affinché i risultati finali possano considerarsi generalizzabili correttamente all'intera popolazione o "universo". Anche se intuitivamente è dato di pensare che per evitare errori la strada migliore sarebbe quella di esaminare l'intero universo, questa cosa, oltre a non essere sempre possibile, talvolta potrebbe persino risultare troppo dispendiosa e addirittura da evitarsi perché dannosa. Si pensi di quante alterazioni per "copiatura" - consentita magari da docenti troppo desiderosi di far bella figura - risentirebbero i risultati di indagini conoscitive degli apprendimenti scolastici tipo Pisa e Iea, se fatte sull'intero universo di studenti anziché su campioni statistici. Gli esiti così registrati potremmo davvero considerarli validi e attendibili? Uno studio sul campione può essere condotto molto più accuratamente che sull'universo.

Occorre, allora, tener conto del fatto che vi sono statistiche che possono permetterci solo di descrivere e sintetizzare in forma numerica diversi aspetti dei dati specificamente raccolti sul campione a disposizione; altre che invece ci permettono di allargare all'universo di appartenenza i risultati cui si perviene operando su un campione che lo rappresenti. Le prime sono chiamate, appunto, *statistiche descrittive*: buona parte di quelle fin qui illustrate appartengono a questa tipologia; le altre appena indicate, invece, potendosi basare sulla teoria della probabilità, permettono invece di far compiere delle previsioni affidabili sulla generalizzabilità dei dati raccolti sul campione studiato all'intera popolazione o universo di riferimento.

La *statistica descrittiva* serve dunque a descrivere e far comprendere il significato logico dell'insieme dei punteggi raccolti su un singolo soggetto sperimentale o su diversi gruppi di soggetti, nonché le relazioni tra le variabili di un campione.

La *statistica inferenziale* utilizza la teoria della probabilità per determinare, appunto, la probabilità che un insieme di dati osservato si verifichi o meno semplicemente grazie a variazioni dovute al caso. Il suo impiego può permettere entro

vincoli formali ben precisi, di trarre dall'analisi dei dati raccolti e opportunamente trattati, inferenze di senso e significato logico estensibili all'intero universo di cui fa parte quel campione. Una generalizzazione, questa, di grande impatto conoscitivo. Si pensi agli effetti degli esiti delle sperimentazioni condotte per la messa a punto dei vaccini anti Covid-19.

I metodi di campionamento probabilistico sono abbastanza numerosi. La tecnologia relativa al trattamento dei dati attualmente disponibile, in termini di calcolo e pervasività e semplicità d'uso, permette un continuo arricchimento del quadro teorico e procedurale in ambito statistico. Tuttavia sono fondamentalmente quattro i metodi di campionamento che possono in questa sede interessarci:

- campionamento casuale semplice;
- campionamento casuale sistematico;
- campionamento stratificato;
- campionamento a grappolo.

Nelle composizioni del campione tutti gli elementi, misure o eventi ecc. che fanno parte della popolazione o universo devono avere la stessa probabilità di essere scelti.

Nel *campionamento casuale semplice* l'operazione può compiersi o estraendoli a sorte, senza reimmissione degli elementi scelti, oppure impiegando una tabella dei numeri casuali. Anche se intuitivamente si può ipotizzare che quanto maggiore è il numero della popolazione tanto maggiore dev'essere la dimensione del campione, in effetti non è sempre così. La grandezza del campione deriva soprattutto dalla precisione che si vuole ottenere e dal livello di rischio d'errore che si vuole correre. Se entrambi questi vincoli vengono stabiliti a priori, sarà facile, attraverso calcoli opportuni, definire quale deve essere la percentuale degli elementi che devono comporre il campione.

Con il *campionamento casuale sistematico* il campione viene formato scegliendo dalla lista uno su ogni K elementi. Così se si vorrà avere un campione di 50 casi da un elenco di 1600, dovremo scegliere un elemento ogni 32 dopo aver stabilito casualmente il primo da cui partire. Quando un dato universo presenta distintamente alcune specificità, occorrerà che il campione le possa rappresentare tutte proporzionalmente. A tal fine ci si servirà di un *campione stratificato* tale da rispecchiare i diversi "strati" omogenei della popolazione. Quando gli strati, ovvero le specificità sono molte, potrebbe risultare più utile comporre *campione a grappolo*. In questi, sia i grappoli o i gruppi o gli strati, sia gli elementi all'interno di ciascun grappolo, vengono scelti casualmente (in modalità semplice o sistematica).

Ipotesi nulla (H_0)

Come vedremo più distesamente nel Capitolo 7, soprattutto da un punto di vista statistico e procedurale, quando viene impiantata una ricerca, particolarmente se di tipo sperimentale, risulta molto utile formulare due tipi di ipotesi: quella definita ipotesi nulla (H_0) secondo cui l'intervento di modificazione controllata della realtà allo studio (variabile sperimentale) non modifica significativamente gli esiti del processo ordinario (quello del gruppo di controllo); e quella della ricerca, ovvero l'ipotesi che si vuole provare, indicata con H_1, ben diversa, anzi, alternativa ad H_0.

L'ipotesi nulla postula quindi la non sussistenza del cambiamento della situazione, nonostante l'alterazione delle variabili sperimentali, indipendenti, oppure la non significatività delle differenze tra due campioni. Ovviamente la si formula con l'intento di respingerla a favore dell'ipotesi della ricerca. La si usa come criterio di riferimento comparativo al fine di evitare di attribuire all'intervento sperimentale ciò che invece potrebbe risultare un mero effetto del caso.

A tal fine occorre fare ricorso, a seconda della ricerca e delle scale di misura impiegate oltre che delle *caratteristiche parametriche o non parametriche del campione* (appartenenza o non appartenenza del campione a una popolazione o universo di tipo casuale, rappresentata da una curva gaussiana "pura"), all'impiego di opportuni *test statistici, parametrici* oppure *non parametrici*, in grado di farci verificare o respingere l'ipotesi nulla.

In presenza di campioni non molto estesi (con un numero abbastanza inferiore a 100) appartenenti a un universo non casuale, le statistiche parametriche non sarebbero utili strumenti esplicativi del fenomeno. Tuttavia è sempre l'esperienza del ricercatore (quanto meno auto-verificata) e la sua sensibilità a far decidere, nei diversi contesti di ricerca, quale delle due tipologie di test adoperare. Così, ad esempio, se per cogliere la "significatività delle differenze" di atteggiamento verso lo studio, dei componenti un gruppo-classe, rilevate con un semplice questionario e impiegando una scala nominale si può usare il test del *Chi quadro*, in caso di piccoli numeri, di elementi tra loro assai simili per esperienze pregresse, il ricercatore dovrebbe sapere che allo scopo risulterebbe più affidabile il test non-parametrico di *McNemar*. Essi danno conto della significatività delle differenze tra ciò che si prevedeva di rilevare, sulla base delle attese teoriche, e ciò che si rileva realmente nel campione esaminato. Il risultato attesta perciò la casualità o meno dello scarto registrato tra le frequenze (della variabile presa in esame) osservate nella realtà, e le frequenze, della stessa variabile, teoricamente previste.

Preme ora precisare, però, che l'estensione e il livello di trattazione degli argomenti scelti di massima per questo volume ci esimono dal trattare più analiticamente le questioni teorico-operative relative ai test statistici, sia parametrici che non-paramentrici, da utilizzarsi per il trattamento dei dati di una ricerca al fine di

verificarne la correttezza metodologico-procedurale e il grado di affidabilità dei risultati. A tale riguardo va detto che tutti coloro che per interessi professionali si occupano di problemi e di ricerca educativi dovrebbero di necessità padroneggiare le più elementari procedure di analisi statistica dei dati informativi e conoscitivi raccolti con l'impiego di qualsiasi scala di misura utilizzabile, anche di *scale empiriche* costruite appositamente per l'occasione. Rinviando quindi chi fosse interessato ai numerosi testi di facile lettura oggi disponibili di statistica per educatori, psicologi e sociologi, si indicano qui solo alcuni dei test statistici che trovano maggiore applicazione nel campo della ricerca educativa. Tra questi vanno segnalati, a livello d'impiego della *scala nominale,* oltre ai citati test del *Chi quadro* e di McNemar sulla significatività delle differenze, anche il test binomiale e di Fischer; con la *scala ordinale,* oltre gli indici di posizione, anche quelli rispetto alla mediana, per due o più campioni (impiegando la statistica descrittiva); il test dei segni, correlazioni a ranghi di Sperman e l'analisi della varianza, quando si faccia ricorso a procedure statistiche di tipo inferenziale. Con la *scala a intervalli,* oltre ai molti test o punteggi standard già indicati, come i *punti T* e i *punti Z,* risultano spesso utili altri punteggi standard (stanine, stend) e l'analisi della *differenza e della semidifferenza interquartile,* per stare dentro la statistica descrittiva; *il test T di Student,* l'analisi della varianza, l'analisi fattoriale esplorativa e confermativa, la *statistica S di Sedecor,* e altri test ancora, per le inferenze statistiche.

Riepilogo

In questo capitolo è stato approfondito il concetto di classificazione quale prima attività orientata a raggruppare diversi elementi in funzione di una determinata caratteristica in modo da formare classi che siano il più omogenee possibile rispetto alla proprietà identificata e in base alla quale vengono inclusi i vari elementi.

Sono state presentate inoltre le conoscenze necessarie all'avvio delle procedure di misurazione in ambito educativo: dagli aspetti qualitativi e quantitativi, alle dimensioni misurabili attraverso scale nominali, ordinali o a intervalli, alle operazioni consentite rispetto alla natura delle variabili allo studio quali il calcolo di media, mediana, moda, deviazione standard e punti per rilevare la specifica distribuzione dei dati rilevati. Vengono infine indicate le ragioni di calcolo di alcune statistiche descrittive e inferenziali.

Riferimenti bibliografici

Andreani Dentici O. (1968). *Abilità mentale e rendimento scolastico.* Firenze: La Nuova Italia.

Beckers J., De Ball R., De Landsheere G. (1985). *Come formulare giudizi di valutazione didattica. La costruzione di scale descrittive.*Teramo: Lisciani & Giunti.

Benvenuto G. (2003). *Mettere i voti. Introduzione alla docimologia.* Roma: Carocci.

Bernardi L. (1984). L'informazione statistica per l'istruzione. In: Becchi E., Vertecchi B. (a cura di). *Manuale critico della sperimentazione e della ricerca educativa.* Milano: Franco Angeli.

Calonghi L. (1962). *Reattivi nella scuola.* Zurigo: Pas Verlag.

Calonghi L. (1976). *Valutazione.* Brescia: La Scuola.

Calonghi L. (1978). *Valutazione delle composizioni scritte.* Roma: Armando.

Chiappetta Cajola L. (1998). *Handicap e valutazione. La funzione di controllo e di regolazione nel processo formativo degli allievi disabili.* Roma: Anicia.

Coggi C., Calonghi L. (1992). *Elementi di statistica per la ricerca scolastica.* Teramo: Giunti & Lisciani

Cristante F., Lis A., Sambin M. (1982). *Statistica per psicologi.* Firenze: Giunto-Barbera.

Domenici G. (1981). *Descrittori dell'apprendimento.* Firenze: Giunti & Lisciani.

Domenici G. (1991). *Gli strumenti della valutazione.* Napoli: Tecnodid.

Domenici G. (1992). *Le prove strutturate di conoscenza.* Teramo-Firenze: Lisciani & Giunti.

Domenici G. (2007). *Manuale della valutazione scolastica.* Roma-Bari: Laterza.

Gattullo M. (1967). *Didattica e docimologia. Misurazione e valutazione nella scuola.* Roma: Armando.

Gattullo M., Giovannini M.L. (1989). *Misurare e valutare l'apprendimento nella scuola media,* Milano: Edizioni Scolastiche Bruno Mondadori.

Grandi G. (1977). *Misurazione e valutazione.* Firenze: La Nuova Italia.

Hudson B. (1987). *Introduzione alle tecniche di valutazione.* Bologna: Zanichelli.

Leti G. (1983). *Statistica descrittiva.* Bologna: Il Mulino.

Lombardo E. (1997). *I dati statistici in pedagogia. Esplorazione e analisi.* Scandicci: La Nuova Italia.

Nunnally J.C. (1976). *Misurazione e valutazione nella scuola.* OS: Firenze.

Siegel S. (1978). *Statistica non parametrica per le scienze del comportamento.* Firenze: Organizzazioni Speciali OS

Vertecchi B. (1976). *Valutazione formativa.* Torino: Loescher.

Vertecchi B. (1984). *Manuale della valutazione.* Roma: Editori Riuniti.

Visalberghi A. (1955). *Misurazione e valutazione nel processo educativo.* Milano: Edizioni di Comunità.

5

Testing, ricerca e valutazione[1]

In questo capitolo vengono approfondite questioni relative alla costruzione degli strumenti di misura in educazione. Tenendo conto del fatto che la rilevazione delle variabili significative in ambito educativo spazia dalle dimensioni personali (atteggiamenti, abilità, competenze) a quelle contestuali e situazionali (effetti di strategie didattiche, training formativi ecc.), si sottolinea come la moderna costruzione di modelli esplicativi dei processi educativo-formativi e dei fenomeni correlati abbia una consistente base induttiva centrata su osservazioni empiriche e intuizioni da validare su basi evidenti di dati, per svilupparsi in forma ipotetico-deduttiva.
Si segnala quindi il consistente e sistematico lavoro empirico che è necessario svolgere per identificare e rilevare le variabili-chiave e stabilire altresì relazioni tra tali variabili.
A tale proposito occorre conoscere i principali costrutti teorici e le diverse modalità di costruzione degli strumenti di misura: a prescindere dalla messa a punto di tali dispositivi, infatti, non vi è in sostanza alcuna dimensione che possa essere realmente colta e studiata.
Particolare attenzione sarà rivolta all'attribuzione dei punteggi (o *scoring system*) nella messa a punto di un questionario (procedura detta testing). Tale attribuzione deve essere empiricamente validata e avere dati normativi di riferimento per la popolazione che stiamo valutando (processo di standardizzazione), al fine di permettere l'interpretazione dei dati raccolti su una base affidabile.
Il capitolo si sofferma, infine, sulla descrizione delle principali procedure di verifica e valutazione degli apprendimenti - con attenzione al calcolo del livello di difficoltà e di discriminatività degli item - e sulla loro utilità nei contesti scolastici.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Teorie, costrutti, variabili, fattori, indicatori.
- Definizione e analisi degli item.
- Costruzione e validazione di scale di valutazione.
- Standardizzazione e scoring.
- Try out, livelli di difficoltà e di discriminatività degli item.
- Testing e processi di valutazione. Caratteristiche formali delle prove di valutazione degli apprendimenti.

[1] Il presente Capitolo è frutto del lavoro congiunto dei due autori. V. Biasi ha redatto i paragrafi 5.1, 5.2 e 5.3, G. Domenici il paragrafo 5.4. e relativi sottoparagrafi.

> Per approfittare di più della tripla opposizione planetaria, è indispensabile procurarsi un telescopio. Il signor Palomar, forse perché porta lo stesso nome d'un famoso osservatorio, gode di qualche amicizia tra gli astronomi, e gli viene concesso d'avvicinare il naso all'oculare d'un telescopio da 15 cm, cioè piuttosto piccolo per la ricerca scientifica, ma che, paragonato ai suoi occhiali, fa già una bella differenza
>
> Italo Calvino, *Palomar*

5.1 Teorie, costrutti, variabili, fattori, indicatori

Possiamo definire una teoria come un *sistema di riferimento, una mappa mentale, uno schema concettuale ad alta coerenza interna* - esplicativi di una certa realtà e di specifiche tematiche - in grado perciò sia di organizzare, inquadrare e spiegare molti dati empirici relativi a certi fenomeni noti, in termini di principi generali o di leggi, sia di prevederne di nuovi, finché tali principi non vengano confutati da prove contrarie.

Una teoria rappresenta dunque un insieme di concetti, una costruzione sistematica - generalmente a carattere ipotetico-deduttivo - basata su osservazioni, capace di spiegare determinati fenomeni e di predirne di nuovi.

Una teoria risulta efficace quando soddisfa alcune regole e funzioni fondamentali, consentendo di:

- organizzare molte osservazioni diverse nel numero minore possibile di *variabili*;
- suggerire nuove ipotesi o previsioni specifiche da *verificare* con osservazioni ancora da svolgere (funzione euristica);
- convogliare l'attenzione sulle variabili di maggiore interesse senza perdere di vista la *complessità dei fenomeni* che si vogliono studiare;
- rendere possibile, in condizioni date, la *previsione dei fenomeni* in questione.

Costrutto Partendo dalla suddetta definizione di teoria scientificamente verificabile, possiamo precisare il significato di *costrutto* quale *concetto astratto postulato da una teoria, non direttamente osservabile, ma inferito* (per esempio, il costrutto di autostima). Una distinzione fondamentale che occorre tenere presente è quella tra "indicatore" e "costrutto".

Indicatore Per indicatore si intende una misura empirica - generalmente risultato di un rapporto tra due dimensioni che possono essere quali-quantitative - che sulla base di una corrispondenza viene posta in relazione a un costrutto. È il caso per esempio di un comportamento: la relazione o il rapporto tra frequenza di episodi di *overeating* e successivo vomito auto-provocato vengono segnalati dal ricercatore come *indicatori* del costrutto "disturbo bulimico", e simili. Per fare un altro esempio, un indicatore dell'affollamento delle aule scolastiche di un Paese può essere dato dal rapporto tra il numero degli alunni e il numero dei metri quadrati delle aule disponibili. In sintesi, l'indicatore è costituito da una *variabile osservata* che si ipotizza cogliere in tutto o in parte il costrutto allo studio. In genere vengono forniti molteplici indicatori per ogni costrutto.

Il costrutto si pone a livello di astrazione maggiore rispetto agli indicatori ed è collegato a essi grazie a un principio di *corrispondenza* individuato o anche talvolta inizialmente solo intuito dal ricercatore medesimo. Comunque sul piano empirico gli indicatori contribuiscono a definire il costrutto.

A tale proposito possiamo riportare l'esempio degli indicatori dell'apprendimento scolastico, definiti sulla base dei descrittori di apprendimento (Domenici, 1981, 2003) e relativi al concetto di divisione aritmetica, ovvero il riuscire da parte dell'allievo a ripartire gli oggetti sul banco in due gruppi numericamente equivalenti o in un gruppo maggiore e uno minore di elementi.

Occorre naturalmente a questo punto ricordare il concetto di "Variabile" che consiste, come si evince linguisticamente, in una dimensione che varia o di aspetto misurabile.

Ipotesi Il concetto di "ipotesi" viene definito come l'*articolazione di relazioni fra costrutti formulate in base a una teoria, e passibili di verifica e/o falsificazione.* L'ipotesi pertanto si presenta come una vera e propria proposizione che anticipa la soluzione di una questione problematica, e che in qualche modo suggerisce la procedura da impiegare per la sua verifica.

Al fine di avanzare delle ipotesi per cercare di spiegare e interpretare dei fenomeni allo studio, occorre indagare bene le caratteristiche delle *variabili* o dei *fattori* che si intendono misurare, in modo da definire degli *indicatori* che ne rilevino la presenza e la natura.

Metodi e strumenti La scelta dei metodi, cioè delle procedure seguite dal ricercatore nella raccolta di informazioni, e la scelta degli strumenti da utilizzare per rilevare i suddetti indicatori, costituiscono fasi fondamentali per verificare un'ipotesi formulata.

La *scelta del metodo* dipende dall'obiettivo della ricerca, nel caso si tratti di *descrivere* o, in modo più complesso, di *verificare* una o più ipotesi. Possiamo quindi utilizzare il metodo osservativo, quello clinico (o studio di caso), quello correlazionale, fino a quello sperimentale e quasi-sperimentale.

Per quanto concerne gli *strumenti di misura* abbiamo la possibilità di scegliere tra

le seguenti principali categorie:

- scale di auto-valutazione (questionari self-report);
- scale di valutazione da parte di altri (griglie di rilevazione o check-list);
- interviste libere o guidate (semistrutturate o strutturate);
- colloqui;
- osservazioni dirette del comportamento individuale, anche nel gruppo;
- osservazioni mediate da registrazioni audio-visive o altre tracce;
- focus group e test proiettivi;
- test di competenza o di profitto.

Per quanto concerne la *discrezionalità* che può intervenire nel processo di raccolta dei dati si possono indicare i seguenti strumenti come espressivi di un livello decrescente di soggettività (da un massimo a un minimo):

- colloquio;
- focus group (discussioni di gruppo);
- intervista semistrutturata;
- intervista strutturata;
- intervista standardizzata;
- intervista face to face + computer (CAPI);
- intervista telefonica;
- intervista telefonica + computer (CATI);
- questionario auto-compilato (self-report);
- inchiesta elettronica (web, e-mail surveys).

Di particolare utilità appare comunque la *costruzione dello strumento*, se non già disponibile in letteratura.

Ci occuperemo nel corso del capitolo delle procedure utilizzate per allestire un questionario self-report: a tale proposito occorre avere chiaro il *costrutto* che si vuole indagare, le *variabili* che riteniamo possano essere coinvolte o i fattori eventualmente sottesi, ciò allo scopo di definire gli *indicatori* sui quali poter articolare gli item.

5.2 Costruzione e validazione di scale di valutazione

Vari autori hanno fornito nel tempo diverse definizioni di strumenti, quali scale di valutazione o questionari, definiti anche test o reattivi.

Secondo Goode e Hatt (1952) il questionario è costituito da «un insieme di domande standardizzate presentate sotto forma di formulario che l'intervistato stesso riempie».

Occorre qui precisare che il questionario è uno strumento caratterizzato da espressioni verbali (domande o affermazioni alle quali si chiede risposta) che fa parte dell'ampia famiglia dei *dispositivi di valutazione* chiamati "test" o "reattivi" e con i quali ci si riferisce alla presentazione di una situazione standardizzata nella quale il comportamento (inteso in senso molto ampio: performance, abilità, preferenza o anche opinione) di una persona viene campionato, osservato e descritto producendo una misura oggettiva e standardizzata in riferimento a un campione.

Già Bonaiuto (1971) aveva definito chiaramente il *concetto di test o reattivo* come una prova standardizzata per valutazioni psicodiagnostiche. Si tratta di strumenti per la provocazione e la valutazione di comportamenti e di "vissuti" impiegati in psicologia. Ha la finalità di consentire allo psicologo, con notevoli economie, inferenze e conclusioni su una vasta gamma di atteggiamenti e di tratti della personalità, disponendo di un ambito relativamente ristretto di caratteristiche sintomatiche che possono essere rilevate e valutate direttamente.

Ogni reattivo (o test) è costituito da una prova specifica che viene somministrata ai soggetti secondo modalità rigorosamente standardizzate e che deve comportare una tecnica il più possibile esatta per l'annotazione, l'interpretazione e la valutazione quantitativa dei rendimenti ottenuti.

Gli scopi principali per gli operatori che elaborano strumenti di questo tipo sono: (1) la previsione del successo o dell'insuccesso dell'individuo in una determinata situazione; (2) l'indicazione e il controllo del trattamento terapeutico; (3) il progresso delle conoscenze scientifiche.

I reattivi possono pertanto venire usati per prevedere la probabilità con cui un individuo può aver successo, ad esempio nelle attività connesse con un determinato tipo di scuola (orientamento scolastico), di lavoro (orientamento e selezione professionale), di disciplina sportiva ecc. In simili casi lo scopo positivo e corretto dell'applicazione dei reattivi è di far sì che il maggior numero di persone sia inserito in ambienti e attività dove possano raggiungere il massimo dell'efficienza e dell'auto-realizzazione.

I requisiti ai quali ogni reattivo deve corrispondere, dopo essere stato sottoposto ad apposita "taratura", sono: (1) la capacità discriminativa, (2) la fedeltà o attendibilità, data dalla correlazione fra i valori risultanti in due prove identiche condotte sugli stessi soggetti, (3) la validità, data dalla correlazione tra i risultati ottenuti al reattivo stesso e i valori delle variabili ("criterio") che si vogliono misurare o predire.

Anche secondo Anastasi e Urbina (1997) e Anastasi (2002, p. 49) un test «consiste essenzialmente in una misurazione obiettiva e standardizzata di un campione di comportamento».

Precisiamo inoltre che già nel 1984 Vittorio Rubini definiva il test, intendendo in particolare quello di tipo psicologico, come: «... un insieme di stimoli idonei ad

attivare in situazione controllata un campione di comportamento individuale .... il comportamento viene registrato secondo predefinite procedure, allo scopo di esprimerlo sinteticamente in un indice metrico e di valutarlo confrontandolo con quello di altri individui posti nella medesima situazione».

Inoltre, come precisa Boncori (2006, p. 22) un «test è un termine tecnico, che designa un tipo ben preciso di strumento di misura e che dovrebbe essere usato solo in presenza di una serie di caratteristiche metrologiche particolari. [...] Chi vede un test e non sa il metodo con cui è stato costruito e verificato può però usare questo termine anche per strumenti che del test hanno l'apparenza, ma non le proprietà».

Approfondimento 5.1
TEST

Boncori (1993) definisce il test «una situazione standardizzata nella quale il comportamento di una persona viene campionato, osservato e descritto, producendo una misura oggettiva e standardizzata di un campione di comportamento».
Il termine "test" si riferisce a una prova standardizzata di valutazione, una sorta di stimolo presentato per ottenere una risposta dal soggetto intervistato.
I test misurano quelle variabili che, in base a una teoria di riferimento, si considerano determinanti un fenomeno o processo allo studio.
Si possono costruire test di profitto, cioè di apprendimento o di performance, test di abilità specifiche (per esempio di memoria, di ragionamento o problem solving, di creatività, prove motorie, percettive, fino a test proiettivi di natura psicologica).
Un test è in genere basato su una serie di item che possono essere espressi come domande, affermazioni o prove da risolvere. Viene attribuito un punteggio a ogni item e ciò rifletterà il costrutto allo studio, e si otterrà un punteggio complessivo individuale in genere distribuito secondo una curva normale. Per interpretare tale punteggio occorrono confronti con punteggi medi (cosiddetti dati normativi) ottenuti da ampie popolazioni di riferimento.
I test debbono essere applicati secondo procedure mantenute costanti.
Il test deve essere attendibile, nel senso che la somministrazione dello stesso test allo stesso soggetto nelle stesse condizioni, deve produrre lo stesso risultato e deve essere valido, cioè misurare esattamente ciò che intende misurare.

In sintesi, come riassume Sartori (2008, p. 36): «Riunendo due tra le tante definizioni proposte (Anastasi, 2002; Boncori, 1993), possiamo definire test ... una situazione standardizzata in cui il comportamento di una persona, adeguatamente campionato, viene prima osservato e poi descritto per mezzo della misura, anch'essa standardizzata e, in questo senso, oggettiva, che si sarà ottenuta in questo modo».

Queste definizioni convergono in genere, peraltro, sulla considerazione che il dato ottenuto mediante la somministrazione di un questionario (o test) sia una risposta alla specifica domanda o item, formulati in quel preciso modo e in quel contesto. Nessun questionario sarà quindi davvero interpretabile indipendentemente dalle ipotesi teoriche iniziali sulle quali è stato articolato, e dalla conoscenza della procedura con cui viene somministrato.

Si parla infatti di teorie e tecniche della costruzione dei test, e ciò vale per tutte le diverse tipologie di test approntabili, siano essi test di profitto, test su abilità specifiche (memoria, attenzione ecc.), test attitudinali o test di atteggiamento.

5.2.1 Modelli e fasi per la costruzione di un questionario e/o di un test di profitto

Dobbiamo assumere naturalmente che ogni processo di misurazione porta con sé un margine di errore: quindi il punteggio che otteniamo a un test o questionario è da considerarsi non tanto il punteggio per così dire "vero", ma il punteggio "osservato" che costituisce uno dei possibili punteggi ottenibili dal soggetto secondo una probabilità accettabile che contiene il punteggio vero.

Tornando a parlare di indicatori, occorre distinguere tra:

- *indicatori riflettivi*: sono considerati specchio o effetto del costrutto teorico allo studio. Tale costrutto è latente e i vari indicatori appaiono come variabili o effetti diretti: per esempio, alla domanda "quante volte utilizzi il feedback formativo?" le risposte fornibili ("mai", "qualche volta" o "spesso") sono da considerarsi frutto del costrutto "valutazione formativa";
- *indicatori formativi*: in questo caso il costrutto latente viene espresso dalla somma di una serie di indicatori e diviene funzione di essi: per esempio il costrutto di stress viene considerato funzione della perdita del lavoro, della separazione dai familiari ecc.

Dalla distinzione tra questi due tipi di indicatori deriva la distinzione tra i cosiddetti modelli regressivi e i modelli fattoriali utilizzabili nella costruzione di un test o questionario: nel caso di *indicatori riflettivi* abbiamo modelli correlazionali (vi sarà cioè una corrispondenza di tipo correlazionale tra il singolo indicatore e il costrutto), mentre nel caso di *indicatori formativi* abbiamo modelli fattoriali (si dovrà pertanto fare ricorso all'analisi fattoriale per individuare dimensioni latenti a una serie di item o variabili).

Il modello regressivo si basa invece sulla tecnica statistica della regressione multipla: si calcola cioè il coefficiente di correlazione tra le variabili in gioco nel senso che più è presente l'indicatore più è probabile il suo valore predittivo (nell'esempio sopra riportato, più frequentemente viene fatto ricorso al feedback formativo,

maggiore sarà stata l'applicazione della valutazione formativa).

Il modello fattoriale prevede invece che si calcolino le correlazioni tra gli item (calcolo delle correlazioni multiple) per individuare come correlano tra loro e si raggruppano identificando i cosiddetti fattori latenti. Nell'esempio sopra accennato del costrutto di stress potrà emergere il fattore "ansia lavorativa" (che potrà vedere raggruppati item relativi alla perdita di lavoro e di sicurezza economica) e il fattore "instabilità affettiva" (che potrà invece correlare con item relativi alla separazione dai familiari, al divorzio, alla solitudine emotiva).

Passiamo ora alla vera e propria applicazione delle fasi di costruzione di un questionario. Una volta chiarito il costrutto che intendiamo studiare, occorre seguire una metodologia o procedura secondo i passi di seguito elencati.

1. Preparazione di una prima versione del test, ossia stesura di un primo ampio insieme di item espressivi secondo il ricercatore del costrutto da indagare (modello regressivo o fattoriale).
2. *Try-out* o prova preliminare di somministrazione del questionario a un primo campione ridotto di partecipanti, al fine di far emergere eventuali difficoltà linguistiche, fraintendimenti, ambiguità nella formulazione delle domande.
3. Selezione e perfezionamento degli item sulla base dei risultati preliminari conseguiti.
4. Definizione della versione finale del questionario, con somministrazione a un ampio campione rappresentativo per la taratura sulla popolazione di riferimento.
5. Definizione delle proprietà psicometriche del test nella versione finale (cioè livelli di attendibilità e validità relativamente alle dimensioni rilevate sia di tipo cognitivo, motivazionale, di performance di competenza ecc.) e definizione delle modalità di somministrazione e attribuzione dei punteggi ai singoli item e alle eventuali scale o sotto-scale, ossia ai fattori sottesi (*scoring*).

5.2.2 Dai costrutti agli indicatori

Una corretta definizione operativa dei *costrutti* indagati sarà giustificata in base alle teorie ritenute rilevanti dal ricercatore. Prendiamo come esempio il caso della costruzione di una Scala o Questionario di Atteggiamento.

Occorre prima di tutto esplicitare in maniera chiara la definizione di atteggiamento cui aderiamo. Ciò rappresenta il *costrutto* allo studio, che è un concetto o un insieme di concetti astratti.

Ricordiamo a questo proposito che l'*atteggiamento* può essere in sintesi definito come uno stato disposizionale, un insieme di giudizi, credenze e aspettative che orientano forme di comportamento. Gli atteggiamenti sono strettamente connessi alle strutture più stabili della personalità dell'individuo: esperienze, convinzioni, emozioni e valori. Occorrerà quindi rilevare e misurare tutte e tre queste *dimensioni* o *variabili* che sottendono gli atteggiamenti: cognitive, affettive, comportamentali. Per ognuna di esse si dovranno individuare degli *indicatori* in base ai quali sarà articolata la stesura degli item.

5.2.3 Definizione degli item nei test di atteggiamento

Per definire gli item occorre porre molta attenzione a ciò che si chiede. Nella fase preliminare della definizione degli item è possibile fare riferimento a vari metodi per l'individuazione dei significati, che qui riportiamo. Per la costruzione dei quesiti relativi ai test di apprendimento o di profitto si veda oltre (Paragrafo 5.4.9).

1. *Introspezione*: in base alla capacità introspettiva è possibile individuare a livello consapevole le caratteristiche dei propri vissuti personali in relazione all'oggetto di studio.
2. *Analisi fenomenologica*: sulla base della descrizione del fenomeno così come ci appare alla coscienza (percezione autentica e immediata), è possibile identificare alcune caratteristiche dette appunto "fenomeniche" dell'oggetto di studio.
3. *Focus group*: in base alla riflessione e al confronto interpersonale è possibile conoscere le credenze e le valutazioni degli altri verso l'oggetto di interesse.
4. *Domande aperte*: sulla base di quanto emerso dall'analisi dei temi discussi nei focus group, si possono delineare affermazioni, in forma di domanda, relative all'oggetto di studio.

Try-out

Il termine "try-out" fa riferimento a una somministrazione preliminare della prova in allestimento. In questa fase non eludibile vengono verificate le caratteristiche delle domande al fine di verificarne la chiarezza, la capacità discriminativa, il livello di difficoltà. La procedura di try-out comporta l'eventuale modifica degli item ambigui o poco comprensibili o l'esclusione di item non discriminativi e/o ad altissimo o bassissimo livello di difficoltà..
Ciò al fine di garantire prove di valutazione degli apprendimenti valide e affidabili.

Una volta individuato un significato, un vissuto o, comunque, un elemento che intendiamo analizzare, si procede generalmente stendendo brevi frasi o proposizioni relative alle dimensioni, sia quantitative sia qualitative, che riteniamo sottostanti al fenomeno allo studio.

Questa prima ampia versione di item viene sottoposta a una prima valutazione

(o *try-out*) da parte di partecipanti volontari al fine di controllare la chiarezza e la comprensibilità degli item stessi e, nel caso, anche a procedure di *item analysis* (si veda Cap. 6) per evitare espressioni ambigue o generiche, che andranno eliminate o modificate.
Va tenuto presente che nel corso della costruzione degli item si possono commettere diversi errori metodologici: occorre, come già anticipato, evitare ogni forma di ambiguità o di difficoltà di comprensione della domanda. È inoltre necessario non influenzare chi risponde né in modo diretto né in modo indiretto. È utile costruire domande differenti per ciascuno degli oggetti d'indagine evitando domande "doppie" e per questo poco chiare, o domande costruite con forme sintattiche difficilmente comprensibili, con termini specialistici o di uso poco frequente o comunque di significato poco noto.

Caso 5.1
QUESTIONARIO SELF-REPORT ASM (SCALA DELLA MOTIVAZIONE ACCADEMICA)*

Con la *Scala della Motivazione Accademica* (ASM) (Figura 5.1) viene valutata la motivazione allo studio (Vallerand et al., 1992; adattamento italiano originale in Alivernini e Lucidi, 2008; cfr. anche la versione ridotta: Biasi, 2019).
La scala si compone di cinque sottoscale, ciascuna composta da item che hanno l'obiettivo di identificare le ragioni che hanno motivato la scelta del corso di laurea frequentato.
Lo studente è invitato a rispondere alla domanda sul perché sta frequentando il corso di laurea prescelto, e a esprimere il grado di accordo con ciascuna delle 20 risposte proposte su una scala Likert a 10 punti.

Le cinque sottoscale rilevano (tra parentesi le risposte tipiche):

1. *mancanza di motivazione* indicata come *amotivazione* ("Non c'è un motivo particolare, qualcosa la dovrò comunque fare"; "Non lo so, per me una scelta vale l'altra");
2. *motivazione con regolazione esterna* ("Perché qualcun altro mi spinge a fare questa cosa"; "Perché così faccio contento/a qualcun altro/a");
3. *motivazione con regolazione introiettata* ("Perché terminando questo corso di laurea posso far vedere quello che valgo");
4. *motivazione con regolazione identificativa* ("Perché è utile per raggiungere i miei obiettivi nella vita");
5. *motivazione con regolazione intrinseca* ("Perché mi piacciono le materie e le discipline che si studiano").

*Da :Vallerand et al., 1992; versione italiana adattata da Alivernini e Lucidi, 2008.

Items
Il termine inglese "items" viene utilizzato per indicare i singoli elementi di un elenco, di un insieme organizzato.
Applicato alla procedura del Testing si riferisce alla definizione di proposizioni, locuzioni o brevi frasi atte a svolgere da stimolo (o domanda) nell'ambito di un Questionario, di una Scala di valutazione o di un Test di profitto.

È molto importante anche la consegna o istruzione fornita in apertura del questionario. Anch'essa deve essere chiara e non ambigua. Nell'esempio riportato in Figura 5.1 l'istruzione completa risulta essere la seguente: "Pensa, ora, ai motivi per cui stai frequentando il corso di laurea a cui ti sei iscritto/a". Il rispondente viene invitato a valutare quanto ciascuno dei motivi indicati corrisponde alle sue motivazioni: da 0 (per niente) a 10 (completamente).

Precisiamo peraltro che nel caso dei questionari self-report, se il rispondente dichiara di non comprendere appieno la domanda, viene invitato a rileggerla e, al fine di evitare ogni forma di influenza, anche quando dovesse essere presente un intervistatore, questi non può che ripetere la domanda, o al massimo utilizzare una differente forma verbale precedentemente discussa e concordata con il ricercatore.

Per evitare eventuali effetti distorsivi relativi alle aspettative dei soggetti e/o degli sperimentatori, si prevedono spesso item cosiddetti *"inversi"* (che correlano cioè in modo inversamente proporzionale rispetto al fattore che rilevano) per i quali viene attribuito in fase di *scoring* un punteggio opposto rispetto agli altri item. Sempre a fini di controllo di effetti indesiderati si utilizza spesso anche la *rotazione di sezioni di item* (per esempio, una prima somministrazione sarà svolta anticipando i primi dieci item e una seconda, invece, inserendo gli stessi in successione nella seconda metà del questionario).In sintesi, nel caso della versione italiana dell'AMS, ha completato la scala un campione di 618 studenti delle scuole superiori. I risultati di un'analisi fattoriale confermativa hanno evidenziato la struttura a cinque fattori dell'AMS e hanno mostrato invarianza fattoriale tra genere e tipi di scuole.

Academic Motivation Scale Alivernini e Lucidi (2008) (adattata).
Pensa ora, ai motivi per cui stai frequentando il corso di laurea a cui ti sei iscritto/a...

	0	1	2	3	4	5	6	7	8	9	10
1. Perché qualcun altro mi spinge a fare questa cosa.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
2. Così posso far vedere che sono capace di laurearmi.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
3. Non c'è nessun motivo particolare, qualcosa la dovrò comunque fare.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
4. Perché è importante per quello che ho deciso di fare in futuro.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
5. Perché in fondo, il corso di laurea che ho scelto mi piace.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
6. Perché qualcun altro vuole che io lo faccia.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
7 Per dimostrare che sono in grado di riuscire in questa cosa.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
8. Non lo so, per me una scelta vale un'altra.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
9. Perchè è utile per raggiungere i miei obiettivi nella vita.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
10. Perchè le cose che si fanno nel corso di laurea al quale mi sono iscritto/a mi interessano.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
11. Perchè è quella che gli altri vogliono da me.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
12. Perchè terminare questo corso di laurea mi farebbe sentire orgoglioso/a di me.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
13. Onestamente solo perchè sono a costretto/a, dipendesse da me non lo farei.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
14. Perché mi serve per quello che voglio fare nella vita.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
15. Perché mi piacciono le materie e le discipline che si studiano.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
16. Perché così faccio contento/a qualcun altro/a.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
17. Perché terminando questo corso di laurea posso far vedere quello che valgo.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
18. Ad essere sinceri non lo so, sento che perderò solo il mio tempo.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
19. Perché chè è importante per quello che ho scelto di fare.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
20. Perché è bello imparare cose nuove in questo ambito.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figura 5.1 Scala della Motivazione Accademica (AMS). (Da: Vallerand et al., 1992; versione italiana ridotta adattata da Alivernini e Lucidi, 2008; cfr. anche Biasi, 2019.)

Scoring

Si riferisce all'insieme delle procedure tecniche e operative per classificare in termini quantitativi e numerici le risposte dei partecipanti.

In pratica concerne le modalità di attribuzione dei punteggi agli items di un questionario o comunque di risposte a un test.

La versione italiana dell'AMS ha registrato livelli soddisfacenti di coerenza interna e le correlazioni tra le sotto-scale sono risultate coerenti con il modello proposto dalla Teoria dell'Autodeterminazione (SDT, Self Determination Theory). I dati ottenuti hanno infatti confermato, come già indicato sopra, la validità fattoriale e l'affidabilità dell'AMS, supportandone così l'uso nella ricerca educativa sulla motivazione allo studio e al successo accademico.

Caso 5.2

ITEM CORRISPONDENTI AI FATTORI ESTRATTI A SEGUITO DELL'ANALISI FATTORIALE[2]

	ITEM	FATTORI
3.	Non c'è nessun motivo particolare, qualcosa la dovrò comunque fare	*A-motivazione*
8.	Non lo so, per me una scelta vale l'altra	*A-motivazione*
13.	Onestamente perché sono costretto/a, dipendesse da me non lo farei	*A-motivazione*
18.	Ad essere sinceri non lo so, sento che perderò solo il mio tempo	*A-motivazione*
1.	Perché qualcun altro mi spinge a fare questa cosa	*Regolazione esterna*
6.	Perché qualcun altro vuole che io lo faccia	*Regolazione esterna*
11.	Perché è quello che gli altri vogliono da me	*Regolazione esterna*
16.	Perché così faccio contento/a qualcun altro/a	*Regolazione esterna*
2.	Così posso far vedere che sono capace di laurearmi	*Regolazione introiettata*
7.	Per dimostrare che sono in grado di riuscire in questa cosa	*Regolazione introiettata*
12.	Perché terminare questo corso di laurea mi farebbe sentire orgoglioso/a	*Regolazione introiettata*
17.	Perché terminando questo corso di laurea posso far vedere quel che valgo	*Regolazione introiettata*
4.	Perché è importante per quello che ho deciso di fare in futuro	*Regolazione identificativa*
9.	Perché è utile per raggiungere i miei obiettivi nella vita	*Regolazione identificativa*
14	Perché mi serve per quello che voglio fare nella vita	*Regolazione identificativa*
19.	Perché è importante per quello che ho scelto di fare	*Regolazione identificativa*
5.	Perché, in fondo, il corso di laurea che ho scelto mi piace	*Regolazione intrinseca*
10.	Perché le cose che si fanno nel corso di laurea dove mi sono iscritto/a mi interessano	*Regolazione intrinseca*
15.	Perché mi piacciono le materie e le discipline che si studiano	*Regolazione intrinseca*
20.	Perché è bello imparare cose nuove in questo ambito	*Regolazione intrinseca*

[2] Si evidenziano gli item che hanno saturato rispettivamente ai 5 fattori latenti estratti attraverso l'Analisi Fattoriale Confemativa.

5.2.4 Standardizzazione, attendibilità e validità delle scale di valutazione

Le prestazioni individuali ai vari tipi di test o scale di valutazione hanno significato e sono interpretabili solo se confrontate in base a specifici criteri quali le *prestazioni medie di specifici gruppi di riferimento* (per esempio, confronto di perfomance cognitive per la stessa fascia di età), ottenute in *condizioni uniformi*, ovvero *standardizzate*. Standardizzare o tarare un test significa quindi somministrarlo a un ampio campione rappresentativo della popolazione per la quale esso verrà utilizzato. In questo modo si ottengono le misure medie per la variabile allo studio, le quali serviranno come *norme di riferimento*.

Torniamo qui a precisare che l'obiettivo della misurazione è sostanzialmente quello di generare risultati che siano allo stesso tempo attendibili e validi:

- con il termine *attendibilità* si intende la riproducibilità del risultato (un risultato è attendibile se a partire da condizioni simili si ripeterà in tempi differenti;
- con il termine *validità* si fa riferimento al fatto che l'informazione prodotta dalla ricerca o dai test misura accuratamente la variabile o la qualità che si intende misurare.

Conosciamo diversi tipi di validità:

- *validità interna*: corrisponde al grado in cui le conclusioni di una ricerca sono riferibili alle relazioni tra le variabili considerate;
- *validità esterna*: corrisponde al grado di generalizzazione dei risultati ottenuti ad altri soggetti e ad altre situazioni;
- *validità concorrente*: si riferisce alla correlazione tra i risultati al test in esame e i risultati ad altri test che misurano lo stesso costrutto;
- *validità di costrutto*: si riferisce al livello di conformità tra i risultati e la teoria che è alla base della ricerca svolta;
- *validità statistica*: si riferisce al grado di probabilità con cui si verifica la relazione tra variabile indipendente e dipendente.

5.3 Testing e processi di valutazione degli apprendimenti

Come affermava a chiare lettere Visalberghi già nel 1955:

> «Il testing culturale, non meno di quello psicologico, costituisce un ausilio a comprendere meglio il singolo individuo, la persona umana nella sua originalità peculiare; può certo essere distorto ad altri fini, contro la volontà ed i suggerimenti degli studiosi che se ne occupano, ma se un simile timore autorizza al rifiuto, perché non rifiutare in blocco tutte le tecniche su cui si fonda il mondo moderno?
> Nel caso dei test o prove oggettive di profitto un primo equivoco da dissipare è quello che nasce dalla credenza che, con il loro impiego, si sostituisca la "misurazione" alla "valutazione". Ammesso che il "testing" sia senz'altro misurazione nel senso pieno del termine (ciò è vero soltanto per le prove oggettive tipo, corredate da norme precise per l'assegnazione del punteggio), rimane tuttavia assolutamente infondata la credenza che esso elimini ogni giudizio valutativo. Vale in realtà l'opposto: la pratica del testing mette in luce tali e tanti problemi di autentica valutazione da obbligare a un impegno in questa direzione quale raramente vien dato di trovare in altri casi».E ancora: «... L'atto decisivo è quindi un giudizio, una valutazione, in cui i risultati dei test entrano come dati di fatto molto importanti, ma non esclusivi... non ci si accontenta molte volte del giudizio diretto dei competenti, ma si tenta di determinarne la "validità" studiando la loro correlazione con criteri opportunamente scelti. Si tratta allora di valutare questi criteri. In certi casi, come vedremo, i criteri sono costituiti da medie fra i risultati di un gran numero di prove scritte di tipo comune, scelte con speciale accuratezza e corrette con eccezionale attenzione ed impegno. Le capacità valutative di decine di specialisti servono da fondamento alla validità del *test*. Ma il test come tale non valuta più, misura. Esso rappresenta un momento intermedio, di pura misurazione, inserito fra serie complesse di atti valutativi... La misurazione nasce dalla valutazione e nella valutazione confluisce.»
>
> Visalberghi, 1955, pp. 13-14 e 17-18

A livello internazionale, successivamente, autori come Kubiszyn e Borich (in un volume edito nel 2000, giunto a oggi all'11ª ristampa) tornano a più riprese a sottolineare il ruolo centrale del testing in ambito educativo, compresa la professione docente.
La valutazione in ambito educativo viene quindi a essere un atto di attribuzione di valore basata sulla definizione di obiettivi condivisi a livello intersoggettivo, sulla misurazione formale e sulla comunicazione in chiave formativa.
Alla luce delle riflessioni e delle dimostrazioni sopra riportate, riteniamo quindi che una valutazione condotta in modo scientifico debba essere basata quanto più possibile sulla misurazione svolta con prove valide e affidabili. In caso contrario l'analisi e l'interpretazione dei dati rischiano di essere davvero esposte a una soggettività incontrollabile.

5.4. Test oggettivi di verifica degli apprendimenti

5.4.1 Test oggettivi di profitto o prove strutturate

Sono così definite quelle prove di rilevazione e verifica di abilità, conoscenze e competenze relative a uno o più ambiti disciplinari o a determinate aree conoscitive, costituite da una serie di quesiti, domande o stimoli chiusi, ciascuno dei quali è corredato da due o più risposte predefinite.

Ai soggetti cui vengono somministrate tali prove strutturate si richiede di scegliere, tra le risposte proposte per ogni quesito (o item), quella o quelle (a seconda del numero di soluzioni previste) ritenute esatte. Alcune forme di domande possono richiedere una elaborazione autonoma della risposta: in questo caso occorre che l'impiego delle espressioni linguistiche richiesto sia minimo, e che i quesiti abbiano una soluzione univoca della risposta, come nel caso di problemi, esercizi, calcoli ecc., con procedure risolutive pur diverse tra loro, ma il cui esito sia univocamente predefinibile.

La chiusura dello stimolo e della risposta consente di determinare a priori, cioè al momento della costruzione della prova, ovvero prima che lo strumento venga somministrato, il punteggio da assegnare a ciascuna domanda a seconda che la risposta risulti esatta, sbagliata od omessa. Ne deriva che il punteggio attribuibile a una specifica prova non potrà che essere quello e solo quello, chiunque sia il correttore. L'univocità del punteggio evita che l'espressione del giudizio venga influenzata dagli elementi che caratterizzano l'equazione personale di chi compie la correzione. La soggettività, pur sempre presente al momento della decisione di scelta di molte caratteristiche della prova e al momento della interpretazione dei dati valutativi con essa raccolti, viene perciò oggettivata nella verifica attraverso l'univocità del punteggio attribuibile convenzionalmente, grazie alla sua predeterminazione resa esplicita. È questa la fondamentale ma non unica ragione per cui vengono definite "prove oggettive".

Altre importanti ragioni rimandano al fatto che a tutti i soggetti cui la prova viene somministrata si richiede l'esecuzione dello stesso compito, di rispondere cioè alle medesime domande e, generalmente, ma non sempre, nello stesso lasso di tempo. Inoltre, e soprattutto nei test standardizzati, i quesiti rappresentano un'adeguata campionatura delle conoscenze relative a uno o più ambiti disciplinari o aree conoscitive a un livello di complessità determinato, in modo che i punteggi conseguiti dalla somministrazione della prova rappresentino una vera e propria misurazione delle caratteristiche rilevate, rispetto a quelle dell'universo di riferimento.

Approfondimento 5.2

TIPOLOGIE DI PROVE DI VALUTAZIONE DEGLI APPRENDIMENTI

Le prove di rilevazione e verifica degli apprendimenti possono classificarsi in funzione della chiusura e/o dell'apertura dello stimolo e della risposta passando da strumenti oggettivi a strumenti del tutto soggettivi di rilevazione dei dati. Si avranno perciò prove a:

- **stimolo chiuso e risposta chiusa** (prove strutturate o prove oggettive);
- **stimolo chiuso e risposta aperta** (prove semistrutturate);
- **stimolo aperto e risposta aperta** (prove tradizionali come l'interrogazione e il tema libero);
- **stimolo aperto e risposta chiusa** (un nonsenso "pedagogico").

Le prove oggettive di profitto oltre che test oggettivi, sono denominate anche prove strutturate di conoscenza poiché al momento della loro somministrazione vengono presentate, contestualmente strutturate, sia le domande (item, quesiti) che le risposte, per ognuna delle quali va scelta quella o quelle ritenute esatte.

L'affioramento delle abilità, delle conoscenze, dei saperi e delle competenze di chi deve sostenere la prova viene cioè reso possibile non con la richiesta di una elaborazione e scrittura delle risposte a ciascuna delle domande, che richiederebbe una loro soggettiva interpretazione al momento della correzione, bensì attraverso la scelta della o delle risposte ritenute giuste, cioè esatte, tra quelle proposte per ogni quesito.

Una prova oggettiva si presenta quindi come un insieme di domande, quesiti o item, ciascuno dei quali offre alcune alternative di risposta di cui una o più d'una esatta (che diventano così la *chiave di correzione del test*) e le altre sbagliate poste con lo scopo precipuo di far attivare una funzione di discriminazione, setacciamento, scelta, tra le risposte date, di quella ritenuta esatta. Per questa ragione sono chiamate *distrattori*.

La struttura degli item può essere assai varia. Variando la struttura, come vedremo, cambia la tipologia delle operazioni cognitive, oltre che procedurali e operative, richieste all'allievo per la scelta della risposta e, ciò che più conta, cambiano anche la qualità e la significatività degli obiettivi dell'apprendimento che si possono rilevare e misurare.

5.4.2. Tipologia e struttura dei quesiti

La tipologia dei quesiti varia con il variare delle modalità con cui possono essere strutturate domande e risposte. Sono di seguito indicati i più importanti tipi di

item: vero/falso; completamenti; corrispondenze; scelte multiple a una o a due soluzioni. La loro scelta per l'allestimento della prova complessiva dipenderà dalle variabili in gioco nel contesto in cui si attiverà il processo di verifica, dalle specifiche funzioni valutative previste, e soprattutto dagli scopi valutativi e dalle caratteristiche dell'ambito conoscitivo e degli obiettivi cognitivi da rilevare.

Vero/falso

Sono gli item che presentano la struttura più semplice. Consistono infatti in una affermazione che può essere vera o falsa (indicata dalle lettere V/F). Il soggetto cui la prova viene somministrata deve porre una crocetta sulla V o sulla F a seconda che consideri vera o falsa l'affermazione (nella formulazione di questo tipo di quesito, ma anche di tutti gli altri, è buona norma evitare la forma negativa, in particolare le negazioni di negazioni, che complicherebbero inutilmente la decodifica delle domande con il rischio di rendere poco valida la rilevazione). A questa tipologia di item possono assimilarsi tutti quelli che presentano possibilità dicotomiche di scelta della risposta: giusto/errato; sì/no, e simili.

Una variante volta a incrementare il grado di complessità della procedura di risposta consiste nell'introduzione di una terza alternativa di scelta, che può essere per esempio contraddistinta dall'accoppiamento delle lettere VF, per indicare che se l'affermazione può essere vera in certe circostanze può, ciononostante, risultare falsa in altre (in taluni casi si può chiedere agli esaminandi di giustificare la risposta scelta attraverso opportune e brevi argomentazioni).

Approfondimento 5.3

Il procedimento di risposta negli item V/F è assai semplice; le conoscenze che si possono rilevare sono quasi esclusivamente di tipo riproduttivo. In altri termini è abbastanza difficile, se non impossibile, sollecitare con questi item prestazioni che richiedano l'attivazione di competenze e abilità complesse, per esempio di applicazione in contesti nuovi di capacità di analisi, di messa in relazione di concetti, regole o altro di ambiti disciplinari diversi, per la soluzione di un problema pratico, teorico, o di qualunque altro tipo. Sollecitazioni invece praticabili soprattutto con gli item a scelta multipla.

Inoltre, per ognuno di tali item la probabilità di rispondere bene, ma casualmente, cioè senza padroneggiare l'argomento e senza possedere la specifica conoscenza che con la particolare domanda si vuol sottoporre a verifica, è del 50%, cioè talmente alta da compromettere l'affidabilità dell'informazione che dalla risposta data si vuol ricavare. Anche per tale ragione, quando si faccia uso di tali quesiti nella strutturazione di un test, conviene elaborarne in numero cospicuo e in maniera che rilevino abilità omogenee relative a una specifica area tematica o concettuale. In questo modo, tale probabilità diventa statisticamente poco significativa. Molte delle considerazioni fatte valgono in buona parte anche per gli item di confronto e di completamento

Anche in questo caso l'allievo è chiamato a scegliere tra veridicità e falsità di ciascuna affermazione. Nonostante si possano impiegare molti artifici per l'arricchimento e la complicazione degli item V/F, quando sia possibile conviene evitare la costruzione o l'uso di prove oggettive che comprendano solo tale tipo di quesiti. Nel caso si vogliano comunque utilizzarli insieme ad altre forme di item è utile accorparli all'inizio della prova poiché la loro semplicità facilita il processo di familiarizzazione con la pratica del testing. Se ne consiglia in tal caso l'uso nella verifica *in itinere*, volta a migliorare tempestivamente l'attività di insegnamento e apprendimento, per cui l'analiticità delle informazioni sullo sviluppo delle conoscenze è prevalente rispetto all'insieme più aggregato e complesso di quelle da rilevarsi e da sottoporsi a controllo con la valutazione intermedia o finale di un più o meno esteso periodo formativo.

Corrispondenze o item di confronto

Le "corrispondenze" sono chiamate anche item di confronto perché con tali quesiti vengono offerte due serie di dati, concetti, fatti, avvenimenti ecc., e si richiede, appunto attraverso un'analisi comparativa, di mettere in corrispondenza biunivoca ciascun elemento della prima serie con il corrispettivo elemento della seconda.

Gli studenti devono quindi porre in corrispondenza le coppie di dati, appartenenti alle due serie fornite, individuando la relazione che li lega. Ovviamente l'ordine dei dati della seconda serie non deve rispecchiare quello della prima, giacché la prova consiste, in effetti, nell'ordinare secondo ben definite relazioni gli elementi posti alla rinfusa nella seconda serie (non a caso queste domande vengono denominate anche "item di riordinamento"). Se si rammenta, inoltre, che le informazioni che si vogliono trarre dalla somministrazione di una prova devono essere valide (devono cioè rispecchiare le reali competenze possedute dagli allievi), per evitare che l'ultima coppia di dati venga messa in corrispondenza attraverso una meccanica giustapposizione, ovvero senza possedere la o le conoscenze che ne consentono di cogliere la relazione che li lega, conviene presentare nella seconda serie un numero di dati superiore a quelli della prima: anche per indicare l'ultima corrispondenza si dovrà compiere perciò una scelta tra più elementi disponibili.

La struttura degli item di confronto consente di rilevare con essenzialità un ampio repertorio di elementi conoscitivi di base relativi a cospicue aree di conoscenza. Nella elaborazione di tali quesiti è possibile impiegare un'ampia gamma di materiali iconografici, dalle tabelle ai grafici, ai disegni ecc., il cui ausilio può far variare di molto la tipologia delle abilità che con queste domande si possono sottoporre a controllo, anche se la loro struttura si presta prevalentemente per la rilevazione di conoscenze riproduttive, analogiche, "convergenti". Le abilità verificabili non sono perciò molto complesse, così come non lo è la procedura di scelta delle risposte esatte. La probabilità di rispondere bene, ma casualmente, è comun-

que più bassa che negli item V/F, anche se cresce relativamente via via che si passa dalla prima all'ultima scelta degli elementi corrispondenti.

Va infine detto che per evitare la deconcentrazione degli allievi durante la risoluzione dei quesiti, con conseguente abbassamento dell'affidabilità delle informazioni che se ne ricavano, è consigliabile che il numero dei dati della prima colonna non superi le 8-10 unità.

Completamenti (o *cloze*)

Attraverso questa tipologia di quesiti, si chiede di completare un brano opportunamente scelto in funzione degli obiettivi da verificare, dal quale siano stati preventivamente tolti alcuni termini significativi offerti assieme ad altri in calce al brano stesso. Evidentemente occorre evitare che gli elementi lacunosi del brano impediscano all'allievo di coglierne il contenuto o la questione tematica trattata. In caso contrario verrebbe compromessa gran parte della validità della prova. Occorre inoltre evitare, sempre per lo stesso motivo, che i termini impiegati come distrattori e presentati assieme a quelli sottratti dal brano, siano abbastanza plausibili e grammaticalmente accordati alle frasi in cui potrebbero venir collocati. È necessario quindi evitare l'offerta di indizi, in questo caso soprattutto sintattico-grammaticali, che permettano anche in assenza di competenze specifiche una corretta integrazione dei dati omessi nel brano.

Ad ogni integrazione dovrebbero poter verosimilmente corrispondere, inoltre, almeno due termini: quello corretto e il distrattore omologo. Si impedisce così che l'ultimo completamento venga fatto per mero incastro dell'ultimo termine nell'ultimo spazio disponibile, con la conseguente perdita di informazioni sulle conoscenze (lacune) degli allievi.

Anche con tali forme di stimoli che prevedono una procedura di scelta delle risposte non certo complessa, le conoscenze verificabili sono tra le più semplici e generalmente di tipo convergente. Inoltre risulta abbastanza difficile trovare brani che contengano tutti gli elementi necessari, o almeno quelli più significativi, per verificare un cospicuo numero di conoscenze relative all'area tematica e concettuale trattata durante la formazione. È opportuno infine aggiungere che se tali item non sono ben strutturati, oltre a far rilevare le conoscenze-obiettivo della verifica, finiscono col misurare anche certe capacità di lettura che, pur importanti, non possono che essere escluse dalla specifica rilevazione delle abilità possedute dagli allievi sull'argomento trattato dal brano: ancora una volta si 'inquinerebbero' gli obiettivi particolari del controllo che si vuol compiere. Per verificare le abilità di lettura conviene invece somministrare prove costruite appositamente a tal fine.

Scelte multiple

Tra i diversi tipi di quesiti che si possono impiegare nell'elaborazione di una prova oggettiva di conoscenza, le scelte multiple presentano il più alto livello di flessibilità della struttura. Possono consistere, infatti, in un nucleo tematico centrale in cui viene posto un determinato problema, seguito da alcune alternative di soluzione, o in una domanda con più alternative di risposta, oppure in una affermazione incompleta che va integrata scegliendo tra i completamenti proposti quello ritenuto esatto. Questi item offrono la possibilità di impiegare non solo il linguaggio verbale scritto, ma anche diagrammi, tabelle, disegni, mappe ecc., sia per la presentazione del problema, del quesito o dell'affermazione, sia nella strutturazione delle risposte. Lo schema strutturale è comunque quello di una domanda seguita da più di due alternative di risposta.

Possono costruirsi scelte multiple a una o a due soluzioni. Nel primo caso si richiederà di riconoscere l'unica risposta giusta; nell'altro di identificare le due soluzioni esatte tra quelle offerte. È superfluo dire che l'informazione sul numero delle risposte che vanno scelte deve per correttezza precedere le domande e deve essere data sempre prima della somministrazione della prova.

Per evitare, come nel caso degli item V/F, che la probabilità di rispondere bene tirando a indovinare sia alta, le alternative di risposta devono essere almeno quattro o cinque; i distrattori, cioè le alternative di risposta non esatte, devono non solo essere plausibili, ma anche, se possibile, parzialmente veri. Quest'ultimo accorgimento mira anche a evitare che involontariamente vengano offerti agli allievi che non padroneggiano l'abilità sottoposta a controllo alcuni indizi utili per l'identificazione della risposta giusta. Sempre per quest'ultima ragione, è necessario che ciascuna alternativa sia sintatticamente congruente con la domanda posta, e che abbia la stessa lunghezza e complessità delle altre (quando ciò si renda di difficile attuazione per semplificare il lavoro di stesura degli item si possono strutturare coppie di alternative con identica lunghezza ed equivalente complessità sintattica).

La procedura di scelta della risposta esatta nelle scelte multiple è più articolata di quella generalmente richiesta dagli altri tipi di item. Le domande, come si è già detto, possono assumere di fatto la consistenza di problemi anche molto complessi, di situazioni nuove, di contesti simulati, per la cui presentazione si può fare ricorso a tabelle, grafici e altri materiali illustrativi, tali da consentire la rilevazione di una vasta gamma di obiettivi valutativi, dalla semplice conoscenza o comprensione di fatti, regole, principi, ai cosiddetti processi intellettuali superiori quali, per esempio (con riferimento, anche in questo caso, alla tassonomia degli obiettivi cognitivi di B. Bloom), la capacità di compiere analisi, sintesi, valutazioni, cioè di applicare in contesti nuovi e originali le diverse abilità e conoscenze precedentemente acquisite.

Trattandosi degli item che in effetti posseggono il più alto grado di versatilità, è consigliabile che nell'allestimento delle prove oggettive di profitto si faccia di essi un uso preponderante se non esclusivo.

Caso 5.3
ESEMPI DI ITEM

Item Vero/Falso

La somma di due numeri dispari è sempre uguale a un numero dispari: V/F.

La tassa sul macinato consisteva in una tassa da pagare sul prezzo del pane: V/F.

Item di confronto

Nella colonna di sinistra sono elencati i poteri dello Stato previsti dalla Costituzione italiana; in quella di destra è riportata invece una serie di istituzioni, organizzazioni e organi, alcuni dei quali esercitano quei poteri. Scrivi accanto a ciascuna espressione della colonna di sinistra i termini corrispondenti presenti nell'elenco di destra o i numeri che li contrassegnano.

Potere legislativo	1) Magistratura
Potere esecutivo	2) Sindacato
Potere giudiziario	3) Governo
	4) Polizia
	5) Parlamento

Item di completamento

Completare il brano che segue scegliendo i termini necessari nell'elenco riportato in basso. Al posto dei termini si possono scrivere i numeri che li contrassegnano.

La digestione consiste nelle varie trasformazioni che subiscono nell'apparato ..... i ..... e le bevande per poter passare nel ..... ed essere..... da tutte le cellule dell'organismo.
La ..... provvede con il continuo flusso del sangue a rifornire tutte le ..... del nostro organismo delle sostanze di cui si è arricchito mediante la..... e dell' ..... fornitogli dalla respirazione.

1) arterie; 2) cellule; 3) composti; 4) rifornimenti; 5) siero; 6) sangue; 7) cibi; 8) assimilate; 9) circolazione; 10) assorbite; 11) azoto; 12) ossigeno; 13) respiratorio; 14) intestinale; 15) digerente; 16) digestione.

Scelta multipla

La foce di un fiume è il luogo dove:
a) Cominciano a raccogliersi le acque.
b) Le acque possono straripare.
c) Le acque si dividono in due corsi per poi riunirsi più in basso.
d) La portata d'acqua è maggiore che in tutti gli altri punti.
e) Le acque si versano in mare, in un lago o in altro corso d'acqua.

5.4.3 Le prove di comprensione della lettura (PCL)

Tra le quattro abilità fondamentali che l'educazione linguistica si pone come propri obiettivi generali, definibili, ovviamente, a diversi livelli di complessità (*saper parlare*: codificazione della lingua orale; *saper ascoltare*: de-codificazione della lingua orale; *saper scrivere*: codificazione della lingua scritta; *saper leggere*: de-codificazione della lingua scritta), la capacità di lettura, ovvero il *saper leggere* (non solo in senso meccanico), svolge un ruolo di primo piano nell'acquisizione e nella capitalizzazione delle conoscenze da parte di ogni singolo soggetto. La comprensione della lettura rappresenta, nel contempo, un obiettivo posto a fondamento e coronamento dell'apprendimento scolastico, un obiettivo formativo di tipo trasversale, che interessa perciò, quasi tutti gli ambiti curricolari. Saper leggere un testo scritto costituisce infatti la condizione necessaria sia per accedere a gran parte dei saperi curricolari scolastici, sia per arricchire le conoscenze individuali, anche e soprattutto in forma autonoma, una volta fuori dalla scuola. Non è un caso che nella nostra società saper leggere sia unanimemente considerato un vero e proprio *comportamento normativo*.

Presupposto fondamentale del raggiungimento di tali obiettivi è senz'altro la capacità, che deve sempre e progressivamente essere promossa, di impiego della lingua madre come strumento di mediazione nei processi di acquisizione di nuovi linguaggi, specifici e non (da quello matematico, altamente formalizzato, ad altri a più basso livello di formalizzazione).

Il saper leggere correla infatti fortemente con l'esito in quei campi di conoscenza che, come le scienze, la matematica e i linguaggi e i saperi settoriali, sono, sì, scolasticamente orientati, ma presuppongono l'attivazione di indispensabili processi di transcodifica dal codice linguistico al codice che li caratterizza peculiarmente, da un sistema informativo a un altro.

Studi e ricerche hanno rilevato come la mancata promozione della capacità di convertire un linguaggio in un altro e di non saper cogliere e utilizzare le funzioni dei diversi usi della lingua rappresentano, inesorabilmente, la causa principale dell'insuccesso scolastico. Direttamente o indirettamente, le deficienze dello sviluppo del linguaggio e nel tempo l'incapacità di saper leggere correttamente, funzionalmente, di saper cioè decodificare e ricevere informazioni nuove, pertinenti, aperte, ma tendenzialmente univoche da un testo scritto, determinano tra l'altro un abbassamento delle spinte motivazionali all'apprendimento.

Questi fattori che rallentano - e talvolta impediscono - o favoriscono l'acquisizione e lo sviluppo di nuove conoscenze, e che sono al tempo stesso concausa ed effetto dell'insuccesso o del successo scolastico, sono stati poco considerati nelle metodologie didattiche più diffuse nelle scuole. La progressività e la relativa lentezza con cui si possono promuovere abilità linguistiche di alto livello, e la ogget-

tiva necessità di una immersione dei soggetti interessati in ambienti altamente motivanti anche linguisticamente, hanno fatto pensare a una bassa educabilità linguistica degli allievi.

Per la messa in atto degli interventi necessari per promuovere una più diffusa ed efficace educazione linguistica in ambito scolastico occorre assai spesso impiegare opportuni *strumenti diagnostici e valutativi*, utilizzabili nei processi di formazione come nella ricerca educativa, quali sono le prove di comprensione della lettura che, permettendo di verificare sia in modo analitico che complessivo le modalità e i livelli di comprensione di un testo ad accentuata specifica funzione comunicativa - descrittiva, argomentativa, conativa ecc. - consentono di indirizzare meglio e con maggiore specificità le attività di compensazione, consolidamento e promozione delle relative abilità fondamentali in questo vasto ambito conoscitivo e formativo. Si consideri, peraltro, che come si è accennato sopra, gli esiti della somministrazione di PCL correlano positivamente con quelli rilevabili con test oggettivi di conoscenza relativi a molti altri ambiti disciplinari: come dimostrato dalle già citate indagini internazionali sull'apprendimento scolastico (si vedano le ricerche comparative IEA e PISA), i risultati alle PCL possono a ragione considerarsi come veri e propri *indicatori*, seppur indiretti, *dei livelli di conoscenza generale di un soggetto*.

Questi strumenti sono costituiti da un brano di lettura (che dovrebbe risultare sconosciuto ai soggetti cui si somministra la prova) di lunghezza variabile (15-30 righe o poche di più) e appositamente scelto o elaborato dai docenti e/o dai ricercatori in rapporto alle esigenze e agli obiettivi della verifica e/o alle ipotesi di ricerca, seguito da un numero di item, preferibilmente a scelta multipla a una o più soluzioni, ritenuto sufficiente per la rilevazione delle competenze sottoposte a verifica.

La prova consiste nella lettura attenta del brano da parte di ciascun allievo e, come per i test di profitto, nell'attività di scelta delle risposte ritenute esatte tra quelle proposte per ciascun quesito.

Per verificare un primo livello di comprensione si possono porre quesiti volti a rilevare se gli allievi *conoscono e comprendono il significato di termini d'uso corrente*, di *termini d'uso specifico* e delle *definizioni contestuali* (di termini, fatti, concetti, regole, principî ecc.) enunciate esplicitamente nel brano.

A un secondo livello si possono verificare la *capacità di comprensione dei periodi*, quella di cogliere tanto *lo sviluppo longitudinale del discorso*, cioè la sequenzialità logica, temporale, spaziale, causale ecc. dei fatti e degli eventi presentati nel brano, quanto *lo sviluppo trasversale del discorso*, cioè le relazioni esplicite o implicite tra fatti, eventi ecc presenti nel brano.

Infine, a un livello più complesso di comprensione si possono porre quesiti con i quali registrare se gli allievi sanno anche *decodificare alcune tra le più importanti e diffuse figure retoriche* presenti nel brano, se sanno *cogliere tanto il messaggio esplicito quanto il messaggio implicito del testo* letto, e quindi se sanno esprimere una *valutazione critica dello stesso* collocandolo in contesti culturali, storici, di genere, specifici o altro. È peraltro utile verificare se il procedimento di lettura manifestato sia di tipo *top-down* o *botton-up*, ovvero se si possegga la capacità di anticipazione ipotetica delle informazioni che si trarranno dal testo da leggere ancora, sulla base di quelle già acquisite dalla parte già letta come fa il "lettore esperto".

Dell'ampio ventaglio degli obiettivi che si possono rilevare con una prova di comprensione della lettura, quelli citati costituiscono solo una parte, anche se tra i più significativi.

Una volta scelto il brano in funzione delle necessità valutative, per procedere agevolmente alla stesura delle domande è utile formare un'apposita griglia, scrivendo in testa a ogni colonna gli obiettivi generali appena indicati, o altri ancora, e trascritti in colonna gli obiettivi specifici della verifica.

La griglia così composta consentirà una facile analisi e integrazione dei diversi obiettivi di verifica in modo che più domande non risultino sovrapposte, cioè tendenti a verificare un numero superfluo di volte una medesima abilità, e che nessun obiettivo significativo venga omesso dalla verifica.

I procedimenti di correzione, di analisi e di prima interpretazione dei dati, sono del tutto identici a quelli dei test di profitto.

5.4.4 Strutturazione dei quesiti, validità e attendibilità delle prove

La strutturazione delle domande e delle risposte nelle prove oggettive di profitto è operazione assai delicata. Da essa derivano infatti i due requisiti fondamentali che tali strumenti devono possedere in modo davvero spiccato: la capacità di rilevare dati e informazioni sia *validi* che *attendibili*.

Il concetto di *validità* rimanda - come si è già visto - alla capacità di una prova o di uno strumento di misura, di essere in grado di rilevare esattamente quello che attraverso il suo impiego si vuole di fatto rilevare: per conoscere la lunghezza dello spigolo di una scrivania è necessario impiegare uno strumento di misura di lunghezze lineari. Analogamente, se si vogliono rilevare certe abilità o conoscenze matematiche lo strumento deve poter cogliere quelle e solo quelle e non altre abilità o conoscenze: gli stimoli o le domande del test devono pertanto evitare che la pur necessaria mediazione linguistica non consenta al soggetto di comprendere pienamente le domande o le questioni poste, pena la non validità della rilevazione. In caso contrario si misurerebbero semmai competenze linguistiche non già matematiche, con conseguente scarsa o nulla validità dello strumento. Una strutturazione dello stimolo congruente con le caratteristiche dei soggetti cui la prova verrà somministrata - età, livello scolare, competenze linguistiche modali ecc. - incrementa

la validità della rilevazione.

L'*attendibilità* rimanda invece alla possibilità di interpretare in modo univoco gli esiti della rilevazione, di avere misure costanti in successive rilevazioni o in misurazioni compiute da più operatori, potendosi, con buona approssimazione, associare ai requisiti della *costanza* e della *fedeltà* di uno strumento o di una misurazione. Per tornare all'esempio della misura dello spigolo di una scrivania, oltre all'esigenza di impiegare un metro lineare (garanzia di validità della misura) vi è anche l'esigenza che lo strumento sia rigido. Se dovesse altrimenti essere elastico, la lunghezza dipenderebbe dalle forze applicate ai due estremi dello strumento da chi compie le rilevazioni, con conseguente differenziazione delle misure effettuate e della inattendibilità, incostanza e infedeltà dei dati e delle informazioni con il cui impiego si perverrebbe. È operando soprattutto, ma non solo, sulle caratteristiche relative alla chiarezza delle risposte che diviene possibile facilitare rilevazioni attendibili.

In particolare le risposte previste per i quesiti posti con una prova oggettiva di profitto devono essere tutte univocamente precodificabili, così come deve essere predeterminato in termini di punteggio il valore da assegnare a ogni domanda a seconda che la risposta risulti esatta, sbagliata od omessa.

Si sarà compreso, quindi, che le due qualità indicate - validità e attendibilità - determinano in gran parte le caratteristiche formali dei test di profitto.

5.4.5. Quando è necessario standardizzare le prove

La validità può essere assicurata in vari modi. Nei test di classe - *classroom test* - cioè in quelle prove che i docenti elaborano autonomamente soprattutto in attuazione della funzione formativa della valutazione, essa è in un certo senso garantita per *assunzione*: gli stimoli vengono strutturati sulla base di una diffusa conoscenza degli allievi e del grado di approfondimento con cui si è trattato lo specifico ambito conoscitivo in relazione al quale si compie la verifica.

Quando invece si vuol rendere la prova un vero e proprio criterio di confronto, una "pietra di paragone" per la misurazione di ben determinati apprendimenti, avendo come riferimento uno specifico universo di soggetti - per esempio tutti gli allievi in uscita dalla scuola dell'obbligo o tutti i maturandi di un certo ordine di scuola - è necessario procedere a una serie di operazioni tese a misurare e garantire:

- la *validità di contenuto*: ciascun item deve risultare altamente e positivamente correlato con i risultati complessivamente registrati con la prova;
- la *validità concorrente*: i risutati del test devono correlarsi con quelli rilevabili con altri strumenti che riguardano le medesime conoscenze;
- la *validità predittiva* dello strumento.

Tali misurazioni sono espressive, rispettivamente, della congruenza di quel test con misure delle stesse abilità compiute contemporaneamente e, in tempi successivi, attraverso altri strumenti impiegati a loro volta come criteri di paragone. Per la standardizzazione vanno inoltre rilevati, dopo una serie di controlli preliminari (*try-out*), gli *errori standard*, i *coefficienti di fedeltà* e altre caratteristiche della prova: tempi di somministrazione; punteggi grezzi; punti standard (si veda Capitolo 4), indispensabili per compiere confronti, in modo che, ben tarata, la prova possa considerarsi a tutti gli effetti un affidabile strumento di misura della qualità o della caratteristica che con essa si vuole rilevare.

In entrambi i casi l'analisi degli item, dopo una prima somministrazione, consentirà, quantomeno con il calcolo degli *indici di difficoltà e di discriminatività* di ciascun quesito, di verificare la qualità delle domande e delle alternative delle risposte e di conseguenza di migliorarla attraverso un'opportuna elisione, sostituzione o modificazione.

5.4.6 Alcuni modi di classificare le prove oggettive di profitto

Oltre che in test standardizzati e non, le prove oggettive di conoscenza possono venir classificate in rapporto a un ampio ventaglio di criteri che possono andare dal numero dei soggetti presi volta per volta in considerazione - in modo da poter distinguere tra *prove individuali* e *prove collettive* - alle diverse e delicate funzioni della verifica e della valutazione dell'apprendimento. In tal caso si avranno *test diagnostici* e *test prognostici* a seconda che abbiano una spiccata capacità di rilevare conoscenze, abilità, ma anche lacune, in riferimento ai passati apprendimenti dei soggetti, piuttosto che quella di predire il successo o l'insuccesso futuro in determinati ambiti conoscitivi in funzione degli apprendimenti passati e delle competenze presenti. Si avranno, ancora, *test formativi*, così fortemente analitici da far cogliere se si siano strutturati o meno alcuni "anelli" o connessioni funzionali della ipotetica "rete sinaptica" degli apprendimenti, promossa con il processo di formazione, al fine di attivare all'interno dello stesso processo procedure differenziate e/o compensative ben calibrate. Infine, *test sommativi o di prodotto*, per alcuni versi più sintetici, strutturati con quesiti che consentono di cogliere il possesso o meno delle competenze più rappresentative, ad alto peso specifico formativo, di aree e di settori curricolari abbastanza estesi, fino a comprendere conoscenze e competenze sempre più generali o sottese a interi corsi di studio.

In funzione dell'impiego previsto varieranno anche i criteri di costruzione della prova. In generale si può dire che in rapporto agli obiettivi della valutazione andranno campionati, determinati, classificati e analizzati gli obiettivi cognitivi da sottoporre a controllo, anche con l'ausilio delle tassonomie messe a punto da gruppi di studiosi di varie scuole di pensiero.

5.4.7 Come costruire test oggettivi di conoscenza

Un pregio indiscusso delle prove oggettive di accertamento delle conoscenze, che per la loro articolata struttura devono essere elaborate molto prima che vengano somministrate, sta nel fatto che consentono una più accorta determinazione di tutte quelle abilità reputate, dal o dai docenti esperti disciplinari che le elaborano, come davvero rappresentative della padronanza dell'argomento, dell'area tematica, delle questioni, problematiche connesse ecc., relativi all'ambito conoscitivo o al segmento curricolare in rapporto al quale si dovrà compiere la verifica. In tal modo, a differenza di quanto generalmente consentono le prove tradizionali in gran parte estemporanee, si può avere la ragionevole certezza di rilevare i più significativi elementi peculiari dei processi di apprendimento senza quelle clamorose omissioni che spesso costringono tardivamente a dover scoprire nella preparazione degli allievi la presenza di lacune insospettate, persino di veri e propri buchi neri conoscitivi o di competenze procedurali, ma ormai difficilmente superabili.

Per la costruzione delle prove oggettive l'ambito curricolare di riferimento sarà più o meno esteso a seconda delle funzioni valutative che si vorranno attuare. Ci si potrà così riferire a un quadrimestre, a un anno scolastico, a un intero corso di studi se si tratterà della funzione di bilancio complessivo delle conoscenze; a un segmento temporale e curricolare più breve, assai circoscritto, se la funzione mira invece a migliorare il processo formativo in corso.

In ogni caso, per la messa a punto di una qualsiasi prova oggettiva di profitto sarà necessario compiere, preliminarmente, almeno le seguenti operazioni:

1. *determinazione delle abilità-obiettivo* che con essa si dovranno sottoporre a controllo, in modo che i quesiti elaborati garantiscano la rilevazione "a tappeto" di tutte le conoscenze e competenze ritenute rappresentative del segmento curricolare preso in esame, ovvero dell'area dei saperi presa a riferimento per la verifica. Quanto più questo è breve, tanto minore sarà il rischio di omettere obiettivi significativi dalla rilevazione.
2. *analisi degli obiettivi determinati e classificati*, al fine di escludere quelli ridondanti e di integrare l'elenco con quelli inavvertitamente omessi;
3. *campionamento degli obiettivi da sottoporre a controllo* e determinazione della tipologia e del numero degli item da impiegare, quindi dell'ampiezza del test.

5.4.7.1 Determinazione degli obiettivi di verifica

La prima operazione da svolgere sarà dunque quella di identificare gli obiettivi, espressi in termini di abilità, conoscenze, capacità ecc., da sottoporre a verifica, avendo sempre come punto di riferimento le caratteristiche psicologico-strutturali, in particolare cognitive, degli allievi cui la prova verrà somministrata.

Le ormai numerose *tassonomie degli obiettivi cognitivi* facilitano lo svolgimento di tale delicata operazione, anche se è pericoloso considerarle, come purtroppo assai spesso avviene, esaustive e vincolanti. Più accortamente vanno interpretate come validi schemi di riferimento che agevolano e semplificano di molto il lavoro di identificazione della tipologia delle abilità e delle conoscenze da sottoporre a controllo e la stessa stesura dei quesiti. Da questo punto di vista la classificazione degli obiettivi di verifica dell'apprendimento (Bloom, 1956) relativa al *dominio cognitivo*, cioè alle attività intellettuali e logiche, articolati dal più semplice al più complesso (a. conoscenza; b. comprensione; c. applicazione; d. analisi; e. sintesi; f. valutazione), e la classificazione di Kratwol, Bloom e Masia (1964) degli obiettivi connessi al *dominio affettivo, quindi emotivo, motivazionale e valoriale* legati ai processi di apprendimento (a. ricettività; b. risposta; c. valutazione; d. organizzazione; e. caratterizzazione), opportunamente semplificata e adattata, possono rappresentare un utile schema di riferimento per la messa a punto delle prove di accertamento delle conoscenze.

In una griglia, costituita da tante colonne quante sono le classi di obiettivi della tassonomia di riferimento, si possono così trascrivere le corrispondenti abilità-obiettivo identificate con l'operazione di 'smontaggio' e di analisi dei particolari saperi da prendersi in esame con il processo di verifica e rilevazione dei dati valutativi nell'attività di ricerca o in quella formativa, soprattutto in contesti formali di insegnamento, nei quali è richiesta la strutturazione dei giudizi valutativi per favorire l'apprendimento.

Approfondimento 5.4

Com'è facile inferire dal testo, il numero tendenziale degli obiettivi di verifica decresce progressivamente passando dal primo all'ultimo degli obiettivi presenti nelle classificazioni descritte, cioè dal più semplice al più complesso. Di fatto, nella valutazione formativa o formante, cioè da compiersi *per promuovere l'apprendimento* che è senz'altro quella che assorbe il maggior numero di prove perché accompagna progressivamente le procedure didattiche, facendone anzi parte costitutiva e integrante, i termini che hanno significato specifico, cioè il lessico specialistico nelle ipotetiche aree di apprendimento, generalmente può superare in numero il riferimento alle regole, ai concetti ecc., che occorre padroneggiare e dar mostra di possedere e usare nel contesto valutativo specifico.

È utile ricordare che i primi tre livelli indicati nel testo possono rimandare solo ad abilità o "saperi" che rappresentano prevalentemente conoscenze o capacità riproduttive, analogiche, ridondanti, in grado di rispecchiare solamente il possesso di un "pensiero convergente", ripetitivo. Se nella verifica, come purtroppo spesso accade, ci si limitasse a cogliere abilità afferenti a questi tre livelli, si rileverebbero solamente conoscenze riproduttive, acquisibili meccanicamente attraverso il solo impiego della memoria, non già con l'attivazione di un più articolato processo di apprendimento, quello appunto che si vorrebbe promuovere, capace di modificare significativamente la mappa cognitiva del soggetto e le sue capacità procedurali di analisi e di prova delle ipotesi di soluzione dei problemi dati. Per avere la certezza che tale processo abbia luogo, occorre verificare anche il possesso di quelle capacità che sono espressive dei cosiddetti 'processi intellettuali superiori', oggi ridefiniti "competenze", con più netto riferimento alla consapevolezza delle strategie cognitive ed emotive adottate, cioè la capacità di effettuare trasformazioni e adattamenti dei dati, delle informazioni, dei fatti, delle regole, dei principi acquisiti, ovvero delle conoscenze possedute, di passare dal generale al particolare e viceversa, e la capacità di applicarli in situazioni nuove, originali o comunque diverse da quelle che hanno caratterizzato di fatto, o presumibilmente, il momento della loro acquisizione attraverso l'intervento didattico promosso. Per poter rilevare tali competenze, occorre allora *simulare contesti nuovi entro i quali gli allievi devono compiere, o dar prova di saper compiere operazioni di vaglio e analisi critica delle conoscenze possedute*, ovunque e comunque acquisite dentro e fuori i diversi ambiti disciplinari, di selezione, connessione e sintesi di quelle necessarie per la soluzione del problema posto, di una loro applicazione intelligente in senso letterale, perciò consapevole, al contesto teorico o operativo presentato.

Vanno identificate, infine e possibilmente, le condizioni entro cui devono manifestarsi le corrispondenti abilità, conoscenze, competenze, indicate secondo il seguente esempio tipologico: "saper applicare il teorema di Pitagora in una situazione di vita pratica simulata, in cui dovendosi misurare la distanza tra due punti di un piano non è possibile percorrere quella più breve". Definiti in tal modo, sarà tra l'altro abbastanza semplice passare dagli *obiettivi* ai *quesiti* che struttureranno la prova oggettiva di verifica.

5.4.7.2 Analisi degli obiettivi

Una volta determinati gli obiettivi, sarà necessario sottoporli a una particolare analisi per identificarne l'eventuale ridondanza e le più strette connessioni reciproche, in modo che la prova che verrà elaborata risulti massimamente efficace. Si procederà così alla elisione e/o all'integrazione degli obiettivi definiti.

5.4.7.3 Campionamento degli obiettivi, determinazione degli item e della lunghezza del test

In questa, come nelle precedenti operazioni, occorre distinguere tra prove standardizzate e prove di classe, prove valutative scolastiche e prove per la rilevazione di dati valutativi riferiti alla verifica delle ipotesi di ricerca, oltre che distinguere tra le tante possibile e differenti funzioni della verifica: diagnostica, prognostica, proattiva, ovvero come strumento di apprendimento, orientamento e riorientamento, complessiva, analitica, formativa, sommativa, in contesti di istruzione, di formazione, e contesti di ricerca confermativa o esplorativa, eccetera.

Se lo strumento servirà per rilevare e valutare *in itinere* l'andamento degli apprendimenti, è ovvio che si dovranno cogliere analiticamente tutte quelle abilità capaci di dar conto delle caratteristiche del processo di formazione: il campionamento non è necessario, dato che occorre rilevare tutti gli obiettivi determinati. In tal caso una certa ridondanza degli obiettivi di verifica può risultare addirittura utile. La prova deve essere abbastanza analitica, così come netta deve essere la distinzione del possesso delle diverse tipologie di abilità e conoscenze. Sarà infatti l'analisi differenziata degli esiti della somministrazione a consentire l'attivazione di itinerari diversificati di consolidamento e/o recupero altamente efficaci.

Ben altro è il caso di una verifica complessiva da fare su una sezione curricolare molto estesa, in cui ha invece senso cogliere le abilità più rappresentative che per ovvie ragioni non possono che essere più aggregate delle prime. Altra cosa ancora è procedere alla messa a punto di test standardizzati che assumeranno il valore di veri e propri strumenti di misura delle conoscenze della popolazione sottoposta a controllo, sulla base dello scarto che presentano con quelle standard dell'universo in rapporto al quale sono stati elaborati.

In questi ultimi due casi, segnatamente nell'ultimo, le conoscenze e le abilità, ovvero gli obiettivi della verifica, devono rappresentare un campione *rappresentativo* dell'intero universo delle conoscenze o abilità-obiettivo relative allo specifico campo conoscitivo e ai livelli di età o agli anni di corso dei gradi o degli ordini di scuola presi in esame.

Dato che i referenti privilegiati dalla presente trattazione sono i giovani ricercatori e i docenti che avvertono la necessità di elaborare in proprio, individualmente o in piccoli gruppi, gli strumenti utili alla conduzione del lavoro formativo quotidiano, dall'osservazione sistematica alle micro-sperimentazioni e simili, non già gli ipotetici costruttori di prove standardizzate, è utile illustrare alcuni dei più agili criteri di determinazione della estensione o lunghezza di un test di verifica sommativa.

La procedura più semplice impiega come criterio della determinazione degli obiettivi da sottoporre a controllo - ovvero i quesiti, il loro numero, e perciò l'e-

stensione stessa del test - la quantità di tempo utilizzato per la trattazione delle questioni tematiche e dei contenuti relativi a ciascun obiettivo o a loro insiemi omogenei.

Considerando con buona approssimazione i diversi tempi, espressi in ore (o in altre unità di misura) dedicati anche attraverso l'allestimento di specifici contesti di apprendimento ad ogni argomento o questione concettuale, con i semplici calcoli richiesti dalla seguente formula si può determinare l'importanza assegnata alle singole parti del curricolo o, nel caso di un'indagine scientifica, agli obiettivi di ricerca corrispondenti espressa in valore percentuale:

$$Ip = Tp / Tc \times 100$$

in cui:
Ip = importanza assegnata a ciascuna delle parti;
Tp = tempo dedicato a ciascuna delle parti o peso specifico attribuito a ciascun obiettivo di ricerca;
Tc = tempo complessivo dedicato a tutte le parti e a tutti gli obiettivi.

Come si può notare, la logica sottesa alla formula di calcolo assegna maggiore importanza agli obiettivi più rilevanti per la ricerca o per il raggiungimento dei quali si è dedicato un più lungo tempo nel processo di formazione o perché hanno un maggiore peso specifico nell'economia formativa complessiva dell'area conoscitiva presa a riferimento. Anche se ciò non è sempre vero, lo è, generalmente, in gran parte dei casi, così che è possibile accettare la formula come sufficientemente indicativa del valore relativo delle abilità e delle conoscenze che si vorrebbero/dovrebbero sottoporre a controllo con quello strumento.

La percentuale degli item di accertamento delle abilità corrispondenti a ognuna delle parti considerate rispecchierà quella che verrà dunque assegnata a ciascuna di queste attraverso il calcolo indicato. La successiva scelta dei livelli di complessità tassonomica in rapporto ai quali si struttureranno i quesiti potrà effettuarsi con una procedura empirica, di buon senso, oppure applicando la stessa formula alla griglia degli obiettivi elaborata. Si saprà così quante domande dovranno formularsi per rilevare, in relazione agli obiettivi di ricerca o ad ogni parte del curricolo, o sezione dell'area dei saperi considerati, le differenti conoscenze.

Per costruire prove oggettive occorre, in sintesi:

Definire i contenuti e gli obiettivi della verifica;
Stabilire la durata della prova;
Scegliere il tipo di item;
Scegliere il modello di valutazione: a "*criterio*" o a "*norma*";

Attribuire i punteggi grezzi;
Somministrare la prova nella versione preliminare (*try-out*);
Calcolare *indici di difficoltà, discriminatività, selettività e affidabilità*;
Selezionare (e accettare) gli items validi e attendibili, ovvero con buoni indici di difficoltà e discriminatività.

Approfondimento 5. 5 (di V. Biasi)

CALCOLO DEGLI INDICI DI DIFFICOLTÀ, DISCRIMINATIVITÀ, SELETTIVITÀ E AFFIDABILITÀ DEGLI ITEM DI UNA PROVA DI VALUTAZIONE DEGLI APPRENDIMENTI

Indice di difficoltà (ID) di un item. E' espresso dal rapporto tra i punti ottenuti ad un item da tutti gli allievi e il punteggio massimo ottenibile se tutti rispondessero in modo corretto.:

$$ID = P_{tot}/P_{max}$$

L'indice di difficoltà varia tra 0 a 1:

- se ID = 0 l'item è troppo difficile (nessun allievo è riuscito a dare una risposta corretta);

se ID = 1 l'item è troppo facile (tutti gli allievi hanno risposto correttamente). Si può dire quindi, in modo più semplice, che l'indice di difficoltà è dato dalla percentuale di risposte sbagliate (o omesse) e sale col crescere di tale percentuale.

Indice di discriminatività di un item o potere discriminante (DP). Indica quanto l'item riesce a discriminare tra chi ha conseguito gli obiettivi di apprendimento previsti e chi ancora no. È dato dal prodotto tra il numero di risposte esatte (E) e il numero di risposte sbagliate date all'item (S), rapportato alla metà del numero totale di risposte (N elevato al quadrato).

$$\mathrm{DP} = \frac{E \cdot S}{\left(\frac{N}{2}\right)^2}$$

Il potere discriminante varia tra 0 ed 1:

- se DP = 0 significa che tutti gli studenti hanno risposto in modo corretto o in modo errato, e l'item non discrimina tra preparati e meno preparati;
- se DP = 1 significa che metà degli studenti ha risposto correttamente e metà in modo errato, quindi il potere discriminante è massimo.

Indice di selettività (IS) di un item Indica quanto l'item è coerente con l'insieme degli item della prova.

È espresso dalla differenza tra il numero di risposte esatte date all'item da 1/3 di studenti con i risultati migliori nell'intera prova (N_m) e il numero totale di risposte esatte all'item date da 1/3 di studenti che ha ottenuto i risultati peggiori nell'intera prova (N_p), rapportato a 1/3 del numero degli allievi:

$$IS = \frac{N_m - N_p}{\frac{N}{3}}$$

L'indice di selettività varia tra -1 e +1

- se IS = -1 significa che gli studenti che hanno conseguito alti punteggi hanno risposto tutti in modo errato all'item mentre coloro che hanno conseguito punteggi bassi hanno risposto tutti in modo corretto all'item (selettività rovesciata);
- se IS = 1 significa che chi ha ottenuto punteggi alti nella prova ha risposto correttamente all'item e chi ha ottenuto punteggi bassi nella prova ha risposto in modo errato all'item;
- se IS = 0 significa che l'item non è selettivo: studenti più preparati e studenti meno preparati rispondono in modo corretto all'item nella stessa misura.

Indice di affidabilità (IA) di un item. Mette in evidenza come un buon item debba essere accessibile ed allo stesso tempo discriminante.
E' espresso dal prodotto dell'indice di difficoltà e dell'indice di selettività:

$$IA = ID \cdot IS$$

L'indice di affidabilità varia da -1 a 1:

- se IA < 0 l'indice di selettività è negativo;
- se IA = 0 l'item è scarsamente utile nella prova;
- se IA = 1 l'item è utile nella prova, perché coerente con gli altri e accessibile.

5.4.8 Prove semistrutturate di verifica degli apprendimenti: caratteristiche e tipologia

È fuori discussione che le prove oggettive rispetto a quelle "tradizionali" (con le quali alla soggettività interpretativa degli stimoli offerti con la prova si aggiunge quella di chi la corregge e la valuta) consentano una migliore rilevazione delle informazioni quali-quantitative sui processi e sui prodotti dell'apprendimento, individuali e collettivi, rilevazione che può avvenire peraltro in brevissimo tempo. E tuttavia tutto ciò non basta a giustificarne il loro impiego esclusivo tanto nei processi valutativi in contesti formali, quanto nelle attività di ricerca in campo educativo.

Vi sono abilità, saperi, conoscenze e competenze che possono essere oggetto di studio della ricerca educativa, per rilevare i quali in modo affidabile è necessario

impiegare strumenti differenti non solo da quelli tradizionali, ma anche dai test oggettivi, "carta e matita". Si pensi, e solo per fare qualche esempio, alla comprensione e alla produzione del linguaggio orale, alla capacità di integrare il sapere al saper fare, come si richiede in molte attività di laboratorio, alla costruzione di modelli o di macchine semplici, alla manipolazione di materiali, alla capacità di applicare in contesti del tutto nuovi i saperi ovunque e comunque acquisiti e/o posseduti, e così via.

È quindi facile dedurre che *per ogni caratteristica tipologica delle abilità o conoscenze da rilevare in base alle ipotesi di una ricerca, e/o funzione valutativa da svolgere, è necessario impiegare uno strumento di accertamento delle competenze la cui struttura sia in un certo senso omologa a quella particolare caratteristica e a quella specifica funzione conoscitiva*: una regola aurea docimologica, questa, non sempre ricordata, purtroppo, anche nei più accreditati "trattati" valutativi.

Se ne ricava che ogni particolare tipo di prova, qualunque siano le sue distintive caratteristiche metriche e strutturali, può ben diventare, in particolari contesti formativi e/o di ricerca, il migliore dei possibili strumenti di rilevazione dei dati valutativi quando sia strutturalmente omologa agli obiettivi della verifica. La questione è semmai quella della necessaria consapevolezza delle caratteristiche metrologiche, e quindi del trattamento che lecitamente si è autorizzati a compiere con i dati e le informazioni assunti con il particolare strumento impiegato.

Tra i due opposti poli delle prove tradizionali, a forte componente "metrica" soggettiva, e di quelle oggettive, vi è infatti tutta un'estesa e articolata gamma di strumenti di rilevazione dei dati valutativi assai importanti che potrebbero invece contribuire a differenziare e a rendere sempre *adeguate e pertinenti* le prove agli scopi della verifica e alle caratteristiche dei saperi e delle abilità da rilevarsi volta a volta nelle differenti attività di ricerca.

Tra i tanti strumenti valutativi che si possono proficuamente impiegare assumono una importanza del tutto particolare le *prove semistrutturate* (Domenici, 2005) così chiamate perché al momento della loro somministrazione rispetto alle prove oggettive presentano ben strutturata solo la parte che sollecita la manifestazione delle competenze da rilevare e verificare.

Sono infatti *prove a stimolo chiuso e risposta aperta*, ma aperta in modo che, nella quasi generalità dei casi, l'allievo sia per così dire, costretto, dalla specificità della sollecitazione, a porre in netta evidenza il possesso o non possesso delle conoscenze sottoposte a controllo e a consentire a chi le corregge di attribuire in forma tendenzialmente univoca e oggettiva i punteggi (così assicurando una buona validità e attendibilità della rilevazione).

Approfondimento 5.6

COSA SONO LE PROVE SEMISTRUTTURATE DI CONOSCENZA

Consistono, in generale, in una serie articolata di quesiti che richiedono non già di scegliere - come per le prove oggettive - l'alternativa esatta tra quelle contestualmente offerte, bensì di elaborare autonomamente le risposte, osservando ben precisi *vincoli prescrittivi* indicati esplicitamente. In tal modo le risposte possono confrontarsi con i corrispondenti criteri di correzione opportunamente predeterminati, rendendo agevole l'identificazione dei livelli di accettabilità delle prestazioni, della scala di misura e dei punteggi da assegnarsi.

Quando per le caratteristiche della specifica prova non risulti possibile formulare univocamente le risposte attese per ciascun quesito, in modo che fungano da *criterio oggettivo di riferimento per la verifica e la valutazione*, si dovranno predeterminare i livelli di accettabilità delle risposte, cioè quali e quanti contenuti informativi queste devono necessariamente veicolare per poterle ritenere accettabili. Operazioni che svolte al momento della costruzione della prova evitano che i quesiti definitivi risultino ambigui o generici.

La relativa apertura della risposta rende estremamente flessibili questi strumenti di rilevazione e/o di formazione che più di altri permettono, per esempio, di verificare e/o promuovere i cosiddetti *processi intellettuali superiori, la capacità di porre in relazione ciò che apparentemente è irrelato, di applicare in contesti nuovi, opportunamente simulati, i saperi disciplinari e multidisciplinari acquisiti, di verificare e/o strutturare, in altre parole vere e proprie competenze*, intendendo per competenza la capacità da parte di un soggetto di impiegare schemi d'azione e/o strategie di risoluzione di un problema in un contesto definito, impiegando tutte le conoscenze possedute, ovunque e comunque acquisite, così da compiere - contestualmente - una vera e propria *metacognizione*, ma anche un controllo procedurale della bontà dello schema/strategia adottato, e perciò una *metavalutazione* tale da permettere l'assunzione di decisioni di intervento risolutivo anche grazie alla riflessione critica della pertinenza o adeguatezza della strategia prescelta al contesto dato (*metadecisione*) per poterla evidentemente cambiare al fine di giungere a un risultato accettabile (Domenici, 2000).

Tra le più importanti prove semistrutturate (di alcune delle quali si illustrano di seguito le principali caratteristiche) vanno annoverate senz'altro le *domande strutturate*, i *riassunti*, i *saggi brevi*, i *rapporti di ricerca*, il *colloquio strutturato*, la *riflessione parlata*, la *simulazione di contesti*. Occorre tuttavia far presente che ogni possibile prova di verifica quando possegga i requisiti formali, ovvero strutturali metrici, sopra indicati, può ad ogni effetto definirsi semistrutturata.

Ovviamente, *l'apertura delle risposte e la necessità di predeterminare il criterio di accettabilità di ciascuna di esse, impone che le domande sollecitino la manifestazione di abilità, saperi, competenze ben definibili*, che rimandino a un ventaglio delle risposte possibili ristretto e prevedibile almeno nei contenuti informativi essenziali, che siano chiare, semplici, specifiche, univocamente interpretabili, e che vengano indicati i vincoli delle risposte (lunghezza massima consentita; contenuti informativi da veicolare ecc.). In tal modo si potranno facilmente interpretare e confrontare le risposte degli allievi con le risposte-criterio formulate in fase di costruzione della prova.

Ben oltre la banale rilevazione di abilità riproduttive o convergenti, le domande strutturate andrebbero impiegate per verificare soprattutto capacità "divergenti" di pensiero, poiché per la loro struttura facilitano la simulazione di veri e propri scenari problematici inusuali, quindi la richiesta di impiegare tutte le conoscenze possedute per risolvere questioni specifiche entro i vincoli dati.

La *validità della rilevazione* discende dalla precisione con cui si specifica il compito richiesto. Lo stimolo chiuso evita l'arbitrarietà interpretativa della sollecitazione permettendo di valutare le risposte come davvero rappresentative delle competenze possedute. L'*attendibilità* cresce con il crescere dell'univocità della prestazione richiesta, ovvero con l'incremento della possibilità di predeterminare le caratteristiche peculiari essenziali, che le risposte devono possedere per poterle considerare adeguate o non adeguate al criterio valutativo.

Sono perciò le condizioni "imposte" attraverso le indicazioni di "lavoro" che i soggetti cui si somministrano devono svolgere e la connessa predeterminazione dei criteri e dei punteggi attribuibili ai diversi aspetti della prestazione che permettono di classificare tale prova come semistrutturata.

In rapporto alle "consegne" si elaborerà - come si è detto - uno schema analitico dei punti essenziali, o di uno spettro coerente di essi, che devono comunque essere presenti nelle risposte per considerarle accettabili e pertinenti, e per ciascuno di essi si stabiliranno i punteggi da attribuire.

Senza lasciarsi influenzare dallo stile espositivo, la correzione consisterà nel controllo della presenza in ogni prova delle componenti e dei punti predeterminati come essenziali o vincolanti per ognuna delle indicazioni date al momento dello stimolo, e nell'assegnazione dei relativi punteggi e di quello complessivo. In stretto rapporto con il valore attribuito a ciascuno dei criteri generali sopra indicati, e ad altri ancora, si potrà determinare una vera e propria *scala descrittiva dei punteggi* da assegnarsi sia a ciascuna delle componenti dei criteri valutativi sia alla prova considerata nel suo complesso, *integrando*, così, *metodi analitici e olistici di correzione*.

5.4.9 Colloquio strutturato, semistrutturato e riflessione parlata come strumenti di ricerca

Tra le tante possibili prove semistrutturate che si possono ideare e utilizzare soprattutto a fini di ricerca, assumono particolare rilievo, assieme alle domande strutturate, il colloquio strutturato e la riflessione parlata.

Le *domande strutturate*, sono costituite da:

- un'introduzione che può contenere, quando necessario, grafici, tabelle materiali audio-video e simili, nonché i vincoli formali da rispettare per la stesura per iscritto delle risposte all'intera prova;
- una serie di domande principali;
- sottodomande che articolano le domande principali, accompagnate entrambe dalle indicazioni dei particolari vincoli di risposta.

L'area tematica oggetto della verifica può essere disciplinare o interdisciplinare, dipendente o indipendente dai contenuti specifici d'insegnamento in contesti formali.

L'articolazione dei quesiti deve essere di tipo gerarchico, ma non tale da produrre il cosiddetto "effetto a cascata" della qualità delle risposte alle domande o sottodomande successive da quella delle risposte date alle domande precedenti. In altri termini, il livello di accettabilità di una risposta non deve ripercuotersi sulle successive.

Ovviamente, l'apertura delle risposte e la necessità di predeterminare il criterio di accettabilità di ciascuna di esse, impone che le domande sollecitino la manifestazione di abilità ben definite, che rimandino a un ventaglio di possibili risposte ristretto e prevedibile almeno nei contenuti informativi essenziali, che siano chiare, semplici, specifiche, univocamente interpretabili, e che indichino i vincoli delle risposte (lunghezza massima consentita; contenuti informativi da veicolare ecc.). In tal modo si potranno facilmente interpretare e confrontare le risposte date con le risposte-criterio formulate in fase di costruzione della prova.

Ben oltre la banale rilevazione di abilità riproduttive o convergenti, le domande strutturate andrebbero impiegate per verificare soprattutto capacità "divergenti", poiché per la loro struttura facilitano la simulazione di veri e propri scenari problematici inusuali, quindi la richiesta di impiegare tutte le conoscenze possedute per risolvere questioni problematiche specifiche.

Caso 5.4

Alcune tipologie di prove semistrutturate

Domande strutturate (Quesiti a trattazione sintetica)
Saggio breve
Riassunto
Prove di simulazione di contesti
Colloquio orale
Riflessione parlata
Relazioni di laboratorio

Principali caratteristiche delle prove semistrutturate

Stimolo chiuso / Risposta aperta
Vincoli prescrittivi
Predeterminazione dei criteri di correzione
Tendenziale maggiore oggettività del giudizio
Valorizzazione delle abilità di problem solving divergente, cioè uso creativo delle conoscenze di cui si è in possesso

La *validità* della rilevazione discende dalla precisione con cui si specifica il compito richiesto. Lo stimolo chiuso evita l'arbitrarietà interpretativa della sollecitazione permettendo di valutare le risposte come davvero rappresentative delle competenze possedute dal soggetto.

L'attendibilità cresce con l'incremento della possibilità di predeterminare le caratteristiche peculiari che le risposte devono possedere per poterle considerare accettabili o non accettabili.

Il *colloquio strutturato*, il *colloquio semistrutturato* e la *riflessione parlata* sono strumenti che, impiegando il linguaggio verbale orale, offrono la possibilità di: (a) cogliere il punto di vista dell'allievo o saggiarne le specifiche competenze su una particolare questione; (b) conoscere le strategie e i processi logici messi in atto nei processi di analisi e soluzione di determinati problemi in contesti ben definiti, nonché la capacità di riflessione metacognitiva e metavalutativa; (c) determinare il grado di flessibilità cognitiva e la capacità di impiego delle conoscenze possedute e delle informazioni disponibili, o volta a volta offerte nel contesto e nella soluzione di problemi posti in modo estemporaneo.

Per perseguire ciascuno degli obiettivi generali di verifica appena indicati, occorre poter scegliere in ogni contesto di ricerca la forma-strumento ritenuta più opportuna.

Occorre ricordare che a differenza di molti strumenti di rilevazione dei dati valutativi o di molte altre prove di verifica dell'apprendimento, il *colloquio* è comunque caratterizzato da un'interazione diretta tra ricercatore e ognuno dei membri appartenenti agli eventuali gruppi o campioni di ricerca, o tra docente e allievo, spesso senza altra mediazione che non siano la parola e i gesti. Questo suo elemento di forza è però anche elemento di debolezza, in quanto il contesto diventa per così dire saturo degli elementi caratteristici della comunicazione verbale e non verbale. In altri termini, il colloquio si svolge e si risolve in stretto rapporto con un numero di fattori estemporanei più alto e imprevedibile che in ogni altra forma di rilevazione di dati e/o di verifica, fattori che, a loro volta, diventano nel contempo concausa ed effetto di variabili concatenate, non sempre preventivamente determinabili (*variabili parassitarie*). Occorre perciò molta maestria per sollecitarlo e condurlo in modo efficace e produttivo.

Soprattutto per tale ragione, il *colloquio*, analogamente all'*intervista* - alla quale è per molti versi assimilabile - è considerato uno strumento di rilevazione poco valido e poco attendibile. Tuttavia il grado della validità, e in qualche modo persino dell'attendibilità delle rilevazioni, cresce con il crescere del livello di strutturazione degli stimoli e delle modalità di codifica degli esiti processuali della "conversazione", che ovviamente variano col variare delle finalità d'impiego dello strumento.

Se lo scopo prevalente del colloquio non è il controllo delle conoscenze relative a una particolare questione tematica, bensì quello di rilevare capacità descrittive, narrative, argomentative, di ragionamento in contesti concettuali od operativi simulati, di cogliere il grado di flessibilità e di originalità d'impiego delle più diverse conoscenze comunque acquisite e in qualsiasi tempo, in uno o più ambiti disciplinari, di formulare ipotesi risolutive di un qualsiasi problema i cui termini vengono variati nel contesto di svolgimento, allora il grado di strutturazione delle domande dovrà essere inferiore a quello del colloquio strutturato e potrà persino tendere al libero scambio di opinioni su una certa questione (limite che può però trasformarsi in una interazione verbale del tutto casuale). Il colloquio diventerà, allora, *semistrutturato*, con una componente, perciò, relativamente libera, non predeterminata.

In una apposita griglia, oltre a poche domande principali, si potranno riportare gli aspetti descrittivi delle qualità delle prestazioni attese, mentre si registreranno sinteticamente durante il dialogo, o immediatamente dopo, i dati informativi su aspetti significativi ma non previsti del colloquio. Quest'ultimo aspetto della registrazione risulta particolarmente importante poiché molti degli obiettivi da verificare possono o risultare indipendenti dagli argomenti di studio o essere metadisciplinari, e perché in questi casi l'interazione tra docente e allievo, ricercatore e componente di un campione di ricerca, diventa assai complessa e articolata per effetto di chiarificazioni, modificazioni estemporanee di ipotesi, tesi, punti di vista.

Altrettanto importante risulta la modalità di conduzione del colloquio. È quasi pleonastico ricordare che non di meno in ambito scolastico anche il dialogo apparentemente disinteressato, o assai prossimo alla più genuina spontaneità, avviene in una situazione di disparità tra chi lo conduce (docente/ricercatore) e il suo interlocutore (allievo/componente di un campione di ricerca), e che per renderlo davvero tale è necessario dar corso a una comunicazione a due vie che a sua volta presuppone un continuo controllo della ricezione delle informazioni trasmesse da ciascuno degli attori.

A tal fine è necessario: chiarire preliminarmente il tema e gli scopi del colloquio; iniziare con una domanda generale, ma non generica; evitare che l'eventuale scarto tra competenza linguistica orale passiva (comprensione) e attiva (produzione) inibisca l'espressione della risposta; ascoltare con attenzione e disposizione positiva, ma sempre in modo intelligentemente critico; evitare di assumere atteggiamenti mimici e verbali autoritari o di dissenso netto verso quanto si afferma erroneamente; evitare, ancora, pause troppo lunghe in caso di mancata risposta offrendo, anzi, agganci concettuali per la prosecuzione del discorso, anche in presenza di divagazioni troppo vistose; chiarire senza impazienza eventuali incomprensioni verbali, evitando ripetizioni pedisseque di domande non comprese e, come purtroppo talvolta accade, con tono di derisione, esattamente come ci si deve comportare quando si conduce una intervista.

Con la *riflessione parlata* si è, invece, in presenza di una vera e propria tecnica diagnostica dei modi e dei modelli di ragionamento attivati da ciascun allievo/componente di un campione di ricerca, nel tentativo di risolvere un ben determinato problema.

Con la riflessione parlata è possibile sia indagare e conoscere le modalità prevalenti attraverso cui ciascun soggetto posto di fronte a particolari situazioni problematiche, soprattutto di carattere pratico, usa e verifica schemi concettuali, elabora strategie, organizza conoscenze e informazioni, cerca soluzioni; ma è altresì possibile promuovere e valorizzare anche le competenze connesse.

È una prova semistrutturata *perché* - seppur in parte - *sono predefinibili le sollecitazioni iniziali, e si può registrare la qualità delle risposte date, o più precisamente, del ragionamento ad alta voce che il soggetto compie via via, potendo quindi risalire alle caratteristiche dei processi cognitivi attivati.*

Si comprende, allora, l'importanza che assume la strutturazione di situazioni e contesti entro cui attivare la riflessione parlata, senz'altro artificiali, ma dentro i quali il soggetto possa muoversi come nella più spontanea delle situazioni possibili, poiché acquisiscono senso e significato non già i concetti, le teorie, le conoscenze possedute, ma il modo con il quale sono organizzati nella struttura cognitiva del soggetto e come vengono utilizzati negli schemi di ragionamento manifestati in situazione.

Riepilogo

In questo capitolo si sono descritte le fasi di costruzione di alcuni strumenti di valutazione o test, attività indicata appunto come "testing" e applicata nel nostro caso a variabili di natura educativo-formativa.

Sono stati presentati i concetti di item, indicatore, fattore, costrutto, teoria, e descritte le principali tecniche di costruzione e validazione di strumenti quali test o questionari.

Particolare attenzione è stata rivolta alla definizione degli item, all'attribuzione dei punteggi (o *scoring system*), alla identificazione di fattori o dimensioni di una scala di valutazione o questionario, operazione empiricamente validata grazie ad analisi statistiche pertinenti quali l'analisi fattoriale e all'intero processo di standar-

dizzazione di un test.

È stato inoltre ricordato, in uno sguardo storico, lo sforzo condotto in modo continuativo da studiosi italiani a partire dagli anni Cinquanta (e ripreso anche più recentemente da colleghi stranieri), nel mettere a fuoco il ruolo centrale del testing applicato ai settori pedagogici e formativi.

Il capitolo presenta infine le caratteristiche formali nelle prove semistrutturate di verifica e valutazione degli apprendimenti, prove di grande rilievo sia nelle attività di ricerca, sia in quelle didattico-formative.

Riferimenti bibliografici

Alivernini F., Lucidi F. (2008). The Academic Motivation Scale (AMS): Factorial structure, invariance and validity in the Italian context. *Testing, psychometrics, methodology in applied psychology*, 15(4), 211-220.

Anastasi A. (1999). *I test psicologici*. Milano: Franco Angeli.

Anastasi A., Urbina S. (1997). *Psychological testing*, 7th ed. Upper Saddle River (NJ): Prentice Hall.

Antiseri D. (1986). *Introduzione alla metodologia della ricerca*. Torino: SEI.

Biasi V. (a cura di) (2019). *Counselling universitario e orientamento. Strumenti e rilevazioni empiriche*. Milano: LED.

Bloom B.S., Engelhart M.D., Furst E.J et al. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York: David McKay Co.

Bonaiuto P. (a cura di) (1971). *Concetti e termini per la ricerca psicologica*. Università degli Studi di Bologna.

Boncori L. (1993). *Teoria e tecniche dei test*. Torino: Boringhieri.

Boncori L. (2006). *I test in psicologia. Fondamenti teorici e applicazioni*. Bologna: Il Mulino.

Debesse M. Mialaret G. (1969). *Traité de sciences pedagogiques*. Paris: PUF (trad. it. *Trattato delle scienze pedagogiche*. Roma: Armando, 1980).

Domenici G. (1981). *Descrittori dell'apprendimento*. Firenze: Giunti & Lisciani.

Domenici G. (a cura di, 2000). *La valutazione come risorsa*. Napoli: Tecnodid.

Domenici G. (a cura di, 2005). *Le prove semistrutturate di verifica degli apprendimenti*, UTET, Torino.

Goode W., Hatt P.K. (1952). *Methods in social research*. New York: McGraw-Hill.

Kratwol D.R., Bloom B.J., Masia B.B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: the affective domain*. New York: David McKay.

Kubiszyn T., Borich G.D. (2020). *Educational testing and measurement*. 11th ed. New York: Wiley & Sons.

Lowell K., Lawson K.S. (1972). *La ricerca nel campo educativo*. Firenze: Giunti-Barbera.

Rubini V. (1987). *Test e misurazioni psicologiche*. Bologna: Il Mulino.

Sartori R. (2008). *Tecniche proiettive e strumenti psicometrici per l'indagine della personalità. Approccio idiografico e approccio nomotetico a confronto*. Milano: Led.

Vallerand R.J., Pelletier L.G., Blais M.R. et al. (1992). The academic motivation scale: a measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, 52, 1003-1017.

Vertecchi B. (1986). *Manuale della valutazione*. Roma: Editori Riuniti

Visalberghi A. (1955). *Misurazione e valutazione nel processo educativo*. Milano: Edizioni di Comunità.

6

Item analisi e misure di tratti latenti[1]

In questo capitolo affronteremo il problema delle tecniche statistiche utilizzate per la verifica del funzionamento di prove basate su domande a risposta chiusa. Prenderemo in considerazione la teoria classica dell'item analisi e faremo riferimento alle tecniche utilizzate *dall'Item Response Theory*.
L'item analisi consente di conoscere sia le caratteristiche e la capacità di misurare di una prova nel suo complesso e sia quelle dei singoli item che la compongono.
Anche i modelli teorici più avanzati, comunque, costituiscono una approssimazione alla realtà che intendiamo misurare, per questo è necessario utilizzare con rigore le tecniche statistiche, ma è altrettanto indispensabile farlo con la consapevolezza dei loro limiti.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- come misurare abilità, conoscenze e atteggiamenti attraverso prove basate su domande;
- quali sono i principi su cui si basa la teoria classica dei test;
- quali sono i principi su cui si basa l'Item Response Theory;
- come condurre l'item analisi di prove che misurano conoscenze o abilità;
- prove adattative

[1] Il presente contributo è frutto del lavoro congiunto dei tre autori. In particolare i paragrafi dal 6.1 al 6.5. sono stati redatti a due mani da P. Lucisano e A.M. De Luca. E. Botta ha redatto i paragrafi 6.6. e 6.7.

"La costruzione d'un modello era dunque per lui un miracolo d'equilibrio tra i princìpi (lasciati nell'ombra) e l'esperienza (inafferrabile), ma il risultato doveva avere una consistenza molto più solida degli uni e dell'altra. In un modello ben costruito, infatti, ogni dettaglio dev'essere condizionato dagli altri, per cui tutto si tiene con assoluta coerenza, come in un meccanismo dove se si blocca un ingranaggio tutto si blocca. Il modello è per definizione quello in cui non c'è niente da cambiare, quello che funziona alla perfezione; mentre la realtà vediamo bene che non funziona e che si spappola da tutte le parti; dunque non resta che costringerla a prendere la forma del modello, con le buone o con le cattive"

Palomar Italo Calvino

6.1 Misurare con misura

Da diversi anni in educazione e soprattutto nella didattica scolastica e universitaria si è diffuso l'uso di test, cioè prove costituite da domande a risposta chiusa, quali ad esempio quelle a risposta multipla, per la misurazione e la valutazione delle conoscenze e delle abilità, ma anche delle attitudini e degli atteggiamenti degli studenti. Dalla fine dello scorso millennio questo uso, facilitato dalle nuove tecnologie è dilagato in una sorta di mania di misurare tutto e di poter usare le misure come criterio unico di decisione.

Decine di migliaia di studenti ogni anno si vedono aperte o precluse le porte dell'università dai risultati di un test, così come con un test si decide quali candidati potranno accedere al concorso per dirigenti scolastici. Questa "cultura della valutazione", è frutto spesso d'improvvisazione e costituisce un rischio per la ricerca educativa, poiché la fretta e il decisionismo con cui, purtroppo, opera la tecnocrazia manageriale non solo genera risultati inadeguati, ma produce effetti di sfiducia sulla possibilità di operare in modo scientifico e al tempo stesso critico nel nostro settore di ricerca.

Nel criticare l'impianto dei test costruiti dal MIUR per selezionare i dirigenti scolastici Bottani (2011) ha dovuto ricordare che per fare bene queste cose occorrono specialisti[1]. Rafforziamo questa affermazione: per costruire test ben fatti occorre una conoscenza esperta, un sapere attraversato da molta pratica, che non può essere sostituito da una rapida lettura di istruzioni per l'uso. Cercare di insegnare a lavorare scientificamente alla rilevazione e alla misura di abilità e fare manutenzione di questa competenza, assomiglia alla fatica di fare acquisire buone abitudini. L'arte di fare test richiede una competenza da artigiani, ma anche la presenza di un progetto complessivo, una architettura e, dunque, di un architetto, perché in questa pratica il *perché* conta quanto il *come*. Spesso docenti che hanno solo una formazione disciplinare vengono chiamati a improvvisarsi nel costruire prove alle quali viene attribuito un valore "oggettivo" per la loro forma ma che nella sostanza non sono in grado di misurare né di offrire informazioni utili per la valutazione. I singoli esperti disciplinari chiamati a proporre domande hanno però anch'essi dei torti: il primo di esporsi ad utilizzare strumenti che non si sanno padroneggiare, il secondo di pensare che la sola conoscenza dei contenuti di una disciplina consenta di costruire strumenti di misura intelligenti attraverso la formalizzazione di prove a domande chiuse. In questo capitolo affronteremo quelle che sono le attenzioni necessarie per verificare il funzionamento di prove basate su domande attenzioni che sono sintetizzate nella parola item analisi, analisi cioè degli item, delle domande che costituiscono una prova.

Approfondimento 6.1

Visalberghi nel suo *Misurazione e valutazione nel processo educativo* (1955) scriveva «La parola misura ha due significati principali, che non sono scollegati come taluni aspetti della vita moderna potrebbero far temere.

Non c'è nessuna ragione di fondo per cui la misura intesa come operazione di conteggio o confronto non debba accompagnarsi alla misura intesa come abito di equilibrio e di discrezione.

Si potrebbero fare, è vero, sottili analisi circa l'origine classica dei due significati ed il loro uso rinascimentale, ma non crediamo che i risultati sarebbero in contrasto con la semplice osservazione di buon senso che l'abito stesso del misurare, implicando l'attitudine a vedere un più ed un meno dove il giudizio affrettato scorge qualità assolute, è esso stesso un abito di riflessività, di moderazione e di prudenza.

Come si spiega allora che la tendenza contemporanea a misurare tutto si accompagni spesso con atteggiamenti completamente opposti?»

In generale è evidente che il lavoro di misura di conoscenze e abilità sia parte del discorso più ampio di costruzione di informazioni che poi confluiscono anche in aspetti di valutazione. In questa stessa prospettiva la valutazione viene assunta come momento finale di un procedimento complesso e che richiede tutta una fase istruttoria di ricerca di informazioni.

È vero ciò che afferma Visalberghi (1955): «la misurazione nasce dalla valutazione e nella valutazione confluisce». Nel caso di misure di conoscenze e abilità, a monte c'è un giudizio, sulla abilità o sulle conoscenze che si intendono misurare, a questo segue la costruzione e l'uso di uno strumento di misura, e successivamente un altro

giudizio che verifica sia la validità sia l'affidabilità dello strumento. Solo dopo queste fasi è possibile dedicarsi a analizzare le misure che abbiamo ricavato.

Le misure su cui ci fermeremo, quelle realizzate attraverso prove basate su domande a scelta multipla, sono solo un mezzo di raccolta delle informazioni e non sono sinonimo di valutazione.

La misura diretta è quella operazione che si effettua confrontando la grandezza da misurare con un'altra grandezza a essa omogenea, presa come campione, cioè assunta come unità di misura: quindi, misurare una grandezza significa trovare un numero che dica quante volte la grandezza campione è contenuta in quella da misurare. Per quanto riguarda le conoscenze e le abilità a cui si riferiscono le prove utilizzate in ambito educativo nella maggior parte dei casi non si tratta di misure dirette, ma di misure indirette che si ricavano dalle risposte a un insieme di domande. La comprensione di un testo, la conoscenza di un argomento di storia o di scienze, non è considerabile in senso stretto come una grandezza direttamente misurabile, ciò che si cerca di misurare sono i tratti latenti che si ipotizza sottostiano al costrutto che si intende misurare e ne siano quindi rappresentativi; si cerca dunque di ricavare informazioni sintetiche sul tratto latente operando su una serie di indicatori empirici o osservazioni.

La misurazione di un fenomeno non osservabile direttamente richiede considerazioni sia di tipo teorico, sia di tipo empirico. Dal punto di vista empirico, l'attenzione deve essere portata alle domande e al modo in cui ci danno informazioni; dal punto di vista teorico è necessario approfondire il rapporto tra l'abilità da misurare, ad esempio la comprensione di un testo (che non è direttamente osservabile), e le domande le cui risposte dovrebbero darne conto. Poiché il fondamento di questa operazione, cioè la relazione tra indicatori empirici e abilità, rimane difficile da esplorare, la semplice analisi di indicatori empirici può portare a inferenze non corrette o a conclusioni fuorvianti.

Prima di descrivere le procedure statistiche utili per verificare che una prova misuri in modo corretto, valido e affidabile il tratto che intende misurare è utile considerare i presupposti teorici dell'analisi statistica.

Se il complesso di conoscenze o l'abilità o l'atteggiamento che intendiamo misurare è stato ben definito e le domande formulate rappresentano un campione ben costruito di indicatori del tratto in questione, e se la popolazione di riferimento si distribuisce in modo normale rispetto al tratto, o comunque in modo abbastanza disomogeneo, allora ci attendiamo che le persone che ottengono nella prova i punteggi migliori rispondano alla maggior parte delle domande comprese le più difficili, quelli che raggiungono i punteggi medi superino le domande più facili e quelle di difficoltà media e non rispondano alle domande più difficili, quelli che nel complesso della prova hanno punteggi più bassi superino solo le domande più facili. Queste assunzioni possono essere ricondotte a un modello teorico di riferimento, che è alla base di tutti i calcoli che seguono. Nel modello (Tabella 6.1) i risultati ottenuti sono riportati in una tabella in cui vi è una riga per ogni soggetto e una colonna per ogni item e in cui i soggetti sono ordinati da quello con il punteggio più alto a quello con il punteggio più basso.

Tabella 6.1. Modello teorico delle risposte attese

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Punti grezzi
Anna	1	1	1	1	1	1	1	7
Biagio	0	1	1	1	1	1	1	6
Carla	0	0	1	1	1	1	1	5
Dario	0	0	0	1	1	1	1	4
Elena	0	0	0	0	1	1	1	3
Fabio	0	0	0	0	0	1	1	2
Gino	0	0	0	0	0	0	1	1
Risposte corrette all'item	**1**	**2**	**3**	**4**	**5**	**6**	**7**	

Se le domande misurano tratti latenti diversi, sono mal formulate, prive di senso, tali da non individuare una sola risposta corretta, tali da implicarsi a vicenda, qualsiasi ragionamento statistico sulle risposte è privo di senso.

Per lavorare in modo assennato è necessario procedere in questo modo:

1. costruire un modello teorico che contenga la definizione operazionalizzata dell'abilità (del tratto latente) e rappresenti le relazioni tra gli eventuali costrutti che sono alla base del concetto di cui vogliamo prendere misure;
2. definire le relazioni attese tra abilità e indicatori;
3. individuare un metodo per misurare gli indicatori.

Approfondimento 6.2
SUGGERIMENTI PER COSTRUIRE LE DOMANDE

La costruzione di domande a scelta multipla richiede numerose attenzioni, riportiamo una serie di suggerimenti per la formulazione di domande a scelta multipla con una sola alternativa corretta.
A ogni domanda segue una serie di alternative di risposta (in genere quattro o cinque) di cui una costituisce la risposta corretta mentre le altre hanno la funzione di «distrattori». Il quesito, l'item, va considerato composto da due parti: il corpo della domanda, o stimolo e le diverse alternative di risposta proposte (risposte). La prima parte va intesa come posizione del problema, la seconda come il «luogo» in cui vengono proposte le possibili soluzioni.
Per la stesura delle domande vanno tenuti presenti i seguenti aspetti:

- a. posizione e soluzione del problema: indicazioni sulla struttura generale del quesito;
- b. formulazione linguistica: indicazioni sul tipo di linguaggio da adoperare nel formulare il quesito;
- c. organizzazione degli item: indicazioni sulla collocazione delle risposte corrette nelle diverse domande;
- d. convenzioni di scrittura;
- e. collocazione delle domande nella prova.

a. Posizione e soluzione del problema

1. Ogni domanda deve focalizzare un solo problema o concetto, esprimendolo con precisione per evitare incertezze e confusioni nella scelta della risposta. Gli errori più frequenti in cui si incorre sono tre:
   - non definire il problema,
   - formulare una domanda doppia, che pone più richieste non correlate
   - formulare problema e soluzione direttamente nelle diverse alternative.
2. Usare distrattori plausibili rispetto al problema considerato. Ogni distrattore deve risultare collegato per contenuto alla domanda. I distrattori dovrebbero rispondere a possibili errori da parte di chi risponde. La difficoltà nell'ipotizzare distrattori plausibili può portare i compilatori di test a utilizzare distrattori poco opportuni (troppo o troppo poco attrattivi) o costruiti male, tali cioè da fornire indicazioni non coerenti con l'abilità indagata.

b. Formulazione linguistica

3. Formulare la domanda in forma semplice ed esplicita. La domanda deve essere formulata con chiarezza, con un linguaggio adeguato al contesto e non può essere ridotta a poche parole, non può cioè contenere assunzioni implicite che non possano essere individuate facilmente dalle persone che devono rispondere. Il corpo della domanda dovrebbe essere composto da una frase in forma interrogativa o affermativa.
4. Usare un linguaggio che per difficoltà è omogeneo al tipo di competenze che si vogliono verificare. La domanda infatti non può essere più difficile di quanto non lo sia nella formulazione dei testi di riferimento utilizzati da chi deve rispondere. Per questo è utile usare nella formulazione della domanda e dei distrattori un lessico di base.
5. Formulare i distrattori e la risposta corretta in modo omogeneo per struttura sintattica e per lunghezza della frase.
6. Evitare che le risposte esatte e i distrattori riprendano letteralmente termini o formulazioni usate nei testi di riferimento e nella domanda.
7. Evitare, nella costruzione delle alternative avverbi come *sempre, tutti, ogni, dappertutto, quasi, spesso, qualche volta*. Questa regola è importante in particolare nella costruzione di prove di atteggiamento che prevedono una graduazione nelle risposte in forma di scala che potrebbe confliggere con gli avverbi citati.
8. Evitare domande con negazioni semplici e composte. Se è proprio indispensabile mantenere una negazione nella domanda, evidenziarla in neretto o con sottolineatura.
9. Quando si tratta di proporre risposte che implicano cifre/date, disporre le risposte in ordine decrescente o viceversa.
10. Evitare alternative del tipo: nessuna di queste, sia a) che c), tutte le precedenti.
11. Evitare che elementi grammaticali (concordanze, consecutio, eccetera) o la struttura della frase favoriscano l'individuazione della risposta esatta.

c. Organizzazione degli item

12. Fare in modo che le domande siano tra loro indipendenti, per evitare che l'una suggerisca una risposta all'altra.
13. In molti casi, come nelle prove di comprensione del testo, è opportuno non usare alcuna risposta precedente come distrattore o come risposta corretta in altre domande.
14. Prevedere la collocazione della risposta corretta in modo che, nell'insieme delle domande, le risposte esatte non corrispondano tutte alla stessa lettera.

d. Convenzioni di scrittura

15. Se la domanda termina con il punto interrogativo i distrattori devono cominciare con la maiuscola.
16. Alla fine di ciascun distrattore usa il punto.
17. Le domande a completamento non terminano mai con i due punti o i puntini sospensivi.

e) collocazione delle domande nella prova

Una volta identificate le domande è necessario disporle secondo un ordine. È possibile decidere di considerare ciascun item un problema a sé e dunque disporre le domande in ordine casuale. È preferibile, però, collocare le domande in relazione alla difficoltà. Ad esempio, si possono mettere all'inizio domande più facili e poi via via domande più impegnative, riservando alcune domande più facili nella parte finale del test. Per questo sono di grande utilità gli indici

ricavati nella item analisi del try out della prova.
L'uso oggi diffuso del proporre test con le domande in ordine diverso e in subordine con i distrattori in ordine diverso, se evita forme di collaborazione tra gli studenti, trasgredisce l'assunto che la prova proposta sia identica. Infatti, lo studente che inizia la prova con le domande più difficili farà un test diverso da quello che inizia la prova con le domande più facili e potrebbe assumere un atteggiamento diverso nell'affrontare la prova nel suo insieme.
È importante assicurarsi, in particolare nelle prove *computer based* che lo studente possa navigare liberamente all'interno della prova, ritornando su domande già viste ed eventualmente modificando le risposte.

6.2 Possiamo misurare le stesse cose in modo diverso

In questo capitolo ci fermeremo solo sul terzo punto dell'elenco precedente, cioè sui diversi metodi o modelli di misurazione, sui loro punti di forza e limiti, riflettendo in particolare sul rapporto tra le misure ottenute e la realtà sottostante, ricordando tuttavia che le osservazioni empiriche sono necessarie ma mai sufficienti se non rimandano ad una base teorica consistente.

Avviamo la riflessione confrontando due diversi modelli di item analisi nella misura dello stesso tratto latente: il modello della *Classic Test Theory* (CTT) e il modello della *Item Response Theory* (IRT). Per spiegarne le differenze immaginiamo di lavorare su un piccolo set di dati. Vogliamo subito chiarire che sia la CTT, sia la IRT non forniscono risultati attendibili su insiemi di dati così ridotti e che dunque gli esempi che seguono hanno solo lo scopo di far comprendere le dinamiche operative dei due modelli.

Nella Tabella 6.2 abbiamo inserito i risultati di una prova costituita da 13 item a scelta multipla con 4 alternative di risposta somministrata a 21 soggetti. Nella seconda riga è riportata la chiave di risposta (la risposta corretta all'item) e in grassetto sono evidenziate le risposte errate date da ciascun esaminato.

Tabella 6.2. Esempio della matrice delle risposte di una prova a scelta multipla

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13
Keys	C	A	B	C	A	D	C	A	C	D	D	A	B
A_01	C	**C**	B	C	A	D	C	**C**	**B**	D	D	**D**	**D**
A_02	**D**	A	B	C	**D**	D	C	**B**	**B**	**C**	D	**D**	B
A_03	**D**	A	B	C	A	D	C	A	C	D	D	A	B
A_04	C	**C**	B	C	**B**	**C**	C	A	C	D	**A**	**D**	B

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13
Keys	C	A	B	C	A	D	C	A	C	D	D	A	B
A_05	**D**	A	B	C	**C**	D	C	A	C	D	D	A	**D**
A_06	C	**D**	B	C	A	**B**	**D**	**C**	**B**	D	**A**	A	B
A_07	**D**	A	B	C	A	**C**	C	A	**D**	**B**	**A**	**B**	B
A_08	C	A	**D**	**D**	A	D	C	**B**	**B**	**C**	**C**	**D**	**C**
A_09	C	A	**A**	C	A	**C**	C	A	**D**	D	D	A	B
A_10	**D**	**B**	**A**	**A**	**D**	D	**B**	**C**	C	**C**	D	**B**	**D**
A_11	C	**C**	B	**B**	**C**	D	C	A	**B**	D	D	**D**	B
A_12	**D**	A	B	**B**	A	**A**	C	**C**	**B**	D	D	**D**	**A**
A_13	C	A	B	C	A	D	C	A	**B**	D	D	**D**	B
A_14	**D**	A	B	C	A	D	C	A	C	D	D	**D**	B
A_15	**D**	A	B	C	A	**C**	C	**C**	**B**	D	**A**	**O**	B
A_16	**A**	A	B	C	A	D	C	A	C	D	D	**D**	B
A_17	**D**	**C**	B	C	A	**C**	C	**B**	**D**	D	D	**D**	**D**
A_18	C	**C**	B	C	**C**	D	C	A	**B**	**C**	D	A	B
A_19	C	A	B	C	A	D	**A**	A	C	D	D	A	B
A_20	**D**	**C**	B	C	A	**A**	**D**	A	**B**	**C**	D	**C**	B
A_21	**A**	A	B	**B**	**B**	D	**B**	**C**	**B**	**C**	**A**	A	B

Dopo aver individuato le risposte esatte e quelle errate dobbiamo decidere come assegnare un punteggio ad ogni risposta. La via canonica è quella di assegnare un punto ad ogni risposta esatta e zero punti alle risposte errate o omesse.

Nella Tabella 6.3 è riportato il punteggio grezzo ottenuto dai 21 esaminati, assegnando un punto alle risposte esatte e zero a quelle sbagliate e sommando i punteggi dati.

Tabella 6.3. Correzione dicotomica e relativi punteggi grezzi

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13	Punteggi
A_01	1	0	1	1	1	1	1	0	0	1	1	0	0	8
A_02	0	1	1	1	0	1	1	0	0	0	1	0	1	7
A_03	0	1	1	1	1	1	1	1	1	1	1	1	1	12
A_04	1	0	1	1	0	0	1	1	1	1	0	0	1	8
A_05	0	1	1	1	0	1	1	1	1	1	1	1	0	10
A_06	1	0	1	1	1	0	0	0	0	1	0	1	1	7

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13	Punteggi
A_07	0	1	1	1	1	0	1	1	0	0	0	0	1	7
A_08	1	1	0	0	1	1	1	0	0	0	0	0	0	5
A_09	1	1	0	1	1	0	1	1	0	1	1	1	1	10
A_10	0	0	0	0	0	1	0	0	1	0	1	0	0	3
A_11	1	0	1	0	0	1	1	1	0	1	1	0	1	8
A_12	0	1	1	0	1	0	1	0	0	1	1	0	0	6
A_13	1	1	1	1	1	1	1	1	0	1	1	0	1	11
A_14	0	1	1	1	1	1	1	1	1	1	1	0	1	11
A_15	0	1	1	1	1	0	1	0	0	1	0	0	1	7
A_16	0	1	1	1	1	1	1	1	1	1	1	0	1	11
A_17	0	0	1	1	1	0	1	0	0	1	1	0	0	6
A_18	1	0	1	1	0	1	1	1	0	0	1	1	1	9
A_19	1	1	1	1	1	1	0	1	1	1	1	1	1	12
A_20	0	0	1	1	1	0	0	1	0	0	1	0	1	6
A_21	0	1	1	0	0	1	0	0	0	0	0	1	1	5

È opportuno, prima di procedere, fare una breve digressione sulla modalità di assegnazione del punteggio (*scoring*). Talvolta per l'assegnazione del punteggio si ricorre a una modalità che prevede la penalizzazione, cioè l'assegnazione di un punteggio negativo a una risposta errata. Ad esempio, nell'approfondimento 6.3, si possono osservare i risultati ottenuti con questo tipo di *scoring* nel caso di un test composto da 80 item e somministrato a due persone. La Tabella 6.3.1 - inserita nel riquadro di approfondimento 6.3 - riporta nella parte superiore il numero di risposte esatte, errate e omesse di ciascun esaminando, mentre nella parte inferiore riporta i punteggi calcolati secondo la regola che assegna alle risposte esatte un punto, alle risposte errate -0,5 punti e a quelle omesse zero punti.

Approfondimento 6.3

L'assegnazione dei punteggi con penalizzazioni privilegia, a parità di risposte esatte, chi ha un atteggiamento conservativo che nel dubbio meglio omette la risposta; infatti, l'Esaminando B ottiene un punteggio superiore all'Esaminando A.

Tabella 6.3.1

Risposte	**Esatte**	**Errate**	**Omesse**	**N. Domande**
Esaminando A	72	6	2	80
Esaminando B	72	2	6	80

Punteggi	**1**	**- 0,5**	**0**	**Punteggio Totale**
Esaminando A	72	- 3	0	69
Esaminando B	72	-1	0	71

Questa modalità potrebbe anche essere valida in certi contesti, ad esempio se il test fosse per dei Medici in cui è giusto premiare un atteggiamento prudente che prevede, nel dubbio, la sospensione della diagnosi rimandandola ad un momento successivo dopo aver svolto ulteriori approfondimenti e/o consulti con i colleghi.

Tuttavia, il punteggio con penalizzazioni può dare origine a situazioni paradossali come riportato nella Tabella 6.3.2 in cui L'Esaminando A che ha il maggior numero di risposte esatte ottiene un punteggio inferiore.

Tabella 6.3.2

Risposte	**Esatte**	**Errate**	**Omesse**	**N. Domande**
Esaminando A	72	8	0	80
Esaminando B	70	0	10	80

Punteggi	**1**	**- 0,5**	**0**	**Punteggio Totale**
Esaminando A	72	- 4	0	68
Esaminando B	70	0	0	70

Le nostre osservazioni ora si focalizzeranno, non più sui punteggi degli esaminati, ma sui singoli item della prova. Nella Tabella 6.4 sono riportate la frequenza assoluta con cui è stata scelta ogni opzione di risposta (parte alta) e la corrispondente frequenza percentuale (parte bassa) per ogni singola domanda. Ci accorgiamo subito che non tutte le domande presentano risultati omogenei (in grassetto i dati riferiti alla risposta corretta). Inoltre, osserviamo che la percentuale maggiore di risposte non sempre coincide con la risposta esatta, ad esempio nell'Item_09 più della metà (52%) degli esaminati risponde B mentre la risposta corretta è C (33%).

Tabella 6.4. Numero e percentuali di risposta alle singole alternative della prova

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13
Key	C	A	B	C	A	D	C	A	C	D	D	A	B
A	2	**13**	2	1	**14**	2	1	**12**	0	0	5	**7**	1
B	0	1	**18**	3	2	1	2	3	11	1	0	2	**15**
C	**9**	6	0	**16**	3	5	**16**	6	**7**	6	1	1	1
D	10	1	1	1	2	**13**	2	0	3	**14**	**15**	10	4

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13
Tot.	21	21	21	21	21	21	21	21	21	21	21	20	21
A	10%	**62%**	10%	5%	**67%**	10%	5%	**57%**	0%	0%	24%	**33%**	5%
B	0%	5%	**86%**	14%	10%	5%	10%	14%	52%	5%	0%	10%	**71%**
C	**43%**	29%	0%	**76%**	14%	24%	**76%**	29%	**33%**	29%	5%	5%	5%
D	48%	5%	5%	5%	10%	**62%**	10%	0%	14%	**67%**	**71%**	48%	19%
Tot	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	95%	100%

L'item analisi classica, muove dall'assunto che tutte le domande di un test misurino lo stesso tratto latente, e attribuisce alle risposte corrette lo stesso punteggio. Nella fase di *try out* della prova (fase non eludibile) vengono verificate le caratteristiche delle domande con indicatori quali la facilità e la discriminatività della domanda (*correlazione punto biseriale*).

6.3 La facilità/difficoltà di una domanda

Osservando ciascuna colonna della Tabella 5 possiamo vedere come si sono orientate le risposte dei soggetti sulla singola domanda e considerare le frequenze di scelta ottenute dalla risposta corretta, da ogni singolo distrattore, dalle omissioni e dagli errori. L'indice di facilità di un item *p* si ricava considerando il rapporto tra il numero delle risposte esatte e quello dei soggetti che hanno svolto la prova; p assume valori compresi fra 0 e 1, dove valori bassi indicano che una domanda è molto difficile e valori alti che una domanda è facile. Il suo complementare, q = 1- p, si comporta in modo inverso e può essere calcolato come il rapporto fra le risposte errate e il numero di casi, è indicato in tabella come TID (*Tentative Item Difficulty*), con ODDS è indicato il rapporto fra il numero di risposte errate e il numero di risposte corrette. Il Logit è l'unità di misura che si ottiene passando dalla scala lineare alla scala logaritmica, è definito come il logaritmo naturale dell'ODDS.

Tabella 6.5. Analisi della difficoltà e della abilità

	item_01	item_02	item_03	item_04	item_05	item_06	item_07	item_08	item_09	item_10	item_11	item_12	item_13	Esatte	TSP	ODDS	LOGIT
A_01	1	0	1	1	1	1	1	0	0	1	1	0	0	8	0,62	1,60	0,47
A_02	0	1	1	1	0	1	1	0	0	0	1	0	1	7	0,54	1,17	0,15
A_03	0	1	1	1	1	1	1	1	1	1	1	1	1	12	0,92	12,0	2,48
A_04	1	0	1	1	0	0	1	1	1	1	0	0	1	8	0,62	1,60	0,47
A_05	0	1	1	1	0	1	1	1	1	1	1	1	0	10	0,77	3,33	1,20
A_06	1	0	1	1	1	0	0	0	0	1	0	1	1	7	0,54	1,17	0,15
A_07	0	1	1	1	1	0	1	1	0	0	0	0	1	7	0,54	1,17	0,15
A_08	1	1	0	0	1	1	1	0	0	0	0	0	0	5	0,38	0,63	-0,47
A_09	1	1	0	1	1	0	1	1	0	1	1	1	1	10	0,77	3,33	1,20
A_10	0	0	0	0	0	1	0	0	1	0	1	0	0	3	0,23	0,30	-1,20
A_11	1	0	1	0	0	1	1	1	0	1	1	0	1	8	0,62	1,60	0,47
A_12	0	1	1	0	1	0	1	0	0	1	1	0	0	6	0,46	0,86	-0,15
A_13	1	1	1	1	1	1	1	1	0	1	1	0	1	11	0,85	5,50	1,70
A_14	0	1	1	1	1	1	1	1	1	1	1	0	1	11	0,85	5,50	1,70
A_15	0	1	1	1	1	0	1	0	0	1	0	0	1	7	0,54	1,17	0,15
A_16	0	1	1	1	1	1	1	1	1	1	1	0	1	11	0,85	5,50	1,70
A_17	0	0	1	1	1	0	1	0	0	1	1	0	0	6	0,46	0,86	-0,15
A_18	1	0	1	1	0	1	1	1	0	0	1	1	1	9	0,69	2,25	0,81
A_19	1	1	1	1	1	1	0	1	1	1	1	1	1	12	0,92	12,0	2,48
A_20	0	0	1	1	1	0	0	1	0	0	1	0	1	6	0,46	0,86	-0,15
A_21	0	1	1	0	0	1	0	0	0	0	0	1	1	5	0,38	0,63	-0,47
Errate	12	8	3	5	7	8	5	9	14	7	6	14	6				
TID	0,57	0,38	0,14	0,24	0,33	0,38	0,24	0,43	0,67	0,33	0,29	0,67	0,29				
ODDS	1,33	0,62	0,17	0,31	0,50	0,62	0,31	0,75	2,00	0,50	0,40	2,00	0,40				
LOGIT	0,29	-0,49	-1,79	-1,16	-0,69	-0,49	-1,16	-0,29	0,69	-0,69	-0,92	0,69	-0,92				

In Figura 6.1 è riportata la distribuzione degli item in relazione all'indice di difficoltà espresso in logit e, nella Figura 6.2, la distribuzione degli alunni in relazione all'indice di abilità, anch'esso espresso in logit.

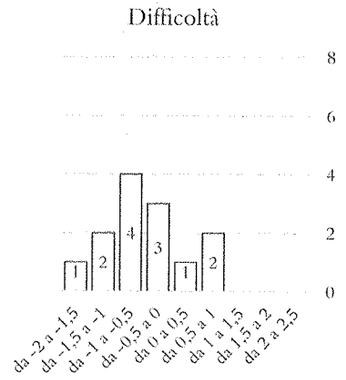


Figura 6.1

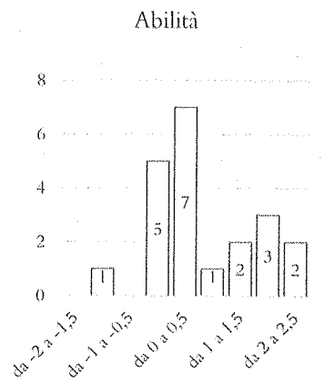


Figura 6.2

6.4 La discriminatività di una domanda

Per discriminatività intendiamo la capacità di una domanda di misurare le differenze nell'abilità considerata tra i soggetti di un dato gruppo in esame: nel nostro caso, la capacità di distinguere livelli di abilità diversi, ovvero separare i coloro che padroneggiano l'abilità in esame da coloro che la possiedono in misura minore.

Un metodo di facile applicazione per ricavare questo indice, D, si basa sul confronto fra il numero di risposte corrette all'item fornite dagli studenti che hanno ottenuto i punteggi migliori nella prova e quello delle risposte corrette all'item fornite da coloro che hanno ottenuto basse prestazioni.

Per fare questo, dopo avere ordinato i risultati della prova in base al punteggio si confrontano i risultati dei due gruppi che si collocano agli estremi della tabella. L'ampiezza dei gruppi può variare in relazione all'ampiezza del campione: Wood (1985) ad esempio suggerisce di usare il 21 % in relazione a campioni più ampi, e nell'uso legato a piccoli gruppi come una classe spesso si ricorre a percentuali più ampie.

La scelta corrisponde al compromesso tra due scopi desiderabili, ma contraddittori: avere due gruppi il più possibile estesi (se si considera solo una classe di 20 studenti i due gruppi da confrontare, usando il 21%, finiscono per essere di 5 soggetti ciascuno), e avere due gruppi il più possibile diversi.

Quindi, item per item, è possibile confrontare il numero delle risposte esatte degli studenti dell'estremo superiore con quello dei ragazzi dell'estremo inferiore

e calcolare il seguente indice:

$$D = (Es - Ei)/n$$

dove D = indice di discriminatività
Es = risposte esatte dell'estremo superiore
Ei = risposte esatte dell'estremo inferiore
n = numero dei soggetti di un estremo.

Questo indice varia da -1 a 1, dove 0 sta per nessuna discriminatività (i soggetti nell'estremo superiore e quelli nell'estremo inferiore rispondono allo stesso modo), e 1 sta per discriminatività massima. In genere si usa scartare gli item con discriminatività inferiore a 0,30.

Per ottenere in modo più completo un indice di discriminatività si può calcolare la correlazione tra il successo/insuccesso nell'item e il punteggio totale del test della prova.

Per calcolare il punto biseriale si applica prevalentemente la formula seguente:

$$r_{punbis} = \frac{Me - Mt}{s}\sqrt{\frac{p}{1-p}}$$

dove
Me = media dei punteggi al test dei soggetti che hanno fornito la risposta corretta all'item (e indica esatta);
Mt = media dei punteggi al test dell'intero campione;
s = deviazione standard dei punteggi dell'intero campione;
p = proporzione di risposte corrette all'item (numero di risposte corrette/numero di soggetti).

Poiché il punto biseriale è una forma particolare di correlazione bisogna tenere presenti alcune cautele nell'interpretarne i risultati. Infatti, se la distribuzione dei punteggi del gruppo è normale, il punto biseriale non può mai andare oltre ± 0,8.

Inoltre, la gamma dei possibili valori dipende dalla percentuale di quelli che rispondono bene o male, e si contrae quando la distribuzione si presenta più irregolare. Il massimo valore di r_{punbis} per p = 0,10 e p = 0,25 quando la distribuzione complessiva dei punteggi è normale, è rispettivamente di 0,58 e 0,73 (0,80 quando p = 0,50).

Un valore del punto biseriale prossimo allo 0 indica che la domanda non discrimina adeguatamente fra chi ha avuto un buon rendimento nella prova e chi non lo ha avuto. Un valore positivo, superiore a .2 indica che la domanda discrimina i soggetti con abilità alta da quelli con abilità bassa, e tanto più questo valore si avvicina a 1 migliore è la capacità della domanda di discriminare fra i diversi gruppi

di soggetti. Un valore negativo indica che fra coloro che rispondono correttamente alla domanda quelli che hanno avuto complessivamente un rendimento basso nella prova sono più numerosi rispetto a quelli che hanno ottenuto complessivamente punteggi più elevati.

Tabella 6.6. Facilità e punti biseriali delle domande proposte in Tabella 6.1

item n.	p (Facilità = Numero Risposte Esatte/ Numero Totale dei soggetti)	Media al test dei soggetti che rispondono correttamente all'item	Punto biseriale
1	0,78	5,86	0,62
2	0,44	5,75	0,25
3	0,22	6,5	0,36
4	0,33	6,33	0,41
5	0,33	6,67	0,53
6	0,67	6,00	0,57
7	0,22	4,50	-0,2
8	0,67	6,00	0,57
9	0,89	5,25	0,04
10	0,67	6,00	0,57

Questi criteri vengono utilizzati per bilanciare la prova ed escludere le domande che non funzionano bene. Dunque, una prova ben fatta deve contenere item di diverso livello di facilità e item capaci di distinguere i soggetti che rispondono complessivamente bene alla prova dai soggetti che invece non lo fanno. L'assunto di misurare lo stesso tratto latente viene controllato teoricamente (*ex ante*) attraverso la validazione dei contenuti e dell'aspetto della prova e (*ex post*), empiricamente, attraverso la verifica di validità del criterio e del costrutto. Un conforto quantitativo viene dall'uso di statistiche come l'*alfa di Cronbach* o il *Kuder-Richardson20 (KR 20)* che sintetizzano la coerenza di comportamento degli item rispetto alla prova complessiva.

Nell'item analisi classica, inoltre, si utilizza l'errore standard della media come stima della precisione delle misure sui singoli soggetti.

Il limite di questo metodo rimane nel fatto che le misure sono somme di punteggi, risposte corrette agli item, ai quali viene attribuito lo stesso valore. Questo, anche quando una prova viene somministrata a molti soggetti, non consente di cogliere bene le differenze interindividuali, tanto che nell'uso popolare quando si correggono prove con questo metodo si ottengono misure discrete che danno luogo a blocchi di candidati con lo stesso punteggio per poi dover intervenire con

correttivi esterni alla misura stessa per tracciare linee di demarcazione tra candidati a pari merito.

Per questi motivi negli anni Sessanta Rasch propose un modello diverso per misurare i tratti latenti. Proviamo a spiegarne rapidamente l'architettura.

Immaginiamo che un insegnante assegni una prova con una domanda difficile, due di media difficoltà e due facili a 6 studenti (Tabella 6.7).

Tabella 6.7. Esempio di risposte ad una prova di comprensione della lettura

		Item1 (difficile)	Item2 (medio)	Item3 (medio)	Item4 (facile)	Item5 (facile)	Punti
1	Anna	1	1	1	1	1	5
2	Biagio	0	1	1	1	1	4
3	Carla	0	0	1	1	1	3
4	Dario	0	0	0	1	1	2
5	Elena	0	0	0	0	1	1
6	Fabio	1	1	0	0	0	2
	difficoltà	4/2=2	3/3=1	3/3=1	2/4=0,5	1/5=0,2	

Se confrontiamo il risultato di Dario e Fabio ci rendiamo conto che i due studenti hanno ottenuto lo stesso punteggio 2 rispondendo a domande di diverso livello di difficoltà (definendo la difficoltà come funzione del rapporto tra risposte errate e risposte esatte). Rasch introduce un modello in cui sono messi a confronto l'abilità dei soggetti con la difficoltà delle domande e prende in considerazione la probabilità di rispondere correttamente ad un item a partire dall'abilità dei soggetti.

Nella Tabella 8 abbiamo evidenziato e inserito sia la difficoltà degli item sia l'abilità degli studenti e ricavato la probabilità di risposta esatta attesa per ciascuno studente ad ogni singola domanda.

Tabella 6.8. Probabilità di rispondere correttamente all'item (valori %)

	Item 9	Item 1	Item 6	Item 8	Item 10	Item 2	Item 4	Item 5	Item 3	Item 7	Abilità
BELINDA	47	28	18	18	18	8	5	5	3	3	-2,2
BARBARA	84	70	57	57	57	35	25	25	16	16	-0,41
DORINA	89	78	67	67	67	44	33	33	22	22	0
CARLA	89	78	67	67	67	44	33	33	22	22	0
ALFONSO	89	78	67	67	67	44	33	33	22	22	0
BEA	92	84	75	75	75	55	43	43	30	30	0,41

ELENA	95	89	82	82	82	65	54	54	40	40	0,85
ALFIO	95	89	82	82	82	65	54	54	40	40	0,85
ALESSIA	95	89	82	82	82	65	54	54	40	40	0,85
Difficoltà	-2,08	-1,25	-0,69	-0,69	-0,69	0,22	0,69	0,69	1,25	1,25	

Abilità e difficoltà vengono calcolate attraverso la seguente formula:

$$P_i(\theta, b_i) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

dove
i indica il numero dell'item e varia fra 1 e 10
θ è l'abilità del soggetto
b_i è la difficoltà dell'item i
$P_i(\theta - b_i)$ è la probabilità di un soggetto con abilità di rispondere correttamente all'item i
e è il numero di Nepero pari a circa 2,71

Nella Figura 6.3 viene rappresentata la Curva Caratteristica che il modello di Rasch associa a ciascun item nel caso di un item di difficoltà media (b=0,5). Sull'asse delle x è riportata l'abilità del soggetto e sull'asse delle y la probabilità che il soggetto risponda correttamente all'item.
Possiamo, tuttavia, osservare come ancora si verifichino situazioni in cui i valori attesi secondo il modello si discostano dai valori osservati empiricamente. Confrontando la Tabella 6.8 con la Tabella 6.9 possiamo vedere come i singoli casi non sempre si comportano secondo il modello teorico. Infatti, alcuni soggetti rispondono a domande più difficili della loro abilità e altri sbagliano a rispondere a domande con difficoltà inferiore alla loro abilità. In particolare, nella Tabella 6.9, le celle in grigio rappresentano l'area relativa all'attesa di errore, in grassetto le risposte diverse da quanto atteso.

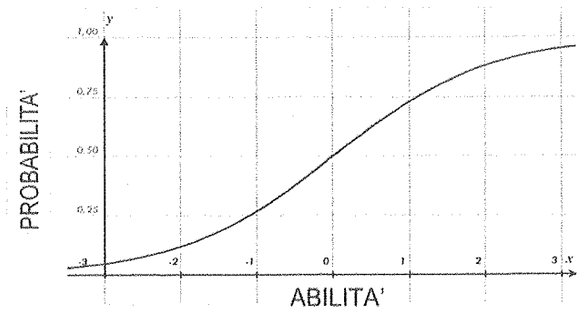


Figura 6.3 Curva caratteristica di un item di difficoltà media.

Tabella 6.9. Item risolti e item sbagliati in relazione all'abilità dei soggetti

	Item 9	Item 1	Item 6	Item 8	Item 10	Item 2	Item 4	Item 5	Item 3	Item 7	Abilità
BELINDA	**1**	0	0	0	0	0	0	0	0	0	-2,2
BARBARA	1	1	**0**	1	**0**	0	0	0	0	1	-0,41
DORINA	1	**0**	1	**0**	1	**1**	**1**	0	0	0	0
CARLA	1	1	1	**0**	1	**1**	0	0	0	0	0
ALFONSO	**0**	1	1	1	1	0	0	0	0	**1**	0
BEA	1	1	**0**	1	**0**	1	0	**1**	**1**	0	0,41
ELENA	1	1	1	1	1	**0**	**0**	1	**1**	0	0,85
ALFIO	1	1	1	1	1	**0**	1	1	0	0	0,85
ALESSIA	1	1	1	1	1	1	1	**0**	0	0	0,85
Difficoltà	-2,08	-1,25	-0,7	-0,69	-0,69	0,22	0,69	0,69	1,25	1,25	

Per migliorare il rapporto tra valori attesi e valori osservati si adotta una procedura ricorsiva che stima i valori del parametro di difficoltà degli item e del parametro di abilità dei soggetti avvicinando il più possibile il modello teorico al dato empirico (*Item calibration*). L'esito di questa procedura è una migliore stima dei parametri del modello. Il processo parte da una stima qualsiasi del parametro di abilità procede per approssimazioni successive del valore del medesimo parametro e si conclude quando raggiunge il valore che presenta la migliore approssimazione possibile in relazione alla funzione di approssimazione che è stata scelta. Esiste una

vasta letteratura che evidenzia i vantaggi dell'uso di questo metodo, che è stato perfezionato nel tempo aggiungendo alla funzione per il calcolo della probabilità di rispondere correttamente a un dato item altri parametri quali la discriminatività e il *guessing* (una stima del rischio che la risposta corretta possa essere individuata rispondendo in modo casuale - Birnbaum 1968), raggiungendo risultati sempre più vicini alla realtà.

Anche per la calibrazione vengono utilizzati procedimenti diversi dal *Maximum Likelyhood* (Metodo della massima verosimiglianza) al *Bayesian Modal* (Modello moda Bayesiana); ciascun metodo cerca di raggiungere una migliore approssimazione della misura del tratto latente ai comportamenti dei soggetti osservati.

Nella Tabella 10 si rendono visibili i calcoli ricorsivi che consentono un'approssimazione delle misure di abilità di un soggetto, sulla base delle sue risposte a tre item.

Tabella 6.10. esempio di calcolo del valore di abilità utilizzando il Maximum Likelihoodsulla base delle sue risposte a tre item

		b-difficoltà	item1 -1,0	item2 0,0	item3 1,0	
theta	u-risposta	P - probabilità	Q = 1-P	P*Q	u-P	□ (u-P) / □ (P*Q)
1,000	1,0	88%	12%	0,105	0,119	
-0,203	0,0	73%	27%	0,197	-0,731	
0,797	1,0	50%	50%	0,250	0,500	
new-theta				0,552	-0,112	**-0,203**
theta	u-risposta	P - probabilità	Q = 1-P	P*Q	u-P	□ (u-P) / □ (P*Q)
0,797	1,0	86%	14%	0,122	0,142	
0,006	0,0	69%	31%	0,214	-0,689	
0,803	1,0	45%	55%	0,247	0,551	
new-theta				0,584	0,003	**0,006**
theta	u-risposta	P - probabilità	Q = 1-P	P*Q	u-P	□ (u-P) / □ (P*Q)
0,803	1,0	86%	14%	0,121	0,141	
0,000	0,0	69%	31%	0,214	-0,691	
0,803	1,0	45%	55%	0,248	0,549	
new-theta				0,583	0,000	**0,000**
theta	u-risposta	P - probabilità	Q = 1-P	P*Q	u-P	□ (u-P) / □ (P*Q)
0,803	1,0	86%	14%	0,121	0,141	
0,000	0,0	69%	31%	0,214	-0,691	
0,803	1,0	45%	55%	0,248	0,549	
				0,583	0,000	**0,000**

6.5 Le stesse cose misurate in modo diverso danno risultati diversi

È abbastanza curioso che, mentre i metodi descritti vengono applicati nella ricerca, sono disattesi quando si tratta di prendere decisioni delicate come l'ingresso in università o la scelta dei dirigenti scolastici. Tuttavia, lo scopo di questa rapida rassegna non è quello di riflettere sugli aspetti migliorativi che i successivi sviluppi che sono sati apportati al metodo di Rasch introducendo nell'impianto nuovi parametri hanno apportato alla misura di tratti latenti, quanto di suscitare attenzione sul fatto che comunque si tratta di approssimazioni e che ogni approssimazione contiene margini di errore.

Il tema dell'errore di misura sembra rimosso da tutti coloro che vogliono utilizzare la misura a sostegno delle loro decisioni. Invece, proprio in funzione dell'uso che si vuole fare delle misure, è necessario che queste vengano assunte in primo luogo in modo accurato, ma poi anche con la consapevolezza dei limiti che si frappongono tra i nostri modelli e la realtà. Qualsiasi misura è in sostanza una approssimazione. La storia è vecchia fu Pitagora a scoprire che qualsiasi valore attribuito alla diagonale di un quadrato di lato 1 è un'approssimazione, anche se decise di tenerlo nascosto, perché metteva in crisi le sue teorie.

Nella Tabella 6.11 sono riportate le misure di abilità sulla base delle stesse risposte relative all'esempio in Tabella 6.8, calcolate con il metodo CTT e con il modello di della Item Response Theory a 1, 2 e 3 parametri. Come si vede anche nella figura 4 in alcuni casi, cambia anche la posizione relativa dei soggetti. Con campioni più ampi l'effetto è ancora più evidente (Lucisano 2010, p. 44).

È dunque la scelta del modello più che le risposte dei soggetti a determinare la misura ottenuta.

Tabella 6.11. Confronto tra le misure calcolate sulla base di modelli diversi

	Punti grezzi	Abilità calibrata 1 Parametro	Abilità calibrata 2 Parametri	Abilità calibrata 3 Parametri
BELINDA	1	-2,88	-2,15	-3,31
BARBARA	4	-0,52	-0,49	-0,42
DORINA	5	0,04	-0,05	0,09
CARLA	5	0,04	-0,09	0,11
ALFONSO	5	0,04	-0,07	0,29
BEA	6	0,61	0,35	0,45
ELENA	7	1,19	0,84	1,15
ALFIO	7	1,19	0,84	1,2
ALESSIA	7	1,19	0,79	1,42

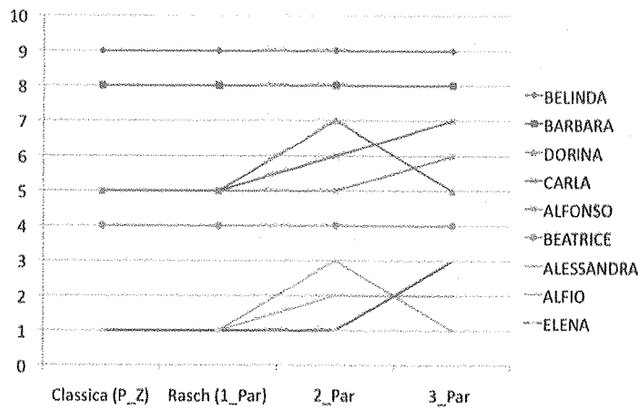


Figura 6.4 spostamenti di rango tra soggetti sulla base dei risultati dello stesso test.

6.6 Chi è senza errore scagli la prima pietra

Abbiamo detto che ogni misura contiene un margine di errore, dunque è necessario avere consapevolezza dell'incidenza dell'errore nelle nostre misure. Come afferma Culligan (2011) l'errore standard cerca di rispondere alla domanda: "Se darò di nuovo questa prova allo stesso studente che punteggio avrà?", o ancora alla domanda "Siamo sicuri che Beatrice (della Tabella 6.1) sia effettivamente più brava di Alfonso, di Carla e Dorina che hanno avuto solo un punto in meno alla prova?"

L'errore standard della media (*Standard Error Measure*, SEM) si calcola utilizzando il coefficiente di attendibilità (KR20) e la deviazione standard (s) dei punteggi attraverso la formula:

$$SEM = s\sqrt{1 - KR20}$$

L'errore standard di misura calcolato in questo modo considera le fonti di errore che sono state usate nel calcolo del coefficiente di affidabilità.

È possibile stimare a partire dall'errore standard l'effetto probabile dell'errore rispetto al punteggio osservato. Per far questo si fa riferimento alla distribuzione dei punteggi attesa sotto una curva normale. Se accettiamo che la nostra stima abbia un errore del 5%, possiamo utilizzare la misura dell'intervallo nel quale sotto la curva normale cade il 95% dei casi e cioè ±1,96 s. Moltiplicando il SEM per 1,96 otterremo la stima di quanto la misura che abbiamo rilevato potrebbe variare in più o in meno.

Calcoliamo ora il SEM del TCL1 (tabella 6.2) abbiamo il valore di s che è 2,108 e KR20 = 0,59.

$$SEM = s\sqrt{1 - KR20} = 2{,}108 \cdot \sqrt[2]{1 - 0{,}59} \cong 1{,}35$$

A questo punto possiamo calcolare l'intervallo di confidenza delle nostre misure moltiplicando il SEM 1,35 per ±1,96, e otteniamo il valore di 2,65. Scopriamo dunque, con una probabilità del 95%, che le nostre misure contengono un errore di più o meno 2,65. Dunque la risposta è che non sappiamo se Beatrice è effettivamente più brava di Alfonso, di Carla e Dorina. Se si osserva la Figura 6.5 in cui abbiamo rappresentato le nostre misure vedrete che in effetti la prova ci dice solo che Alessandra è certamente più brava di Barbara e Belinda, ma per il resto ci dà poche informazioni.

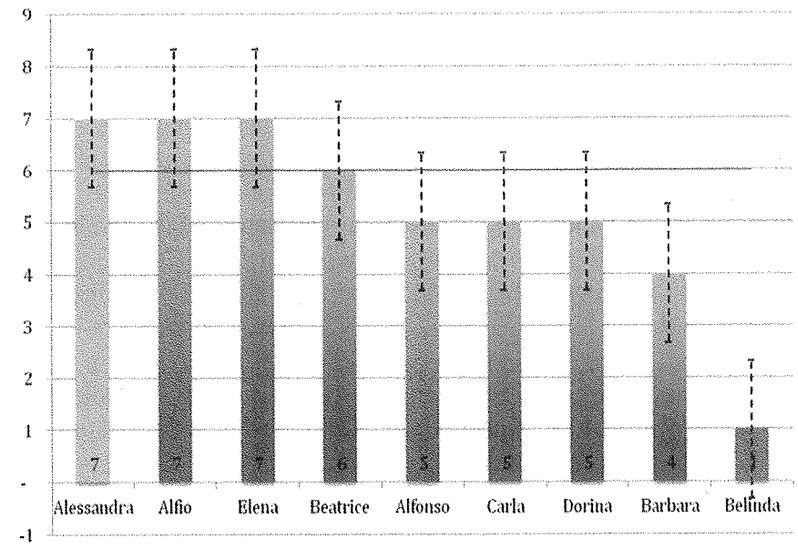


Figura 6.5. Punteggi alla prova TCL1 con intervallo di confidenza.

Anche con la IRT è possibile valutare l'errore standard di misura. In questo caso la procedura è più complessa e fornisce indicazioni sull'errore attribuibile alla misura di ciascun singolo soggetto, con esiti in termini di ampiezza dell'errore non dissimili dal metodo tradizionale. Quello che emerge è che l'errore rimane una componente significativa delle misure che realizziamo. Abbiamo visto come gli stessi dati possono dare luogo a misure diverse in relazione al modello utilizzato per misurare. E dunque è necessario a monte scegliere il modello più adatto al tipo di informazioni che vogliamo ricavare dai dati. Inoltre abbiamo osservato che non è sufficiente limitarsi a interpretare le informazioni, ma è necessario comprendere l'entità e il significato dell'errore che ciascuna informazione si porta appresso.

L'errore standard di misura per il modello di Rasch a un parametro si calcola con la formula:

$$SEM(\theta) = \sqrt{\frac{1}{\sum_i P_i(\theta, b_i) Q_i(\theta, b_i)}}$$

θ è l'abilità del soggetto b_i è la difficoltà dell'item i
P è la probabilità di rispondere correttamente all'item di difficoltà b_i del soggetto di abilità θ
Q è la probabilità di rispondere in modo sbagliato all'item di difficoltà b_i del soggetto di abilità θ

Per chi volesse approfondire l'argomento è utile consultare l'articolo di Ivailo Partchev, *A visual guide to item response theory (2004)* (www.metheval.uni-jena.de/irt/VisualIRT.pdf)

In altre parole, abbiamo imparato che fare calcoli non è poi così difficile. Ciò che è difficile è sapere bene che cosa si vuole fare con le misure e che è necessario dotarsi di strumenti che abbiano una forte coerenza interna. Abbiamo anche verificato che non si possono fare test con poche domande e tararli su pochi soggetti. E, ancora, soprattutto che prima di trarre conclusioni è necessario ripetere le misure più volte.

Se accettiamo, infatti, di giocare con le regole della scienza non possiamo usarne una parte e rimuovere la parte scomoda e dunque dobbiamo imparare a non avere fretta, a fare bene i conti e a convivere con l'errore e a trovare tutti gli accorgimenti per ridurne la portata. E se la scienza con tutti i suoi dubbi non fa notizia, è meglio per chi fa ricerca ritirarsi che cercare in ogni modo l'ascolto dei politici e dei giornalisti rinunciando al rigore.

6.7 Dalle prove standardizzate alle prove su misura

È utile osservare che tutto quanto esposto in questo capitolo fa riferimento prevalentemente alla costruzione di prove lineari tradizionali, in cui a tutti gli studenti vengono somministrate le stesse domande e la lunghezza della prova è predefinita. Negli ultimi anni, grazie anche al progresso della tecnologia, si è ampiamente sviluppata la ricerca in relazione alla costruzione di prove adattative o su misura. La caratteristica principale di questo tipo di prove è che, a partire da una banca di item appositamente costruita (Choppin, 1976), la prova si adatta allo studente, infatti si usano le potenzialità computazionali e di interattività offerte dal computer per somministrare un insieme di item che viene determinato nel momento in cui si effettua la prova. Il vantaggio è che il soggetto si troverà a rispondere ad item che hanno un livello di difficoltà via via sempre più adeguato al suo livello di abilità, massimizzando l'informazione e riducendo significativamente l'errore. Esistono diverse tipologie di test adattivi, le principali sono i test adattativi item per item (CAT) e i test adattativi multilivello (MST).

Nei classici test adattivi l'adattamento della prova all'abilità dello studente avviene dopo la somministrazione di ogni singolo item, spesso a partire da un livello di abilità stimato sulla base di altri parametri. Il disegno di un test multilivello (MST) è caratterizzato invece dal fatto che ciascun livello è rappresentato da un insieme di item, detto modulo o *testlet*, di difficoltà predefinita; l'adattamento del test all'abilità dello studente avviene quindi sulla base del rendimento cumulativo su un insieme di item piuttosto che sul risultato ottenuto in ogni singolo item. In entrambi i casi, CAT e MST, se lo studente sta svolgendo bene la prova gli verrà somministrato un item o un insieme di item più difficili, viceversa, un item o un insieme di item più facili. I test multilivello offrono inoltre rispetto ai test adattivi item per item il vantaggio di un maggior controllo sull'assemblaggio delle forme e sulla validità di contenuto.

Generalmente in questi modelli l'abilità dello studente e la difficoltà degli item vengono misurati sulla stessa scala utilizzando il modello di Rasch o, più raramente, modelli IRT a più parametri. Le prime prove adattative sono state costruite nell'ambito della teoria classica dei test (CTT), ma le possibilità di costruire una banca con caratteristiche misuratorie robuste, e quindi prove adattative, si sono rafforzate nell'ambito dell'Item response theory (IRT) (Birnbaum, 1968; Hambleton e Swaminathan, 1985; Hambleton, Swaminathan e Rogers, 1991), che ha permesso di superare alcuni dei limiti posti dalla CTT (Weiss, 1985), in particolare la dipendenza della misura dell'abilità dal gruppo di soggetti a cui la prova viene somministrata.

Approfondimento 6.4

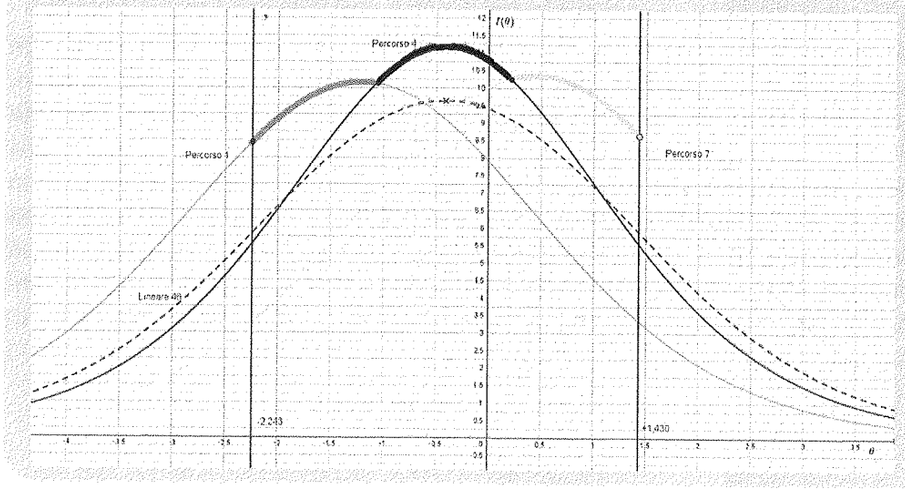
In una recente ricerca (Botta, 2021) è stata sperimentata la costruzione di una prova adattiva MST per la stima delle abilità in matematica degli studenti del grado 10 in Italia. La prova, costituita da 3 livelli e 7 possibili percorsi permette di stimare con accuratezza l'abilità degli studenti che ricadono in tre distinti intervalli, abilità bassa, media o alta, coprendo un ampio range del tratto latente.

La ricerca illustra nel dettaglio il laborioso percorso di costruzione della prova in tutte le sue fasi, dalla definizione del costrutto e conseguente costruzione della banca di item e verifica della sua adeguatezza al modello IRT a un parametro, fino all'assemblaggio e alla somministrazione della prova stessa. Nell'ambito della ricerca si mostra con chiarezza che i limiti delle prove lineari e di quelle adattative classiche, item per item sono effettivamente superati.

Oltre alle verifiche teoriche si è visto, con confronti diretti, che la prova risulta essere significativamente più informativa di una prova lineare costruita sulla medesima banca di item, sia in generale sia guardando solo alla sua parte adattativa. Permette dunque di stimare con maggiore attendibilità l'abilità degli studenti entro un ampio intervallo del continuo, migliorando le stime in generale per tutti gli studenti e in particolare per coloro che si collocano agli estremi dell'intervallo. Abbiamo visto che la prova riesce al contempo ad offrire a studenti con abilità scarse un congruo numero di item a cui sono effettivamente in grado di rispondere e agli studenti con abilità molto alte item che risultano ancora sfidanti ed effettivamente in grado di indagare all'interno della zona di sviluppo prossimale (Vygotskij, 1974 e 1987).

In fase di costruzione e assemblaggio della prova è stato possibile porre e rispettare stringenti vincoli di copertura del costrutto che ne hanno garantito la validità di contenuto.

La figura riporta i grafici delle funzioni informative della funzione lineare e dei tre percorsi principali della prova MST, ed in essa è messo in evidenza l'intervallo di abilità usato come riferimento per la costruzione della prova MST. Si può osservare che i percorsi della prova MST sono sempre più informativi della prova lineare, ciascuno nel proprio intervallo di riferimento.



Riepilogo

In questo capitolo abbiamo esaminato le cautele necessarie per poter mettere a punto prove basate su domande con risposte a scelta multipla. Si tratta di metodologie che consentono di migliorare la qualità delle misure. Sono stati presentati i due modelli prevalenti di item analisi quello legato alla Classical Test Theory e quello della Item Response Theory che prende le mosse dal lavoro di Rasch. Si è comunque voluto sottolineare come le misure ottenute da una prova, sia pur ben costruita, rappresentano solo una approssimazione, basata su modelli teorici di quello che stiamo cercando di misurare. Per questo mentre in relazione ai nostri obiettivi è necessario utilizzare con rigore le tecniche statistiche che consentono di migliorare le misure, spetta al ricercatore colmare con la sua intelligenza, l'osservazione e "attenzione educativa" gli spazi che restano tra la misura e la dimensione dell'esperienza di chi ha risposto alla prova. In questa prospettiva si è presentato un approfondimento sulle prove adattative, che oggi costituiscono la migliore soluzione per la costruzione di test capaci di restituire misure che meglio rispondono ad una descrizione corretta delle conoscenze o delle abilità che si intendono misurare.

Riferimenti bibliografici

Anastasi A. (1981). *I test psicologici* (1976). Milano: Franco Angeli.

Birnbaum A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading: Addison-Wesley.

Botta, E. (2021). *Sperimentazione di un modello adattativo multilivello per la stima delle abilità in matematica nelle rilevazioni su larga scala,* Roma: Nuova Cultura

Bottani N. (2011). Concorso per dirigente, non si scherza sui questionari. *Tuttoscuola,* 13 settembre 2011. From: http://www.tuttoscuola.com/cgi-local/disp.cgi?ID=26489.

Calonghi L. (1978). *Statistiche d'informazione e di valutazione* (2 vol.). Roma: Bulzoni.

Carmines E. G., Zeller R. (1979). *Reliability and validity assessment.* London: Sa ge.

Choppin B. (1976). *Developments in item banking.* Windsor U.K.: Monitoring National Standards of Attainment in School, 1976

Cronbach L. J., Meehl P. E. (1955). Construct validity in psichological tests. *Psychological Bulletin,* 52, 281-302.

Cronbach, L. J., (1974). *Essential of Psychological Testing.* London: Harper (trad. it. *I test psicologici,* Firenze, Giunti-Barbera, 1977).

Culligan B. (2011). *Item response theory, reliability and standard error.* From: http://www.wordengi ne.jp/resear ch/pdf/IRT_reliability_and_standard_error.pdf

FARR R., CAREY R. F. (1986). *Reading. What can be measured?* Newark: IRA.
GATTULLO M. (1961). *Didattica e docimologia.* Roma: Armando.
HAMBLETON, R.K. AND SWAMINATHAN H. (1985). *Item Response Theory: Principles and applications.* Boston, MA: Kluwer-Nijhoff
HAMBLETON R. K., SWAMINATHAN H., ROGERS H. J. (1991). *Fundamentals of Item Response Theory.* London: Sage Publications Inc.
HOGAN T. P., AGNELLO J. (2004). An empirical study of reporting practices concerning measurement validity. *Educational and Psychological Measurement,* 64, 802-812.
LIVINGSTON S. A. (1985). Reliability of test results. In T. HUSEN, N. POSTLETWHITE, *The International Encyclopedia of Education* (vol. VII, pp. 201-210). Oxford: Pergamon Press.
LORD F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale: Erlbaum.
Lucisano P. (1982). *Lettura e comprensione.* Torino: Loescher.
LUCISANO P. (2003). Validità e affidabilità delle pratiche valutative: a proposito del progetto Pilota 2. *Cadmo,* 11, 2, 37-56.
LUCISANO P. (2010). Una prova di abilità linguistiche per l'uscita dai corsi di formazione professionale. *Educational, Cultural and Psichological Studies,* 1, 25-52.
NUNNALLY J. C. (1978). *Psychometric theory.* New York: McGraw Hill.
PUMFREY P. D., (1985). *Reading: tests and assessment techniques.* Sevenoaks: Hod der & Stoughton.
PYRCZAK F. (1979). Definition of Measurement Terms. In R. Schreiner, *Reading test and teachers. A Practical guide,* Newark: IRA.
RASCH G. (1993). *Probabilistic models for some intelligence and attainment tests,* with a forward and afterword by B. D. Wright. Chicago: Mesa Press (1960, Danish Institute for Education Research).
THISSEN D. (2000[2]). Reliability and measurement precision. In H. Wainer (Ed.), *Computerized adaptive testing. A primer* (pp. 159-183). Mahwah: Lawrence Erlbaum Associates.
VISALBERGHI A. (1955). *Misurazione e valutazione nel processo educativo.* Milano: Edizioni di Comunità. Wright B. D., Masters G. N. (1982). *Rating scale analysys.* Chicago: Mega Press.
VYGOTSKIJ, L. S. (1974). *Storia dello sviluppo delle funzioni psichiche superiori e altri scritti,* Giunti Barbera, Firenze 1974, pp. 201-202.
VYGOTSKIJ L. S. (1987). *Il processo cognitivo.* Torino: Universale Scientifica Boringhieri
Weiss, D. J. (1985). Adaptive testing by computer. *Journal of consulting and clinical psychology,* 53(6), 774

7

Disegni sperimentali di ricerca, microsperimentazione didattica e osservazione sistematica[1]

In questo capitolo vengono presentate, principalmente, le caratteristiche procedurali della ricerca sperimentale e dell'osservazione sistematica in campo educativo. Si prendono così in esame il modello sperimentale classico, a uno o più gruppi di controllo, i disegni di quasi sperimentazione con particolare riferimento alle procedure formali della loro conduzione e alla messa a punto del piano della ricerca. L'analisi delle modalità di verifica delle ipotesi di ricerca attraverso l'ideazione e la conduzione di un esperimento in senso stretto, si arricchisce con la disamina di un particolare tipo di indagine didattica, assai circoscritta, "condotta su punti specifici e esattamente delimitati, esauribili in un lasso di tempo ragionevolmente breve", proposta da Domenici e Vertecchi agli inizi degli anni Ottanta, a circa un secolo dalla nascita della pedagogia sperimentale, e da loro denominata "microsperimentazione". Il micro viene inteso non già come fattore riduttivo dell'importanza dell'osservazione e/o della sperimentazione svolte secondo quel modello e delle sue ricadute conoscitive, bensì come rinvio alla dimensione contenuta della proposta al fine di garantire un miglior controllo delle variabili allo studio. Come è noto tale controllo è necessario ad ogni ortodossa sperimentazione scientifica, compresa quella di tipo educativo caratterizzata da un elevato livello di complessità.
Vengono, così, presentati alcuni esempi di microsperimentazioni diversificate relative a varie tematiche: dalle strategie didattiche per la gestione del tempo a scuola, alla formazione dei docenti, all'atteggiamento degli stessi, alle ricadute sull'apprendimento degli allievi, alla valutazione dell'apprendimento online.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Il modello sperimentale classico, a uno o più gruppi di controllo e i disegni sperimentali;
- I disegni di quasi-sperimentazione come evenienza conoscitiva straordinaria e procedure formali della loro conduzione;
- I disegni di micro-sperimentazione didattica;
- Il piano della ricerca in ambito sperimentale.
- Analisi e protocolli di una Microsperimentazione per la formazione dei docenti e di una Microsperimentazione per lo studio dei processi di apprendimento degli allievi;
- L'Osservazione sistematica. Caratteri generali

[1] Il presente contributo è frutto del lavoro congiunto dei due autori. In particolare V. Biasi ha redatto i paragrafi 7.2, 7.3 e 7.4. G. Domenici ha scritto i paragrafi 7.5 e 7.6. Il paragrafo 7.1 è stato redatto in collaborazione.

Insomma, non si può osservare un'onda senza tener conto degli aspetti complessi che concorrono a formarla e di quelli altrettanto complessi a cui essa dà luogo. Questi aspetti variano continuamente, per cui un'onda è sempre diversa da un'altra onda; ma è anche vero che ogni onda è uguale a un'altra onda, anche se non immediatamente contigua o successiva; insomma ci sono delle forme e delle sequenze che si ripetono, sia pur distribuite irregolarmente nello spazio e nel tempo. Siccome ciò che il signor Palomar intende fare in questo momento è semplicemente vedere un'onda, cioè cogliere tutte le sue componenti simultanee senza trascurarne nessuna, il suo sguardo si soffermerà sul movimento dell'acqua che batte sulla riva finché potrà registrare aspetti che non aveva colto prima; appena s'accorgerà che le immagini si ripetono saprà d'aver visto tutto quel che voleva vedere e potrà smettere.

Italo Calvino, *Palomar*

7.1 Disegni sperimentali: modello sperimentale classico e quasi-sperimentazioni

Dall'analisi dei principali *approcci alla ricerca* e delle caratteristiche peculiari che dovrebbe possedere una *ipotesi*, sia essa fattuale che teorica, così come indicato nel Capitolo 2, è facile desumere che, con buona approssimazione, le pur diverse modalità di compiere indagini scientifiche in ambito educativo possano ben articolarsi in due grandi tipologie generali. Quelle nelle quali il ricercatore ha la possibilità di modificare volontariamente una parte della realtà di studio per poter cogliere, attraverso la conduzione della stessa ricerca, i mutamenti eventualmente prodotti dal suo intervento; quelle, invece, nelle quali tale possibilità non è data, sia per evidenti limitazioni "materiali", fattuali, della realtà di studio, sia per limitazioni deontologiche, se si vuole, strettamente connesse all'etica professionale dei ricercatori, storicamente determinate. Nel primo caso avremo *ricerche di tipo sperimentale*; nell'altro *ricerche-non sperimentali*.

Un carattere a parte, o se si vuole ibrido, posseggono le situazioni che permettono al ricercatore di studiarle come se fossero veri e propri esiti di una modificazione da lui stesso indotta (attraverso l'introduzione di fattori o variabili indipendenti, cioè sperimentali), sia in contesti nei quali quel tipo di modificazione è di fatto impossibile, sia in quelli nei quali è possibile in linea di principio ma non praticabile o inibita sul piano etico-professionale. In entrambi questi casi si parlerà

di ricerche *quasi-sperimentali*.

Escludendo dall'analisi, in questo capitolo, tutte le ricerche non-sperimentali – che, occorre ricordare, hanno contribuito e contribuiscono non poco all'arricchimento dei quadri di conoscenza in area educativa – esamineremo le principali caratteristiche delle ricerche sperimentali e quasi-sperimentali, che peraltro risultano le meno praticate anche a causa dei forti vincoli formali, d'ordine soprattutto quantitativo oltre che qualitativo, cui non solo si debba, ma si possa anche sottostare padroneggiando tra l'altro sofisticate tecniche statistiche.

Sperimentare significa, dunque, alterare intenzionalmente alcuni elementi o fattori di una certa realtà, controllandone con continuità gli effetti procedurali e quelli finali in modo che si possano cogliere le relazioni tra modifiche introdotte e risultati registrati (Calonghi, 1977, 1984, 1994; Domenici, 2006; Domenici, Lucisano e Biasi, 2017).

Una sperimentazione può quindi nascere e può essere attuata quando la situazione o l'ambito generale nel quale si decide di operare risultino in parte "conosciuti", "indagati", ma dei quali non siano ancora note le più importanti variabili che interagiscono nel contesto specifico di riferimento. Ciò presuppone che si sia ragionevolmente certi che i risultati dell'iniziativa sperimentale non saranno quantomeno peggiori (o più dannosi) di quelli normalmente conseguiti nella situazione ordinaria, condizione, com'è facile intuire, certo inaccettabile sul piano deontologico.

Ciò presuppone che si sia ben studiato il *problema*, talvolta anche riproducendolo in laboratorio in modo un po' artefatto. Sussiste quindi una circolarità fra problema e letteratura (ovvero conoscenze) sull'argomento.

Soprattutto nella ricerca sperimentale, infatti, una progressiva messa a fuoco degli elementi che caratterizzano il contesto di riferimento, delle soluzioni prima e/o altrove tentate per risolvere lo stesso problema o questioni affini, l'eventuale svolgimento di osservazioni e indagini preliminari, diventano condizioni necessarie, ancorché non sufficienti per una sua piena riuscita.

Sulla base di tutti gli elementi conoscitivi, raccolti talvolta anche dall'esperienza diretta e soprattutto dalla *ricognizione della letteratura sull'argomento*, si potrà meglio definire descrittivamente e analiticamente l'ipotesi inizialmente formulata.

In forza di questa operazione intellettuale complessa e delicata verrà alla fine messo a punto il *piano sperimentale*, e saranno ideati e costruiti gli strumenti per il controllo-confronto della situazione finale (in uscita) rispetto a quella (iniziale) d'ingresso.

Ogni attività sperimentale dovrebbe infatti rappresentare una modificazione intenzionale e controllata di alcuni fattori – chiamati *variabili indipendenti* –, in una situazione data, che presenta ovviamente alcuni vincoli che non possono essere alterati, *variabili assegnate* – di cui bisogna tener conto per "vedere" come mutano

le *variabili dipendenti.*

Ma mutano, se mutano, rispetto a che cosa? E quanto? Ecco dunque la necessità di formulare, nell'*ipotesi,* lo scarto che si vuol registrare in questo mutamento come prova della validità dell'ipotesi stessa e della sua conferma empirica.

L'ipotesi dovrebbe in altri termini contenere in sé il *criterio di accettabilità* o meno, sul piano quantitativo e qualitativo, dei risultati della sperimentazione.

Il metodo sperimentale si articola sostanzialmente in quattro fasi successive: la prima si riconosce naturalmente nell'*osservazione* del fenomeno allo studio, ovvero del problema, seguono il *ragionamento* (con problematizzazione e messa a punto dell'ipotesi da testare), la *pianificazione* e l'attuazione dell'esperimento e della rilevazione di dati, l'*interpretazione* dei risultati conseguiti.

Dal punto di vista procedurale, oltre che statistico, è utile formulare - come si è peraltro giù accennato nel Capitolo 4 - due tipi di ipotesi: quella definita *ipotesi nulla* (H_0), secondo la quale l'intervento sperimentale (l'alterazione operata nel gruppo sperimentale) non modificherà significativamente gli esiti che si registrano o si registreranno in condizioni ordinarie (rappresentate dal gruppo di controllo nel processo sperimentale), ovvero che non vi siano differenze tra i due diversi campioni osservati; e l'*ipotesi* H_1, alternativa ad H_0, che invece si vuole sottoporre a verifica proprio attraverso la ricerca.

L'ipotesi nulla postula quindi la non sussistenza del cambiamento della situazione, nonostante l'alterazione della o delle variabili sperimentali, oppure la non significatività delle differenze tra H_0 e H_1, cioè tra i due campioni. Ovviamente essa viene formulata con il proposito di respingerla - proprio attraverso lo svolgimento dell'indagine - a favore dell'ipotesi alternativa. Ovviamente occorre usare opportuni test statistici capaci di farci accettare o respingere l'ipotesi nulla. L'*ipotesi,* dunque, risulta di grande rilievo, poiché ci permette di stabilire se lo scarto registrato al termine della sperimentazione sia attribuibile all'influenza delle *variabili indipendenti, sperimentali,* cioè alla situazione intenzionalmente modificata e determinata nel processo attivato, oppure a *variabili non controllate o sconosciute,* casuali o addirittura *parassitarie o intervenienti.*

Per compiere un esperimento possiamo operare in diversi modi (Calonghi, 1984, 1994; Domenici, 2006).

Il più diffuso consiste nel porre a confronto due situazioni, l'una per così dire usuale, ordinaria; l'altra nella quale si siano volontariamente alterati alcuni aspetti proprio per conoscere gli effetti dell'alterazione introdotta. Potremmo, per esempio, fare riferimento a due gruppi di allievi: in questo caso uno rappresenterebbe il gruppo sperimentale, cioè quello in cui alteriamo volutamente alcuni fattori del contesto (variabili indipendenti), l'altro, il gruppo di controllo, così come è con-

templato dal cosiddetto *disegno sperimentale classico.*

Lo schema che segue, rappresenta visivamente l'insieme delle operazioni da compiere nel caso del *disegno sperimentale classico*:

$$G_1 \longrightarrow T_1 \longrightarrow R_1$$

$$G_2 \longrightarrow T_2 \longrightarrow R_2$$

Legenda: G_1 =gruppo di controllo; G_2 = gruppo sperimentale; T_1 = trattamento "normale" o ordinario o assenza di intervento; T_2 = trattamento sperimentale; R_1 = risultato conseguito dal gruppo di controllo; R_2 = risultato conseguito dal gruppo sperimentale.

Quando si voglia fare riferimento ai problemi scolastici, per poter rilevare quei dati, bisogna poter osservare sistematicamente comportamenti cognitivi e non, utilizzare test di verifica dell'apprendimento, e altri strumenti di rilevazione dei dati. Innanzitutto occorre fare in modo che i due gruppi (G_1 e G_2) siano identici o equivalenti, non devono, cioè, presentare diversità significative rispetto alle variabili prese in considerazione. Per questa ragione, quando è possibile la scelta dei componenti dei gruppi deve essere random, cioè casuale.

Il successivo trattamento, ovviamente, deve essere differenziato. Nel G_1 opereremo normalmente, nel G_2 opereremo in funzione dell'ipotesi della ricerca, ovvero introducendo il fattore o variabile sperimentale (T_2).

Ora va detto che non sempre si può avere la possibilità di fare confronti con gruppi di controllo. Se ci troviamo nell'impossibilità di agire con due gruppi contestualmente, possiamo allora utilizzare lo stesso gruppo prima come gruppo di controllo e, poi, come gruppo sperimentale. Opereremo, cioè con un *gruppo unico ricorrente,* ovvero agiremo sullo stesso gruppo con due trattamenti diversi e in tempi diversi, per rilevare come mutano i risultati dell'attività che abbiamo volta a volta messo in atto.

In un certo periodo di tempo agiremo cioè sul gruppo senza modificare intenzionalmente la situazione e registreremo gli esiti; successivamente gli interventi verranno alterati ovvero saranno quelli previsti dal trattamento sperimentale, per confrontare, infine, i rispettivi risultati (R_1; R_2), cosi come è illustrato dallo schema precedente.

Nelle situazioni, poi, nelle quali si paventano influenze di variabili parassitarie, connesse a fattori non ben conosciuti, per essere sicuri dell'effetto sperimentale, può rivelarsi utile operare - sempre che sia possibile - su tre o addirittura quattro gruppi. In quest'ultimo caso, al terzo gruppo non si rileveranno i dati iniziali, ma

si introdurrà la variabile sperimentale; nel quarto gruppo si ometterà la variabile sperimentale, ma si rileveranno i dati in uscita dalla sperimentazione.

Si andranno, quindi, a rilevare e confrontare i risultati successivi al trattamento per coglierne l'eventuale significatività. Un problema di non poco conto, come più volte è stato ribadito, è *disporre di strumenti di misura e controllo delle caratteristiche iniziali e in uscita* (oltre che processuali) che ci forniscano dati e/o informazioni altamente affidabili, cioè valide e attendibili, il cui trattamento avvenga attraverso l'impiego di statistiche adeguate.

Dati, informazioni (e protocolli di ricerca) devono essere affidabili, in quanto devono rispecchiare di fatto quello che è avvenuto ed essere perciò costanti e fedeli. Nelle stesse situazioni, con gli stessi trattamenti dovremmo registrare gli stessi risultati. La ricerca sperimentale dovrebbe essere infatti *nomotetica*, dovrebbe cioè farci pervenire a leggi generali o almeno tendere verso di esse.

Gli esiti cui si perviene con tale ricerca, più che con altre tipologie di indagini scientifiche, possono rappresentare conclusioni esprimibili, infatti, se non proprio come leggi generali, in forma appunto nomotetica, almeno come proposizioni generali. "Proposizioni generali e nessi tra le variabili (cioè tra gli esiti educativi e i fattori che vi influiscono) sono 'verità di media', per usare un'espressione di A. Binet, sono affermazioni cioè che riguardano (probabilisticamente) la generalità dei casi, ma non sono necessariamente valide per ogni caso" (Calonghi, 1994).

Il problema fondamentale è, quindi, quello di avere *strumenti* adeguati per la verifica delle situazioni di *ingresso* e di *uscita*. Per esempio, nell'ambito della valutazione degli apprendimenti, gli strumenti che ci permettono di avere queste informazioni sono di tipo strutturato (test oggettivi) o semistrutturato, comunque strumenti che permettono di svolgere "osservazioni sistematiche" affidabili, in modo diretto e/o mediato.

Come già detto, il *metodo sperimentale* comporta la manipolazione di una variabile indipendente per verificarne l'effetto sulle variabili dipendenti. Le conclusioni saranno però accurate se non verranno influenzate da distorsioni (o *bias*). Questi bias possono essere a carico:

- del ricercatore;
- dei partecipanti o soggetti sperimentali;
- del disegno di ricerca.

Per evitare queste distorsioni occorre operare in *condizioni controllate*, costanti e uniformi, ed è quindi necessario ricorrere a:

- *standardizzazione delle procedure*: ossia all'utilizzo di procedure uniformi e costanti in ogni fase della sperimentazione e per ogni partecipante;

- *standardizzazione delle norme*: ossia all'uniformità nell'attribuzione dei punteggi e nella valutazione dei dati.

La standardizzazione del significato di un concetto implica darne una *definizione operazionale:* vale a dire formulare un concetto nei termini delle procedure utilizzate per determinarlo.

Nelle realizzazione delle procedure sperimentali sono però sempre necessarie alcune accortezze. Negli esperimenti può infatti intervenire, come si è accennato, una *variabile confondente o interferente*, ovvero un fattore che modifica il comportamento dei partecipanti a un esperimento in maniera non prevista dal disegno sperimentale. In sintesi, esistono due tipi di variabili confondenti:

- *effetto dell'aspettativa* (effetto Rosenthal): distorsione dei risultati provocata dall'aspettativa che lo sperimentatore o i soggetti sperimentali hanno in merito ai risultati stessi;
- *effetto placebo*: distorsione dei risultati provocata quando i partecipanti a un esperimento modificano le loro risposte in assenza di qualunque tipo di manipolazione sperimentale (per esempio in funzione di effetti di suggestione positiva o negativa - effetto nocebo - sul verificarsi dei risultati attesi).

Metodo sperimentale: disegni di ricerca

Con il termine "Disegno di ricerca" si indica la messa a punto di una procedura scientifica che consente di escludere una interpretazione ambigua dei risultati, tenendo sotto controllo le possibili variabili interferenti. Si distinguono i seguenti modelli.

Disegno tra i soggetti (*between-subjects*): disegno sperimentale in cui i partecipanti sono assegnati casualmente alla condizione sperimentale o alla condizione di controllo. In questo modo si hanno gruppi sperimentali e gruppi di controllo, in base all'esposizione o meno alle condizioni della variabile dipendente.

Disegno entro i soggetti (*within-subjects*): utilizza ogni partecipante come controllo di se stesso ed esamina per esempio il comportamento o la performance di uno dei partecipanti prima del trattamento e dopo il trattamento.

Quasi-esperimenti: l'ipotesi allo studio viene verificata avvalendosi di gruppi cosiddetti sperimentali già presenti in natura. I quasi-esperimenti sono infatti condotti in un contesto naturale o sul campo. Occorre tenere presente che le inferenze basate sui dati di un quasi-esperimento non sono molto *affidabili*, poiché non si possono tenere sotto controllo tutte le variabili confondenti: ci si dovrà quindi avvalere di molteplici osservazioni condotte anche in contesti diversi per poter innalzare la generalizzabilità dei risultati conseguiti.

Attuare *disegni pre-sperimentali* (necessari per dissodare il terreno in vista della sperimentazione vera e propria) o le cosiddette *"quasi- sperimentazioni"* è per alcuni versi più agevole, talvolta rappresenta l'unica via conoscitiva praticabile.

Infatti, a causa di alcuni limiti deontologici, cioè di vincoli in qualche modo imposti dall'etica professionale, non sarebbe lecito, per esempio, sperimentare oltre un certo limite gli effetti dell'aggressività sull'apprendimento poiché essa potrebbe diventare irreversibile così producendo sul campione coinvolto risultati eticamente inaccettabili. Possiamo invece rilevare tali effetti in particolari contesti dove per qualche ragione vi sono soggetti particolarmente aggressivi, operando perciò come se la situazione fosse stata intenzionalmente prodotta dal ricercatore. Avremo, allora, una *quasi-sperimentazione*: utilizzando lo stesso modulo classico della sperimentazione, cioè il confronto tra due gruppi, impiegando strumenti di misura e valutazione e altro registreremo gli esiti dell'aggressività, ma senza creare artatamente una situazione pericolosa o non accettabile per verificarne sperimentalmente l'ipotesi sottesa.

Nel campo delle scienze dell'uomo, la maggior parte di quelli che i profani chiamano "esperimenti", sono in realtà delle "quasi-sperimentazioni": si impiegano cioè quelle situazioni che si verificano in natura, ovvero in contesti particolari che possono apparire "naturali", usuali, nella nostra esperienza diretta e/o indiretta, per poter rilevare come mutano le variabili dipendenti in rapporto alle variabili indipendenti, che in questo caso sono per così dire, appunto, naturali, mentre nella sperimentazione in senso stretto esse vengono intenzionalmente prodotte. In altre circostanze, attraverso l'osservazione sistematica, lo studio dei casi o altro potremmo identificare e strutturare vere e proprie situazioni quasi o pre-sperimentali, cioè tali da consentire, successivamente, sempre che ci siano le condizioni, l'attuazione piena di una sperimentazione.

Metodo sperimentale: le procedure di controllo

Si tratta di strategie utilizzate dai ricercatori per mantenere costanti tutte le variabili e le condizioni non legate all'ipotesi che deve essere verificata.

Controllo a singolo cieco e a doppio cieco: nel primo caso, i partecipanti non sono a conoscenza delle condizioni sperimentali in cui si trovano; nel secondo caso, anche lo sperimentatore non ne è a conoscenza.

Controllo placebo: Condizione sperimentale in cui non è somministrato il trattamento o non vi è una manipolazione della variabile indipendente.

Compiere la *verifica delle ipotesi* con una attività sperimentale presuppone e comporta che le ipotesi stesse siano state preventivamente formulate in modo che siano identificabili univocamente gli esiti attesi, e che gli stessi siano espressi in termini quantitativi.

Infatti senza un preciso *riferimento a indicatori capaci di mettere in collegamento la tipologia delle variazioni delle misure*, prima e dopo l'intervento sperimentale, risulta assai difficile se non impossibile trarre delle conclusioni dagli esiti registrati.

La mancata definizione descrittiva delle ipotesi da provare (detta *operazionalizzazione*) o la mancata programmazione delle fasi e delle operazioni da svolgere per il controllo dei risultati sperimentali conducono, o possono condurre, a una reale impossibilità di verificare rigorosamente le ipotesi sottoposte alla prova dei fatti.

La verifica non può fare riferimento solo ad aspetti qualitativi che in sé sono certo di grandissimo rilievo, ma che tuttavia per essere attentamente esaminati presuppongono riferimenti a dati affidabili, cioè validi e attendibili sull'andamento delle variabili e dei fattori entrati in gioco nel processo sperimentale. In altri termini, occorre tener presente che l'oggetto della verifica è sempre costituito dall'esito del confronto tra variabili indipendenti e variabili dipendenti, tra la situazione sperimentale e quella di controllo.

Per un verso è dunque indispensabile, come si è già accennato, che l'ipotesi formulata contenga espliciti riferimenti qualitativi e quantitativi alle variabili indipendenti (dette anche sperimentali) e a quelle dipendenti (dette anche risultato), per altro verso è necessario che le operazioni della verifica siano state programmate ancor prima dell'attivazione del vero e proprio processo sperimentale.

Senza la sussistenza di queste due condizioni le inferenze assumibili sulla base dei dati raccolti sarebbero scorrette in quanto, come abbiamo già fatto cenno, sui risultati potrebbero aver giocato un ruolo rilevante variabili estranee (parassitarie) alle ipotesi, senza peraltro avere nessuna possibilità di verifica. Si vuol dire che senza tali accortezze di rigore non si saprà mai se quegli stessi esiti si sarebbero potuti raggiungere senza le modificazioni introdotte in via sperimentale.

Vanno dunque previste procedure di intervento processuale (che si svolgeranno durante l'attività sperimentale) e procedure e strumenti da impiegarsi per l'analisi dei dati in uscita tali da neutralizzare, per così dire, i possibili effetti delle *variabili estranee* (*variabili di disturbo o parassitarie*) su quelle strettamente sperimentali (variabili indipendenti).

Approfondimento 7.1

Un esempio di immediata comprensione dell'incidenza che possono assumere le variabili impreviste e/o parassitarie in ambito di ricerca, è dato dagli esiti della "Indagine sul profitto scolastico nelle scuole secondarie superiori sperimentali" svolta in Italia quando le classi della prima consistente sperimentazione nazionale di ordinamenti curricolari fortemente innovativi terminarono, con gli esami di maturità nel 1980, il loro percorso quinquennale.[9] Svolta con i test IEA di matematica, scienze, comprensione della lettura, letteratura educazione civica e lingua straniera, inglese o francese, interessò pressoché l'universo delle classi sperimentali e un equivalente campione di controllo, della stessa numerosità, costituito da classi a regime didattico ordinario. Anche se il gruppo sperimentale riportò complessivamente risultati più elevati

del gruppo di controllo (ciascun campione era costituito da 1200 studenti), gli esiti significativamente più alti (eccettuati quelli relativi a una delle tre prove di matematica calibrate per tipo di scuola) si rilevarono casualmente - durante l'analisi e il trattamento dei dati - in 200 studenti iscritti in classi a regime ordinario allocate prevalentemente in istituti con una prevalenza di classi sperimentali. Docenti e studenti di questo gruppo, definito "misto", opportunamente intervistati rivelarono che la loro richiesta di far parte delle classi sperimentali non era stata accolta a causa delle limitazioni imposte dall'apposita normativa ministeriale (si consideri che tale limitazione, congiunta alla grande attesa dell'innovazione sperimentale, aveva costretto molti genitori a pernottare davanti alle scuole per cercare di iscrivere, non sempre con successo, i propri figli alle classi sperimentali a numero limitato). Per questa ragione - vera e propria variabile, interveniente imprevista perché non se ne aveva conoscenza durante la composizione del campione - quegli allievi e quei docenti, avevano informato atteggiamenti e disposizioni verso l'insegnamento e l'apprendimento a un impegno inusuale e per molti versi sfidante, con risultati che è facile desumere dai dati riportati nella tabella che segue (gli esiti di queste classi di controllo - definite miste - sono riportati nella terza colonna).

Tabella riassuntiva dei risultati registrati*

*Fonte: *La Documentazione Educativa*, "Indagine sul profitto scolastico nelle scuole secondarie superiori sperimentali", n. 1, MPI-Istituto della Enciclopedia Italiana, Roma, 1981.

Prove IEA discipline		Sperim.	Contr.	Miste
Mat. X	M	≠ 8.28	7.89	14.85
	d.s.	5.95	5.50	8.29
Mat. Y		≠ 9.87	7,42	—
		5.80	5.19	
Mat. Z		10.47	≠ 12.02	9.61
		7.14	5.65	6.54
Scienze		≠ 18.71	16.77	26.93
		10.01	8.58	7.88
Comprens. lettura		≠ 26.79	26.75	32.24
		8.15	8.87	7.79
Letteratura		22.63	≠ 22.95	28.52
		6.66	6.95	5.99
Educazione civica		22.87	≠ 23.16	32.21
		8.22	8.60	7.73
Inglese		36.04	≠ 39.29	47.89
		15.00	14.87	12.71
Francese		47.75	≠ 49.41	56.32
		15.07	13.39	11.01

*La ricerca affidata dall'Ufficio Studi e Programmazione del MPI al Centro Nazionale Italiano Tecnologie Educative (CNITE) è stata effettuata da M Laeng, G. Domenici e dall'analista E. Nervegna, con la collaborazione di B. Vertecchi. (Cfr. MPI-CNITE, *Indagine sul profitto scolastico nelle scuole secondarie superiori sperimentali*, Roma, CNITE,1980, ora in *La Documentazione Educativa*, n. 1, MPI-Istituto dell'Enciclopedia Italiana, Roma, 1981)

7.2 La sperimentazione a scuola

Il termine "sperimentale" nall'ambito della ricerca educativa è diventato spesso - nella realtà della scuola - sinonimo di "transitorio" o "provvisorio" rispetto alle norme ministeriali vigenti in un dato periodo.

Si trattava quindi più di una "messa alla prova", come nel caso di questioni relative all'introduzione del tempo pieno o di forme particolari di sostegno, di nuovi curricoli educativi, in rapporto sia a strutture e ordinamenti sia alla didattica (artt. 3 e 4 del DPR 419/1977), che di vere e proprie sperimentazioni.

Spesso si finiva per burocratizzare le nuove procedure perdendo di vista il controllo rigoroso delle componenti in gioco e quindi rendendo difficile o arbitrario trarre conclusioni sull'efficacia dei nuovi ordinamenti e/o delle metodologie messi, appunto, alla prova.

Per garantire il controllo necessario a ogni ortodossa sperimentazione scientifica - compresa quella di tipo educativo che è caratterizzata da un elevato livello di complessità delle variabili coinvolte - già nel 1980 Domenici e Vertecchi proposero non più grandi progetti di cambiamento, ma iniziative circoscritte "condotte su punti specifici ed esattamente delimitati, esauribili in un lasso di tempo ragionevolmente breve". Queste iniziative furono definite "microsperimentazioni" laddove "micro" non veniva inteso come fattore riduttivo dell'importanza dell'osservazione svolta e delle sue ricadute conoscitive, bensì riferito alla dimensione contenuta dell'indagine. La ricerca viene così rivolta a singoli aspetti del fare scuola e a problemi dell'educazione ben definiti.

Occorre partire a tal fine dalla definizione operativa del problema riscontrato e procedere all'osservazione sistematica delle sue manifestazioni e al controllo degli effetti prodotti da eventuali interventi formativi o didattici proposti.

Caso 7.1
IL CASO DELLA PRIMA MICROSPERIMENTAZIONE SUL TEMPO A SCUOLA

Domenici (1980) ha messo a punto una procedura di auto-analisi dell'utilizzazione del tempo scolastico da parte di un gruppo di insegnanti (si tratta di una ricerca descrittiva in quanto il comportamento dei docenti partecipanti all'indagine non è stato modificato nell'ambito della ricerca).

Ogni microsperimentazione si sviluppava attraverso una ricostruzione della letteratura sull'argomento da studiare, seguita dall'individuazione di un problema da chiarire, la definizione di un'ipotesi di intervento osservativo o sperimentale, la progettazione di una procedura e la costruzione della strumentazione necessaria. Seguiva la conduzione dell'esperienza e la rilevazione dei dati, la loro discussione e valutazione.

Il tempo medio richiesto dall'esecuzione di tutte queste fasi era di circa due mesi durante i quali si svolgevano riunioni settimanali del gruppo di lavoro. Furono coinvolti piccoli gruppi di 3-6 insegnanti.
Una prima indagine fu condotta da Domenici e Vertecchi (1980) attraverso l'osservazione sistematica dei comportamenti in classe degli allievi durante diversi tipi di attività didattica. Attraverso una griglia di osservazione il docente poteva registrare ogni cinque minuti il comportamento collettivo manifestato. Le osservazioni sono state condotte dagli stessi docenti per un tempo complessivo di 100 ore. Le osservazioni hanno riguardato lo svolgimento di attività didattiche in relazione a diverse materie. I dati sono stati raggruppati e analizzati al fine di cogliere le principali tendenze. Sono stati quindi suddivisi in due gruppi di tendenza del comportamento collettivo degli allievi: una tendenza positiva caratterizzata dalla registrazione di comportamenti di attenzione e interesse; e una negativa caratterizzata da fenomeni di brusio e disturbo.
I dati rapportati a cento e relativi a varie attività hanno messo in evidenza come il massimo coinvolgimento degli allievi (espresso da attenzione e interesse) si registrava in quelle fasi della didattica in cui le attività svolte erano fortemente strutturate, sia che si riferissero a contenuti curriculari para- o extra-curriculari (lettura di quotidiani, visione di filmati ecc.). Bassa partecipazione e scarso interesse si sono registrati invece nel caso di attività poco strutturate o poco incidenti sul totale del tempo scolastico (come nel caso delle discussioni e dello studio in classe).
Queste rilevazioni sul grado di coinvolgimento in funzione del tempo e del significato attribuito alle varie attività didattiche, hanno rappresentato innovative microsperimentazioni volte a raggiungere una migliore organizzazione didattica.

Riportiamo di seguito le fasi di attuazione di una microsperimentazione:

1. *rilevazione di una situazione problematica e definizione dello status questionis:* contiene i riferimenti di massima agli studi già effettuati sulla questione problematica specifica e altre affini;
2. *problema*: definisce la situazione operativa, la variabile sulla quale si intende intervenire, l'ipotesi che si vuole verificare;
3. *procedura*: descrive le fasi di attuazione della microsperimentazione e gli strumenti da utilizzare;
4. *risultati*: sono i dati rilevati tenendo conto dell'ipotesi;
5. *discussione*: esame dei dati tenendo conto dell'ipotesi e della letteratura;
6. *bibliografia*: sono elencati i titoli fondamentali di cui si è tenuto conto nel lavoro.

Microsperimentazione

Si tratta di una sperimentazione svolta su piccoli numeri di partecipanti e con un numero ridotto di variabili le quali vengono così sottoposte meglio a stretto controllo nel confronto con gruppi di controllo.
Si definisce gruppo sperimentale (GS) il gruppo dei partecipanti sottoposti a intervento e gruppo di controllo (GC) il gruppo ove non si somministra alcun trattamento. Come è noto, sarà dal confronto dei risultati registrati dai due gruppi che si potrà evincere se vi è stato o meno un effetto positivo dell'intervento educativo proposto e quindi se l'ipotesi formulata può essere ritenuta confermata o confutata.

Caso 7.2
MICROSPERIMENTAZIONE E RICERCA-FORMAZIONE

Il caso del Progetto di Ricerca di Interesse Nazionale - PRIN 2011-2016 per la formazione dei docenti

Nell'ambito dell'ampio progetto PRIN (Progetto di Ricerca di Interesse Nazionale del 2011-2016 - Responsabile scientifico: Gaetano Domenici) dal titolo "Successo formativo, inclusione e coesione sociale: strategie innovative, ICT e modelli valutativi", *è stata condotta* anche una indagine empirica per la formazione dei docenti a una innovativa strategia didattica modulare flessibile, rilevando poi a fine anno scolastico gli effetti sui livelli di apprendimento degli allievi, e confrontando tali risultati con quelli registrati da un gruppo parallelo di controllo costituito da docenti non ancora formati all'applicazione di tale didattica innovativa.
Tale microsperimentazione è stata compiuta con il coinvolgimento di docenti della scuola primaria e secondaria opportunamente "formati in servizio" (per lo sviluppo di specifiche competenze professionali), attraverso l'impiego delle ICT su piattaforma e-learning.
La strategia didattica integrata ha tenuto conto dei seguenti fattori: la flessibilizzazione della proposta formativa, come adattamento didattico-curricolare non solo alle diversità interindividuali, ma anche intraindividuali con attenzione a modelli didattico-curricolari orientati alle competenze; l'impiego di strumenti standardizzati di autovalutazione delle strategie di apprendimento e delle competenze strategiche; l'inclusione degli allievi e degli studenti con disabilità o con Bisogni Educativi Speciali attraverso l'applicazione del modello ICF-CY; il trattamento delle difficoltà di apprendimento di origine psicogena; l'allestimento di contesti di inclusione di allievi con differente background culturale e/o figli di immigrati; la promozione dei processi di costruzione dell'identità di genere.
Il progetto ha ipotizzato che gli esiti formativi quali-quantitativi in uscita dai Gruppi-classe Sperimentali (nei quali sia stata impiegata la strategia didattica integrata dai docenti così formati), relativi ai differenti livelli e gradi di istruzione presi in esame, rilevati attraverso strumenti standardizzati, risultassero significativamente più alti rispetto ai corrispondenti esiti dei Gruppi-classe di Controllo.
I risultati pubblicati in un ampio volume curato da Domenici (2017) hanno confermato le ipotesi allo studio.

Approfondimento 7.2
MICROSPERIMENTAZIONI E MACROSPERIMENTAZIONI: PRINCIPALI DIFFERENZE E ASPETTI COMPLEMENTARI

Si precisa che le microsperimentazioni, caratterizzate dallo studio di un numero ridotto di variabili alla volta e da un periodo di tempo di svolgimento dell'indagine generalmente contenuto, permettono un maggior livello di controllo sia delle procedure sia degli effetti allo studio con una riduzione del peso di variabili interferenti eventuali. Esse necessitano di particolare attenzione in sede di generalizzabilità dei risultati e a tale fine vi è la raccomandazione di svolgere ripetute indagini con relativi diversi campioni e rilevare la stabilità dei risultati, confermati naturalmente da un solido impianto statistico.
Le macrosperimentazioni, invece, beneficiano in genere di campioni molto ampi, ma rilevano contemporaneamente molteplici variabili che non è facile tenere sotto stretto controllo metodologico. Vi sono quindi minori limiti nella generalizzabilità dei risultati, ma maggiori rischi di non riuscire a evitare l'effetto interferente di dimensioni cosiddette parassitarie.
Si mette a questo punto in evidenza una certa complementarità di queste due modalità di sperimentazione, e laddove possibile in termini di costi e livelli di impegno, si può raccomandare lo studio di un fenomeno attraverso entrambe queste due procedure separate.

7.3 Microsperimentazione per lo studio dell'atteggiamento del docente e degli effetti sull'apprendimento degli allievi. Procedure e risultati di una ricerca

Il progetto PRIN (Domenici, 2017) ha comportato un confronto tra i risultati degli apprendimenti degli allievi provenienti da classi dove è stata applicata una didattica innovativa con i risultati conseguiti in classi di controllo parallele che hanno seguito una didattica tradizionale, nell'ipotesi di ottenere, nel primo caso, un incremento significativo dei livelli di apprendimento.

Precisiamo che il primo momento della sperimentazione ha riguardato la formazione dei docenti a tale didattica innovativa (cioè alla strategia didattica flessibile integrata), formazione che è stata centrata su quattro aree scientifiche principali: "Didattica e valutazione", "Didattiche speciali", "Intercultura e differenze di genere", "Psicologia dell'apprendimento".

Solo la reale attuazione nelle classi della didattica innovativa a cui i docenti erano stati formati permetteva di testare l'ipotesi allo studio e perciò, in prima istanza, è stata verificata la *portata del grado di attuazione della didattica innovativa da parte dei docenti* coinvolti nel periodo di formazione in presenza e on-line, ossia del gruppo sperimentale.

È stato quindi chiesto direttamente ai docenti nell'ambito di un "Questionario sugli atteggiamenti professionali dei docenti: didattica e valutazione" somministrato all'inizio (pre-sperimentazione) e al termine di tutte le attività proposte (post-sperimentazione) e da loro attuate nelle rispettive classi, di fornire una misura indicativa del *grado di applicazione del modello didattico integrato*, attraverso la domanda "Indichi, infine, con buona approssimazione, in quale misura ha applicato nelle sue classi il Modello Didattico Integrato proposto nella ricerca a cui ha partecipato" (Tabella 7.1).

Tabella 7.1 Gradi di applicazione (e relative frequenze percentuali) della Strategia Didattica Integrata indicato dai docenti del gruppo sperimentale (Domenici, 2017)

Aree formative	Dal 50% al 100%
Ambito relativo all'organizzazione flessibile della didattica, alla valutazionee all'orientamento formativo	54,3%
Ambito relativo alle difficoltà di apprendimento degli allievi e alla didattica speciale	68,8%
Ambito relativo alle differenze interculturali e di genere degli allievi	33,3%
Ambito relativo ai processi psicologici dell'apprendimento	75,0%

È stato così possibile rilevare il grado di attuazione percepito di una programmazione didattica integrata grazie a strategie formative flessibili basate sulla "modularità": le quali, a nostro avviso, permettendo l'adattamento dell'offerta didattica alle caratteristiche cognitive e affettive degli allievi, facilitano l'individualizzazione della proposta di istruzione, condizione necessaria per innalzare la qualità sia dei processi sia dei risultati formativi.

In sintesi, la percentuale di applicazione delle indicazioni fornite tramite il progetto PRIN per la realizzazione di una strategia didattica flessibile e realmente integrata è risultata soddisfacente.

In particolare, il tasso di applicazione ossia "la misura con cui il docente ha applicato nelle sue classi il Modello Didattico Integrato proposto nella ricerca a cui

ha partecipato" è risultata molto alta, con una frequenza percentuale superiore al 50% e addirittura talvolta pari al 75%, quindi decisamente affidabile.

I dati sopra illustrati mettono infatti in evidenza come i docenti abbiano applicato in modo differenziato, comunque accettabile, per le quattro aree formative, il modello di intervento didattico proposto, dedicando peraltro notevole attenzione all'applicazione dei contenuti psicologici. Tutto ciò ha permesso di confermare l'affidabilità della variabile indipendente, ossia l'efficacia della formazione dei docenti del gruppo sperimentale, i quali sono stati poi capaci di tradurre nella pratica didattica reale la strategia integrata proposta.

7.3.1 Auto-efficacia e atteggiamento motivante del docente

Una gamma interessante di risultati si è registrata grazie alla somministrazione nell'ambito del suddetto progetto - nel periodo di formazione, oppure prima e dopo l'attività didattica a scuola - di alcune *misure di atteggiamento dei docenti* (Biasi e Patrizi, 2017).

A tale scopo è stato utilizzato, tra gli altri, uno strumento validato denominato "Scala di AutoEfficacia del Docente" (SAED; adattamento italiano di Biasi, Domenici, Capobianco, Patrizi, 2014), per la rilevazione della percezione di auto-efficacia del docente. Tale strumento è stato somministrato sia ai docenti appartenenti al gruppo sperimentale (formati alla didattica innovativa, che hanno realmente applicato nelle rispettive classi) sia ai docenti appartenenti al gruppo di controllo, per le necessarie comparazioni.

Dall'analisi dei dati ottenuti si sono evidenziati i risultati che di seguito riassumiamo, precisando anche il costrutto su cui si basa lo strumento utilizzato al fine di meglio interpretarne gli esiti.

Attraverso la *Scala di AutoEfficacia del Docente (SAED)* - tradotta da Biasi e Domenici (2013) e adattata alla popolazione italiana da Biasi, Domenici, Capobianco, Patrizi (2014) - si misura la *percezione di auto-efficacia del Docente* che corrisponde al giudizio sulle *capacità personali di riuscire a realizzare i risultati desiderati in termini di impegno e di apprendimento da parte degli studenti.* Questo costrutto ha mostrato, anche a livello internazionale, di essere fortemente legato a molti risultati educativi significativi tra cui: la persistenza, l'entusiasmo, l'impegno e il comportamento didattico del docente; e, come effetto secondario decisamente importante, ha mostrato un valore predittivo sui risultati degli studenti negli apprendimenti, sulla loro motivazione e sulle loro convinzioni in tema di auto-efficacia.

La percezione di *auto-efficacia del docente riguarda, in sintesi, la sua capacità di ottenere l'impegno degli studenti,* la possibilità di *scegliere strategie di insegnamento adeguate* e di *gestire la classe in modo funzionale,* e ciò - come già detto - ha un valore predittivo per il grado di successo scolastico (Sewell e St. George, 2000; et al., 2003).

È in particolare la sotto-scala denominata "Efficacy for student engagement" ("Efficacia nel motivare lo studente") ad avere la maggiore influenza sul livello di apprendimento dell'alunno (Gowrie e Ramdass, 2013). In sintesi, se il docente ritiene di essere capace di motivare lo studente, se cioè ha fiducia di riuscire a motivarlo, percepisce di esserne capace - questa auto-percezione si basa naturalmente anche sui feedback derivanti da precedenti esperienze - ottiene in genere il suo coinvolgimento e il suo impegno nello studio che naturalmente incideranno sui risultati.

Secondo Albert Bandura (1977, 1997), l'auto-efficacia corrisponde alla «convinzione delle proprie capacità di organizzare e realizzare il corso di azioni necessarie per gestire adeguatamente le situazioni che si incontreranno in un particolare contesto, in modo da raggiungere gli obiettivi prefissati» e, come anche altri studiosi hanno dimostrato, può avere un impatto sul comportamento e sulla motivazione.

Nella letteratura internazionale l'auto-efficacia rappresenta quindi un argomento rilevante sia per gli psicologi sia per gli educatori.

La *Scala di AutoEfficacia del Docente* (SAED, versione italiana della Teacher Self-Efficacy Scale elaborata da Tschannen-Moran e Woolfolk Hoy nel 2001), di cui qui si riferisce, si articola in tre fattori: "Autoefficacia nel coinvolgimento degli studenti" (8 item), "Autoefficacia delle strategie di insegnamento" (8 item), e "Autoefficacia nelle tecniche di gestione della classe" (8 item). L'adattamento italiano ha confermato l'importanza dei tre fattori e le relative analisi statistiche ne hanno mostrato la validità.

Nell'ambito della microsperimentazione di cui si riferisce, 18 docenti appartenenti al gruppo sperimentale e 14 appartenenti al gruppo di controllo hanno compilato la SAED prima e dopo le attività previste dal progetto PRIN. I dati del campione appaiato (quindi in totale 32 docenti; si veda Tabella 7.2) sono stati elaborati attraverso l'applicazione del test Anova per misure ripetute per valutare le differenze tra gruppo sperimentale e gruppo di controllo, in entrata e in uscita. È stato valutato anche l'effetto di interazione "Tempo" X "Gruppo" che è risultato statisticamente significativo ($F_{1,10} = 4.13$; $p < 0.05$).

In sintesi, è stato mostrato come sia il fattore "Autoefficacia nel coinvolgimento degli studenti" ("Efficacy in Student Engagement") (Figura 6.1) a ricoprire un ruolo significativo in quanto viene rinforzato grazie al percorso di formazione innovativa dei docenti realizzato.

Tabella 7.2 Sono riportate le medie e le deviazioni standard del gruppo sperimentale e di controllo sia in entrata (T1) sia in uscita (T2) per il fattore E: "Autoefficacia nel coinvolgimento degli studenti" (SAED)

	Gruppi	Media	Deviazione standard	N
Saed_E_T1	Sperimentale	58,28	4,75	18
	Controllo	58,43	7,14	14
Saed_E_T2	Sperimentale	59,78	5,62	18
	Controllo	55,64	6,32	14

Fonte: Biasi e Patrizi, 2017; Biasi, 2017.

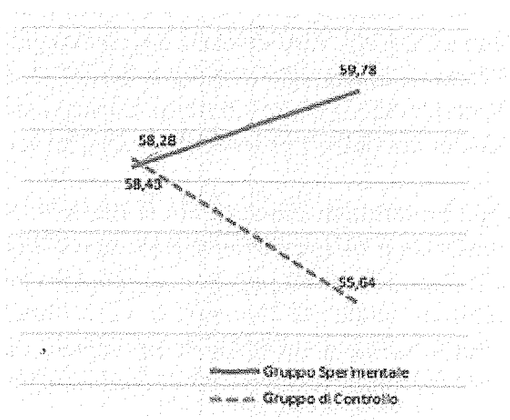


Figura 7.1. Sono riportate le medie del gruppo sperimentale e di controllo sia in entrata sia in uscita per il fattore E: "Autoefficacia nel coinvolgimento degli studenti" (SAED). Si osserva un effetto di interazione statisticamente significativo tra Gruppo Sperimentale/Controllo e Entrata T1/uscita T2 ($F_{1,10} = 4.13$; $p < 0,05$) (Biasi e Patrizi, 2017; Biasi, 2017).

I dati hanno quindi messo in evidenza per i docenti del gruppo sperimentale – grazie alla formazione ricevuta e all'applicazione effettiva della nuova didattica integrata – lo sviluppo della capacità di motivare lo studente, testimoniato dalla crescita dell'auto-percezione di efficacia nell'ottenere il coinvolgimento degli allievi nelle attività didattiche.

Tale risultato acquista validità in particolare nel confronto con i punteggi registrati dai docenti appartenenti al gruppo di controllo: i quali vedono invece decrescere in modo speculare la percezione di "Autoefficacia nel coinvolgimento degli studenti", cioè nel tempo indicato come T2 (post-attività didattica tradizionale) rispetto al tempo T1 (pre-attività didattica tradizionale).

Possiamo interpretare questa decrescita registrata nell'ambito del gruppo di controllo come una percezione di perdita di controllo del gruppo-classe o come un non avere più il polso della situazione o la piena capacità di promuovere interesse e attenzione per ottenere i necessari progressi nell'apprendimento.

La capacità di coinvolgere lo studente vede al contrario una fioritura significativa grazie all'applicazione da parte dei docenti della strategia didattica flessibile integrata, cui essi sono stati sensibilizzati nel corso del periodo di formazione e che hanno poi adeguatamente tradotto e adattato nell'ambito delle attività delle classi scolastiche in cui sono stati impegnati.

Il contributo della formazione psicopedagogica (Biasi, 2017) offerta nell'ambito del progetto PRIN ha particolarmente sollecitato i docenti ad agganciare *le operazioni cognitive agli interessi specifici del singolo alunno e all'assetto motivazionale medio del gruppo-classe in relazione alla fascia di età considerata.* A tale fine i docenti sono stati formati a:

- rilevare con interviste aperte e colloqui gli interessi tipici dello specifico allievo e del gruppo-classe di riferimento;
- considerare analiticamente gli interessi e l'assetto motivazionale medio relativamente alla fascia di età in cui si opera, per adattare contenuti, procedure, esercizi ecc.;
- adattare la programmazione didattica trimestrale al mondo degli interessi e all'assetto motivazionale tipico della fascia di età degli allievi della propria classe;
- consolidare la costruzione dei concetti attraverso il *rafforzamento delle abilità di astrazione e generalizzazione* con l'impiego di appositi esercizi di problem solving convergente e divergente (da adattare naturalmente alla disciplina insegnata e al gruppo-classe di riferimento).

L'intervento didattico, sostenuto anche dalla considerazione degli aspetti psicopedagogici illustrati, è risultato in definitiva funzionale all'innalzamento dei livelli

quali-quantitativi degli apprendimenti dei discenti come sintetizzato nel paragrafo che segue.

7.3.2 La ricaduta positiva sugli apprendimenti degli allievi

A questo punto è stato possibile esaminare la ricaduta, statisticamente significativa, di tale didattica innovativa sulle variabili dipendenti allo studio, cioè sui vari livelli di apprendimento conseguiti dagli allievi delle classi sperimentali rispetto alle classi di controllo dove è stata mantenuta una didattica tradizionale (Domenici, 2017; Biasi, 2017).

Si sono evidenziati complessivamente i seguenti risultati:

1. nella *scuola primaria*, per la quale sono state coinvolte in particolare le classi V, si sono ottenuti miglioramenti significativi dopo il periodo di didattica innovativa proposta rispetto ai gruppi di controllo appaiati, impiegati per la comparazione (i quali hanno invece seguito la didattica tradizionale per un equivalente periodo di tempo), in particolare nelle varie prove di apprendimento nei confronti pre-post del gruppo sperimentale;
2. per la *scuola secondaria di primo grado*, di cui sono state coinvolte in particolare le classi III, sono stati registrati risultati significativi sia dopo il periodo di didattica innovativa proposta sia a seguito della didattica tradizionale per un equivalente periodo di tempo, nonché miglioramenti sistematici nelle varie prove di apprendimento nei confronti pre-post;
3. per la *scuola secondaria di secondo grado*, di cui sono state coinvolte in particolare le classi III, si sono ottenuti miglioramenti significativi dopo il periodo di didattica innovativa proposta rispetto ai gruppi di controllo appaiati (i quali hanno seguito la didattica tradizionale per un equivalente periodo di tempo), in particolare nelle prove di lettura e nelle prove semistrutturate di apprendimento logico-matematico e nelle prove logico-scientifiche.

In sintesi, l'applicazione del modello didattico flessibile integrato a cura dei docenti delle classi facenti parte dei Gruppi Sperimentali, ha prodotto una facilitazione nei processi di apprendimento rinvenibile nell'incremento dei livelli di apprendimento registrati nelle varie prove somministrate, e nel miglioramento nella qualità degli apprendimenti per fascia scolastica, intesa anche come fascia di età dei discenti coinvolti: nella scuola primaria si è registrata un'evidente maturazione in particolare delle abilità logico-linguistiche, nella scuola secondaria di secondo livello uno sviluppo peculiare delle abilità logico-matematiche e scientifiche.

Nella fascia scolastica secondaria di primo grado sono stati registrati effetti singolari: si sono cioè ottenuti miglioramenti statisticamente significativi sia "pre/post trattamento" sia "pre/post controllo". Ciò conferma, da un lato, l'ancora forte plasticità dei processi di apprendimento in questa fascia di età, per cui sia la didattica innovativa sia quella tradizionale sortiscono effetti incoraggianti nonostante la problematicità riconosciuta in letteratura dell'età puberale e pre-adolescenziale.

Si può oggi ipotizzare che, con una ancor più sistematica formazione dei docenti all'applicazione della suddetta strategia didattica integrata, che ponga attenzione alle esigenze cognitive e affettive dei ragazzi, e che preveda magari tempi più lunghi di intervento, si potrebbe innalzare il grado di individualizzazione dei processi di insegnamento incrementando ulteriormente i livelli di apprendimento.

Particolare rilievo hanno acquistato poi i risultati registrati nella scuola secondaria di secondo livello in quanto una didattica che favorisce lo sviluppo delle abilità logico-matematiche e scientifiche può rappresentare una vera risposta alle difficoltà attualmente incontrate nel potenziamento del pensiero scientifico, questione importante sia nella transizione al mondo del lavoro sia per il progresso economico del Paese già indicati dalla LIE come propri obiettivi per il 2030.

Si precisa per correttezza metodologica che per ogni prova di apprendimento sono stati calcolati i confronti tra Gruppo Sperimentale in entrata *vs* Gruppo di Controllo in entrata: tali confronti non hanno evidenziato differenze statisticamente significative confermando la bontà dell'omogeneità dei livelli di partenza degli apprendimenti dei discenti inseriti nei diversi gruppi allo studio, e quindi garantendo la comparabilità dei campioni utilizzati.

Concludendo, sulla base dei riscontri empirici rinvenuti, si può affermare che la formazione dei docenti a una didattica flessibile integrata, debitamente arricchita anche dalle *conoscenze e competenze psicologiche che facilitano i processi di apprendimento*, ha loro permesso una maggiore percezione di autoefficacia in particolare per quanto riguarda la *capacità di motivare lo studente*, coinvolgendolo nelle varie attività didattiche.

7.4 Auto-valutazione nella didattica laboratoriale universitaria on-line attraverso prove semistrutturate. Sintesi di un'apposita microsperimentazione.

7.4.1 Obiettivi dell'indagine

Nell'ambito dell'obiettivo generale volto a identificare e potenziare alcuni fattori facilitanti, i processi di apprendimento in modalità *e-learning*, attraverso l'*individualizzazione* dell'apprendimento e la *promozione di un atteggiamento attivo verso i contenuti da apprendere*, abbiamo inteso testare l'ipotesi secondo la quale le prove semistrutturate di auto-valutazione nell'ambito dell'attività laboratoriale migliorano significativamente i livelli di apprendimento conseguiti.

7.4.2 Struttura e caratteristiche delle attività del laboratorio on-line

Nell'ambito del Corso di Laurea in Scienze dell'Educazione dell'Università "Roma Tre", *è stato attuato un laboratorio didattico on-line* basato sull'applicazione di prove semistrutturate di auto-valutazione.

La piattaforma *formonline* (http://formonline.uniroma3.it) con modalità *Open Source* per *dell'apprendimento on-line* ha permesso l'attuazione delle prove semistrutturate grazie alla risorsa offerta secondo la modalità "compito" da compilare in riferimento a parti specifiche del programma di studio e da inviare poi ai tutor in attesa di un feedback.

I "compiti" prevedevano domande (con stimolo chiuso e risposta aperta) su specifici capitoli e paragrafi da preparare e prevedevano i vincoli spaziali nel numero di 10 o 12 o 15 righe di testo da scrivere (centrati quindi in particolare su capacità cognitive di tipo convergente). Si proponeva inoltre l'applicazione dei contenuti allo studio in contesti nuovi e diversificati, per i quali necessitano di modalità di revisione e di adattamento a cura dello studente (in particolare per lo sviluppo delle capacità cognitive divergenti). Complessivamente sono stati proposti 12 compiti articolati nel corso del secondo semestre di lezioni, costituenti appunto dell'attività laboratoriale in questione.

Ogni compito veniva pubblicato *on-line* e veniva "ritirato", cioè doveva essere consegnato in piattaforma, nel giro di 7, 10 o al massimo 15 giorni (a seconda del numero di domande da completare).

I tutor inviavano una risposta di avvenuta ricezione e, quando necessario, orientavano lo studente verso un adeguato approfondimento o chiarimento di alcune risposte, senza dare la soluzione ma orientandolo sulle parti del programma da approfondire e stimolando così un atteggiamento attivo e critico di ricerca ed elaborazione dei contenuti disciplinari.

Le attività del Laboratorio on-line non erano obbligatorie ma fortemente consigliate.

Riportiamo nella Tabella 7.3 la ripartizione dei gruppi seguiti dai tre tutor.

Tabella 7.3 - Ripartizione di gruppi e tutor con distribuzione delle frequenze degli studenti che hanno o meno completato le prove del Laboratorio on-line

	Tutor 1	Tutor 2	Tutor 3	Frequenze totali	Frequenze percentuali
Studenti che hanno portato a termine le attività on-line	136	84	51	271	76,8 %
Studenti che "non" hanno concluso le attività on-line	31	16	35	82	23,2 %
Totale studenti	167	100	86	353	

Precisiamo ulteriormente la natura dell'attività laboratoriale svolta partendo dalla definizione delle *prove semistrutturate* elaborata da Domenici (2005; 2009):

> "Sono prove a stimolo chiuso e risposta aperta, ma aperta in modo che, nella quasi generalità dei casi, l'allievo sia, per così dire, costretto, dalla specificità della sollecitazione, a porre in netta evidenza il possesso o non possesso delle conoscenze sottoposte a controllo e a consentire a chi le corregge di attribuire in forma tendenzialmente univoca e oggettiva i punteggi (così assicurando una buona validità ed un'alta attendibilità della rilevazione" (Domenici, 2005, p. 24).

Tra le caratteristiche specifiche di queste prove vi è la capacità:

> "di far cogliere taluni dei saperi posseduti dagli studenti e soprattutto la loro abilità di selezionarli per l'occorrenza; di far conoscere le procedure valutative impiegate a tal fine e le modalità di revisione e adattamento al contesto delle stesse..." (Domenici, 2005, p. 14).

Tali prove applicate nella didattica laboratoriale on-line orientano gli studenti non solo per l'auto-valutazione degli apprendimenti ma, in particolare, riteniamo, nella ricerca e nello sviluppo di efficaci metodologie di studio, proprio in quanto la loro struttura accentua il valore auto- ed etero-formativo dei processi di verifica e di valutazione.

Nella Tabella 7.4 si riporta a titolo esemplificativo un modello di prova utilizzata (Domenici e Biasi, 2012).

Tabella 7.4 Esempio di prova semistrutturata

Argomento: Metodo sperimentale. Definizione e applicazioni"

Istruzione: Voi preparerete la vostre risposte consultando, dopo aver studiato, i testi del programma, e le confronterete con le risposte dei docenti, che verranno pubblicate in piattaforma dieci giorni dopo. Buon lavoro e a presto,

Si indichino: a) la differenza tra variabile dipendente e indipendente; b) due situazioni sperimentali ipotetiche nelle quali la variabile dipendente dell'una diventi variabile indipendente dell'altra.

a) _____

b) _____

Descrivere brevemente: a) in che cosa consiste, sul piano formale, "il metodo sperimentale"; b) il diverso "ruolo" del ricercatore nelle indagini scientifiche di tipo "sperimentale" e in quelle di tipo "non sperimentale".

a) _____

b) _____

Indicare quali caratteristiche deve avere un gruppo di controllo rispetto al/ai gruppi sperimentale, e a che cosa serve.

Caratteristiche:

_____

A cosa serve:

_____

Immaginate, ora, di allestire un gruppo sperimentale per l'attivazione di conflitto attraverso la somministrazione del "Disegno di stress" con prove pre- e post-trattamento per rilevare lo stato emotivo dei partecipanti. Descrivete, sinteticamente, almeno due degli esiti che vi attendete. Costruite, poi, una griglia per la registrazione degli esiti emotivi della prova e allegate il file, motivando le ragioni della scelta degli elementi che la caratterizzano.

_____

Immaginate, inoltre, di allestire un relativo gruppo di controllo e di applicare la medesima griglia di rilevazione dello stato emotivo. Che esiti attendete? Elencateli in ordine decrescente di importanza.

_____

Come si può rilevare, il tipo di prova proposto prevede sia la raccolta di informazioni attraverso la consultazione dei testi di studio (operazione cognitiva di tipo prevalentemente convergente: basata sulla capacità di dare risposte esatte a problemi che ammettono un'unica risposta accettabile), sia il confronto e la ristrutturazione di tali conoscenze applicandole a contesti simulati diversificati (operazione cognitiva di tipo prevalentemente divergente consistente nella capacità di fornire risposte diverse a problemi che ammettono più risposte accettabili). Ciò al fine di favorire sia un *transfert* di apprendimento sia una ristrutturazione creativa del problema considerato.

7.4.3 Ricadute sui livelli di apprendimento degli allievi

Le ricadute sui livelli di apprendimento sono risultate buone come indicato nel prospetto riportato nella Tabella 7.5. Il gruppo degli studenti che hanno completato le attività (n = 252) ha conseguito un punteggio medio alla prova oggettiva d'esame pari a 27.7; gli studenti che non hanno portato a termine le attività o che non le hanno mai iniziate, nel senso che non si sono iscritti al Laboratorio (n = 168) hanno conseguito un punteggio medio alla prova oggettiva d'esame pari a 25.7.

Queste differenze risultano statisticamente significative e vengono quindi a sottolineare l'opportunità di questo specifico intervento sui contenuti e sui metodi di studio, grazie all'applicazione di prove semistrutturate di auto-valutazione.

Tabella 7.5 - Risultati medi conseguiti alla prova oggettiva d'esame da studenti che hanno concluso le attività laboratoriali on-line *versus* coloro che non le hanno concluse (Biasi e Domenici, 2014)

	N	Punteggio medio alla prova oggettiva d'esame	Deviazione standard	Significatività statistica
Studenti che hanno portato a termine le attività laboratoriali on-line	252	27.69	3.20	$F_{(1, 449)} = 38.05$
Studenti che non hanno svolto le attività laboratoriali on-line (non concluse o non iniziate)	198	25.65	3.80	$p < 0.001$
Totale studenti	450			

Sulla base dei dati emersi dalla microsperimentazione didattica in questione, possiamo ritenere che le strategie auto-valutative degli apprendimenti condotte con prove semistrutturate facilitino il processo di apprendimento.

Ciò a conferma anche di quanto riportato in letteratura in merito ai vantaggi e punti di forza della didattica on-line basati primariamente sullo sviluppo di un atteggiamento attivo verso i contenuti da apprendere.

Tale atteggiamento attivo, che corrisponde in genere a una specifica attivazione motivazionale, sollecita, a nostro avviso, lo sviluppo del pensiero critico per merito delle procedure di auto-valutazione, e - in particolare - favorisce specifiche abilità cognitive: le prove semistrutturate hanno effetto infatti non solo sull'*acquisizione delle conoscenze* ma - e questo è un aspetto innovativo dell'applicazione di tali strumenti nella didattica laboratoriale - sullo sviluppo e il consolidamento di un efficace *metodo di studio*.

7.5 L'osservazione sistematica come strumento di ricerca e di qualificazione della didattica

Da tutto ciò che è stato fin qui detto sulla ricerca e sulle attività formative, dovrebbe ormai emergere una interpretazione del lavoro scolastico come vero e proprio lavoro di ricerca educativa. Oltre a una preparazione specialistica in specifici ambiti disciplinari e metodologici, l'insegnante dovrebbe possedere a tale scopo un ampio ventaglio di conoscenze relativo alle scienze dell'educazione che gli consen-

ta di affrontare con facilità e con rigore scientifico i problemi che via via deve affrontare nel complesso e articolato rapporto di insegnamento-apprendimento.

Va però ribadito che proprio in campo educativo la ricerca nella maggioranza dei casi non può essere ricerca sperimentale. E non già o non solo per questioni formali, inerenti cioè la sua pur elevata complessità organizzativa e di conduzione, la messa a punto e l'impiego di procedure e di strumenti talvolta assai sofisticati e altro ancora, quanto, se non soprattutto, per ragioni deontologiche, di etica professionale. In relazione a queste ragioni, ogni accettabile sperimentazione dovrebbe garantire preventivamente, come più volte si è voluto ribadire, esiti che nel peggiore dei casi non risultino meno buoni di quelli che si registrerebbero in mancanza dello stesso intervento sperimentale. E ciò, come si può ben intuire, non è questione facilmente risolvibile senza un robusto apporto teorico e tecnico-procedurale, senza l'impiego di una enorme quantità di dati conoscitivi e di un livello organizzativo tali da consentire se non veri e propri processi di simulazione almeno una accettabile previsione sia dei processi sia dei risultati dello specifico intervento sperimentale progettato.

Purtroppo però la maggiore dignità attribuita a livello di senso comune alla ricerca sperimentale rispetto ad altri tipi di indagine (si pensi alle "sperimentazioni" ovunque e comunque ostinatamente così definite nel mondo scolastico persino per designare situazioni che nulla hanno a che fare non solo con tali procedure, ma nemmeno e persino con semplici innovazioni controllate) ha fatto assumere ai docenti e persino ad alcuni ricercatori comportamenti dicotomici: per un verso di vere e proprie fughe in avanti nel mondo delle cosiddette sperimentazioni e quindi di impegno spesso assai oneroso in senso lato in situazioni impropriamente definite sperimentali; per altro verso di accettazione, talvolta supina, dello *status quo* se non di ritorno al passato, con veri e propri atteggiamenti di regressione, nella rassicurante roccaforte del fare scuola tradizionale, quasi che l'innovazione o possa essere solo di tipo sperimentale o non la si possa considerare degna d'esser introdotta. Viceversa, dato che non è sempre possibile intervenire nella alterazione intenzionale di ben definite *variabili indipendenti* per verificarne la qualità e il peso specifico attraverso l'analisi di quelle *dipendenti*, non essendo cioè facile prefigurare come e quanto mutano certi fattori in conseguenza del mutamento volontario di alcuni altri che in modo significativo contribuiscono a determinare una situazione, altri tipi di ricerche non sperimentali meglio si presterebbero a risolvere in campo educativo situazioni problematiche per le quali si vuole sottoporre a *controllo obiettivo* la soluzione ipotizzata.

Tra queste assume particolare rilievo la *ricerca osservativa* e in particolare, tra le diverse forme che essa può avere, l'*osservazione sistematica*. La sua relativa semplicità di conduzione le assegna un posto di grande significato nei processi di acquisizione e accumulazione di nuove conoscenze, attraverso la ricerca scientifica in senso stretto, ma anche attraverso lo svolgimento dell'attività professionale dei docenti.

Nel processo stesso di indagine scientifica, mosso da una reale situazione problematica, l'osservazione sistematica costituisce un momento assai delicato e importante sia nella raccolta preliminare di quei dati conoscitivi dalla cui analisi e interpretazione dipenderà la qualità e messa a punto dell'ipotesi, sia, dopo, durante lo svolgimento dell'indagine, come contributo conoscitivo utile per la verificazione e/o la falsificazione della ipotesi formulata e la sua stessa modificazione.

Si potrebbe addirittura affermare, onestamente, che la via aurea del processo formativo quotidiano, e non solo di esso, sia senz'altro costituita dalla messa in atto dell'osservazione sistematica e, ovviamente, dalle scelte operative che da essa conseguono, almeno quando il comportamento professionale voglia essere fondato su basi di dati conoscitivi affidabili.

Certo la sperimentazione consente interventi più celeri, anzi immediati, del mutamento che si vuole apportare al contesto stesso in cui si opera: come si è detto, con essa si altera intenzionalmente il quadro operativo attraverso la modificazione sostanziale e calibrata dei fattori ritenuti cruciali (variabili indipendenti) nella determinazione di certi processi e/o di certi esiti (variabili dipendenti).

Tuttavia la sperimentazione, oltre a dover fare i conti, come d'altro canto ogni altra indagine, con certe variabili assegnate (fattori di vincolo), può e deve essere attuata solo in situazioni molto avanzate sul piano conoscitivo, cioè dopo che sia garantita la disponibilità di tantissimi e significativi dati di conoscenza della situazione problematica, che talvolta possono essere acquisiti solo attraverso una osservazione preliminare e *in itinere* condotta con sistematicità e rigore.

Sarà dunque utile considerare seppur in modo essenziale le *tecniche*, i *criteri* e gli *strumenti* con cui condurre l'osservazione sistematica e analizzare i risultati delle registrazioni, sia in contesti formali di ricerca, sia in contesti educativi e didattici.

7.6 Tecniche e strumenti dell'osservazione sistematica nei processi educativi

È utile, preliminarmente, rilevare gli elementi distintivi delle due principali modalità in cui può articolarsi l'osservazione in generale: quella casuale e quella appunto sistematica.

La prima, pur assai preziosa per la conoscenza, per esempio di forme di comportamenti dei soggetti che sarebbe difficile poter rilevare in situazioni altamente formalizzate come quella scolastica (si pensi alla rilevazione casuale fuori dalla scuola o persino in essa, quando gli allievi credono di non essere osservati o godono del massimo di "libertà", di comportamenti aggressivi altrimenti non resi manifesti), non è intenzionale, né controllata, non consentendo così un'analisi rigorosa delle variabili che hanno rilevanza in specifici contesti o in relazione alle caratteristiche del processo formativo che si vuole attuare. Non si possono infatti attendere certe manifestazioni spontanee, "naturali", e coglierle occasionalmente per poter decidere il tipo di processo di istruzione e formazione da attuare in funzione dei fatti rilevati.

L'osservazione sistematica consente invece di rilevare, in relazione a una teoria esplicativa di un fenomeno, o a una ipotesi o, ancora, agli esiti di precedenti studi osservativi ecc., e quindi in rapporto a ben precise categorie descrittive, elementi significativi del fenomeno che si vuole osservare con un alto grado di validità e attendibilità. L'intenzionalità e il controllo dei dati osservativi derivano dalla specificità e dalla delimitazione del progetto di ricerca in cui essa si inserisce.

A seconda dei criteri di classificazione, l'osservazione sistematica può assumere forme diverse.

Se si impiegano criteri temporali, potremo avere *osservazioni longitudinali* (diacroniche) e *trasversali* (sincroniche); se si fa riferimento al ruolo dell'osservatore si avranno osservazioni *con o senza partecipazione del ricercatore* nella situazione osservata; se si fa riferimento alla variabile spazio, in senso lato, ci saranno osservazioni *estensive* e *intensive* e così via, fino a classificazioni che possono dipendere dall'oggetto stesso della osservazione.

Ovviamente, tra i poli estremi appena indicati vi è tutta un'ampia gamma di forme differenziate con cui è possibile condurre l'osservazione sistematica, la quale "…. è una procedura che va ben oltre il generico impiego della vista per vedere o guardare qualcosa, poiché con essa si osservano e si rilevano dati e informazioni relativi all'ambito della realtà scrutata, che hanno rilevanza significativa rispetto all'ipotesi-guida della stessa osservazione" (Domenici, 2006, p. 75).

Con essa si tende in altri termini a dare organizzazione sistematica, accuratezza ed efficacia alla raccolta di dati che in rapporto a una teoria esplicativa di un fenomeno o a una ipotesi, assumono significati specifici.

I problemi posti dalla difficoltà di rendere univoci i contenuti della osservazione e della comunicazione in senso lato, di definire nettamente i confini di chi osserva e di chi è osservato, ovvero la difficoltà di rendere espliciti i giochi interattivi delle relative equazioni personali (si tratta in effetti di una vera e propria "doppia equazione personale"), di ridurre al massimo la fluttuazione e l'incostanza dei comportamenti umani e le distorsioni delle osservazioni, e altri problemi ancora, devono trovare soluzioni accettabili quando si vuole che i dati rilevati con l'osservazione sistematica siano validi, attendibili e obiettivi.

Per rendere oggettiva l'osservazione, è però quanto meno necessario "definire accortamente che cosa, come e quando osservare, e fare in modo che i dati registrati rispecchino fedelmente la realtà che rappresentano." (Domenici et al., 2017, p. 113).

Qualunque sia la forma scelta, sempre in relazione alle necessità conoscitive individuate e rese esplicite, è indispensabile stabilire il contenuto dell'osservazione e le modalità con cui essa dovrà essere compiuta. Si tratta:

- di *scegliere o di definire la situazione da indagare* in modo che si possano descrivere gli elementi significativi della realtà che ci si propone di osservare: è una vera e propria scelta di campo in funzione dell'ipotesi-guida. Vanno cioè definiti e descritti quali segmenti unitari di comportamento o di manifestazione significativa di un certo fenomeno (durata, intensità ecc.), vanno osservati;
- di *decidere i modi, i criteri e gli strumenti (videocamere, griglie carta e matita ecc.) con cui registrare gli elementi e/o le unità osservate*, affinché la registrazione sia uno specchio fedele di ciò che è avvenuto o avviene nella situazione;
- di *definire*, quindi, le modalità di raccolta, trattazione, analisi e interpretazione dei dati, perciò *i test statistici da impiegarsi* a tal fine.

In estrema sintesi vanno identificati e descritti i tratti essenziali con cui si manifestano certi fenomeni e che bisogna perciò cogliere, e vanno classificate le categorie descrittive del fenomeno e le modalità di registrazione delle stesse al fine di consentire una analisi rigorosa dei dati assunti.

Gli strumenti di rilevazione, di registrazione, e di analisi dei dati osservativi variano a seconda dell'ipotesi-guida, e possono essere talvolta assai sofisticati come si vedrà nel corso del Capitolo 10 (per un rigoroso approfondimento delle questioni si rimanda ai testi indicati nella bibliografia).

Griglie di osservazione o check-list

Si tratta di liste di azioni e comportamenti che vengono registrati tramite un modulo da compilare a cura dell'osservatore (versione carta e matita).

Vengono corredate da scale per rilevare la presenza/assenza di un comportamento, oppure la frequenza o l'intensità dell'azione osservata.

Con riferimento sia all'attività di ricerca che didattica, si può dire in generale che è possibile utilizzare facilmente delle griglie di osservazione (*check-list*) o anche delle semplici tabelle a doppia entrata (Domenici, 1981), in cui in ordinata siano elencati i tratti distintivi oggetto dell'osservazione, e in ascissa i soggetti osservati o il periodo dell'osservazione o entrambi tali variabili (più colonne per ciascun allievo). Tali griglie consentono di registrare agevolmente una gran quantità di dati che con semplici analisi statistiche (per lo più calcoli di frequenza) offrono rilevanti e significative informazioni sul problema indagato.

Queste tabelle possono assumere la forma di vere e proprie *check-list*, ovvero di un elenco di descrizioni di ciò che va osservato in cui nella casella di incrocio «comportamento descritto-soggetto», con dei semplici segni è possibile registrare in situazione la presenza o l'assenza, per ciascun soggetto (e/o per l'intero gruppo di gruppo di soggetti osservati), della qualità descritta. Se la variabile rilevata è l'apprendimento, magari relativa ai singoli studenti che compongono una classe, si possono usare

vere e proprie *liste dei descrittori dell'apprendimento* che hanno tutte le caratteristiche degli strumenti di registrazione dei dati valutativi prima indicati (Tabella 7.6).

Tabella 7.6 Griglia per la rilevazione dei comportamenti

Descrizione dei comportamenti Attività: dibattito sui pericoli della guerra batteriologica	Allievi				
	1	2	3	4	...
Interviene nel dibattito con contributi pertinenti	/				
Interviene nel dibattito con contributo di idee originale				/	
Interviene con ripetizione del "già detto"					
Interviene con battute spiritose		/			
Coglie relazioni pertinenti tra fatti apparentemente irrelati			/		

Oltre alla presenza o meno di una certa caratteristica, si può registrare il grado con cui questa si manifesta impiegando una scala ordinale, possibilmente a un numero pari di livelli, per evitare di subire l'effetto di tendenza centrale nella codifica del dato osservato (Domenici, 1993). Per esempio:

L'alunno interviene nelle situazioni problematiche per prefigurarne la soluzione.
☐ 1. Mai
☐ 2. Qualche volta
☐ 3. Spesso
☐ 4. Sempre

L'alunno ricerca la collaborazione dei suoi compagni.
☐ Costantemente
☐ Frequentemente
☐ Saltuariamente
☐ Mai

Oppure impiegando una scala grafica:

Mai	Qualche volta	Spesso	Sempre

Altri modelli di osservazione sistematica saranno esaminati nel Capitolo 10.

Come si è già accennato, operando sulle registrazioni compiute su tali strumenti con banali calcoli di frequenza e simili, si possono trarre importanti indicazioni sull'andamento sincronico e diacronico di certe caratteristiche del comportamento cognitivo, affettivo-motivazionale, psicomotorio, sociale ecc., proprie di un solo soggetto e dell'intero gruppo preso in esame.

L'osservazione sistematica può però riguardare non solo gli allievi, ma anche gli stessi docenti, il loro "comportamento che insegna", al fine di poterlo modificare positivamente.

Oltre alle tecniche di *Team teaching* e simili, di ripresa audio-video in situazione, con telecamera o altri mezzi, del comportamento insegnante ecc. se ne possono impiegare di più semplici, del tutto simili a quelle già indicate con riferimento agli allievi. Gruppi di docenti possono cioè elaborare, sulla scorta dell'esperienza acquisita, delle liste descrittive di certe unità comportamentali e registrare con un segno, allo scadere di intervalli di tempo regolari, il tipo dominante di attività e/o comportamento messo in atto.

Nonostante gli "errori" di registrazione - d'altro canto in agguato in ogni processo di osservazione sistematica, dovuti principalmente ai pregiudizi di chi osserva, a distorsioni interpretative di certi comportamenti ecc. - gli esiti pongono spesso in evidenza profonde discrasie tra l'immagine professionale che si ha di se stessi e i comportamenti concreti attivati nella classe. Conoscerle e rimuoverle su base razionale è condizione necessaria, ancorché non sufficiente, per l'avvio di un reale processo di ottimizzazione del rapporto formativo.

Approfondimento 7.3 (di V. Biasi)

OSSERVAZIONE OCCASIONALE *VERSUS* OSSERVAZIONE SISTEMATICA

L'osservazione sistematica costituisce un momento delicato e importante nella raccolta dei dati conoscitivi dalla cui analisi e interpretazione dipende la verifica delle ipotesi formulate per la spiegazione di un fenomeno allo studio.

Tale metodologia comporta la ripetizione di osservazioni nel tempo accompagnate preferibilmente dalla loro rilevazione attraverso strumenti specifici (dalle griglie di osservazione alle tecniche di registrazione).

Si diversifica dall'osservazione occasionale di un fenomeno la quale può costituire un momento molto proficuo di avvio di riflessione sullo stesso, originando anche forme di *insight* utili alla problematizzazione successiva. Come dice però la parola, si tratta di una occasione, una rilevazione casuale, fortuita, che deve essere ripetuta e meglio focalizzata per dare avvio a riflessioni sistematiche.

Se non si passa dall'osservazione occasionale a quella sistematica risulterà difficile cercare di avanzare ipotesi interpretative di un fenomeno evitando che abbiano gioco ipotesi preconcette o veri e propri pregiudizi distorcenti.

L'approccio scientifico comporta quindi il passare da un'iniziale osservazione occasionale alle procedure dell'osservazione sistematica per garantire la verificabilità delle ipotesi esplicative del fenomeno osservato.

Riepilogo

Nel capitolo sono state presentate le caratteristiche principali e le fasi procedurali della sperimentazione, della quasi-sperimentazione e della microsperimentazione nella ricerca educativa.

È stato esplicitato il fondamentale ruolo delle operazioni di controllo delle variabili su cui si focalizza lo studio sperimentale allestito e sono state evidenziate le pecularità delle indagini condotte attraverso microsperimentazioni.

È stata inoltre ricordata l'utilità di mantenere uno sguardo sul confronto tra micro- e macrosperimentazioni considerandone sia le differenze sia gli aspetti complementari che possono arricchire la comprensione del fenomeno indagato.

Sono stati infine forniti diversi esempi di microsperimentazioni condotte in ambito educativo in tema di strategie didattiche, formazione dei docenti, atteggiamento degli stessi, livelli di apprendimento degli allievi.

Specifica attenzione è stata diretta alle caratteristiche e alle funzioni dell'osservazione occasionale rispetto all'osservazione sistematica considerando che quest'ultima costituisce un momento importante nella rilevazione dei dati conoscitivi (condotta in particolare attraverso griglie di osservazione) in vista di successive possibili operazioni di misurazione, trattamento, interpretazione e modificazione, intenzionale della realtà cui essa ha fatto riferimento.

Riferimenti bibliografici

Bandura A. (1977). Self-efficacy: Toward a unifying theory of behavioral change, *Psychological Review*, 84, 191-215.

Bandura A. (1997). *Self-efficacy. The exercise of control.* New York: W.H. Freeman, (trad. it., *Autoefficacia. Teoria e applicazioni*, Erickson, Milano, 2000).

Biasi V., Domenici G. (2013). *Scala sull'Auto-Efficacia dei Docenti (S.A.E.D.). Traduzione della Teacher Efficacy Scale di Tschannen-Moran & Woolfolk Hoy (2001)*, Laboratorio di Didattica e Valutazione degli apprendimenti e degli atteggiamenti, Università "Roma Tre", Roma.

Biasi V., Domenici G. (2014). Motivational Processes in Online Learning: The Role of Tutorship for Laboratory Activities through the Semistructured Self-Evaluation Test. *Education Research International*, 1-7. http://dx.doi.org/10.1155/2014/242417

Biasi V., Domenici G., Capobianco R., Patrizi N. (2014). Teacher Self-Efficacy Scale (Scala sull'Auto-Efficacia del Docente - SAED): adattamento e validazione in Italia (Teacher Self-Efficacy Scale: Adaptation and Validation in Italy). *Journal of Educational, Cultural and Psychological Studies*, 10, 485-509.

Calonghi L. (1977). *Sperimentazione nella scuola*. Torino: SEI.

Calonghi L. (1984). *I disegni sperimentali nella ricerca scolastica*. In: E. Becchi e B. Vertecchi, a cura di, *Manuale critico della sperimentazione e della ricerca educativa*. Milano: Franco Angeli.

Calonghi L. (1994). *Sperimentale, Pedagogia*, Enciclopedia Pedagogica, vol. VI, pp. 11042-11052. Brescia: Editrice La Scuola.

Caprara G., Barbaranelli C., Borgogni L., Steca P. (2003). Efficacy beliefs as determinants of teachers' job satisfaction. *Journal of Educational Psychology*, 95, 821-32.

De Landsheere G. (1973). *Introduzione alla ricerca in educazione*. Scandicci: La Nuova Italia.

De Landsheere G. (1985). *La ricerca sperimentale in educazione*. Scandicci: La Nuova Italia.

Domenici G. (a cura di) (2005). *Le prove semistrutturate di verifica degli apprendimenti*. Totino: Utet.

Domenici G. (2006). *Metodologia della ricerca educativa. Corso introduttivo*. Roma: Monolite.

Domenici G. (a cura di) (2009). *Valutazione e auto-valutazione per la qualificazione dei processi formativi e-learning*. Lecce: Pensa Multimedia.

Domenici G. (a cura di) (2017). *Successo formativo, inclusione e coesione sociale: strategie innovative. Strategie didattiche integrate per il successo scolastico e l'inclusione*, vol. 1. Roma: Armando.

Domenici G., Biasi V. (2012). Semistructured tests as didactic tools in distance learning: cognitive and affective processes. In: AAVV. *VIII International Conference on Education. Samos Island (Greece). Research and Training Institute of the East Aegean*, 161-168.

Domenici G., Lucisano P., Biasi V. (2017). *La ricerca empirica in educazione. Elementi introduttivi*. Roma: Armando Editore.

Domenici G., Vertecchi V. (1980). Disegni di microsperimentazione. *Riforma della Scuola*, 26(1), 52-55

Gowrie G., Ramdass M. (2013). Dimensions of teacher efficacy and student academic achievement in selected primary schools in Trinidad and Tobago. *BritishJournal of Education, Society & Behavioural Science*, 4(10), 1358-1371.

Postic M., De Ketele J.M. (1993). *Osservare le situazioni educative. I metodi educativi nella ricerca e nella valutazione*. Torino: SEI.

Sewell A., St. George A. (2000). Developing efficacy beliefs in the classroom. *Journal of Educational Enquiry*, 1(2), 58-70.

Tschannen-Moran M., Woolfolk Hoy A. (2001). Teacher efficacy: Capturing and elusive construct. *Teaching and Teacher Education*, 17, 783-805.

Zanniello G. (a cura di) (1997). *La prepedagocità della sperimentazione*. Palermo: Palumbo.

8

La sperimentazione didattica. Un dialogo possibile con la ricerca-azione e la ricerca-formazione

Il capitolo, attraverso alcuni iniziali cenni storico-normativi sulla sperimentazione nella scuola italiana, apre una riflessione sul concetto di sperimentazione didattica, ne discute gli aspetti connessi al paradigma quantitativo-sperimentale e alle scelte di tipo metodologico-procedurale. Approfondisce poi la riflessione critica sulle possibilità di sperimentazione dentro la scuola e affronta il tema complesso della professionalità dell'insegnante come "ricercatore".
Da qui, lungo il capitolo ci si interroga sulle possibilità di un dialogo problematico e critico tra sperimentazione didattica e ricerca-azione, esplicitando opportunità, condizioni e limiti di tale impianto di ricerca all'interno dei contesti scolastici.
Viene infine approfondito il concetto di ricerca-formazione, quale reale opportunità di "fare ricerca" dentro le scuole e con gli insegnanti, con lo scopo di promuovere le competenze professionali del docente in questo campo per il miglioramento delle pratiche didattiche. L'attenzione è posta sulla responsabilità scientifica, e metodologica più in particolare, di chi conduce gruppi e percorsi di Ricerca-Formazione anche con intenti di sperimentazione didattica.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Aspetti normativi connessi alla sperimentazione didattica in Italia.
- Concetto di sperimentazione nella ricerca educativa e specifica riflessione sui disegni quasi-sperimentali.
- Concetti di validità esterna e interna nella ricerca e riflessioni specifiche sulla validità interna delle sperimentazioni.
- Importanza della sperimentazione didattica per l'insegnante-ricercatore.
- Concetto di ricerca-azione, le caratteristiche di questi disegni di ricerca e gli aspetti problematici e critici di un loro dialogo con la sperimentazione didattica.
- Concetto di Ricerca-Formazione come opportunità per fare ricerca *nella* scuola e *con gli* insegnanti.

La soluzione di continuità tra le generazioni dipende dall'impossibilità di trasmettere l'esperienza, di far evitare agli altri gli errori già commessi da noi. La vera distanza tra due generazioni è data dagli elementi che esse hanno in comune e che obbligano alla ripetizione ciclica delle stesse esperienze, come nei comportamenti delle specie animali trasmessi come eredità biologica; mentre invece gli elementi di vera diversità tra noi e loro sono il risultato dei cambiamenti irreversibili che ogni epoca porta con sé, cioè dipendono dalla eredità storica che noi abbiamo trasmesso a loro, la vera eredità di cui siamo responsabili, anche se talora inconsapevoli. Per questo non abbiamo niente da insegnare: su ciò che più somiglia alla nostra esperienza non possiamo influire; in ciò che porta la nostra impronta non sappiamo riconoscerci.

Italo Calvino, *Palomar*

8.1 La sperimentazione nella normativa scolastica italiana

Il concetto di sperimentazione in ambito educativo appare propriamente nella normativa scolastica italiana a partire dall'inizio degli anni Settanta del secolo scorso, anni di grande rinnovamento per il nostro sistema scolastico in termini di apertura a istanze scientifiche e culturali internazionali.

Fu proprio la Legge n. 477 del 30 luglio 1973 ("Delega al Governo ad emanare norme sul riordinamento dell'organizzazione della scuola e sullo stato giuridico del personale direttivo, ispettivo, docente e non docente della scuola") a dare l'avvio all'emanazione di quelli che ancor oggi sono conosciuti come i "decreti delegati" (DPR nn. 416, 417, 418, 419, 420) del 1974, applicati a tutte le scuole statali di ogni ordine e grado. Essi, dando voce nella scuola a ideali democratici e progressisti derivanti in gran parte dai fermenti innovativi del Sessantotto, contengono disposizioni sullo stato giuridico degli insegnanti, sugli organi collegiali e sulla sperimentazione e la ricerca educativa nella scuola. Tuttora in vigore, furono capaci di definire in modo del tutto nuovo il ruolo del docente, caratterizzandolo come professionista all'interno di una comunità professionale che agisce collegialmente, competente nel progettare le azioni educative e didattiche e proteso verso un *fare* che è sempre anche riflessione, ricerca, sperimentazione di soluzioni innovative e rispondenti ai bisogni degli alunni (Vannini, 2009; Ciani, Ferrari, Vannini, 2020).

In particolare fu il DPR 419 ("Sperimentazione e ricerca educativa, aggiornamento culturale e professionale e istituzione dei relativi istituti") a formalizzare l'importanza della *sperimentazione nella scuola*, definendola:

Art. 1 ...espressione dell'autonomia didattica dei docenti, [essa può] può esplicarsi: a) come ricerca e realizzazione di innovazioni sul piano metodologico-didattico; b) come ricerca e realizzazione di innovazioni degli ordinamenti e delle strutture esistenti.

Come si osserva, già da questa prima indicazione normativa la sperimentazione scolastica veniva associata al concetto di "ricerca". Il decreto tuttavia – richiedendo una pianificazione strutturata di un disegno di sperimentazione solo laddove si volessero innovare ordinamenti e strutture (opzione b)[1] – creerà fin da subito un sottile malinteso sul piano delle sperimentazioni metodologico-didattiche. Per queste ultime non veniva infatti menzionata alcuna richiesta di pianificazione progettuale di ricerca.

Gattullo (1984, *cfr.* nota p. 68) – riferendosi al testo del DPR 419 – evidenziava come questa teorizzazione suggerisse un rapporto ambiguo e confuso tra *innovazione* e sperimentazione, perché solo nel caso di sperimentazioni di ordinamenti e strutture veniva esplicitamente richiesta agli insegnanti una reale procedura sistematica e rigorosa di ricerca, includendo in essa la necessità di una *verifica delle ipotesi*. Sul piano invece delle cosiddette sperimentazioni metodologico-didattiche, lo studioso affermava:

«...di fatto, nel linguaggio ufficiale ogni innovazione è definita come sperimentazione [pur se] mancano indicazioni e richieste riguardo alle ipotesi da formulare e alle procedure di verifica da compiere».

Se mancava, da parte del legislatore, un'attenzione a connettere la sperimentazione didattica a effettive prospettive di ricerca scientifica dentro le scuole, molto più esplicita era invece la connessione tra sperimentazione e autonomia didattica dei docenti, così come la necessità di saldare qualsiasi progetto sperimentale all'interno di decisioni collegiali assunte dagli attori della scuola, offrendo così un primo respiro normativo a un'idea di professionalità docente che comprendesse il *fare ricerca* all'interno del *corpus* di competenze dell'insegnante, e degli insegnanti nella loro collegialità.

A una lettura a distanza di ormai molti decenni, ciò che emerge dai decreti delegati e dalle esperienze che ne seguirono è in generale un'idea di sperimentazione un po' zoppicante, capace sì di aprire alle scuole e agli insegnanti una prospettiva di ricerca "per l'innovazione", senza tuttavia supportarla e accompagnarla con mi-

[1] Il 419/1974, unicamente per le innovazioni di ordinamenti e strutture richiedeva (all'articolo 3): "Ogni proposta o programma di sperimentazione deve contenere: la identificazione del problema che si vuole affrontare con la relativa motivazione; la formulazione scientifica dell'ipotesi di lavoro; la individuazione degli strumenti e delle condizioni organizzative; il preventivo di spesa; la descrizione dei procedimenti metodologici nelle varie fasi della sperimentazione; le modalità di verifica dei risultati e della loro pubblicizzazione".

sure di sistema e richieste adeguate. Nei fatti, il contesto normativo non aprì mai spazi effettivi a processi di ricerca scientifica *dentro* le scuole, né tantomeno a una formazione esplicita degli insegnanti su competenze teorico-metodologiche nella ricerca empirica e sperimentale in campo educativo.

Occorrerà in parte attendere la normativa sull'autonomia degli istituti scolastici (art. 21 della Legge 59/1997), e in particolare il DPR 275 del 1999[2], per avere una chiara esplicitazione di come le scuole siano effettivamente da considerare enti di ricerca, potendo esercitare autonomia su tre ambiti fondamentali: didattico, organizzativo e *di ricerca, sperimentazione e sviluppo*. All'art. 6 possiamo leggere:

> «Le istituzioni scolastiche, singolarmente o tra loro associate, esercitano l'autonomia di ricerca, sperimentazione e sviluppo tenendo conto delle esigenze del contesto culturale, sociale ed economico delle realtà locali e curando tra l'altro:
> - la progettazione formativa e la ricerca valutativa;
> - la formazione e l'aggiornamento culturale e professionale del personale scolastico;
> - l'innovazione metodologica e disciplinare;
> - la ricerca didattica sulle diverse valenze delle tecnologie dell'informazione e della comunicazione e sulla loro integrazione nei processi formativi;
> - la documentazione educativa e la sua diffusione all'interno della scuola;
> - gli scambi di informazioni, esperienze e materiali didattici;
> - l'integrazione fra le diverse articolazioni del sistema scolastico e, d'intesa con i soggetti istituzionali competenti, fra i diversi sistemi formativi, ivi compresa la formazione professionale.»

Purtroppo, tuttavia, anche qui non si fa cenno alcuno all'irrinunciabilità di momenti di verifica e valutazione delle innovazioni didattiche, se non specificamente per progetti di innovazioni che richiedano modificazioni strutturali. In questi casi, potendo fare richiesta di finanziamento diretto al Ministero, essi «devono indicare con chiarezza gli obiettivi; quelli attuati devono essere sottoposti a valutazione dei risultati, sulla base dei quali possono essere definiti nuovi curricoli e nuove scansioni degli ordinamenti degli studi...» (art. 11, c. 2).

E così, nuovamente, le innovazioni metodologico-didattiche - rientrando nell'ordinaria autonomia funzionale delle scuole - non sono tenute a garantire un *controllo* durante i processi di innovazione e in merito ai risultati raggiunti. Quasi che tali questioni - essendo tipiche di un ricercatore professionista - non riguardassero invece gli insegnanti, considerati nei fatti - dentro le realtà scolastiche (e nonostante un dibattito pedagogico di grande rilievo sul ruolo del docente ricercatore) - dei *manovali* dell'innovazione, senza effettive intenzionalità di sperimentazione scientifica.

[2] Che su questi aspetti riprendeva in gran parte il decreto 297/1994 (artt.277 e 278).

Falanga e Treccani, cercando di approfondire la norma dell'art. 6 del DPR sull'autonomia degli istituti, ribadiscono come l'attività di ricerca nella scuola dell'autonomia debba essere

> «... presentata in termini di sistema e di progettualità. Di sistema perché la ricerca è strettamente interconnessa con la sperimentazione, l'innovazione e lo sviluppo delle finalità istituzionali; di progettualità perché la ricerca deve radicarsi e trarre sostanza da un progetto, cioè da un quadro di legittimazione e razionalizzazione scientifica a sostegno delle azioni.
> La ricerca educativa a scuola non è meno scientifica della ricerca in altri ambiti della conoscenza; ma perché ciò avvenga è necessario che essa si esplichi secondo criteri scientifici collaudati e tratti dall'esperienza nonché corroborati dalla letteratura di settore (Falanga e Treccani, 2020, pp. 68-69).»

Se la normativa sull'autonomia ha aperto dunque scenari importanti nel nostro Paese per avvicinarlo a un'idea di ricerca educativa *dentro* le scuole, essi non sono stati pienamente esplicitati né supportati o accompagnati a una loro effettiva realizzazione, soprattutto sul fronte della scuola *quale ente di ricerca*. Le scuole, infatti, nonostante il Regolamento del 1999, continuano oggi a sentirsi «più una sotto-articolazione dell'amministrazione pubblica, non sfruttando pertanto appieno il riconoscimento della loro personalità giuridica autonoma (Falanga e Treccani, 2020, p. 102)».

Nel 2015, la Legge 107 riprende esplicitamente il quadro dell'autonomia scolastica e sottolinea - nell'articolo 1 che delinea i fini democratici del nostro sistema di istruzione - come la scuola debba essere un «laboratorio permanente di ricerca, sperimentazione e innovazione didattica». Nulla di più viene detto sul ruolo dell'insegnante come ricercatore, né tantomeno sullo statuto scientifico e le responsabilità che il fare ricerca a scuola dovrebbe comportare.

Non appaiono dunque superate le questioni che Gattullo mise in luce fin dal 1984 sul rapporto tra innovazione e sperimentazione «inteso e praticato nel sistema scolastico italiano in modo tutt'altro che corretto» (1984, p. 67). Mai pienamente sostenuto da una normativa che lo portasse su un piano di concreta attuazione.

Ciò che deve essere sperimentato, affermava Gattullo (pp. 66-67), è l'innovazione. È appunto la sperimentazione che "dà la vista all'innovazione" in quanto qualsiasi idea innovativa dovrebbe:

- innanzitutto, poter assumere una forma sistematizzata di ipotesi (e dunque di possibili connessioni causali tra variabili da sottoporre a controllo o verifica), saldamente ancorata a un quadro teorico significativo, che ne illumini i concetti e li renda - in qualche modo più o meno imperfetto - dicibili, condivisibili e finanche misurabili;

- in seguito, essere sottoposta - tramite impianti, procedure e strumenti che ne garantiscano un'accettabile coerenza interna - a controllo sul campo, dove le variabili potranno essere misurate, pur tramite un processo di riduzione/semplificazione;
- infine, condurre a un'analisi e un'interpretazione seria e rigorosa delle misure effettuate e delle relazioni di influenza che alcuni fattori possono avere su esiti educativi che si ritengono rilevanti, significativi per il miglioramento della scuola.

Si tratta dunque di procedure non spontanee o di buon senso, bensì da supportare con solide competenze metodologiche, capaci di garantire una validazione rigorosa di quanto viene proposto come *innovazione*.

L'ambiguità di cui nel nostro Paese è sempre stato contraddistinto il concetto di sperimentazione (Becchi, 1997, p. XIV; Corda, Costa, Vertecchi, Visalberghi, 1975) è dovuta a profonde ragioni culturali e storico-politiche che hanno reso arduo il cammino di un "pensiero sperimentale" legato alla pedagogia. Nonostante la normativa abbia, fin dal fiorire della scuola degli anni Settanta, indicato come cogente la via della sperimentazione, essa non ha avuto la forza di attecchire nei contesti se non in una sua versione puerile e semplificante. L'idea è stata piuttosto quella di un "provare a cambiare", di un tentativo di rinnovamento soprattutto in tanti e diversificati micro-contesti didattici, certamente con risultati di valore, e un insieme di spinte appassionate dei docenti ad agire per un insegnamento sempre più efficace. Ciò che tuttavia è mancato è stato un reale accompagnamento teorico e metodologico alle azioni di innovazione degli insegnanti dentro le scuole, sia come richiesta della normativa sia come effettiva disponibilità politica e istituzionale a supportare la professionalità docente. Quasi a disconfermare silenziosamente - più volte, nella pratica - la possibilità che le scuole possano davvero essere luogo di ricerca e sperimentazione.

Come già si è discusso anche nel Capitolo 7 di Biasi e Domenici, la sperimentazione - per essere davvero tale - necessita di condizioni esterne e interne favorevoli: sia per quanto concerne dunque i contesti organizzativi disponibili ad accogliere impianti di ricerca sperimentale, sia ancor più per quanto concerne l'*abito mentale* di chi "sperimenta", degli attori che condurranno i processi di sperimentazione.

Tale *habitus* è da ricercarsi nella "logica" del pensiero riflessivo di stampo deweyano (Dewey, 1938; trad. it. 1949), nel processo di induzione-ipotesi-deduzione di chi fa ricerca empirica, ma in particolare nell'impegno e nella passione di un pensiero coraggioso e umile (si veda Vannini, 2012) che interviene per innovare e controllare i risultati, che avanza ipotesi e verifica la loro validità cercando sempre più le condizioni di migliore efficacia nei processi di insegnamento e apprendimento.

È dunque, come scrive Becchi:

> ....la passione personale di chi - impegnato nella scuola e soprattutto a far conoscere ai giovani come si possa intervenire, in bene o in male, modificandolo, sul comportamento di piccoli e grandi - si chiede continuamente come rendere più efficace il suo insegnamento.
> Questo impegno significa studiare, scegliere e far conoscere i criteri migliori dell'esperienza e dell'apprezzamento educativo, convinti che un approccio in termini di sperimentalismo non agnostico, ma socialmente e politicamente impegnato, ai problemi dell'insegnare e dell'educare, dell'organizzare e del rinnovare, sia nello stesso tempo fra le modalità pedagogiche più gratificanti perché più ricche di implicazioni teoriche e operative (Becchi, 1968, in Prefazione p. XII).

8.2 Innovare e sperimentare nella didattica

Come approfondito da Biasi e Domenici nel Capitolo 7, la sperimentazione in campo educativo si caratterizza come un impianto di ricerca quantitativa con intervento; essa può essere condotta attraverso disegni sperimentali veri e propri, quasi-esperimenti e pre-esperimenti (Campbell e Stanley, 1963). Si colloca dunque propriamente all'interno di un paradigma positivista-sperimentale, all'interno del quale è possibile comprenderne pienamente il significato, non in termini meramente tecnico-procedurali, bensì a partire dalla postura del ricercatore "di fronte" all'oggetto di studio e dalla natura del suo sguardo che, dall'esterno, cerca di definire variabili e ipotizzare relazioni fra esse (Vannini, 2012; 2019).

In questo sforzo di oggettivazione - dei cui limiti deve essere consapevole in termini di riduzione e semplificazione della complessità della realtà -, il ricercatore impegnato in una sperimentazione fa ancora qualcosa di più oltre a immaginare correlazioni fra variabili: egli ipotizza relazioni di causalità, la possibilità cioè che intervenendo attivamente su una variabile, e che essa possa essere modificata, avanzando così anche previsioni per il futuro.

Ritornano qui gli elementi chiave, annunciati inizialmente, della sperimentazione: la possibilità cioè di utilizzare procedure sistematiche per il controllo delle innovazioni. E, d'altro canto, la possibilità di "provare a innovare", a cambiare in meglio la realtà educativa, dandosi gli strumenti per verificarne l'efficacia.

L'idea principale della sperimentazione in ambito educativo, e insieme forse il suo fascino, si basa appunto sulla deliberata possibilità del ricercatore di manipolare e controllare le condizioni che determinano certi eventi didattici a cui siamo particolarmente interessati (Cohen et al., 2007). Essendo difficilmente concepibile una sperimentazione didattica in laboratorio - dove quella rete di relazioni umane che connotano l'azione educativa potrebbe essere snaturata, modificando sostanzialmente le caratteristiche stesse di un intervento formativo - essa potrà essere

dunque realizzata solo *sul campo*. Introducendo in esso variabili indipendenti, si potranno definire disegni sperimentali che permettano di osservare gli esiti dei nostri interventi cercando il più possibile di neutralizzare tutte le altre variabili; ciò è possibile attraverso la costruzione randomizzata di gruppi equivalenti sperimentali e di controllo, l'organizzazione di misure rigorose in ingresso e in uscita, tutti elementi per i quali si rimanda al Capitolo 7.

Come si intuisce, non si tratta affatto di disegni di ricerca facilmente implementabili in ambito educativo, data proprio la complessità dei contesti nei quali si realizzano i processi d'insegnamento-apprendimento e che costituiscono "il campo" ineliminabile dalla struttura di qualsiasi sperimentazione. Abbiamo pertanto di fronte, in qualche misura, un disegno di ricerca quasi utopico che, nella pratica, si presenta sempre imperfetto: da un lato esso mira a realizzare le condizioni di controllo sperimentale delle cosiddette "scienze dure", dall'altro cerca di garantire uno *spazio di complessità* nel quale osservare e agire sulle variabili. Sono appunto tutte queste difficili condizioni che portano spesso a realizzare disegni di quasi-esperimento piuttosto che esperimenti veri e propri, più adatti invece a situazioni laboratoriali.

Gli stessi Cook e Campbell (1979) evidenziarono la necessità di un realismo critico per riflettere sulla natura e sui limiti degli esperimenti (si veda Becchi, 1997), che possono fare luce su connessioni causali fra variabili che sono sì "reali", ma possono essere percepite solo in modo imperfetto dai nostri sensi (Cook e Campbell, 1979, p. IX). Pur dalle posizioni di un paradigma quantitativo e positivistico, per cui l'oggetto esiste al di fuori dell'esperienza umana, i due studiosi vedono i limiti dell'esperimento e la sua fallibilità nella pretesa di spiegare quanto è governato da relazioni tra variabili molteplici difficilmente isolabili. Potremmo dire dunque che la sperimentazione didattica è possibile *solo nel modo in cui è percepibile dalla mente umana*, la quale - immersa essa stessa nella complessità del reale - riesce a costruire (operativizzare) variabili e a definire relazioni di causa-effetto comunque incompiute. A tutto ciò si associa la riflessione sul fatto che gli impianti sperimentali, anche i più perfetti, "non confermano o falsificano una teoria e, nella fattispecie, una spiegazione causale, al massimo eliminano ipotesi alternative" (Becchi, 1997, p. 180).

Nonostante il quadro di vincoli che è stato qui delineato, resta essenziale ancora oggi - per la vitalità delle scienze pedagogiche - poter parlare di sperimentazione didattica. Sono ancora rilevanti a questo proposito le inattuali parole di Egle Becchi sull'importanza della sperimentazione all'interno di una razionalità limitata (1968, p. IX):

> A tutto si intende dare verifica, a nulla si vuole attribuire significato assoluto e definitivo; e questo non tanto per la consapevolezza, particolarmente acuta nell'educatore attento e responsabile, della labilità imponderabile della situazione in cui opera, né per una scelta teoretica di tipo relativistico, quanto piuttosto per la presenza nella cultura pedagogica odierna, di modelli di natura scientifica che persuadono appunto all'impostazione di ricerca aperta, positiva e rigorosa nello stesso tempo, della problematica educativa (Becchi, 1968, prefazione).

Entro questa cornice trova certamente ragione d'essere, nel tempo odierno, un impegno della ricerca educativa nell'ambito della sperimentazione didattica, nell'uso cioè di un metodo che - seppur imperfetto - riesca a dare evidenza di quali possano essere interventi didattici più o meno efficaci per l'apprendimento degli alunni. Rinunciare a una postura *sperimentale* della ricerca educativa e didattica vorrebbe infatti dire rinunciare a partecipare a una riflessione internazionale su ipotesi di tipo esplicativo e predittivo, abdicare a fornire orientamenti di ampio respiro sull'opportunità o meno di specifiche pratiche didattiche e formative, capaci di sollecitare l'attenzione di insegnanti ed educatori che - in modo critico - possano mettere in dialogo le evidenze scientifiche e l'applicazione in situazioni contestuali, così da aprire ancora nuovi problemi e nuove ipotesi.

All'interno di un quadro ideologico e valoriale di una scuola equa e di qualità, che punti all'emancipazione intellettuale di un numero sempre più elevato di persone (e con risultati sempre più democratici), la sperimentazione didattica diviene un mezzo necessario e irrinunciabile per condurre ricerche e dare fondatezza - seppur limitata e aperta a procedure di pensiero riflessivo e critico e a sempre nuove disconferme - alle innovazioni. Quali le metodologie didattiche che sono, nel tempo e grazie a controlli rigorosi, risultate più efficaci, al di là della particolarità e della contingenza delle situazioni? Quali le pratiche didattiche in classe più utili per un apprendimento motivato e significativo?...

Queste - si ritiene - sono alcune domande che conferiscono valore di per sé stesse alla sperimentazione e all'uso - impegnato e sempre problematico - di disegni di ricerca di tipo sperimentale.

Si possono qui riprendere le parole di Vertecchi (1992, p. 16) quando afferma che in ambito educativo abbiamo comunque sempre bisogno di "buone misure", che liberino il discorso pedagogico dal respiro corto dell'esperienza contestuale e dall'impressionismo.

8.2.1 La questione della validità interna nelle sperimentazioni

I temi della validità interna ed esterna di una ricerca sperimentale si pongono come questioni centrali, sia a livello procedurale sia - ancor prima - a livello teorico-metodologico.

Il concetto di validità di una sperimentazione attiene in generale al suo grado di coerenza; nello specifico molta letteratura internazionale si sofferma sulla validità

interna di un esperimento - *condicio sine qua non*, senza la quale cioè qualsiasi esperimento non è interpretabile (Campbell e Stanley, 1963, p. 5) - considerandola come il grado di coerenza tra le differenze osservate sulla variabile dipendente (differenze tra gruppo di controllo e gruppo sperimentale) e la manipolazione della variabile indipendente. A questo si associa poi il concetto di validità esterna, in ragione del quale i risultati di un esperimento possono essere estesi (generalizzati) anche ad altri soggetti, casi e situazioni.

Le questioni, già trattate in altri capitoli, appaiono di notevole complessità. Qui, nello specifico, ci si vuole in particolare soffermare sul concetto di *validità interna*. Abbiamo visto che esso, nella sua versione più specifica, pone la questione di quanto i cambiamenti osservati sulla variabile dipendente siano effettivamente attribuibili all'intervento e non ad altre variabili (Adams e Lawrence, 2019, p. 89-90; Gay et al., 2009, p. 242). Si comprende fin da subito come questa condizione specifica di validità interna - sulla quale rimane sempre fondamentale l'impegno e la cura del ricercatore - sia difficile da garantire in modo completo, se non in un esperimento perfetto che elimini tutte quelle "minacce alla validità" già elencate in precedenti capitoli. Ancor più impegnativo e denso di ripercussioni diviene il concetto di validità interna del disegno di una ricerca sperimentale nel suo significato più ampio: quanto i risultati che sono stati trovati sono coerenti con gli obiettivi iniziali della ricerca? Ma, anche, con le sue ipotesi e i costrutti delle variabili che sono stati definiti, con l'intervento didattico che è stato messo a punto, con gli strumenti e le procedure di misurazione della variabile dipendente, ecc...? I risultati di una sperimentazione, insomma, dovrebbero descrivere accuratamente il fenomeno indagato (Cohen et al., 2007, p. 135). Ancora: la validità interna di una sperimentazione dovrebbe dare garanzie sulla sua "capacità di sottoporre a verifica un'ipotesi in funzione della quale essa è stata concepita" (Viganò, 1996, p. 159).

Tutto questo discorso ci conduce a riflettere sulla necessità di "agganci" coerenti e ben pensati fra le varie fasi di una sperimentazione. Come per qualsiasi ricerca, ciò che ne caratterizza l'organicità, e dunque anche la credibilità, è un atteggiamento di attenzione costante del ricercatore a ritornare sistematicamente sulle domande iniziali della ricerca, riflettendo sulle sue motivazioni iniziali e il quadro teorico e valoriale nel quale la ricerca si innesta.

Tali questioni hanno trovato, nel nostro Paese, un'importante riflessione teorico-epistemologica soprattutto in Lucia Lumbelli, che ha approfondito - fin dagli anni '80 del secolo scorso (Lumbelli, 1979; 1980; 1984; 2006) - il tema delle connessioni tra impianto teorico e validità dei dati di una ricerca quantitativa, soffermandosi sull'importanza di una fase esplorativo-qualitativa che possa appunto conferire credibilità al processo della ricerca empirica. Tanto più se si tratta di disegni sperimentali, dove la tensione a un forte controllo del piano di sperimentazione (con conseguente riduttività della complessità del reale) può rischiare di far dimenticare aspetti di validità interna fondamentali.

Il ricercatore sperimentale, infatti, opera continui passaggi verso una semplificazione della complessità dei contesti. Fin dalla definizione del problema della ricerca, egli inizia una sua traduzione in un linguaggio comprensibile agli altri estrapolando gli interrogativi principali e formulandoli secondo un codice linguistico il più possibile condivisibile a livello intersoggettivo (Vannini, 2019); ciò conduce a fare un primo piccolo passo verso l'*oggettivazione* e l'*astrazione*. Come afferma Lumbelli, riprendendo Dewey e "Le fonti di una scienza dell'educazione":

> «...l'astrazione dai contesti problematici concreti deve essere argomentata in quanto è corretta ed auspicabile solo a determinate condizioni: l'astrazione deve garantire la rilevanza dei concetti dal punto di vista del problema di partenza; deve essere provvisoria e funzionale alla fase di ricerca conoscitiva, i cui risultati devono essere poi confrontati in modo critico e riflessivo con l'originaria esperienza pratica concreta» (Lumbelli, 2006, p. 27).

Queste indicazioni ci rammentano che, nella ricerca pedagogica, la pratica educativa non è solo il punto di partenza, ma altrettanto deve essere il punto di arrivo, deve cioè costituire l'elemento primo e ultimo di confronto della validità dei risultati a cui si perverrà.

> «...lo scopo di "metter alla prova di realtà" i concetti o variabili di ipotesi da controllare poi sperimentalmente. Si potrebbe con qualche cautela parlare della componente induttiva del processo di astrazione di cui è stata finora ricostruita la componente teorico-deduttiva. Una componente di indagine empirica che è qualitativa nel senso di preliminare a, o preparatoria della, ricerca sperimentale o "quantitativa"» (Lumbelli, 2006, p. 45).

Il compito della ricerca esplorativa diviene dunque - secondo un modello più specificamente quantitativo-sperimentale - quello di individuare e definire la fenomenologia delle situazioni e dei comportamenti nei contesti educativi, in un processo continuo di deduzione-induzione, che va dai fatti all'elaborazione teorica e viceversa. In questo modo, viene pienamente valorizzato quel momento creativo di definizione delle ipotesi così importante nel metodo scientifico, così come lo descrive Dewey. La fase esplorativo-qualitativa tende quindi a definire variabili da controllare e categorie di osservazione da utilizzare, si può dire - con Lumbelli - che "garantisce la qualità della quantità" in una fase in cui si va attuando la progressiva *operazionalizzazione* dei concetti che andranno poi a costituire le vere e proprie variabili della ricerca quantitativa e sperimentale.

8.2.2 L'importanza di un approccio sperimentale nell'ambito della ricerca didattica

Come osserviamo, un atteggiamento di razionalità *limitata* (Lipari, 2002) appare il più corretto quando ci si approccia oggi ai temi della sperimentazione in ambito educativo. Nessun senso di assolutezza del metodo, dunque, bensì una costante cautela e cura metodologica in vista di un fine di particolare rilevanza scientifica. Vediamo meglio che cosa intendiamo con tutto ciò.

Possiamo ragionevolmente affermare che la sperimentazione costituisce ancora oggi un metodo fondamentale in ambito educativo; di certo essa diviene irrinunciabile soprattutto nella *ricerca didattica*. Proprio in questo campo così rilevante per gli studi pedagogici, la sperimentazione assume un ruolo prioritario nel conferire uno statuto scientifico alle innovazioni didattiche, tale da spostare le riflessioni sull'efficacia delle metodologie didattiche dal piano del senso comune a quello delle considerazioni teoriche ponderate, delle variabili da costruire attraverso terminologie appropriate e condivisibili, delle ipotesi da verificare attraverso impianti misurativi ben pensati, delle interpretazioni aperte al dubbio e a sviluppi di ricerca successivi. Si tratta dunque di un approccio che sposta l'innovazione dal piano degli entusiasmi del "qui e ora", al livello più credibile della riflessività scientifica, che vede soprattutto nei tempi lunghi, nell'analisi sistematica dei risultati - anche e soprattutto in una dimensione longitudinale - i suoi punti di forza e di validità.

Insomma la sperimentazione didattica, come si è detto, richiama fortemente al bisogno di misure, all'importanza di esercitare un pensiero scientifico nel campo pedagogico, e chiede a chi si occupa di didattica di assumere la postura cognitiva di chi, coraggiosamente, innova e al contempo, umilmente, verifica. Con la disponibilità a un'analisi rigorosa dei risultati e l'apertura a interpretazioni capaci di costruire nuove ipotesi, migliori e più coerenti con i contesti reali.

Ovviamente, come abbiamo visto, ciò comporta rischi e difficoltà: i disegni di ricerca sperimentali sono complessi nella loro pretesa di controllo delle variabili e pertanto spesso poco sostenibili all'interno delle realtà scolastiche che - pur bisognose di disporre di informazioni sull'efficacia dei processi didattici che mettono in atto - pongono vincoli organizzativi e strutturali alla realizzazione di disegni propriamente sperimentali.

Pur riconoscendo dunque l'irrinunciabilità di disegni di sperimentazione - quale forma di ricerca empirica specificamente orientata a controllare l'efficacia (in termini di qualità ed equità) dei processi di insegnamento-apprendimento - occorre considerare la necessità di forme di sperimentazione realmente sostenibili all'interno dei contesti educativi. In particolare, come prima accennato, può essere utile a questo proposito fare riferimento a disegni di tipo *quasi-sperimentale*. Essi, rinunciando alla equivalenza fra gruppi randomizzati sperimentali e di controllo e, alle volte, rinunciando anche alla presenza dello stesso gruppo di controllo, propongono impianti per la verifica degli effetti della variabile indipendente spesso imperfetti, ma particolarmente interessanti nel "negoziare" tra rigore del controllo e attenzione ai contesti. La loro sostenibilità all'interno delle situazioni scolastiche li rende infatti particolarmente interessanti per lo sviluppo della sperimentazione didattica nel senso di costruire via via dei momenti esplorativi circa la validità delle ipotesi in campo. Come affermavano Campbell e Stanley (1963), tali disegni ci mettono pienamente dentro la logica di un atteggiamento sperimentale, che non consente in realtà di confermare o disconfermare una ipotesi, bensì di "sondare il terreno" e raccogliere sempre nuove informazioni sulla validità dell'ipotesi che abbiamo formulato, così da vedere se proprio tale ipotesi possa sopravvivere a numerose esplorazioni.

> In a very fundamental sense, experimental results never "confirm" or "prove" a theory-rather, the successful theory is tested and escapes being disconfirmed. The word "prove," by being frequently employed to designate deductive validity, has acquired in our generation a connotation inappropriate both to its older uses and to its application to inductive procedures such as experimentation. The results of an experiment "probe" but do not "prove" a theory. An adequate hypothesis is one that has repeatedly survived such probing, but it may always be displaced by a new probe (Campbell e Stanley, 1963, p. 35).

È fondamentale, affermano i due studiosi, avere una buona consapevolezza delle debolezze dei disegni sperimentali e, ancor più, di quelli quasi-sperimentali (Campbell e Stanley, 1963). Molte sono le minacce alla validità dei loro risultati ed è la consapevolezza del ricercatore (e la fondatezza teorica e pratica delle sue ipotesi, vogliamo sottolineare) che può aiutarlo a illuminare le sue interpretazioni dei risultati, nel senso di proseguire o meno sulla strada intrapresa, per saggiare, continuare a esplorare e raccogliere ulteriori prove sulla credibilità della propria ipotesi.

In questo senso, diventa qui importante vedere i vantaggi dei disegni quasi-sperimentali all'interno del campo della sperimentazione didattica. In particolare, ne cogliamo qui tre aspetti che paiono rilevanti per supportare l'innovazione nella scuola, tutti fortemente connessi alla loro "imperfezione".

Innanzitutto il vantaggio della imperfezione del disegno quasi-sperimentale, che lo pone nella posizione di un impianto utile alla ricerca esplorativa, dove è possibile curare particolarmente l'aggancio tra i reali bisogni di conoscenza scientifica dei contesti scolastici e le ipotesi da costruire, tra le ipotesi operative e i quadri teorici di riferimento, in una parola: tra la teoria e la pratica.

In secondo luogo va rimarcata l'enfasi sulla diacronicità. In un disegno quasi-sperimentale è maggiormente possibile dedicare tempi distesi alla sperimentazione; la consapevolezza che le variabili saranno comunque più blandamente tenute sotto

controllo, può spingere a volte a dare maggiore attenzione alla ripetizione di controlli nel tempo (si pensi ad esempio al disegno su gruppo unico) e all'aggiustamento del piano sperimentale *in itinere*.

Infine, va evidenziato che gli impianti quasi-sperimentali, particolarmente adatti a innestarsi entro contesti di scuola reale, hanno la possibilità di coinvolgere maggiormente gli "attori" dei contesti stessi, per costruire insieme con loro il disegno di ricerca adeguandolo ai vincoli del contesto: in termini di strumenti di misurazione delle variabili, di realizzazione dell'intervento, di tempistiche da mettere in atto ecc.

Di questa necessità di coinvolgimento degli attori nella ricerca sul campo si parlerà maggiormente nei prossimi paragrafi; basti qui ricordare alcune parole particolarmente eloquenti di Luigi Calonghi:

> I piani d'esperimento dovrebbero essere in grado di adeguarsi sia al complesso intarsio di fattori che intervengono, sia al variare del loro dosaggio, del loro articolarsi.
> [...] La sperimentazione deve rispondere alle esigenze della didattica, deve poter seguire, tra l'altro, i processi di formazione, di apprendimento e i contesti individuali e sociali sempre più ampi (Calonghi, 1984, p. 85).

8.3 Nella scuola, l'insegnante... "sperimenta"

Allo scopo di entrare ancor più nel merito del coinvolgimento degli insegnanti nella ricerca "sperimentale", possiamo qui considerare la sperimentazione come un vero e proprio "modo di pensare" la ricerca, come un "abito mentale" inteso in senso deweyano che richiama aspetti di rigore, di attenzione alla misura, di intervento didattico controllato, di analisi precisa delle pratiche didattiche in termini di ricadute sugli apprendimenti degli studenti, ecc.; ma anche di attenzione alla verifica di ipotesi "credibili", nate dentro i contesti, sostenibili, capaci di coinvolgere responsabilmente coloro che la didattica la praticano da protagonisti...

La postura dell'insegnante-ricercatore che *sperimenta* è quella di chi sa immaginare interventi innovativi efficaci, ma li realizza e li controlla solo dopo averli attentamente progettati, dotandosi di estrema cautela nella fase di implementazione, di strumenti atti a monitorare il processo e, via via, i risultati. Si tratta di rallentare l'uso di un pensiero intuitivo per fare appello a un pensiero riflessivo, sistematico, che pone il dubbio come atteggiamento di fondo.

In questo senso, si può forse parlare di un *atteggiamento sperimentale* che sappia permeare la cultura delle scuole e coinvolgere gli insegnanti in quanto anch'essi ricercatori. Lo stesso Calonghi (1984, p. 86), fin dagli anni '80 ribadiva l'importanza di una sperimentazione didattica capace di mettersi in discussione con gli specifici interessi dei soggetti che dentro la scuola agiscono e "rinnovano"; di una sperimentazione dunque come occasione di formazione *alla* e *della* professionalità docente.

Sulla complessa questione della funzione dell'insegnante come ricercatore (Stenhouse, 1997; De Landsheere, 1978; Pontecorvo, 1984), la riflessione sarebbe qui molto ampia e articolata e avrebbe necessità di toccare una lunga serie di fattori che hanno condotto - almeno fino a questo tempo storico presente - a constatare il fallimento di un progetto professionale che sarebbe invece stato centrale e dirimente per la costruzione di una scuola italiana rinnovata, realmente aperta a un atteggiamento scientifico e sperimentale, capace di orientare le proprie scelte sulla base di un disegno pedagogico autonomo da spinte politico-culturali neo-liberiste o pseudo-aziendaliste. In questa sede non si entrerà propriamente nel merito di tali questioni; bensì si intende ora solo tenerle sullo sfondo dei discorsi che si andranno a delineare, allo scopo di riflettere su come la sperimentazione didattica possa - ancora oggi - costituire una via virtuosa per progettare dentro le scuole percorsi di crescita professionale dei docenti impegnandoli dentro veri e propri impianti di ricerca empirica che risultino per loro "interessanti" e significativi.

8.3.1 Le vie aperte dalla Ricerca-Azione

La possibilità di portare la sperimentazione *dentro* i contesti, per coinvolgere i contesti stessi in veri e propri percorsi di ricerca al fine di affrontare situazioni insoddisfacenti e usare la ricerca come leva per il cambiamento, è una sfida che parte da lontano e apre un dibattito importante e molto articolato che riporta alla questione tra *ricerca di base*, orientata alla conoscenza, e *ricerca applicata* (od operativa), orientata all'azione e alla presa di decisone (De Landsheere, trad. it., 1985). Un dibattito che, attraverso le prime ricerche partecipative del primo '900[3], porta a immaginare il disegno di ricerca "con intervento" (specificamente sperimentale e nomotetico, aderente a un paradigma quantitativo) come una nuova architettura metodologica, capace di unire direttamente teoria e pratica e divenire ricerca-intervento.

Attraverso una storia complessa (si veda in particolare Barbier, 2007), prende così avvio una tradizione di *ricerca partecipante* (International Council for Adult Education, 1977) che apre moltissimi dubbi sulla sua validità, proprio perché "non rientra nella ricerca scientifica propriamente detta" (De Landsheere, op. cit., p. 19), ma persegue quale principale obiettivo quello sociale e politico, quello cioè "di far

[3] Barbier (trad. it., 2007, p. 19ss), a questo proposito, afferma che le radici più profonde "vanno ricercate nei metodi d'investigazione proposti da alcuni ricercatori in scienze sociali del XIX secolo e del primo quarto del XX secolo. Penso alle inchieste operaie di Karl Marx che a suo tempo portò gli operai delle delle manifatture a riflettere sulle proprie condizioni di vita attraverso le risposte a un'inchiesta in un questionario concepito come strumento militare..."

funzionare al meglio il gruppo implicato nello sviluppo della stessa ricerca".

Come afferma Nigris (1998, p. 166), questo tipo di ricerca:

> «si prefigge [...] di superare il modello di conoscenza che vede ricercatori e operatori impegnati su fronti diversi (uno teorico e l'altro pratico), senza la possibilità di verificare congiuntamente i risultati della ricerca stessa negli specifici contesti educativi. La Ricerca Partecipante intende cercare nuove forme di collaborazione fra chi opera e chi fa ricerca, che garantiscano al tempo stesso l'aderenza ai problemi e ai contesti concreti e la loro verifica in campo da un lato, e la rigorosità dei procedimenti e dei risultati della ricerca dall'altro».

Le questioni epistemologiche che questa nuova forma di ricerca apre sono molteplici; si tratta in effetti dell'irrompere di un paradigma interazionista e qualitativo *dentro* una tradizione positivistico-sperimentale, che scardina alcuni aspetti basilari della sperimentazione (in particolare l'idea del controllo delle variabili) e che porterà ben presto alla definizione di una forma di ricerca "con intervento" del tutto qualitativa e contestuale: la ricerca-azione (R-A). L'*action-research*, così come venne originariamente definita da Lewin (1946) nell'ambito delle scienze sociali, ha l'intento di presentarsi come un dispositivo di ricerca che è al contempo *leva politica di cambiamento* all'interno dei contesti sociali.

Inizialmente, afferma De Landsheere (op. cit, p. 18), negli anni '30, il proposito di Lewin (*che negli aspetti essenziali aderiva a principi del modello empirico-analitico*) «... era quello di collegare il metodo della ricerca sperimentale classica a un obiettivo di cambiamento sociale specifico». In seguito, a partire dal dopoguerra, la R-A andrà definendosi come metodo "in aperta opposizione" al paradigma sperimentale, come una «protesta contro la separazione tra il pensiero e l'azione» (Barbier, op. cit., p. 29).

In questo senso:

> «...la R-A si definisce come un approccio che ingloba l'attore in un progetto, vale a dire in una politica, in una intenzionalità dei contratti, dei ruoli, delle aspettative e, nello stesso tempo, lo coinvolge in un processo di riflessione e di analisi che noi diciamo essere armato, vale a dire operante grazie a dispositivi di raffinamento dei dati» (Pourtois, 1984, p. 139).

Non si tratta pertanto più di una ricerca sperimentale che semplicemente tenta di ampliare il suo essere "operativa" o applicata, ma della costruzione di una forma di ricerca nuova, dove l'attore stesso del contesto "è in ricerca". La ricerca diviene ricerca "in azione", collocata *nel farsi* delle prassi degli attori sul campo, dentro i contesti. Ardoino (1980, p. 18) afferma che è: «scienza dell'azione, o, più esattamente, logica dell'azione, orientata essenzialmente verso la conoscenza raffinata della pratica, in vista della sua ottimizzazione».

Rimane certamente l'idea generale di "ricerca con intervento" (Coggi e Ricchiardi, 2005), ma così contestualizzata, *dentro* la situazione, da divenire più strumento politico che di ricerca: una nuova strategia di ricerca "sperimentale", ma capace di distinguersi da essa per il rifiuto del paradigma che la sostiene e che «comporta intrinsecamente una logica artificiale in rapporto alla realtà vivente» (Barbier, trad. it. 2007, p. 25).

Insomma, se la R-A nella sua fase iniziale mantiene la forma di una ricerca controllata e strutturata in fasi ("una metodologia sperimentale in vista dell'azione", come afferma Barbier), secondo un'idea di ricerca applicata dove il ricercatore - pur coinvolgendo gli attori - mantiene le redini delle procedure di indagine; nella sua fase avanzata la R-A quasi sospinge "dall'interno" il paradigma sperimentale a uscire dal campo e a dare spazio a una strategia di ricerca che ideologicamente rifiuta qualsiasi elemento di razionalismo. Il cosiddetto "intervento" (caratterizzato dall'introduzione della variabile indipendente nel disegno sperimentale classico) si trasforma allora nel coinvolgimento totale del ricercatore sul campo che «partecipa a ogni fase dell'evoluzione del progetto, allo stesso titolo che i soggetti della ricerca, cioè gli attori. In altre parole la R-A costituisce un progetto sociale, rivestito da un progetto scientifico» (Pourtois, 1984, p. 135).

Se è vero che si possono distinguere oggi diversi modelli di R-A (Sorzio, 2019), possiamo dire che è tra i più recenti che si giunge a distinguere il fare ricerca dall'«*essere in ricerca* [...]. In questo caso l'intervento è, senza dubbio, quello di *appropriarsi di un sapere*, ma anche di *arricchire personalmente* gli attori» (Pourtois, 1984, p. 139). Barbier enfatizza il ruolo emancipatore e critico (op. cit., p. 45) di questo modo di fare ricerca, dove sono le pratiche degli attori e del ricercatore a essere esse stesse oggetto di indagine. Si tratta in questo caso - afferma ancora l'autore - di un orientamento esistenzialista, aperto alle influenze della psicosociologia clinica, dove è soprattutto lo strumentario dell'etnografia a orientare l'apprendimento nel contesto degli attori-ricercatori che agiscono, studiano e apprendono la prassi nel loro farsi sul campo, negoziando rappresentazioni attraverso un agire comunicativo di fiducia reciproca tra gli attori (Habermas, 1986).

Nel manuale di metodologia di Cohen, Manion e Morrison del 2007, la R-A in ambito educativo viene descritta come uno *strumento potente per il cambiamento e il miglioramento all'interno dei contesti educativi* (p. 297), così come ampiamente approfondito in Italia da Losito e Pozzo (2005). Il suo scopo è quello di essere utilizzata in situazioni dove vi sia una esplicita richiesta di intervenire per la risoluzione di problemi e/o perseguire un cambiamento, e richiede sempre un coinvolgimento - se all'interno di un istituto scolastico - di insegnanti o gruppi di insegnanti che costituiscano, insieme ai ricercatori, un gruppo di ricerca che lavorerà in modo collaborativo, eventualmente coinvolgendo anche altre componenti dell'istituzione interessata.

Tra le caratteristiche che McKernan (1991, pp. 32-33) elenca per definire lo statuto della R-A, si possono evidenziare le capacità della R-A: di aumentare la conoscenza a partire dalla soluzione di problemi pratici e fortemente contestualizzati; di interessarsi di problemi centrali per gli attori del contesto, con lo scopo di promuovere cambiamento e migliorare la qualità delle condizioni di lavoro e delle pratiche; di promuovere collaborazione e partecipazione tra gli attori; di mettere in atto processi ciclici di raccolta dati, feedback e progettazione di fasi successive della ricerca; di utilizzare in modo eclettico diverse strategie di raccolta dei dati comunque sempre distanziandosi da un paradigma di tipo positivista-sperimentale e di controllo delle variabili; di promuovere auto-valutazione, dialogo, narrazione, riflessività.

Sebbene non si possano definire fasi strutturate per la conduzione di una R-A, il processo è sempre connotato da momenti che si connettono fra loro in una relazione dialettica e "a spirale" (Gay et al., 2009, p. 489) che ha come premessa (Lewin, 1946) la costituzione di un gruppo di ricerca collaborativo, dove ricercatori e attori siano ugualmente coinvolti e condividano conoscenze e competenze (Coggi e Ricchiardi, 2005, p. 69).

In particolare, il processo ciclico della R-A prevede sempre (Gay et al., 1989, p. 489):

- l'identificazione di un focus problematico (pregnante e significativo) all'interno del quale gli insegnanti desiderano promuovere un cambiamento;
- la raccolta di dati, più spesso attraverso tecniche di tipo qualitativo, ma anche attraverso strumenti strutturati che - pur non compresi all'interno di disegni di ricerca orientati a un paradigma quantitativo e nomotetico - restituiscono dati utili ad analisi di tipo statistico;
- l'analisi e l'interpretazione dei dati, attraverso procedure analitiche miste, che tengono conto di dati narrativi e dati numerici, promuovendo una triangolazione fra strumenti e fonti atta a restituire descrizioni coerenti e complesse al gruppo di attori della ricerca;
- la progettazione e lo sviluppo di un piano d'azione per il cambiamento della situazione di partenza: si può forse parlare di un "intervento non controllato", ma dal quale nuovamente si potrà partire per valutare ricadute delle azioni compiute e problemi ancora da risolvere, così da apprendere i nuovi passi di ricerca che dovranno essere compiuti successivamente.

Si tratta appunto di un processo dinamico e ciclico, che mette in gioco insegnanti e ricercatori impegnati insieme *nell'azione e nella riflessione sull'azione*, in una logica *responsiva* dell'imparare a imparare.

Va rimarcato che tale processo - di pieno coinvolgimento del ricercatore nella ricerca e nell'intervento e di rifiuto di un paradigma più specificamente sperimentale - «presenta un rischio elevato di veicolare delle imperfezioni metodologiche» (Pourtois, 1984, p. 151). In questo senso, la R-A non può essere utilizzata per il controllo di ipotesi, dato l'impianto prettamente qualitativo e contestuale; essa tuttavia - afferma Pourtois - «è il luogo elettivo dell'applicazione pratica della conoscenza scientifica». De Landsheere (1979, p. 228) parla, in questa prospettiva, di una R-A come "stadio intermedio" tra la scoperta della conoscenza e l'utilizzazione quotidiana.

8.3.2 La sperimentazione: un dialogo difficile con le istanze della R-A

Ritornando ora al discorso sulla sperimentazione in ambito educativo dal quale si è inizialmente partiti, la domanda che si pone è: "come connettere questa logica della R-A ai temi della sperimentazione didattica?"

La risposta la si può forse trovare attraverso due piste.

La prima è quella della necessità di aprire "spazi" durante i percorsi della ricerca educativa di tipo sperimentale che lascino possibilità di immersioni sul campo, di contaminazioni e coinvolgimenti all'interno degli specifici contesti educativi dove il ricercatore possa esplorare connessioni tra ipotesi teoriche e ipotesi operative, provare strumenti, fare esperienza di "oggetti" della ricerca che prima o poi, forse, potranno trasformarsi in costrutti operativi e in variabili da controllare entro impianti più quantitativi. In questa prospettiva, fondamentali sono le posizioni di Lucia Lumbelli sulla *ricerca esplorativo-qualitativa* di cui si è precedentemente parlato; ma anche le riflessioni dello stesso Calonghi di trovare possibilità nella sperimentazione didattica per garantire - attraverso *disegni quasi-sperimentali* - spazi per mantenere la complessità dei contesti.

La seconda pista da seguire è quella della R-A come possibilità e luogo privilegiato per la formazione degli insegnanti alla *ricerca*, e alla metodologia della ricerca nello specifico.

La questione qui apre a ulteriori numerose questioni secondarie, che chiamano in causa nuovamente la distanza tra i paradigmi di riferimento della sperimentazione e quelli della R-A, la necessità, da un lato, di garantire validità e affidabilità della ricerca e, dall'altro, di mantenere la complessità dei contesti e l'impegno/la responsabilità politica dei ricercatori dentro i contesti.

Può la R-A offrire l'opportunità di una formazione degli insegnanti alla ricerca attraverso la ricerca?

A questa domanda si tenta qui di dare una risposta problematica e critica, a partire soprattutto da una riflessione su quanto avvenuto nel nostro Paese con un'infatuazione, a volte troppo entusiastica, dei ricercatori pedagogisti per la R-A, quasi un'adesione non sempre consapevole per una forma di ricerca di particolare fasci-

no, ma capace facilmente di condurre verso percorsi di indagine empirica privi di qualsiasi attenzione per la validazione interna del disegno di ricerca e l'affidabilità dei dati raccolti.

È noto che un uso semplificato e non consapevole della R-A porta al rischio di confondere la ricerca empirica e sperimentale con un empirismo (Visalberghi, 1989; 1993) che nulla ha a che fare con la ricerca scientifica. Il modello della R-A fa una scelta di campo ben precisa tra i paradigmi della scienza e, consapevolmente, sceglie di rifiutare la ricerca sperimentale e nomotetica. Difficile pensare che "attori" dei contesti educativi non preparati alla ricerca possano altrettanto consapevolmente fare questa scelta.

Allo stesso tempo, l'impegnativa pariteticità richiesta tra insegnanti e attori all'interno del gruppo di ricerca e il pieno coinvolgimento del ricercatore (anche dal punto di vista esistenziale, come afferma Barbier [trad. it. 2007]) nel contesto sociale ed educativo - aspetti fondanti di un autentico approccio di R-A - sono elementi connotativi che, in alcune esperienze un po' troppo semplificanti, sono stati confusi con meri percorsi di formazione attiva degli insegnanti in servizio, dove i ruoli ben distinti di formatori, da un lato, e di insegnanti-discenti, dall'altro, si sono mantenuti saldamente distinti.

La risposta alla domanda iniziale può essere dunque forse rintracciata nella più umile e pragmatica constatazione che la sperimentazione nel campo della didattica ha sì necessità di avvicinare i grandi paradigmi scientifici, senza tuttavia pensare di confonderli in un "inganno sul campo" agli insegnanti che necessitano invece prima di tutto di apprendere la postura, i metodi e gli strumenti a disposizione del ricercatore, così da poterli usare per costruire via via nel tempo la propria professionalità. In questo senso forse la logica della R-A non è la risposta più adatta per coinvolgere insegnanti in gruppi di ricerca senza che prima ci sia stata una loro autentica e impegnativa *formazione* alla ricerca empirica, capace di offrire loro strumenti critici per valutare e decidere se e come - da esperti e protagonisti - si voglia successivamente partecipare a gruppi di R-A.

8.4 Fare ricerca e sperimentazione con la "ricerca-formazione"

Come già accennato, la problematica dell'insegnante-ricercatore (e delle difficoltà strutturali della scuola italiana a dare reale riconoscimento a questa funzione docente) resta qui in filigrana come una questione che orienta la riflessione su come e quanto la sperimentazione possa divenire strumento di ricerca nei contesti scolastici.[4]

[4] Vengono qui riprese alcune riflessioni tratte da: Ira Vannini (2018) "Fare ricerca educativa per promuovere la professionalità docente": "Qui ed ora" del Centro CRESPI. In: G. Asquini (a cura di), *La Ricerca-Formazione. Temi, esperienze, prospettive*, Milano: Franco Angeli.

Nei contesti internazionali del nord Europa o del dibattito statunitense, la cosiddetta *teacher research* (Lankshear e Knobel, 2004) dà risposte importanti a questi interrogativi, mettendo appunto al centro dello sviluppo della professionalità l'*insegnante come ricercatore*. Non mancano infatti esempi di percorsi di ricerca, anche con intervento sperimentale, che chiamano direttamente in campo il protagonismo degli insegnanti nel *fare ricerca*, competenti nelle metodologie di indagine, spesso favoriti all'estero con specifici programmi, borse di studio, ... per analizzare e riflettere sulle pratiche di insegnamento "dal di dentro", all'interno dei contesti (Craig, 2009). La risposta che propongono questi percorsi è l'opportunità di entrare all'interno della unicità dei contesti scolastici, di studiarli a partire da ipotesi che risultano particolarmente valide proprio perché fortemente agganciate a problemi reali, vissuti dagli stessi insegnanti-ricercatori. I processi della ricerca sono poi gestiti con competenza metodologica: si tratta di ricerche che mettono tendenzialmente in campo metodologie qualitative, perché fortemente contestuali, ma che prestano attenzione anche alle analisi quantitative. Esse sono inoltre preferenzialmente rivolte alla messa in campo di "nuove" pratiche didattiche in classe, ma non mancano anche percorsi sui processi di innovazione istituzionale della scuola. La *teacher research*, in questo senso, costituisce un'occasione fondamentale per sviluppare la professionalità dei docenti e per acquisire un sapere che consenta loro di prendere decisioni autonome rispetto ai processi di insegnamento-apprendimento che vanno gestendo nelle classi.

In questa prospettiva, occorre dunque riflettere su quali siano le condizioni contestuali affinché una ricerca "sperimentale" *degli insegnanti* sia realmente possibile da implementare all'interno delle scuole: con quale quadro normativo di cornice? con quali supporti istituzionali interni agli istituti scolastici? con quale riconoscimento di queste attività a livello sociale e in termini di carriera individuale del docente? ecc. È facile comprendere come solo in presenza di tali condizioni positive risulterebbe possibile agire in un'ottica di *empowerment professionale* (Fetterman, 2000) e di effettiva promozione di una ricerca empirica interna alle scuole, sotto la diretta responsabilità dei docenti.

Nell'attuale cornice sociale e politico-istituzionale, la realizzazione di vere e proprie attività di ricerca all'interno degli istituti scolastici evidenzia la necessità (e l'urgenza) di dover essere ancora pazientemente e tenacemente costruita. A questa costruzione dovrebbero partecipare i molteplici attori del mondo della scuola - dai responsabili politici ai referenti di governo delle scuole, dalle associazioni degli insegnanti al mondo della ricerca e della formazione - e le parole chiave dovrebbero essere *progettualità, lungimiranza* e *tempi lunghi*. In nessun caso è possibile pensare a una ricerca dentro le scuole che sia vissuta come una indicazione calata dall'alto (che richieda pratiche esecutive di pseudo-ricerca); come si è detto, l'interrogativo è invece quale debba essere il progetto che orienta la progressiva co-

struzione di un'autentica *cultura della ricerca e della sperimentazione* dentro le scuole.

La professionalità docente ha oggi più che mai necessità di supporto e accompagnamento. Troppo spesso, nei vissuti recenti con gli insegnanti all'interno delle scuole (Vannini, 2016), emerge infatti una richiesta potente di aiuto in termini formativi che, se da un lato si esplicita come esigenza di precise indicazioni su strategie e tecniche didattiche e docimologiche, dall'altro lato implicitamente rimanda a una generica domanda di sostegno alla funzione docente nella scuola (e forse nella società). Quasi l'espressione di un bisogno di ritrovare il senso del proprio *essere insegnanti* e di fiducia nel proprio agire in classe in modo efficace.

È dunque forse nel supporto a un pensiero riflessivo (Dewey, 1949) e critico, a un pensiero sistematico e progettuale, a un "abito mentale" orientato alla ricerca e alla sperimentazione che possiamo sostenere lo sviluppo della loro professionalità?

È a partire da queste domande, contingenti e urgenti per il sistema scolastico del nostro Paese, che nasce l'esigenza di discutere e delineare un approccio alla ricerca empirica che sia *per* gli insegnanti e *con* gli insegnanti; che sia a un tempo di coinvolgimento attivo e partecipante, ma anche di vera e propria formazione, con il sostegno di ricercatori esperti.

Se è assolutamente vero che le ipotesi della ricerca educativa non possono che nascere sempre dentro i contesti reali e durante autentiche fasi esplorativo-qualitative di ricerca (Lumbelli, 2006), vero è anche che lo stare nei contesti insieme agli insegnanti sposta e ri-modella via via obiettivi e disegni metodologici, richiamando la necessità di forme di indagine che coinvolgano gli insegnanti più a lungo, in più fasi della ricerca, preoccupandosi anche e soprattutto dei benefici che essi stessi possono ricevere (in termini di crescita professionale) durante il percorso; e non solo al termine.

È così che la ricerca empirica e i processi di sperimentazione didattica, pur sotto la responsabilità metodologica del ricercatore esperto, si contaminano via via di elementi contestuali, socio-relazionali, anche profondamente politici; perché più forte diviene la richiesta della scuola di intervenire per la trasformazione dei contesti, più le scelte metodologiche del ricercatore devono trovare spazio per il confronto, la negoziazione, la revisione condivisa di obiettivi e risultati attesi. Le condizioni di un disegno sperimentale si trovano dunque a doversi modificare, a rallentare, a dare spazio a un pensiero problematico e critico in molteplici momenti del percorso di indagine, assumendo la responsabilità di una ricaduta formativa rilevante sui contesti scolastici e sugli principali attori della scuola, gli insegnanti. Si tratta di dare spazio a modalità di Ricerca-Formazione (Asquini, 2018).

Il termine "Ricerca-Formazione" (R-F) è utilizzato nel dibattito pedagogico italiano pur senza aver acquisito una definizione condivisa tra le varie correnti di ricerca. A livello internazionale, non esiste la possibilità di una sua effettiva traduzione letterale. In inglese potremmo parlare di *Teacher Professional Development Research.*

In questo senso,[5] la R-F è considerata come una vera e propria ricerca empirica che si realizza nel campo della professionalità docente, al fine di promuoverne lo sviluppo. È una ricerca che si svolge all'interno delle istituzioni scolastiche, fortemente ancorata al contesto, dove ricercatori e insegnanti condividono itinerari di ricerca, ma soprattutto scopi di cambiamento istituzionale. Piuttosto che una scelta di metodologia di indagine (ogni R-F è attenta a garantire validità e affidabilità dei dati raccolti, adottando procedure di validazione specifiche in funzione dei disegni - qualitativi, quantitativi o misti - che vengono applicati), è un *modo di porsi* dei ricercatori nei confronti dei contesti scolastici individuati e degli insegnanti che in essi agiscono.

Chi si impegna all'interno di un percorso di R-F ha come primo obiettivo quello di poter incidere efficacemente sullo sviluppo professionale degli insegnanti coinvolti e, allo stesso tempo, del contesto scolastico, esplicitando quadri politico-valoriali che orientino il cambiamento. L'azione di ricerca si svolge a seguito di decisioni deliberate nel gruppo: preferibilmente una comunità interdisciplinare di studiosi in collaborazione con gli insegnanti. La responsabilità metodologica e della ricaduta formativa è prioritariamente dei ricercatori, ma il coinvolgimento dei docenti costituisce un orientamento continuo delle azioni del ricercatore.

Di seguito nell'Approfondimento 8.1 sono presentati gli elementi fondamentali connotativi della Ricerca-Formazione; essi non intendono esaurire la questione complessa dello sviluppo professionale degli insegnanti, bensì aprire altrettante questioni e interrogativi che - in ogni specifico percorso di R-F - devono essere presi in esame e discussi all'interno del gruppo di ricerca.

Approfondimento 8.1

LA RICERCA-FORMAZIONE SECONDO IL CENTRO DI RICERCA CRESPI

Ricerca-Formazione (R-F) significa fare ricerca empirica avvalendosi di metodologie di ricerca differenziate e proponendosi di promuovere la professionalità degli insegnanti (e degli operatori educativi all'interno delle istituzioni scolastiche e dei servizi per l'infanzia) attraverso la costruzione di percorsi comuni di ricerca, in un quadro di collaborazione inter-istituzionale. Possiamo considerarla come una scelta politico-metodologica per fare ricerca in collaborazione con gli insegnanti, per lo sviluppo professionale degli insegnanti e per una effettiva ricaduta dei risultati nella realtà scolastica e formativa. Una scelta del ricercatore che caratterizza, accompagna e sostanzia (nelle sue finalità e nelle procedure applicative) le specificità e il rigore dei vari approcci metodologici della ricerca empirica in campo educativo. Tale opzione si pro-

[5] Sono qui riprese le definizioni di Ricerca-Formazione del Centro di ricerca CRESPI (https://centri.unibo.it/crespi/it).

pone come trasversale ai molteplici e più specifici approcci metodologici di ricerca prescelti dal ricercatore e li accompagna senza snaturarli, adattandosi ed esplicitando le diverse procedure che sono peculiari di quell'approccio e garantendo il necessario rigore metodologico per tutto il percorso comune di ricerca.

Alcuni impegni imprescindibili che connotano la R-F e che costituiscono altrettante questioni aperte per il ricercatore:

1. una esplicitazione chiara della finalità della ricerca in termini di crescita e sviluppo della professionalità degli insegnanti direttamente coinvolti e un'attenzione a documentare e analizzare le ricadute in termini di cambiamento;
2. la creazione di un gruppo di R-F di cui facciano parte ricercatore/i e insegnanti, nel quale vengano chiariti i diversi ruoli del partecipanti e in cui vengano negoziati e chiariti obiettivi e oggetti, scelte valoriali e metodologiche della R-F;
3. la centratura sulle specificità dei contesti - istituzionali e non - in cui si svolge la R-F, che si concretizza in tutte le fasi della ricerca attraverso un'analisi dei vincoli e delle risorse in essi presenti;
4. un confronto continuo e sistematico fra i partecipanti alla ricerca sulla documentazione dei risultati e dei processi messi in atto nei contesti scolastici e in quelli della formazione;
5. l'attenzione alla effettiva ricaduta degli esiti nella scuola, sia per l'innovazione educativa e didattica, sia per la formazione degli insegnanti.

Riepilogo

Il capitolo ha permesso di riflettere sul tema della sperimentazione in ambito educativo e delle questioni teorico-metodologiche connesse all'attuazione di disegni sperimentali dentro le scuole per l'innovazione della didattica. Ha toccato tematiche complesse quali la validità interna di una sperimentazione, la necessità di fasi di tipo esplorativo-qualitativo, l'opportunità di implementare disegni quasi-sperimentali per garantire sostenibilità e adesione ai contesti.

Ha aperto la questione del dialogo tra sperimentazione e ricerca-azione, ripercorrendo alcuni tratti salienti del dibattito sulla ricerca-azione e aprendo riflessioni problematiche su aspetti di validità e sostenibilità di questi disegni di ricerca nella scuola italiana, in particolare in relazione alla più ampia questione della professionalità dell'insegnante come ricercatore.

Ha infine aperto sulla proposta della Ricerca-Formazione come opportunità di fare ricerca oggi dentro le scuole, insieme agli insegnanti, per lo sviluppo della professionalità docente e l'innovazione - metodologicamente supportata e controllata - della didattica.

Riferimenti bibliografici

Adams K.A., Lawrence E.K. (2019). *Research methods, statistics, and applications*, 2nd ed. Thusand Oaks: Sage Publications.

Asquini G. (a cura di) (2018). *La ricerca-formazione. Temi, esperienze, prospettive.* Milano: Angeli.

Barbier R. (2007). *La ricerca-azione.* Roma: Armando.

Becchi E. (1968). *Problemi di sperimentalismo educativo.* Roma: Armando.

Becchi E. (1997). *Sperimentare nella scuola. Storia, problemi, prospettive.* Firenze: La Nuova Italia.

Calonghi L. (1984). *I disegni sperimentali nella ricerca scolastica.* In: Becchi E., Vertecchi B. (1984). *Manuale critico della sperimentazione e della ricerca educativa.* Milano: Angeli.

Campbell D.T., Stanley J.C. (1963). *Experimental and quasi-experimental design for research on teaching.* Houghton: Mifflin.

Ciani A., Ferrari L., Vannini I. (2020). *Progettare e valutare per l'equità e la qualità nella didattica.* Aspetti teorici e indicazioni metodologiche. Milano: Angeli.

Coggi C., Ricchiardi P. (2005). *Progettare la ricerca empirica in educazione.* Roma: Carocci.

Cohen L., Manion L., Morrison K. (2007). Research methods in education. 6th ed. New York: Routledge.

Cook, T. D., & Campbell, D. T. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin.

Craig C.J. (2009). *Teacher research and teacher as researcher.* In: Saha L.J., Dworkin A.G. (eds.), *International handbook of research on teachers and teaching. Part One.* Cham (Svizzera): Springer International.

De Landsheere G. (1979). *Dictionnaire de l'évaluation et de la recherche en éducation.* Paris: Presse Universitaire de France (Puf).

De Landsheere G. (1978). *La formazione degli insegnanti domani.* Roma: Armando Editore.

De Landsheere G. (1985). *La ricerca sperimentale nell'educazione.* Teramo: Giunti & Lisciani.

Dewey J. (1938). *Logica, teoria dell'indagine.* Torino: Einaudi (trad. it. di Aldo Visalberghi, 1949).

Falanga M. Treccani L. (2020). *Manuale di normativa scolastica.* Brescia: Scholé - Editrice Morcelliana.

Fetterman D.M. (2000). *Steps of empowerment evaluation: from California to Cape Town.* In: Stufflebeam D.L. et al. (eds.), *Evaluation models.* Boston, Dordrecht, London: Kluwer Academic Publisher.

Gattullo M. (1984). *Sperimentare e decidere.* In: Becchi E., Vertecchi B. (1984).

Manuale critico della sperimentazione e della ricerca educativa. Milano: Angeli.

GAY L.R., MILLS G.E., AIRASIAN P. (2009). *Educational research. Competencies for analysis and application.* Upper Saddle River (New Jersey): Pearson International.

HABERMAS J. (1986). *Teoria dell'agire comunicativo, I. Razionalità nell'azione e razionalizzazione sociale.* Bologna: Il Mulino.

INTERNATIONAL COUNCIL FOR ADULT EDUCATION (1977). *Status report on the Participation Research Project.* Toronto (Canada).

LANKSHEAR C., KNOBEL M. (2004). *A handbook for teacher research: from design to implementation.* New York: Open University Press.

LEWIN K. (1946). Action research and minority problems. *Journal of Social Issues,* 2(4), 34-46.

LIPARI D. (2002). *Progettazione e valutazione nei processi formativi.* Roma: Edizioni Lavoro.

LOSITO B., POZZO G. (2005). *La Ricerca-Azione. Una strategia per il cambiamento nella scuola.* Roma: Carocci Faber.

LUMBELLI L. (1979). *Una proposta metodologica: l'approccio fenomenologico-sperimentale.* In O. Andreani, *Processi di insegnamento-apprendimento,* (pp. 23-38). Firenze: La Nuova Italia.

LUMBELLI L. (1980). *La ricerca esplorativa in pedagogia. Ricerche pedagogiche.* In AA.VV., *Educazione e qualità della vita,* 59-77, Atti del Convegno di Pedagogia di Parma, 25-27 ottobre 1979.

LUMBELLI L. (1984). *Qualità e quantità nella ricerca empirica in pedagogia.* In Becchi E., Vertecchi B. (a cura di), *Manuale critico della sperimentazione e della ricerca educativa.* (pp. 101-133). Milano: Franco Angeli.

LUMBELLI L. (2006). *Costruzione dell'ipotesi ed astrazione nella Pedagogia sperimentale.* In Bondioli A. (a cura di), *Fare ricerca in pedagogia. Saggi per Egle Becchi,* (pp. 25-60). Milano: FrancoAngeli.

MCKERNAN J. (1991). *Curriculum action research.* London: Kogan Page.

NIGRIS E. (1998). Un nuovo rapporto tra ricerca e innovazione. La ricerca-azione. In: Mantovani S., *La ricerca sul campo in educazione. I metodi qualitativi* (pp. 163-201). Milano: Bruno Mondadori.

PONTECORVO C. (1998). *Qualitativo/quantitativo oppure descrittivo/sperimentale?* Presentazione al Dottorato di Pedagogia sperimentale. Università di Roma La Sapienza.

PONTECORVO C. (1984). *La formazione di insegnanti/ricercatori.* In E. Becchi, B. Vertecchi (a cura di), *Manuale critico della sperimentazione e della ricerca educativa,* (pp. 376-386). Milano: Franco Angeli.

POURTOIS J.P. (1984). *La ricerca-azione in pedagogia.* In E. Becchi, B. Vertecchi. *Manuale critico della sperimentazione e della ricerca educativa* (pp. 134-155). Milano: Franco Angeli.

SORZIO P. (2019). *La ricerca-azione.* In: L. Mortari, L. Ghirotto, *Metodi per la ricerca educativa* (145-162). Roma: Carocci.

STENHOUSE L. (1977).*Dalla scuola del programma alla scuola del curricolo.* Roma: Armando.

VANNINI I. (2012). *Come cambia la cultura degli insegnanti.* Milano: Franco Angeli.

VANNINI I. (2016). Accompagnare la professionalità dell'insegnante per una scuola di qualità. In: IPRASE, *Costruire un nuovo curricolo. Saggi in onore di Ennio Draghicchio* (173-183). Trento: IPRASE - Provincia Autonoma di Trento.

VANNINI I. (2019). L'approccio quantitativo-sperimentale. In: L. Mortari, L. Ghirotto, *Metodi per la ricerca educativa* (187-244). Roma: Carocci.

VERTECCHI B. (1992). *Quantità e qualità nella ricerca didattica.* Firenze: La Nuova Italia.

VIGANÒ R. (1996). *Pedagogia e sperimentazione.* Milano: Vita e Pensiero.

VISALBERGHI A. (1989). Ricerca empirica e sperimentazione. *Scuola e Città,* 5-6, 209-215.

VISALBERGHI A. (1993). La ricerca pedagogica: empirica e no. *Cadmo,* 2, 3-4.

9

Studio di caso e focus group

In questo capitolo parleremo del metodo di ricerca "studio di caso" e della tecnica "focus group" (che può essere utilizzata anche all'interno dello studio di caso), in combinazione con altri strumenti, attraverso tecniche di triangolazione alle quali accenneremo.
Definiremo i concetti su cui si basano i due approcci e spiegheremo in quali casi optare per l'uno o per l'altro. Del focus group, affronteremo il tema della preparazione, conduzione, raccolta, codifica e analisi dei dati, dando consigli pratici riguardo agli aspetti più difficili da gestire. Verranno, inoltre, forniti dei consigli per l'organizzazione di focus group in modalità telematica.
Infine, termineremo dando rilievo alla restituzione dei risultati, tappa fondamentale dell'indagine in qualsiasi disegno di ricerca.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Studio di caso.
- Triangolazione.
- Focus group.
- Restituzione.

"Silenzio e parola? Il signor Palomar spera sempre che il silenzio contenga qualcosa di più di quello che il linguaggio può dire. Ma se il linguaggio fosse davvero il punto d'arrivo a cui tende tutto ciò che esiste? O se tutto ciò che esiste fosse linguaggio, già dal principio dei tempi? Qui il signor Palomar è ripreso dall'angoscia.
Dopo aver ascoltato attentamente il fischio del merlo, egli prova a ripeterlo, più fedelmente che può.
Segue un silenzio perplesso, come se il suo messaggio richiedesse un attento esame; poi echeggia un fischio uguale, che il signor Palomar non sa se sia una risposta a lui, o la prova che il suo fischio è talmente diverso che i merli non ne sono affatto turbati e riprendono il dialogo tra loro come nulla fosse. Continuano a fischiare e a interrogarsi perplessi, lui e i merli."

Italo Calvino, *Palomar*

9.1 Studio di caso

9.1.1 Cos'è lo studio di caso?

Un ricercatore in ambito sociale, psicologico o pedagogico può, basandosi su una lunga tradizione di ricerca, scegliere di concentrarsi su un "caso" ed essere guidato da un'abbondante letteratura che descrive il metodo (Ragin e Becker, 1992; Gerring, 2004; Yin, 2009, 2012).

Sono due le principali correnti epistemologiche che lo hanno ispirato. La prima è proposta da autori come Stake (1995) e Merriam (2009) e si colloca nel paradigma socio-costruttivista o interpretativo. Secondo tale visione, il ricercatore è l'interprete di una realtà multipla e soggettiva e partecipa allo studio che genera una conoscenza strettamente relativa al momento e al contesto in cui si trova (Harrison et al., 2017).

Il secondo approccio, sviluppato tra gli altri da autori come Yin (2009) o Eisenhardt (1989), si inserisce maggiormente nel paradigma post-positivista, che prevede lo sviluppo di un rigoroso protocollo di studio e considera attentamente la validità dei risultati ottenuti e i suoi possibili bias. Tratteremo di questo ultimo approccio, il quale implica un costrutto concettuale preliminare e/o una fase esplorativa, in modo da garantire che tutti gli elementi del caso siano descritti e analizzati adeguatamente.

Di fatto, lo studio di caso è spesso oggetto di critiche riguardo all'impossibilità di generalizzare i risultati ottenuti, in quanto si studiano casi unici e irriproducibili e perché c'è meno controllo sul rischio di mancanza di oggettività del ricercatore. Tuttavia, molti studiosi utilizzano questo metodo con tutta la cura e il rigore scientifico richiesti.

Cosa si studia in un caso?

Secondo Yin (2009), lo *studio di caso* analizza "un fenomeno contemporaneo nel suo contesto di vita reale, soprattutto quando i confini tra fenomeno e contesto non sono chiaramente evidenti" (p. 59). La definizione di Gerring (2004) non è molto diversa quando dice che si tratta di uno "studio approfondito di una singola unità (un fenomeno relativamente circoscritto)". L'idea sottostante queste definizioni è che si deve rendere conto della totalità del caso attraverso una descrizione esaustiva, tracciando un confine attorno a un fenomeno. Si tratta, tuttavia, di un obiettivo chimerico (Dumez, 2010; 2013a).

Esiste anche una definizione, molto diversa, di un ricercatore educativo che sostiene che "i casi sono storie con un messaggio" (Herreid, 1997, p. 92). L'enfasi qui non è più posta sul confine, ma sull'essenza narrativa della nozione di caso, ossia il messaggio.

Infine, Abbott (1988) propone una definizione che concilia quelle precedenti e offre un quadro riflessivo interessante, secondo cui i casi sono delle "realtà sfocate con proprietà complesse definite in modo autonomo [...] impegnate in un continuo dialogo con il loro ambiente, un dialogo di azione e costrizione che chiamiamo trama" (p. 65). In questa definizione, il caso ha delle caratteristiche proprie, un'autonomia, ma non è circoscritto all'interno di confini precisi e dialoga costantemente con il contesto e la narrazione permette la descrizione di questo dialogo.

Lo *studio di caso* è una metodologia di ricerca atta a studiare, nei loro contesti, fenomeni complessi non ancora indagati, oppure ad ampliare conoscenze acquisite su fenomeni già indagati. È un approccio metodologico che mira a raccogliere informazioni, utilizzando una varietà di fonti, su un numero circoscritto di argomenti (persona, evento, gruppo di individui, istituzione o organizzazione), al fine di consentire al ricercatore un'analisi dettagliata e approfondita di come funziona o si comporta il fenomeno studiato e di come interagisce con il contesto (Berg, 2000; Yin, 2009; Stake, 2008). Si raccolgono e si analizzano storie di vita, documenti scritti, biografie, interviste o osservazioni che servono per la decostruzione e la ricostruzione dei complessi fenomeni studiati (Yin, 2009).

Attraverso lo studio di caso si cerca di comprendere un fenomeno - di apprezzarne l'unicità e la complessità, il modo in cui è radicato in un contesto e come interagisce con esso (Stake, 1995, p. 16). Secondo Stake, i casi raramente esistono da soli e, se sono caratterizzati da fenomeni, probabilmente saranno osservabili da qualche altra parte. Tuttavia, l'obiettivo del metodo è la particolarizzazione, non la generalizzazione.

Sebbene non ci sia una definizione unanime di questo metodo (VanWynsberghe e Khan, 2007), tutti gli autori sottolineano l'importanza di definire i confini del

caso, che permettono di delimitare il fenomeno che verrà studiato ("Bounded system") (Creswell, 2007; Merriam, 2009). Per poter parlare di un "caso", infatti, esso deve essere sufficientemente circoscritto nello spazio e nel tempo per determinare il percorso metodologico adeguato a una raccolta di dati dettagliata e approfondita (Creswell, 2013).

Siggelkow (2007) identifica tre usi principali dello studio di caso: la motivazione a studiare un importante quesito di ricerca alla luce di un caso particolarmente interessante; l'ispirazione di nuove idee generate dall'immersione in un caso ricco di sfaccettature, attraverso un approccio induttivo; e, infine, l'illustrazione di una teoria, a partire da un caso, che potrebbe condurre a nuove intuizioni.

L'obiettivo principale dello studio di caso è cercare di comprendere molte delle variabili che influenzano la situazione concreta studiata e come interagiscono tra loro. Sebbene questa metodologia non consenta di stabilire relazioni causali (quando non accompagnata dall'uso di strumenti di ricerca ti tipo quantitativo), presenta i seguenti vantaggi:

- aiuta a definire delle prime ipotesi per facilitare indagini future;
- permette di studiare in profondità un fenomeno, per trarne informazioni ed eventualmente proporre la formulazione di una teoria;
- consente di stimare la plausibilità di una teoria (Eckstein, 1973) o, all'opposto, proporre la sua confutazione (Koenig, 2009).

Secondo Yin (2009), l'uso del metodo è legittimo quando sono soddisfatte determinate condizioni:

- lo studio deve rispondere a domande di ricerca sul "cosa?", "come?" e "perché?";
- il ricercatore non può orientare o condizionare il comportamento dei soggetti della ricerca;
- il ricercatore si interessa ai fattori contestuali del fenomeno studiato;
- i limiti tra il fenomeno studiato e il suo contesto non sono chiari.

Questa forma di ricerca può essere particolarmente adatta a un insegnante-ricercatore, in quanto consente di studiare un fenomeno in modo più approfondito rispetto a quanto consentito dall'indagine su larga scala o dalla ricerca sperimentale (Dufour, Fortin e Hamel, 1991).

9.1.2 Quali sono le diverse categorie di studio di caso?

Esistono diverse categorie di studio di caso, la cui scelta è guidata dalla domanda di ricerca e dall'ambito dello studio. Le tipologie più utilizzate sono probabilmente quelle di Yin (2009) e Stake (1995), riassunte nella Tabella 9.1.

Tabella 9.1 *-Le diverse categorie di Studi di caso*

Yin (2009)		
Explanatory	**Descriptive**	**Exploratory**
Affronta una problematica finalizzata alla spiegazione di presunti nessi causali tra fenomeni complessi	Viene utilizzato per descrivere un fenomeno e il suo contesto	Esplora un fenomeno di interesse del ricercatore che mira alla scoperta di nuovi nessi causali e/o risultati.
Stake (1995)		
Intrinsic	**Instrumental**	**Collective**
Mira a una migliore comprensione del fenomeno rappresentato da un caso, considerato particolare o unico dal ricercatore, senza cercare, a priori, di generalizzare i risultati	Ha per obiettivo di risolvere un problema o contribuire alla teoria. Il caso è, qui, un mezzo che serve soprattutto a perseguire un interesse esterno al caso stesso.	Un approccio che prevede lo studio di più casi contemporaneamente o in sequenza, al fine di produrre risultati generalizzabili a una popolazione più ampia.

9.1.3 Organizzazione del lavoro

La prima cosa da fare è trovare un caso che sia rilevante per il ricercatore. Il caso scelto deve essere definito nel modo più accurato possibile, in relazione alla sua natura e al suo perimetro temporale (inizio e fine del caso) e spaziale (area geografica, organizzazione, gruppo sociale ecc.).

Uno studio di caso, contrariamente a quanto suggerisce l'uso del singolare, consiste in un processo di confronto sistematico, ed è dunque sempre "multiplo". È per questo motivo che Dumez (2013b) sostiene che la comune distinzione tra casi singoli (con una sola unità di analisi), olistici (che formano un'unità), multipli (con più casi) e incorporati (con delle sotto-unità di analisi), non è né adeguata né utile. Secondo lo studioso, le tre *domande fondamentali* da porsi all'inizio di uno studio di caso, sono:

1. *Di cosa tratta il mio caso?* Si tratta di identificare la categoria (oppure le diverse categorie) in cui può essere incluso il caso studiato e, quindi, confrontare questo caso con altri simili. Il caso viene definito come l'esemplificazione (l'istanziazione) di una classe di fenomeni (George & Bennett, 2005, pp. 17-18).
2. *Di cosa il mio caso è fatto?* Ogni caso è multiplo e incorporato (embedded multiple case study), poiché è composto da altri casi scoperti sul percorso di indagine, sui quali può essere necessario lavorare.
3. *Cosa fa il mio caso?* In assenza di un inquadramento teorico di partenza, il caso può consentire l'elaborazione di un'ipotesi (Lijphart, 1971) capace di porsi come punto di partenza per definizioni e teorie nuove. Quando, invece, lo studio si basa su un quadro concettuale ben definito, il ricercatore può essere confrontato a un caso deviante, che si discosta, cioè, dalla teoria, permettendo di esaminarla, arricchirla o, eventualmente, di rimetterla in discussione per formularne un'altra. Mentre gli statistici tendono a escludere i casi devianti, lo Studio di caso consente di prenderli in considerazione e farne la base di un ragionamento originale.[1]

Una volta selezionato e definito il caso (o i casi) (1ª fase), è necessario formulare una serie di domande che motivano lo studio e l'uso del metodo, rispondendo alla domanda: "Cosa voglio verificare o analizzare con lo 'Studio di caso'?". Insieme alle domande di ricerca, devono essere definite le ipotesi e teorie rivali (2ª fase), prima di determinare quali metodi di raccolta (3ª fase) e di analisi (4ª fase) verranno utilizzati per documentarli adeguatamente. La scelta, qui, è molto ampia: analisi di documenti, osservazione, interviste, questionari, analisi clinica, tematica ecc. Dopodiché si procede allo studio del caso (5ª fase), seguito dalla raccolta e dall'analisi dei dati. Poiché questo metodo non consente una spiegazione causale, l'analisi dei dati si concentrerà sul confronto delle domande e delle ipotesi iniziali con i dati raccolti. A quel punto, il ricercatore può decidere se ritiene che i dati ottenuti possano essere estrapolati ed estesi ad altre situazioni o meno, e indicare i possibili percorsi di ricerca per approfondire il fenomeno studiato. L'analisi dei dati porta all'interpretazione dei risultati e all'evidenziazione degli aspetti più distintivi e/o più istruttivi del caso (6ª fase).

Il volume del materiale raccolto è generalmente considerevole nello Studio di caso e, come osserva Eisenhardt (1989), "c'è il costante pericolo di essere 'soffocati dai propri dati'". Si suggerisce di seguire le pratiche per la raccolta e l'analisi dei dati raccomandate da Yin (2009) e Stake (1995, 2010).

Lo studio termina con la descrizione delle implicazioni teoriche e/o pratiche dei risultati (7ª fase). Nel report finale, il ricercatore è tenuto a descrivere l'insieme del processo di ricerca e a comunicare quanto appreso dal caso, le sue conclusioni e la loro validità.

[1] Hervé Dumez (2013) sostiene l'importanza maggiore dell'abduzione rispetto ai casi devianti. Ricorda come il legame tra approccio abduttivo e studio di caso sia stato sottolineato da Koenig (1993), David (2000), Dubois e Gadde (2002). Tuttavia, è importante tenere in mente che le idee create dall'abduzione da un singolo caso non hanno una reale validità scientifica, anche quando sono consolidate da un processo di triangolazione (Jick, 1979; Flick, 2009, p. 444). È solo attraverso la verifica, tramite l'uso di metodi quantitativi, che queste idee acquisiranno o meno una reale validità scientifica.

9.1.4 La triangolazione

Generalmente, nello Studio di caso si utilizza la "triangolazione" di diversi metodi e fonti di dati, attraverso cui si confrontano misurazioni indipendenti dello stesso fenomeno, mostrando che portano allo stesso risultato, o almeno non si contraddicono (Miles e Huberman, 2003, p. 480). L'obiettivo è di raggiungere una maggiore comprensione del fenomeno, aggiungere validità allo studio e pervenire a una interpretazione coerente dei dati (Patton, 2002; Creswell, 2009; Yin, 2009; Stake, 2010).

All'interno delle forme di triangolazione, quella maggiormente utilizzata è legata alla diversificazione degli strumenti (Savoie-Zajc, 2019).

I ricercatori seguono disegni di ricerca che combinano due o più metodi di raccolta dati, in cerca di complementarità e di conferma, per ottenere un'analisi più fine e un'interpretazione più affidabile. I metodi più frequentemente abbinati sono l'intervista individuale con l'osservazione, il focus group con l'intervista individuale oppure l'analisi documentaria in seguito all'intervista individuale o al focus group.

In effetti, il focus group è un metodo generalmente utilizzato in una prospettiva di triangolazione (Barbour e Kitzinger, 1999; Merton, Fiske e Kendall, 1956), in aggiunta ad altri metodi qualitativi, come le interviste individuali, l'osservazione o i questionari.

Tipologie di triangolazione

Triangolazione teorica, consistente nell'utilizzare più quadri teorici o punti di vista disciplinari per approcciare i dati con prospettive e ipotesi diverse.

Triangolazione degli strumenti di raccolta in cui vengono utilizzati più strumenti che portano, ciascuno, a una specifica ricostruzione dell'oggetto di studio e vengono utilizzati con una particolare prospettiva teorica.

Triangolazione dei ricercatori, dove vengono adottati i punti di vista di più ricercatori per arricchire prospettive e chiavi di interpretazione (Bourgeois-Guérin e Beaudoin, 2016).

Triangolazione delle fonti, quando i dati vengono raccolti da diverse fonti/informatori per confrontare varie testimonianze e interpretazioni.

Triangolazione "ecologica", dove le analisi e le interpretazioni sono soggette a verifica da parte dei partecipanti alla ricerca, al fine di convalidare le interpretazioni ed, eventualmente, raccogliere nuovi dati.

Approfondimento 9.1
TRIANGOLAZIONE E FOCUS GROUP

Due abbinamenti sono molto frequenti, quello del focus group (FG) con interviste individuali e quello del FG con metodi di tipo quantitativo. Nel primo caso, l'uso di interviste individuali fornisce informazioni che completano le dimensioni sociali e collettive, esplorate attraverso il FG, e permette di raccogliere informazioni molto approfondite a livello individuale. Nel secondo caso, i FG vengono, generalmente, condotti nella fase esplorativa della ricerca per affinare le ipotesi di ricerca, identificare le variabili da analizzare e preparare la costruzione (o scelta) del questionario e delle sue dimensioni. Inoltre, aiuta a formulare le domande secondo un approccio linguistico e sintattico coerente con quello utilizzato dalla popolazione in esame.
A volte, il FG può essere utilizzato dopo aver impiegato metodi standardizzati per approfondire un aspetto emerso nel corso della ricerca. Secondo Fern (2001, p. 142-143) e Morrison (1998, p. 159), il FG permette di "decristallizzare" il rigido abbinamento, nella formulazione delle domande di un questionario standardizzato, tra un termine e il suo significato e può fornire informazioni più sfumate e articolate.

9.2 Focus group

9.2.1 Cos'è il Focus group?

La tecnica del Focus group (FG) è "un modo per raccogliere dati qualitativi, che - essenzialmente - implica il coinvolgimento di un piccolo numero di persone in una discussione informale di gruppo, "focalizzata" su un particolare argomento o insieme di questioni" (Wilkinson, 2004, p. 177). È una tecnica di raccolta delle informazioni che emergono da una discussione di gruppo e dalle interazioni che si producono tra i partecipanti, intorno a un numero circoscritto di domande.

Il FG ricrea, in miniatura, un ambiente sociale temporaneo per i partecipanti che "crea una dinamica di gruppo in cui le dichiarazioni fatte da un individuo possono generare reazioni e portare altri partecipanti a partecipare alla discussione. Gli argomenti presentati a favore o contro un punto di vista possono aiutare alcuni partecipanti a formarsi un'opinione su un argomento per il quale non mostravano particolare interesse prima. Come nella società, i partecipanti a volte cambiano idea quando ascoltano le parole di altri partecipanti" (Geoffrion, 1993, p. 312). Tuttavia, il FG è un gruppo artificiale che è stato convocato da una persona, in condizioni organizzate per i propri fini, che non può, di conseguenza, essere assimilato a una conversazione informale. Se, a volte, viene considerato come "una società pensante in miniatura" (Farr e Tafoya, 1992, citato in Markova, 2003), ne differisce, tuttavia, per il fatto di essere organizzato in base a criteri prestabiliti che riguardano il reclutamento dei partecipanti, gli argomenti da discutere e le domande aperte da

porre secondo una determinata logica.

Il FG, a prima vista, può sembrare relativamente facile da realizzare, ma in realtà, per realizzare un lavoro rigoroso richiede tante accortezze.

Com'è nato il Focus group?

È nato dalle tecniche di marketing del dopoguerra negli Stati Uniti, per raccogliere le aspettative dei consumatori e quindi rendere più attraenti i prodotti destinati alla vendita. Si ispira alle tecniche di dinamica di gruppo utilizzate da Carl Rogers, leader della corrente psicologica umanista. È stato adoperato negli anni '80 dalla ricerca universitaria in vari campi (istruzione, sanità pubblica, ambiente, scienze sociali).

9.2.2 Quando utilizzare il Focus group?

Sono diverse le situazioni in cui il FG può risultare utile. In particolare, quando si cerca di:

- farsi un'idea complessiva della materia studiata;
- identificare le caratteristiche di un problema e i fattori che concorrono alla sua comparsa;
- ottenere un ritratto della realtà studiata, così come viene vissuta;
- studiare la varietà di opinioni e sentimenti dei partecipanti su un determinato argomento e evidenziare le differenze di prospettiva tra gruppi di individui (insegnanti vs. studenti, dirigenti scolastici vs. insegnanti ecc.);
- analizzare i fattori associati alle opinioni, alle motivazioni e ai comportamenti dei partecipanti;
- studiare le aspettative, le esigenze e i bisogni di una particolare popolazione, al fine di formulare interventi e/o migliorarne la qualità.

Più specificamente, all'interno del disegno di ricerca, è utile per:

- raccogliere informazioni utili alla definizione del disegno di ricerca o all'interpretazione dei risultati;
- identificare le aree di ricerca da approfondire attraverso la ricerca, in modo da permettere la triangolazione dei metodi di raccolta dei dati;
- ottenere una serie di idee su un argomento, in vista dello sviluppo di strumenti standardizzati;
- costruire studi paralleli sullo stesso tema, ma con campioni diversi (insegnanti, alunni, studenti di diverse fasce di età, ecc.).

Il FG non può avere come scopo, invece, il costruire consenso all'interno del grup-

po dei partecipanti o di educare i partecipanti, e non può essere utilizzato per raccogliere informazioni sensibili (che potrebbero recare danno alla persona che lo condivide in gruppo) o relative a un argomento contrastante su cui le opinioni si radicalizzano. Tuttavia, può offrire, se gestito adeguatamente, un ambiente favorevole all'espressione e alla discussione di opinioni diverse.

Inoltre, non si può evincere dai FG la frequenza o la distribuzione di credenze o comportamenti all'interno di una popolazione. Essi possono, tuttavia, permettere di identificare delle tendenze e capire meglio le credenze e gli atteggiamenti alla base di determinati comportamenti.

Questo strumento di ricerca non richiede un campione rappresentativo della popolazione di riferimento - che ne abbia, cioè, le stesse caratteristiche e la stessa distribuzione - e può rappresentare per gli insegnanti, formati alla sua conduzione, un'opportunità di studiare alcuni aspetti della loro pratica.

9.2.3 Quali sono le difficoltà e i limiti del focus group?

È importante considerare i rischi della tecnica, e in particolare di alcune delle sue fasi, per prepararsi in modo adeguato e venire incontro alle difficoltà quali, in particolare quelle di seguito elencate.

- La quantità dei dati raccolti può essere tale da renderne difficile l'analisi e l'interpretazione. Inoltre, i contenuti delle discussioni devono essere interpretati in riferimento al particolare contesto sociale creato dal gruppo, alle caratteristiche della sua composizione e all'andamento della riunione.
- Il campione può essere non rappresentativo e, di conseguenza, la selezione dei partecipanti (che sono volontari e quindi motivati e disponibili) può portare a interpretare in modo erroneo la realtà studiata. Di fatto, il campionamento non probabilistico, utilizzato per il FG, non consente di generalizzare i risultati ed estenderli alla popolazione di riferimento.
- La conduzione della riunione, da parte del moderatore/facilitatore (si veda paragrafo 5.2.5), richiede una preparazione adeguata a consentire a tutti i partecipanti di esprimere le proprie opinioni ed evitare i rischi di:
  - prevaricazione di alcuni partecipanti sugli altri;
  - "blocco" della discussione provocato dalla timidezza o riluttanza a esprimere idee personali in pubblico;
  - superficialità dovuta a una tendenza a rimanere nei binari delle idee considerate "corrette" dal punto di vista socio-culturale, ossia socialmente accettabile in una determinata comunità;
  - condizionamenti del pensiero dei partecipanti indotti dalla presenza di un leader carismatico o da persone con mansioni professionali di livello superiore o ritenute depositarie di "potere" nei loro confronti;
  - fraintendimento del moderatore, del ricercatore o di un partecipante, generato dall'interpretazione erronea di un concetto formulato da un membro del gruppo. I significati devono, dunque, essere chiariti e concordati;
  - orientamento della discussione e delle risposte nella direzione desiderata (anche inconsapevolmente) dal moderatore, ossia suggerimento involontario delle opinioni;
  - disagio e imbarazzo dei partecipanti confrontati a parlare di argomenti che rappresentano, per loro, un'eccessiva intrusione nel loro vissuto.

I FG non sono sempre il metodo migliore per incoraggiare una discussione libera e aperta quando il tema affrontato è delicato (Kitzinger e Farquhar, 1999). Tuttavia, non si deve concludere che i FG inibiscano i partecipanti o siano, per forza, inappropriati per determinati argomenti. A volte, avviene l'esatto contrario. Il fatto di trovarsi in gruppo può facilitare una discussione su argomenti tabù, quando i partecipanti meno inibiti conducono gli altri in una dinamica che permette di superare timidezza e diffidenza. La partecipazione congiunta può essere fonte di sostegno reciproco e consentire l'espressione di sentimenti che possono discostarsi dalla norma culturale o il racconto di esperienze stigmatizzanti (bullismo, molestie, violenza sessuale ecc.). In quel caso, le reazioni verbali e non-verbali del gruppo all'introduzione di informazioni "tabù" vanno considerate dimensioni importanti del FG (Kitzinger, Markova e Kalampalikis, 2004).

Queste riflessioni ci portano a parlare dell'organizzazione del FG, cominciando con la presentazione della sua struttura, per proseguire, poi, con la descrizione del suo svolgimento.

9.2.4 Focus group: dimensione, composizione e durata

Il FG è composto dal ricercatore, dal moderatore, da un aiuto-moderatore e dai partecipanti. Il moderatore ha il ruolo complesso di porre le domande, gestire/moderare la conversazione, coinvolgere i partecipanti nell'analisi dei problemi in esame e raccogliere le informazioni utili alla ricerca (Zammuner, 2003, p. 189). Deve aver ricevuto una preparazione in merito, tuttavia, per ragioni di mancanza di fondi, il ricercatore è spesso costretto a ricoprire questa funzione, insieme a quella dell'aiuto-moderatore. Sebbene lo svolgere questa doppia (o tripla) funzione ha di positivo che permette di seguire tutte le fasi della ricerca, Krueger (1994) ha insistito sull'importanza che ci sia un aiuto-moderatore insieme al moderatore.

Generalmente, il moderatore presenta ai partecipanti una serie di domande, tuttavia potrebbe scegliere di presentare un materiale (per esempio, articolo, video o audio) che stimoli la conversazione. In alternativa ancora, il moderatore potrebbe chiedere ai membri di impegnarsi in un'attività specifica (ad esempio, esercizio di

team building o di brainstorming) per favorire e raccogliere l'espressione di opinioni relative a una data situazione. L'aiuto-moderatore si occupa di registrare (in modalità audio e/o video) la sessione e aiuta il moderatore a prendere appunti, creare un ambiente favorevole alla discussione di gruppo (per esempio, accogliere i ritardatari, assicurarsi che tutti abbiano un posto a sedere, organizzare un rinfresco) e, eventualmente, di partecipare all'analisi e all'interpretazione dei dati raccolti (Krueger e Casey, 2000).

Nel FG non direttivo (in cui viene lasciata ampia libertà di espressione ai membri del gruppo), la traccia delle domande è semi-strutturata. Nel rispondere alle domande, i partecipanti hanno, dunque, la possibilità di esprimersi liberamente e il moderatore interviene solo per porre domande, stimolare, canalizzare, re-orientare, approfondire i discorsi e distribuire il tempo di parola.

La discussione raggruppa da 6 a 12 partecipanti (Baumgartner, Strong e Hensley, 2002; Bernard, 1995; Johnson e Christensen, 2004; Krueger e Casey, 2000; Langford, Schoenfeld e Izzo, 2002; Morgan, 1997; Onwuegbuzie, Jiao e Bostick, 2004) e dura normalmente da 1 a 2 ore (Morgan, 1997; Vaughn et al., 1996) al fine di garantire l'intervento di tutti.

La discussione in un piccolo gruppo può avviarsi più facilmente, tuttavia un numero di partecipanti inferiore a 6 può limitare la diversità delle informazioni raccolte. All'opposto, formare un gruppo di più di 12 persone potrebbe inibire i partecipanti, rendere difficile l'ascolto dei diversi punti di vista e il coinvolgimento di tutti nella conversazione. In generale, i grandi gruppi sono più difficili da guidare: richiedono più direttività da parte del facilitatore che deve controllare la rotazione dei turni di parola.

Sono riportati in letteratura l'uso di "mini-Focus group" (Krueger, 1994, p. 17), che includono tre (Morgan, 1997) o quattro (Krueger, 1994) partecipanti, generalmente con un alto e specifico livello di competenza (Hague, 2002).

La selezione delle persone che partecipano ai colloqui di gruppo deve contribuire a facilitare una conversazione fluida all'interno del gruppo, motivo per cui la maggior parte dei ricercatori raccomanda l'omogeneità per favorire scambi produttivi attraverso la condivisione di opinioni ed esperienze analoghe (Babour e Kitzinger, 1999; Duchesne e Haegel, 2005). Tuttavia, è importante che non siano esageratamente simili tra loro, in modo da ottenere prospettive diverse all'interno di un gruppo e, quindi, una visione più ampia e articolata (El Hadj, 2010).

Il reclutamento dei partecipanti

È un'operazione delicata perché, non solo si costruisce un campione secondo criteri che avranno un peso sulle analisi, ma anche un gruppo le cui interazioni saranno decisive per la qualità del materiale raccolto. Certo, le relazioni che si svilupperanno all'interno del gruppo sono sempre in parte imprevedibili, di conseguenza un reclutamento che consentirebbe di anticipare e controllare tutto ciò che potrebbe avere luogo durante la discussione è illusorio (Bloor et al., 2001). Tuttavia, questo margine di imprevedibilità non deve esimere il ricercatore dal prestare molta attenzione ad alcuni parametri essenziali come il modo in cui si entra in contatto con i partecipanti, il loro grado di conoscenza reciproca, la dimensione e l'omogeneità dei gruppi.

Inoltre, è importante esaminare l'asimmetria delle relazioni sociali, ossia le gerarchie all'interno del gruppo studiato, in quanto queste differenziazioni sociali saranno riprodotte nel FG (differenze di genere, differenze di status ecc.) e avranno un'influenza sullo svolgimento degli scambi e, quindi, sui dati raccolti (Kitzinger et al., 2004). Un individuo che fa parte della stessa "comunità interpretativa" degli altri membri del suo gruppo, ossia che interpreta il discorso sociale allo stesso modo e usa gli stessi "codici" degli altri (Breton e Proulx, 2002, p. 231), si sente a proprio agio e si esprime più liberamente, ma lascia impliciti gli aspetti del suo pensiero connessi alla sfera di significati condivisi. In quel caso, il ricercatore deve essere consapevole del fatto che, sebbene la conversazione sia più spontanea e agevole, alcuni assunti rischiano di non essere esplicitati (Breton e Proulx, 2002, pp. 231-232).

L'obiettivo è quello di costruire gruppi che permettano di cogliere situazioni differenziate rispetto al tema in esame, non è di riprodurre, in miniatura, la struttura della popolazione di riferimento. Il principio è quello della massima diversificazione dell'unità di analisi, sulla base di pochi criteri particolarmente rilevanti (Duchesne e Haegel, 2004). Ogni gruppo dovrebbe essere formato in funzione di due criteri: le caratteristiche sociali (socioprofessionali, etnici, di età e genere) e la caratteristica comune che si ritiene determinante per la comprensione dell'argomento (in genere riguarda un tipo di esperienza e non di opinione).

Spesso i FG coinvolgono delle persone estranee per tutelare l'anonimato. Tuttavia, in alcuni casi, la scelta della popolazione target richiede di prendere persone che si conoscono. In questo caso le dinamiche di gruppo sono più facili da instaurare, ma è difficile garantire la riservatezza delle informazioni in quanto i partecipanti non sono vincolati dalle stesse regole di privacy del ricercatore.

È possibile organizzare più FG sullo stesso tema se è necessaria un'ulteriore raccolta di dati. Lo svolgimento di più FG ha, infatti, per obiettivo di valutare in che misura è stata raggiunta la saturazione dei dati (Flick, 1998; Lincoln e Guba, 1985; Morse, 1995; Strauss e Corbin, 1990), ossia quando i contenuti sono ricorrenti da un'intervista all'altra per cui la raccolta di più dati non può portare a nessuna novità (Sandelowski, 2008; Saumure e Given, 2008), o la saturazione teorica, ossia quando le informazioni raccolte non modificano le spiegazioni e i concetti già elaborati (Sandelowski, 2008).

Più la discussione è strutturata, minore è la necessità di moltiplicare il numero di FG; al contrario, meno è strutturata, maggiore è la variabilità delle dinamiche

all'interno dei gruppi e più numerosi devono essere i FG (in modo da evitare che le differenze osservate siano dovute soltanto alle differenze nelle dinamiche di gruppo).

9.2.5 La scelta del luogo di svolgimento del focus group e i focus group virtuali

L'organizzazione del FG richiede, prima di tutto, un luogo adatto per un incontro di 5-10 persone, che sia piacevole, spazioso, tranquillo (di certo, non deve essere un luogo di passaggio), neutro (che non influenzi i partecipanti e non li metta a disagio) e di facile accesso. In alcuni studi (Stewart e Shamdasani, 1990) è stato osservato come ambienti più piccoli potessero facilitare le interazioni e la concentrazione.

Oltre all'accoglienza, sono importanti l'ambiente fisico e l'allestimento dello spazio, perché influenzano il comportamento. La disposizione circolare, ad esempio, promuove la partecipazione e pone tutti i partecipanti su un piano di parità: il fatto che tutti possano vedersi favorisce le interazioni.

A volte, non risulta possibile far spostare le persone, e dato gli ultimi sviluppi in termini di lavoro a distanza, è importante parlare dei FG organizzati in modalità virtuale. Nel caso di un *FG telematico*, i partecipanti possono iscriversi e firmare il consenso (a partecipare ed essere registrati) compilando un Google Form appositamente redatto. Il ricercatore, fino all'ultimo momento, dovrà ricordare l'appuntamento ai partecipanti (via sms, mail, chiamate) così da avere meno defezioni possibili. Può sempre, comunque, prevedere una lista alternativa di persone in caso di disdetta.

Prima della riunione, il materiale di presentazione della ricerca e di spiegazione dello svolgimento del FG deve essere spedito ai partecipanti. Con questo materiale informativo, o nel Form, è necessario assicurarsi che tutti i partecipanti abbiano accesso a un dispositivo (telefono, computer, tablet ecc.) con funzionalità wireless, chiedere loro di essere presenti prima dell'inizio della riunione per risolvere eventuali problemi tecnici e di scegliere uno spazio in cui non saranno disturbati e dove potranno parlare liberamente.

All'inizio della riunione, si chiede a tutti di accendere la propria videocamera e di spegnere il microfono quando non parlano. Per facilitare lo scambio, può essere introdotta la proposta di usare la funzione "alzare la mano" per partecipare alla discussione. Può risultare più difficile ottenere la partecipazione online di tutti ed è, dunque, essenziale stimolare chi rimane in silenzio. Possono essere usate frasi come, ad esempio: "Per coloro che non hanno parlato molto, cosa ne pensate... ?". Infine, è bene ricordare che le piattaforme per videoconferenze offrono la possibilità di registrare le riunioni, facilitando così il lavoro del moderatore e del suo aiutante.

Entriamo, ora, nel merito della preparazione delle domande e dello svolgimento del FG.

9.2.6 Preparazione della guida all'intervista

Prima di formulare con precisione i temi da affrontare, è necessario cominciare con una fase di ricerca bibliografica attinente alle nostre domande di ricerca, aprendosi alle teorie e ai concetti rilevanti che sono stati utilizzati dai ricercatori precedenti, così da definire le domande da formulare durante il FG. Le domande sono vincolate agli obiettivi ed è, dunque, essenziale verificare che siano definiti in modo molto chiaro e circoscritto. In questa ottica, può essere di aiuto cercare di rispondere alle seguenti domande (Krueger, 1994, p. 43):

- "Quali sono le ragioni per cui ho deciso di realizzare questo studio?";
- "Quali sono le informazioni importanti che voglio ottenere?";
- "A chi saranno utili/chi utilizzerà queste informazioni?";
- "Come dovranno essere usate queste informazioni?

Ricerca bibliografica

Lo studio della letteratura ci consente di ampliare la nostra conoscenza dell'argomento scelto, raccogliendo informazioni sui concetti e su quali metodi di indagine sono stati utilizzati per rispondere alle domande di ricerca. Inoltre, ci permette di scoprire quali domande sono state poste in studi simili al nostro, quali informazioni sono state raccolte e come sono state categorizzate/organizzate in unità di informazione.

In Tabella 9.2 sono presentate alcune tappe del lavoro da svolgere. La ricerca bibliografica è utile, in particolar modo, nelle quattro prime tappe e nell'ultima per un confronto tra i risultati ottenuti e la letteratura più aggiornata.

Tabella 9.2 Tappe dell'organizzazione di un Focus group (FG)

1	Identificare il problema e gli obiettivi della ricerca
2	Specificare gli obiettivi del FG
3	Fare un elenco delle domande da discutere con i membri del FG
4	Elaborare la guida per lo svolgimento del FG (Topic guide/Questioning route)
5	Procedere al reclutamento dei partecipanti
6	Condurre i Focus group
7	Analizzare i dati raccolti durante gli incontri
8	Presentare le conclusioni e tracciare le linee guida per uno studio successivo o per proporre interventi

Per guidare la conversazione, è necessario preparare in anticipo un elenco completo di domande relative ai problemi di ricerca, il *Questioning route* (griglia strutturata), che sarà utilizzato per la conduzione del FG (non parleremo del Topic guide che è molto più generico, in quanto riassume soltanto i temi chiave da trattare).

Le domande dovranno seguire una logica cha andrà dal più generale al più specifico, dal meno importante al più rilevante per il problema studiato, con un inizio facile e una parte finale che possano permettere ai partecipanti di aggiungere osservazioni.

Questioning route

È la lista ordinata delle domande da formulare durante il FG, favorisce la coerenza del discorso e, di conseguenza, aiuta l'analisi delle informazioni raccolte. Per costruirlo, il moderatore e il suo aiutante sono costretti a pensare in anticipo alle frasi e alle parole da utilizzare e prevedere quali discorsi potrebbero generare ognuna di queste domande.

Il moderatore deve utilizzare questa guida con flessibilità per consentire alla conversazione di svilupparsi al ritmo dei partecipanti, sebbene i suoi contenuti debbano essere sufficientemente strutturati da assicurare uno scambio sugli specifici temi della ricerca. La guida non deve contenere più di 5-6 domande e le risposte dovranno essere ottenute entro un tempo che non superi le due ore. Se le domande di ricerca sono numerose, è preferibile organizzare più FG e concentrarsi su poche domande in ognuno di essi.

Per elaborare le domande, si seguono diverse fasi: il brainstorming, la definizione delle domande e del loro ordine, la stima del tempo di discussione per ognuna di esse, la richiesta di feedback di esperti della materia sulla loro formulazione e la verifica della comprensione e dell'esaustività delle domande, attraverso un FG di prova con persone corrispondenti alla tipologia dei partecipanti. In seguito a questa verifica, converrà modificare o eliminare le domande che non sono state comprese o che non hanno stimolato l'interazione tra i partecipanti.

Le domande non devono indirizzare la risposta, devono essere formulate in modo aperto senza risultare ambigue. È importante utilizzare lo stesso lessico degli interlocutori, formulare frasi facili da capire, univoche (focalizzate su un singolo argomento) e preferibilmente brevi. Sono da evitare le domande che inducono risposte binarie, come "sì/no", "vero/falso", "questo/quello" ecc.

Esempio 9.1

ELENCO DI DOMANDE RELATIVE ALLA FORMAZIONE DEGLI INSEGNANTI, PER UN FG RIUNENDO INSEGNANTI CHE ESERCITANO DA MENO DI CINQUE ANNI

Seppur si tratta di informazioni già in possesso del ricercatore, per iniziare potrebbero essere utili delle domande che rompano il ghiaccio tra i partecipanti. Del tipo:

Da quanti anni lavorate come insegnanti e da quanti anni insegnate nella vostra istituzione attuale?
Che tipo di incarico avete?
Cosa insegnate in questo anno scolastico e per quale livello?

Mi piacerebbe avere il vostro punto di vista sulla formazione che avete ricevuto, se avete la sensazione che ci siano delle cose che vi sono mancate quando avete iniziato o, al contrario, delle cose che vi sono state molto utili. In pratica, un parere un po' in generale sulla vostra formazione...

(A tal proposito,) vorrei sapere se avete delle opportunità di seguire delle formazioni durante l'anno e, se sì, se corrispondono ai vostri bisogni e se siete nelle condizioni pratiche di poter seguirle.

Ad esempio, quali tipi di formazione avete ricevuto/vorreste ricevere?
(Se si sono presentate delle difficoltà pratiche) quali difficoltà avete incontrato specificamente?

(A questo riguardo/avevo giustamente una domanda relativa a questo...) Quando avete iniziato, avete ricevuto un accompagnamento, un aiuto, siete state/i seguiti/e? E in tal caso, come valutereste questa assistenza? (in funzione delle risposte, questa domanda potrebbe essere posta prima.)

Riassumendo, di tutte le cose di cui abbiamo discusso, qual è per voi la più importante?

(Dopo aver riassunto quanto detto) Vi sembra un riassunto adeguato di ciò che è stato detto?

(Dopo avere ricordato lo scopo della ricerca) Secondo voi, rispetto a queste domande siamo passati accanto a qualcosa di importante?

9.2.7 Prepararsi per l'interazione con i partecipanti

È importante preparare una nota introduttiva che spieghi il motivo e l'obiettivo dell'incontro. Inoltre, deve essere spiegata la necessità tecnica di registrare i dibattiti, che è, tuttavia, vincolata alle norme sul rispetto della privacy e dell'anonimato dei partecipanti, così come lo sono tutti i dati raccolti durante la riunione. Queste informazioni saranno ribadite all'inizio della riunione, ma dovranno essere state trasmesse in anticipo a ciascun partecipante. A volte, la complessità dei temi da discutere richiede di preparare i partecipanti, inviando loro una sintesi degli argomenti da discutere.

Prima di iniziare a condurre un FG, e come in ogni ricerca, è indispensabile essere consapevoli delle proprie aspettative, che avranno un'influenza sul modo di guidare la conversazione e di interagire con le persone. Solo essendone a conoscenza si può cercare di evitare di condurre l'incontro in modo parziale (cercando di evidenziare alcune informazioni) e di interpretare in modo falsato quanto osservato e sentito.

Ciò premesso, ci sono specifici atteggiamenti e comportamenti che aiutano la conduzione e il superamento delle difficoltà che possono emergere in seno al gruppo. Il ruolo del moderatore è cruciale ed è più complesso di quello di colui che intervista una sola persona alla volta. Ha la responsabilità di guidare la conversazione senza controllarne eccessivamente l'andamento e senza influenzare le risposte alle sue domande. Non deve mai lasciare trasparire il proprio punto di vista né emettere pareri o giudizi su ciò che viene condiviso. Al contrario, deve mettere a proprio agio i partecipanti (e su questo aspetto i primi 10 minuti possono essere decisivi), facilitare le dinamiche di gruppo, incoraggiare la discussione, richiedere ai membri eccessivamente loquaci di lasciar parlare gli altri o spingere alcuni membri taciturni a intervenire. Deve decidere quando adottare un atteggiamento più distaccato o, invece, più interventista per permettere, per esempio, una prosecuzione del dibattito che si sarebbe fermato senza il suo intervento, per sottolineare le incoerenze emerse dai discorsi o insistere sui disaccordi o, ancora, per incoraggiare i partecipanti a chiarire il loro punto di vista e argomentare la loro opinione.

Una regola importante è anche di non dire mai ai partecipanti che hanno torto o ragione e di non sbilanciarsi, facendo capire che siamo più d'accordo con una persona anziché con un'altra. Quando facciamo un commento, deve servire a stimolare e approfondire la conversazione e le interazioni. Inoltre, non è buona pratica insistere nel chiedere a qualcuno il suo punto di vista, in quanto una persona potrebbe non avere opinione rispetto a una specifica questione.

È importante pensare a definire i termini quando potrebbero essere interpretati in modi diversi (per esempio: un clima di classe piacevole, ossia in cui c'è una buona intesa tra gli alunni). Anzi, quando è possibile, è preferibile utilizzare la definizione che i partecipanti hanno di un termine, piuttosto che la propria. Mentre, nel riassumere quanto è stato detto dai membri del gruppo, è bene usare parole diverse da quelle che hanno usato per verificare di aver inteso nel modo corretto ciò che hanno condiviso.

Approfondimento 9.2

Alcune persone possono richiedere interventi specifici per non inficiare il dibattito o per mantenere un'atmosfera gradevole.

Una persona maldisposta o addirittura ostile potrebbe mettere a disagio gli altri e impedire una condivisione fluida. In questo caso, si deve evitare di stimolarla, cercando, per esempio, di non incrociarne lo sguardo o, se è necessario, chiederle di adottare un atteggiamento di rispetto nei confronti degli altri partecipanti.

Una persona esperta, o che si ritiene tale, potrebbe intimidire e inibire gli altri. In questo caso è dunque importante insistere sul fatto che si è interessati al parere di ciascuno e che tutte le opinioni hanno valore. E se prende uno spazio eccessivo, la si può interrompere gentilmente per permettere a tutti di intervenire.

Una persona timida tende a non intervenire ed è importante valorizzare ogni suo contributo, metterla a proprio agio e, se possibile, assegnarle un posto di fronte al moderatore, in modo che sia maggiormente coinvolta. Ai lati, è preferibile mettere chi parla molto. Al contrario, un interlocutore potrebbe essere logorroico o parlare senza rispondere alle domande. In questi casi, si può distogliere volutamente l'attenzione al primo e, al secondo, riproporre la domanda.

Quando una persona si esprime in modo troppo generico, si può usare la tecnica della caricatura, esagerando o estremizzando quanto detto, in modo da spronare l'interlocutore a precisare il suo pensiero ed esprimerne le sfumature.

Se un partecipante fuorvia la domanda ed esce dal tema, "contaminando" il gruppo che si mette a parlare di un altro argomento, si deve trovare il modo di intervenire garbatamente, di riassumere i diversi punti di vista e riproporre la domanda per orientare il discorso nella giusta direzione.

Attenzione ai piccoli gruppi di persone che si conoscono e potrebbero creare sottogruppi, coalizioni e conversazioni che non coinvolgono tutti, ostacolando le interazioni.

Il FG deve permettere di raccogliere le informazioni utili alla ricerca e, pertanto, il moderatore ha come principale funzione il mantenere i dibattiti all'interno dei temi individuati e di assicurare un tempo di esposizione a tutti (può essere utile l'uso di un timer per annotare il tempo concesso a ciascuno). È sicuramente necessario seguire l'elenco delle domande per non dimenticare nessun punto importante, anche se la flessibilità dell'ordine nel quale porle assicura la fluidità della conversazione. Vediamo, ora come procedere alla raccolta dei dati.

9.2.8 Raccolta dei dati

Si può attingere a molte fonti e tipologie di dati all'interno di un FG (Kruger, 1994), ma la maggior parte dei ricercatori effettua le analisi utilizzando solo il testo delle conversazioni trascritte. L'analisi delle trascrizioni audio e video è quella che richiede più tempo e rigore, e vengono studiate insieme agli appunti presi dal moderatore e dall'aiuto-moderatore. L'analisi può anche essere effettuata sulla base di una trascrizione parziale degli elementi considerati più importanti delle registrazioni. Questo approccio può aiutare il ricercatore a concentrarsi sulle domande di ricerca e scegliere le parti dei discorsi utili a una maggiore comprensione del fenomeno in esame. In questo caso, gli appunti sono il materiale principale e le registrazioni servono a riascoltare i passaggi di interesse e citarli. Infine, la raccolta di dati meno rigorosa e più propensa all'errore è quella basata sulla memoria dei fatti e dei contenuti dei FG, ed è, quindi, altamente consigliato l'uso di registrazioni.

I dati ottenuti possono provenire da tre fonti diverse: il singolo individuo, il gruppo e l'interazione di gruppo (Duggleby, 2005). Non c'è accordo tra gli studiosi su quale delle tre unità di analisi sia più appropriata da esaminare, sebbene, la maggior parte di loro scelga il gruppo come unità di analisi (Morgan, 1997). Spesso i ricercatori presentano trascrizioni di estratti della conversazione a sostegno dei temi che sono emersi, senza, tuttavia, descrivere il loro metodo di analisi dei dati in modo accurato. In tal caso, la presentazione dei temi emersi non rende conto, ad esempio, della specificità dei pareri o del grado di adesione o meno alle opinioni, portando così all'emarginazione dei casi atipici e delle "interazioni argomentative" (Kitzinger, 1994) che possono arricchire la lettura dei dati (Sim, 1998). Inoltre, nell'ottica di raggiungere una buona saturazione dei dati e conseguente validità teorica, descrittiva e interpretativa (Maxwell, 2005), è essenziale analizzare e interpretare le informazioni relative ai dissidenti e ai dati anomali.

Se nel FG l'unità di analisi è il gruppo e non l'individuo (Wilkinson, 1998), è importante che l'analisi delle sue dinamiche e il riepilogo dei temi emersi non prenda in considerazione soltanto le tendenze che hanno contribuito alla definizione delle categorie e non faccia passare sotto silenzio le opinioni dei singoli (Crabtree et al., 1993). Considerare il gruppo, più delle unità che lo compongono, induce a prestare attenzione ai suoi aspetti condivisi, comuni, a fondere le differenze individuali per conferire coerenza ai dati raccolti. Al contrario, dare particolare rilievo alla posizione dei singoli porta a concentrarsi sulle contraddizioni, tensioni e ambivalenze. Ci potrebbero essere voci che non hanno contribuito a determinare una categoria, ma che devono poter essere riconosciute. Inoltre, ci sono voci che non si sono fatte sentire, magari per timidezza, perché non danno peso al loro punto di vista, perché non vogliono rivelare il loro pensiero o perché è meno articolato rispetto al pensiero degli altri. Se il ricercatore non li evidenzia, darà spazio, nel suo resoconto, soltanto alle maggioranze, che non necessariamente esprimono un punto di vista significativo o interessante. Oltre alla descrizione dei temi emergenti è, dunque, opportuno restituire alcune citazioni letterali dei partecipanti e analizzare la proporzione di persone che nel gruppo non ha espresso consenso o si è trovato, addirittura, in disaccordo con il concetto emerso.

Il linguaggio non-verbale dei partecipanti deve essere raccolto. Tali dati non verbali includono cinque sistemi, ossia quello vocale (le variazioni di intensità, intonazione, tono e ritmo della voce), prossemico (il modo di utilizzare lo spazio interpersonale per comunicare, la distanza), cronemico (il ritmo del discorso e la durata delle pause, l'alternanza dei turni), e cinesico (gestualità, posture ed espressioni facciali) (Gorden, 1980). L'aiuto-moderatore si trova in una posizione migliore, perché sgravato dalla conduzione del FG, per raccogliere queste informazioni e può aiutarsi con l'uso di una matrice (Onwuegbuzie et al., 2009), come quella presentata in Tabella 9.3.

Tabella 9.3 Matrice di osservazione del consenso, dissenso o imparziali

Partecipanti	1	2	3	4	5	6
Domande						
1						
1.1						
2						
3						
.....						

Per velocizzare il processo di scrittura si possono riassumere le informazioni, nelle celle, nel modo seguente: A = assenso indicato (verbale o non verbale); D = dissenso indicato (verbale o non verbale); AS = affermazione significativa fornita che suggerisce un assenso; DS = affermazione significativa fornita che suggerisce dissenso; MR = non ha indicato accordo o dissenso (ovvero, mancata risposta). La lettura della matrice permette di contare i membri del FG che rientrano in ciascuna categoria e di rendicontare, non solo le tendenze emerse, ma anche i singoli posizionamenti rispetto ai temi centrali.

I conteggi sono utili non solo a descrivere i consensi e dissensi, ma anche a rendere conto dei modelli di risposta dei partecipanti (Carey, 1995; Kidd e Parshall, 2000; Morgan, 1993, 1997; Silverman, 1985), ma non devono sostituire i dati qualitativi che, insieme ai dati quantitativi, permettono un'analisi mista molto ricca (Sandelowski, 2001; Onwuegbuzie e Teddlie, 2003; Morse, 2003).

Anche la posizione dei partecipanti deve essere trascritta, in modo da ricordare chi si è seduto accanto a quale membro del gruppo. A questi dati devono essere affiancate le informazioni demografiche rilevanti, in modo da analizzare se possono essere associati alla scelta del posto e, in seguito, a modelli di risposta. Può essere utile, ad esempio, realizzare un diagramma di Venn, che permette di illustrare in quale misura i soggetti/gli elementi studiati (individui, istituzioni, caratteristiche sociodemografiche ecc.) interagiscono, si sovrappongono o si influenzano a vicenda.

A differenza dell'intervista, nel FG vanno analizzate, oltre ai contenuti degli specifici interventi, le interazioni tra i partecipanti e tra i partecipanti e il moderatore (Myers, 1998; 2006). Nella conversazione con gli altri membri del FG, ciascun partecipante sceglie di parlare in un certo modo di un concetto e, secondo Myers (2006), non è tanto quello che sostengono quanto in quale modo ne parlano che è interessante analizzare. A questo proposito, andiamo a vedere come analizzare i dati raccolti.

Diagramma di Venn

In tali diagrammi, cerchi di varie dimensioni sono usati per rappresentare istituzioni, per esempio, e più grande è il cerchio, più importante è l'istituzione. La distanza tra i cerchi viene utilizzata per rappresentare il grado di influenza o contatto tra istituzioni e le sovrapposizioni tra cerchi indicano i livelli di interazione. Può essere anche utile per rappresentare la distribuzione di caratteristiche dei soggetti, come il genere, il paese di origine e l'età o altro (eventualmente anche prima del FG per la selezione dei partecipanti).

9.2.9 L'analisi dei dati

È altamente raccomandato di non svolgere da soli il lavoro di analisi, ma di confrontarsi con altri ricercatori, in modo da identificare i preconcetti che condizionano l'interpretazione dei significati e della loro successiva classificazione.

Una prima fase consiste nel trascrivere, parola per parola, i contenuti (verbatim) della discussione subito dopo la sessione affiancando i commenti del moderatore e le osservazioni dei segnali non-verbali emessi dai partecipanti. In seguito a questo, si procede all'analisi che deve essere svolta e descritta in modo da essere riproducibile da altri ricercatori e realizzata con il massimo di oggettività e rigore, rimanendo sempre focalizzati sulle domande e gli obiettivi della ricerca.

In pratica, le trascrizioni vengono ritagliate e codificate in base all'argomento/al concetto che li può sintetizzare. Il codice può essere un'espressione, una frase significativa, un sentimento, una parola chiave o altro. In questa fase, lungi dal ridurre e ordinare il materiale, o cercare di districare ciò che è importante da ciò che non lo è, ciò che è "significativo" da ciò che sembra essere casuale, dobbiamo cercare di coglierne la complessità. Questi verbatim, a seconda dell'idea che trasmettono, vengono classificati in categorie. Man mano che si procede, le prime categorie si affinano, suddividono, mentre ne emergono di nuove. Ovviamente, alcune frasi contengono più concetti diversi e devono, di conseguenza, essere classificate in più categorie. Queste categorie vengono, poi, raggruppate in temi più generali (dimensioni), mantenendo eventualmente una suddivisione in sottocategorie più dettagliata, in relazione alle diverse sfaccettature dell'idea espressa. Questa fase di codifica permette di eliminare gli elementi "fuori tema", di rilevare la frequenza con la quale vengono riproposti alcuni temi e di associarli alle informazioni di carattere non verbale ed emotivo. Tuttavia, l'obiettivo principale non è di valutare l'importanza quantitativa di ciascuna categoria: è l'emergere delle categorie stesse che costituisce il maggiore risultato dell'analisi, perché riflettono concettualmente il contenuto delle discussioni (Duchesne e Haegel, 2004).

Varie tecniche di elaborazione dei dati di un FG sono state descritte da diversi studiosi (in particolare: Leech e Onwuegbuzie, 2007 e 2008; Onwuegbuzie, Dickinson e Leech, 2009): Constant comparison analysis, Classical content analysis, Keywords-in-context e, infine, Discourse analysis.

La *Constant comparison analysis* (Glaser & Strauss, 1967; Glaser, 1978; 1992; Strauss, 1987), ossia l'analisi comparativa dei dati, è caratterizzata da tre fasi (Strauss e Corbin, 1998). Durante la prima fase di "codifica aperta", i dati vengono suddivisi in piccole unità, alle quali viene attribuito un codice. Nella seconda fase di "codifica assiale", questi codici vengono raggruppati in categorie. Infine, nella terza e ultima fase di "codifica selettiva", il ricercatore sviluppa uno o più temi che raccontano il contenuto di ciascuna delle categorie (Strauss e Corbin, 1998). Questo me-

todo è particolarmente utile quando si svolgono più FG nel corso dello stesso studio: permette di verificare il livello di saturazione dei temi o significati sintetizzati dalle categorie e di procedere a un campionamento teorico, ossia di calibrare il campione per approfondire e completare i concetti teoretici emersi (Charmaz, 2000). Inoltre, i gruppi ottenuti dall'analisi del primo FG possono dare luogo a un'ipotesi di classificazione per i dati dei FG successivi.

La *Classical content analysis* (analisi del contenuto classica) è simile alla Constant comparison analysis. Tuttavia, secondo tale approccio, non si tratta di creare nuovi temi, bensì di contare il numero di volte in cui vengono utilizzati singoli codici già identificati (Leech e Onwuegbuzie, 2007, p. 569). In tal caso, è essenziale descrivere in modo dettagliato i significati racchiusi in ogni codice.

Il metodo chiamato *Keywords-in-context* (KWICs), ossia parole chiave in contesto, ha lo scopo di analizzare il modo in cui vengono utilizzate determinate parole in uno specifico contesto, ossia di esaminare il contesto culturale di uso delle parole, per dirlo come Fielding e Lee (1998). L'analisi può focalizzarsi sull'uso delle parole all'interno di un FG, oppure nel confronto tra diversi FG o, ancora, tra individui all'interno del gruppo. Questo approccio contribuisce alla contestualizzazione culturale dello studio dei temi emergenti e delle parole-chiave ad essi associati (Fielding e Lee, 1998).

L'approccio *Discourse analysis* (Analisi del discorso), sviluppato dagli psicologi sociali Potter e Wetherell (1987), è un metodo atto a spiegare come nel FG la parola sia costruita nell'interazione e nella conversazione tra gli interlocutori e, di conseguenza, porta a studiare le interazioni discorsive (Myers e Macnaghten, 1999). Non considera la parola scambiata come un mezzo passivo che rende conto di ciò che accade altrove, bensì come un processo attivo tramite il quale i soggetti interpretano una realtà in gran parte indeterminata (Dore e McDermott, 1982), che risulta da una scelta fatta secondo una gerarchia di priorità di ordine culturale, individuale e situazionale. Parte dal presupposto che le persone partecipino a una discussione attraverso un processo di negoziazione e di costruzione del discorso, durante il quale viene mobilitato un insieme eterogeneo di valori e conoscenze (Salazar Orvig e Grossen, 2004).

Approfondimento 9.2
L'USO DI SOFTWARE

È, oramai, diventata prassi consolidata l'uso di software per l'analisi qualitativa dei dati CAQDAS (Computer Aided Qualitative Data Analysis Systems) - come NVivo, Atlas/TI o MaxQDA - che consentono, ad esempio, di analizzare le relazioni tra i dati e le co-occorrenze lessicali, di verificare gli effetti di una modifica del sistema di categorie sui contenuti già elaborati, di aggregare dei codici in categorie, di analizzare le relazioni tra le categorie. Offrono, inoltre, una visione molto più chiara del sistema di codifica ottenuto.

È anche possibile effettuare un'analisi delle registrazioni, attraverso l'uso di software che consentono di registrare digitalmente il FG ed evidenziare specifici passaggi (per esempio, Sound Forge, AVID DVCPRO, Descreet Combustion). Permette al ricercatore di tornare rapidamente alle citazioni selezionate all'interno della trascrizione e di raggruppare le citazioni per creare dei cluster e analizzare le relazioni tra questi raggruppamenti.

Esistono anche software come Transana che permette di confrontare, suddividere e collegare i dati video e audio con le trascrizioni. Consente, inoltre, di includere le pause e le sovrapposizioni nella conversazione nell'analisi dei dati.

Dopo aver proceduto all'analisi dei dati, giunge il momento della scrittura della relazione, che include una sintesi narrativa e descrittiva, seguita da un'analisi interpretativa dei dati e delle possibili raccomandazioni basate sui risultati ottenuti.

9.2.10 In conclusione, la restituzione

Per concludere, scegliamo di porre l'accento su una fase della ricerca che viene spesso dimenticata, ma che conferisce senso al lavoro svolto ed è indispensabile da un punto di vista etico (Zonabend, 1994; Vidal, 2010): la restituzione, ovvero la condivisione dei risultati delle analisi con i soggetti coinvolti nella ricerca, che ha come scopo di "mettere a disposizione di questi ultimi un'interpretazione situata delle loro pratiche dal punto di vista del ricercatore" (Zucchermaglio et al., 2013).

Questo momento dell'indagine offre la possibilità di confrontarsi con le analisi dal punto di vista di chi, con chi o per chi sono stati condotti e permette di testare, affinare o correggere le proprie analisi e interpretazioni dei dati e di far emergere nuovi percorsi di ricerca.

La restituzione risponde, inoltre, al diritto degli attori del campo studiato di ottenere un ritorno del ricercatore sulle conoscenze acquisite grazie al loro contributo e la loro partecipazione. La restituzione deve contenere i risultati ottenuti in modo da essere loro utili. È tenuta, dunque, a essere di facile comprensione e a proporre interpretazioni che possano portare a formulare proposte e indurre trasformazioni sul campo.

Riepilogo

Nel capitolo sono stati definiti, in particolare, i seguenti concetti: studio di caso, caso, triangolazione, focus group, ricerca bibliografica, constant comparison analysis, classical content analysis, keywords-in-context, discourse analysis, questioning route e, infine, restituzione.

Per ciascuno dei due argomenti - dello Studio di caso e del Focus Group - è sta-

to descritto in quali circostanze vengono utilizzati e con quale fine; quali sono i loro punti di forza e limiti; e quali sono le tappe da seguire per applicarli. Per quanto riguarda il FG, è stato esplicitato il ruolo del ricercatore, moderatore e aiuto-moderatore, la modalità di raccolta e codifica dei dati, e sono stati presentati metodi per la loro elaborazione. Inoltre, sono state descritte le modalità di reclutamento dei partecipanti e di preparazione e svolgimento del FG in presenza e in modalità telematica.

Infine, è stata ricordata l'importanza della restituzione come fase fondamentale dell'indagine.

Riferimenti bibliografici

Abbott A. (1988). *The system of the professions*. London: University of Chicago Press.

Barbour R.S., Kitzinger J. (eds.) (1999). *Developing focus group research*. London: Sage.

Baumgartner T.A., Strong C.H., Hensley L.D. (2002). *Conducting and reading research in health and human performance* (3d ed.). New York: McGraw-Hill.

Berg B.L. (2000). *Qualitative research methods for the social sciences* (4th ed.). Boston: Allyn & Bacon.

Bernard H.R. (1995). *Research methods in anthropology: Qualitative and quantitative approaches*. Walnut Creek (CA): AltaMira.

Bloor M., Frankland J., Thomas M., Robson K. (2001). *Focus group in social research*. London: Sage

Bourgeois-Guérin V., Beaudoin S. (2016). *La place de l'éthique dans l'interprétation de la souffrance en recherche qualitative*. Recherches qualitatives, 35(2),23-44.

Carey M. (1995). *Comment: Concerns in the analysis of focus group data. Qualitative health research*, 5,487-495.

Charmaz K. (2000). *Grounded theory: objectivist and constructivist methods*. In: Denzin N.K., Lincoln Y.S. (eds.). *Handbook of qualitative research*. 2nd ed., (pp. 509-535). Thousand Oaks (CA): Sage.

Crabtree B.F., Yanoshik M.K., Miller W.L., O'Connor P.J. (1993). *Selecting individual or group interviews*. In Morgan D.L. (ed.). *Successful focus groups: Advancing the state of the art*, (pp. 137-149). Newbury Park (CA): Sage.

Creswell J.W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks CA: Sage.

Creswell J.W. (2013). *Qualitative inquiry and research design: choosing among five approaches*. Thousand Oaks (CA): Sage.

David A. (2000). *Logique, épistémologie et méthodologie en sciences de gestion: trois hypothèses revisitées*. In David A., Hatchuel A., Laufer R. (eds.). *Les nouvelles fondations des sciences de gestion*, (pp. 83-109). Paris: Vuibert.

Dubois A., Gadde L.E. (2002). Systematic combining: an abductive approach to case research. *Journal of Business Research* 55(7):553-60.

Duchesne S., Haegel F. (2004). *L'enquête et ses méthodes : les entretiens collectifs*. Paris: Nathan.

Dufour S., Fortin D., Hamel J. (1991). *L'enquête de terrain en sciences sociales*. Montréal: Éditions Saint-Martin.

Duggleby W. (2005). What about focus group interaction data? *Qualitative Health Research*, n. 15, pp. 832-40.

Dumez H. (2010). La description: point aveugle de la recherche qualitative. *Le Libellio d'Aegis* 6(2),28-43.

Dumez H. (2013-a). *Méthodologie de la recherche qualitative*. Paris: Vuibert.

Dumez H. (2013-b). Qu'est-ce qu'un cas, et que peut-on attendre d'une étude de cas? *Le Libellio d'Aegis*, n. 9(2), 13-26.

Eckstein H. (1973). Case study and theory in political science. In Greenstein F.I., Polsby N.W. (eds.). *Handbook of political science*, (pp. 79-137). Reading (MA): Addison-Wesley.

Eisenhardt K.M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.

Farr R., Tafoya E. (1992). *Western and Hungarian representations of individualism: a comparative study based on group discussions and social dilemmas*. Manoscritto non pubblicato.

Fern E.F. (2001). *Advanced Focus Group Research*. Thousand Oaks: Sage Publications.

Fielding N., Lee R.M. (1998). *Computer analysis and qualitative research*. London: Sage.

Flick U. (1998). *An introduction to qualitative research: theory, method and applications*. London: Sage.

Flick U. (2009). *An introduction to qualitative research* (4th ed.). Thousand Oaks (CA): Sage.

Geoffrion P. (1993). *Le groupe de discussion*. In: Gauthier B. (ed.). *Recherche sociale. De la problématique à la cueillette des données* (pp. 311-335). Sainte-Foy: Presses de l'Université du Québec.

George A.L., Bennett A. (2005). *Case studies and theory development in the social sciences*, Cambridge (MA): M.I.T. Press.

Gerring J. (2004). What is a case study and what is it good for? *The American Political Science Review*, 98(2), 341-354.

Glaser B.G., Strauss A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glaser B.G. (1978). *Theoretical sensitivity*. Mill Valley (California): Sociology Press.

Glaser B.G. (1992). *Discovery of grounded theory*. Chicago: Aldine.

Gorden R.L. (1980). *Interviewing strategy, techniques, and tactics*. Homewood, IL: Dor-

sey.
HAGUE P. (2002). *Market research* (3rd ed.). London: Kogan Page.
HARRISON H., BIRKS M.,FRANKLIN R., MILLS J. (2017). Case study research: foundations and methodological orientations. *Forum: Qualitative Social Research,* 18(1), Art. 19.
HERREID C.F. (1997). What is a case? Bringing to science education the established teaching tool of law and medicine. *Journal of College Science Teaching,* 27(2), 92-94.
JICKT.D. (1979), Mixing Qualitative and Quantitative Methods: Triangulation in Action, Administrative Science Quarterly. *Qualitative Methodology,* 24(4), 602-611.
JOHNSON R.B., CHRISTENSEN L.B. (2004). *Educational research: quantitative, qualitative, and mixed approaches.* Boston: Allyn and Bacon.
KIDD P., PARSHALL M. (2000). Getting the focus and the group: Enhancing analytical rigor in focus group research. *Qualitative Health Research,* 10, 293-308.
KITZINGER J. (1994). The methodology of focus groups: The importance of interaction between research participants. *Sociology of Health and Illness,* 18, 103-121.
KITZINGER J., FARQUHAR C. (1999). The analytic potential of "sensitive moments" in focus group discussions. In: R. Barbour, J. Kitzinger (eds.), *Developing focus group research: Politics, theory and practice,* (pp. 156-172). London: Sage.
KITZINGER J., MARKOVA I., KALAMPALIKIS N. (2004). Qu'est-ce que les focus groups? *Bulletin de Psychologie,* 57(3), 237-243.
KOENIG G. (1993). Production de la connaissance et constitution des pratiques organisationnelles. *Revue de Gestion des Ressources Humaines,* n. 9, pp. 4-17.
KOENIG G. (2009). L'étude de cas à visée infirmationniste. *Le Libellio d'Aegis,* 5(4), 7-13.
KRUEGER R.A. (1994). *Focus groups: A practical guide for applied research* (2nd ed.). Thousand Oaks (CA): Sage.
KRUEGER R.A., CASEY M.A. (2000). *Focus groups: A practical guide for applied researchers* (3rd ed.). Thousand Oaks (CA): Sage.
LANGFORD B.E., SCHOENFELD G., IZZO G. (2002). Nominal grouping sessions vs. focus groups. *Qualitative Market Research,* 5, 58-70.
LEECH N.L., ONWUEGBUZIE A.J. (2007). An array of qualitative data analysis tools: A call for qualitative data analysis triangulation. *School Psychology Quarterly,* 22, 557-84.
LEECH N.L., ONWUEGBUZIE A.J. (2008). Qualitative data analysis: A compendium of techniques for school psychology research and beyond. *School Psychology Quarterly,* 23, 587-604.
LIJPHART A. (1971). Comparative politics and the comparative methods. *American Political Science Review,* 65(3), 682-93.
LINCOLN Y.S., GUBA E.G. (1985). *Naturalistic inquiry.* Beverly Hills (CA): Sage.

MARKOVA I. (2003). Les focus groups. In S. Moscovici, F. Buschini (eds.). *Les méthodes des sciences humaines.* Paris: Presses Universitaires de France.
MAXWELL J.A. (2005). *Qualitative research design: An interactive approach* (2nd ed.). Newbury Park (CA): Sage.
MERRIAM S.B. (2009). *Qualitative research: A guide to design and implementation.* 2nd ed. San Francisco (CA): Jossey-Bass.
MERTON R.K., FISKE M., KENDALL P.L. (1956). *The focused interview.* Glencoe (Illinois): Free Press.
MILES M.B., HUBERMAN A.M. (2003). *Research design and management. Qualitative data analysis.* (3rd ed.). Thousand Oaks (CA): Sage.
MORGAN D.L. (ED.) (1993). *Successful focus groups: Advancing the state of the art.* Newbury Park (CA): Sage.
MORGAN D.L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks (CA): Sage.
MORRISON D.E. (1998). *The search for a method. Focus groups and the development of mass communication research.* Luton: University of Luton Press.
MORSE J.M. (1995). The significance of saturation. *Qualitative Health Research,* 5, 147-9.
MORSE J.M. (2003). Principles of mixed methods and multimethod research design. In: Tashakkori A., Teddlie C. (eds.). *Handbook of mixed methods in social and behavioral research,* (pp. 189-208). Thousand Oaks (CA): Sage.
MYERS G. (1998). Displaying opinions: Topics and disagreement in focus groups, *Language in Society,* 27, 85-111.
MYERS G. (2006). "Where are you from?": Identifying place, *Journal of Sociolinguistics,* 10, 320-43.
MYERS G., MACNAGHTEN P. (1999), Can focus groups be analysed as talk? In R.S. Barbour, J. Kitzinger (eds.). *Developing focus group research: Politics, theory, and practice,* (pp. 173-185). London: Sage.
ONWUEGBUZIE A.J., DICKINSON W.B., LEECH N.L., ZORAN A.G. (2009). A qualitative framework for collecting and analyzing data in focus group research. *International Journal of Qualitative Methods,* 8(3), 1-21.
ONWUEGBUZIE A.J., TEDDLIE C. (2003). A framework for analyzing data in mixed methods research. In: A. Tashakkori, C. Teddlie (eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351-383). Thousand Oaks (CA): Sage.
ONWUEGBUZIE A.J., JIAO Q.G., BOSTICK S.L. (2004). *Library anxiety: Theory, research, and applications.* Lanham, Maryland: Scarecrow.
PATTON M.Q. (2002). *Qualitative research & evaluation methods.* Thousand Oaks (CA): Sage.
POTTER J., WETHERELL M. (1987). *Discourse and social psychology: Beyond attitudes and behaviour.* London: Sage.

RAGIN C.C., BECKER H.S. (1992). *What is a case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press.

SALAZAR ORVIG A., GROSSEN M. (2004). Représentations sociales et analyse de discours produit dans des focus groups: un point de vue dialogique. *Bulletin de psychologie*, 57(3), 263-272.

SANDELOWSKI M. (2001). Real qualitative researchers don't count: The use of numbers in qualitative research. *Research in Nursing & Health*, 24, 230-240.

SANDELOWSKI M. (2008). Theoretical saturation. In L.M. Given (Ed.), *The Sage encyclopedia of qualitative methods* (vol. 1, pp. 875-876). Thousand Oaks (CA): Sage.

SAUMURE K., GIVEN L.M. (2008). Data saturation. In L.M. Given (ed.), *The Sage encyclopedia of qualitative methods* (vol. 1, pp. 195-196). Thousand Oaks (CA): Sage.

SAVOIE-ZAJC L. (2019). Les pratiques des chercheurs liées au soutien de la rigueur dans leur recherche: une analyse d'articles de recherches qualitatives parus entre 2010 et 2017. *Recherches qualitatives*, 38(1), 1-161.

SIGGELKOW N. (2007). Persuasion with case studies. *Academy of Management Journal*, 50(1), 20-24.

SILVERMAN D. (1985). *Qualitative methodology and sociology: Describing the social world*. Aldershot (UK): Gower.

SIM J. (1998). Collecting and analyzing qualitative data: Issues raised by the focus group. *Journal of Advanced Nursing*, 28, 345-352.

STAKE R.E. (1995), *The Art of Case Study Research*. London: Sage Publications.

STAKE R.E. (2008). Qualitative case studies. In N.K. Denzin, Y.S. Lincoln (eds.) *Strategies of Qualitative Inquiry*. Los Angeles : Sage Publications.

STAKE R.E. (2010). *Qualitative research studying how things work*. New York: Guilford Press.

STEWART D.W., SHAMDASANI P.N. (1990). *Focus group. Theory and Practice*. New York: Sage.

STRAUSS A., CORBIN J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park (CA): Sage.

STRAUSS A., CORBIN J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks (CA): Sage.

STRAUSS A. (1987). *Qualitative analysis for social scientists*. Cambridge (UK): University of Cambridge Press.

VANWYNSBERGHE R., KHAN S. (2007). Redefining case study. *International Journal of Qualitative Methods*, 6(2), 1-10.

VIDAL L. (2010). Ni politique de l'autruche, ni stratégie du hérisson: pour l'étude de la fabrique de l'anthropologie de la santé. *Anthropologie et santé*, 1.

WILKINSON S. (1998). Focus group in health research: exploring the meanings of health and illness. *Journal of Health Psychology*, n. 3, pp. 329-348.

WILKINSON S. (2004), Focus group research. In D. Silverman (ed.) *Qualitative research: Theory, method, and practice* (pp. 177-199). Thousand Oaks (CA): Sage.

YIN R.K. (2009). Case study research: design and methods (4th ed.). Thousand Oaks (CA): Sage.

YIN R.K. (2012). *Applications of case study research* (3rd ed.). Thousand Oaks (CA): Sage.

ZAMMUNER V.L. (2003). *I focus group*. Bologna: Il Mulino.

ZONABEND F. (1994). De l'objet et de sa restitution en anthropologie. *Gradhiva*, n. 16, pp. 3-14..

ZUCCHERMAGLIO C., ALBI F., Fatigante M., Saglietti M (2013). *Fare Ricerca situata in psicologia sociale*. Bologna: Il Mulino.

10

Scelta degli strumenti per la ricerca educativa[1]

In questo capitolo viene affrontato il problema della coerenza e dell'adeguatezza degli strumenti della ricerca in relazione alle ipotesi formulate dal ricercatore (scopi, obiettivi), ai contesti di ricerca (confermativa o esplorativa), e alle modalità di raccolta dei dati.
Saranno inoltre riportati alcuni esempi di strumenti spesso utilizzati nell'ambito della ricerca educativa: dalle check-list ai questionari di auto-valutazione, dai Diari di bordo al Portfolio, ecc. Vengono quindi esaminate le principali caratteristiche, i punti di forza ed eventuali limiti di tali strumenti.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- coerenza, adeguatezza e pertinenza degli strumenti con le ipotesi della ricerca;
- procedure dell'osservazione sistematica;
- strumenti di rilevazione dei dati osservativi;
- tecniche di registrazione dei dati conoscitivi.

[1] Il presente contributo è frutto del lavoro congiunto dei due autori: G. Domenici ha redatto il paragrafo 10.1, G. Moretti il paragrafo 10.2.

Le preoccupazioni del coltivatore per cui ciò che conta è quella data pianta, quel dato pezzo di terreno esposto al sole dalla tale ora alla tale ora, quella data malattia delle foglie che va combattuta in tempo con quel dato trattamento sono estranee alla mente modellata sui procedimenti dell'industria, cioè portata a decidere sulle impostazioni generali e sui prototipi. Quando Palomar s'era accorto di quanto approssimativi e votati all'errore sono i criteri di quel mondo dove credeva di trovare precisione e norma universale, era tornato lentamente a costruirsi un rapporto col mondo limitandolo all'osservazione delle forme visibili; ma ormai lui era fatto com'era fatto: la sua adesione alle cose restava quella intermittente e labile delle persone che sembrano sempre intente a pensare un'altra cosa ma quest'altra cosa non c'è.

Italo Calvino, *Palomar*

10.1 Coerenza e adeguatezza degli strumenti per la ricerca

10.1.1 Ipotesi, piano della ricerca e scelta degli strumenti

Anche nel quadro degli studi pedagogico-educativi - nei differenti ambiti disciplinari, nelle diverse aree e procedure di indagine, quantitative e qualitative, dei campi di ricerca empirico-sperimentali e non-sperimentali - la conoscenza accreditata come scientifica procede da problema a problema formulando, provando e discutendo criticamente e pubblicamente le diverse ipotesi risolutive ideate, e le connesse teorie interpretative. Un processo, un percorso, che, come ormai si sarà compreso, si estrinseca emblematicamente formulando congetture circa la possibile risoluzione del problema, e sottoponendole a confutazioni da parte dello stesso ricercatore, ancor prima che alla prova dei fatti e della comunità degli esperti di area, e magari di aree vicine (Popper 1934, 1983).

Questo richiamo alla peculiarità del metodo scientifico risulta particolarmente importante per poter cogliere la relazione che intercorre tra strumenti impiegabili per la rilevazione dei dati e l'impianto concettuale della ricerca. Come è stato più ampiamente detto nel Capitolo 3, il pur complesso e articolato processo di indagine si sviluppa essenzialmente attraverso il passaggio che va dallo stato di disagio diffuso causato dalla rottura di un certo equilibrio, alla determinazione del problema e dei suoi elementi costitutivi, della sua posizione all'interno di una più ampia area problematica attraverso l'impiego, magari anche originale, di tutte le informazioni possedute sulla questione che si sta esaminando, sulle vie di soluzione tentate nel tempo - longitudinalmente - e in altri luoghi - sincronicamente. Un processo, questo, che, partendo dalle prime suggestioni, conduce alla incubazione delle possibili soluzioni del problema, alla successiva formulazione di vere e proprie ipotesi esplicative dello stesso, e alle modalità operative che occorre seguire per la loro prova di verificazione e/o falsificazione. Un processo che contiene in sé, prima in forma intuitiva, in nuce, poi in forma sempre più chiara, la prefigurazione di strategie risolutive, quindi di procedure, mezzi e strumenti da impiegarsi per poter sottoporre a controllo l'ipotesi formulata. La formalizzazione di alcuni di questi passaggi necessari, ancorché non sufficienti per poter svolgere correttamente l'indagine, verranno a costituire la base del *Piano della ricerca*. A seconda che ci si trovi ad operare in un contesto di indagine a prevalente o esclusivo carattere confermativo di teorie o di esiti generali di ricerche già consolidate, o, diversamente, a prevalente o esclusivo carattere esplorativo o di scoperta, potremo impiegare tipologie di strumenti, o strumenti di assunzione dei dati, che sono già noti, codificati e provati nella loro struttura formale in precedenti indagini: test, questionari, griglie di osservazione, standardizzati o da sottoporre a try-out, eccetera; oppure, nelle indagini esplorative o di frontiera, strumenti da doversi appositamente ideare, costruire e mettere a punto per lo svolgimento di quella specifica ricerca - magari provandoli preliminarmente in situazioni più ristrette e più facilmente controllabili, e, quando possibile, tararli e standardizzarli, ovvero per provare la o le ipotesi risolutive della questione problematica presa in esame da quella indagine.

Approfondimento 10.1.

INDAGINI CONFERMATIVE E INDAGINI ESPLORATIVE

In generale vengono considerate come confermative quelle indagini volte a conoscere in una situazione problematica il valore assunto da variabili dipendenti facendo leva su quelle indipendenti, o sperimentali, all'interno di un quadro teorico-operativo ormai consolidato che ha reso esplicite le relazioni tra entrambe le tipologie di variabili.

Le ricerche a carattere esplorativo o di scoperta, sono viceversa indagini che vengono ideate e svolte per cogliere le relazioni quali-quantitative tra i fattori che si presume contribuiscano a spiegare un dato fenomeno o evento. Vengono compiute quando il terreno conoscitivo specifico è poco o per niente dissodato, quando cioè una teoria e/o un fatto empirico cozzano con un'altra teoria e/o un altro fatto empirico e occorre comprenderne le ragioni. Sono ovviamente molto più complesse e hanno generalmente bisogno di maggiori e più articolati studi preliminari.

Anche per tali ragioni l'*ipotesi*, che rappresenta in qualche modo l'elemento forse più creativo e divergente dell'intero impianto di ricerca, dovrebbe essere formulata, quando possibile, in forma descrittiva così da rendere evidenti, seppur in sintesi essenziale, i momenti formali più caratterizzanti il processo e il percorso dell'indagine (soprattutto quando questa sia di tipo sperimentale), quindi anche i modi di procedere e, appunto, gli strumenti da impiegare per la sua verificazione. In una frase, l'ipotesi, che dovrebbe descrivere e far affiorare a un livello di interpretazione univoca l'emblematico "se ... allora...", suggerisce implicitamente, quando non espli-

citamente, anche quali dati tra gli infiniti possibili bisogna rilevare, quando e in che modo rilevarli e trattarli, infine la loro dimensione quali-quantitativa che indichi, perciò, le condizioni della sua accettabilità. In altri termini, la descrizione dell'ipotesi fatta adottando la formula appena indicata e resa qui di seguito più esplicita (quando ovviamente vengano riempiti tutti gli spazi descrittivi rappresentati dai puntini di sospensione) suggerisce, se non addirittura indica con precisione, l'oggetto delle rilevazioni e i modi in cui compierle: "date le, o in presenza delle, condizioni ... (*che vanno evidentemente descritte*), se si altera la situazione intervenendo in tal modo... (*ovvero, sulla/e variabile/i indipendente/i, dette anche sperimentale/i*) allora si avrà questa nuova situazione-risultato... (*che va descritta*) ovvero che la/e variabile/i dipendente/i assumerà/anno il o i seguenti valori.... in grado così di confermare la validità dell'ipotesi formulata". Si comprende come in qualche modo sia l'ipotesi a guidare l'azione di determinazione delle forme, dei modi, e perciò delle caratteristiche essenziali degli strumenti di raccolta delle informazioni e dei dati qualitativi e/o quantitativi indispensabili per provare la validità dell'ipotesi stessa.

10.1.2 Strumenti di rilevazione dei dati come costrutti teorici. Il problema della loro coerenza e adeguatezza agli scopi della ricerca

Lo o gli strumenti di assunzione di quei dati - tra i tanti rilevabili - in grado di confermare o respingere l'ipotesi di ricerca, possono a ragione considerarsi, perciò, come veri e propri *costrutti teorici*. Racchiudono in sé, infatti ed essenzialmente, sia il "precipitato" fattuale, il consolidato strumentale, appunto, della o delle teorie che, in modo implicito o esplicito, precedono e accompagnano ogni specifica indagine, soprattutto di tipo sperimentale; sia la conoscenza empirica e teorica sulla cui base sono stati già costruiti e provati o verranno appositamente architettati e costruiti. Uno strumento rappresenta, a ben vedere, il dispositivo tecnico o tecnologico che nel processo di studio della realtà che si vuole osservare e indagare, direttamente o indirettamente, potenzia e moltiplica al meglio le originarie capacità sensoriali umane, permette di farci scrutare e vedere meglio la realtà. Esso è dunque un artefatto in grado di potenziare le principali capacità sensoriali umane attraverso un impiego razionale e finalizzato delle più moderne e avanzate tecniche e tecnologie "ingegneristiche" di settore, disponibili.

La coerenza dello strumento in senso lato, con l'ipotesi formulata univocamente e con la teoria che la sostiene, rappresenta perciò il requisito formale fondamentale che deve possedere un buono strumento di rilevazione dei dati. È utile ricordare, a questo proposito, che, così come le nuove discipline alla loro nascita, durante la loro fase costitutiva - cioè di formalizzazione sia del proprio apparato teorico-concettuale interpretativo della realtà di cui si occupano, sia di quello metodologico-procedurale di accrescimento del proprio dominio conoscitivo - vengono

solitamente denotate e connotate, e non solo dai loro cultori, con quello che essi stessi fanno in quel campo specifico di ricerca, altrettanto avviene nel rapporto tra ricerca e strumenti di raccolta dei dati. Per analogia, in senso niente affatto tautologico, la denotazione e la connotazione di una indagine specifica, soprattutto nelle ricerche di frontiera, vengono spesso compiute facendo riferimento agli strumenti (e ai dati) a tal fine ideati, costruiti, messi a punto e impiegati. Per molti anni l'intelligenza (come il suo costrutto) è stata considerata come ciò che i test di intelligenza rilevavano. Si può pertanto inferire che senza uno strumento adeguato, coerente con gli assunti e le ipotesi della ricerca, si rischia di rilevare altro da quello che si vuole o si crede di cogliere, o di aver rilevato. Coerenza e adeguatezza formali degli strumenti rinviano perciò in buona parte ai requisiti di validità e attendibilità dei dati rispetto agli obiettivi e agli scopi della ricerca, in una parola alla o alle sue ipotesi.

Le anticipazioni delle possibili soluzioni del problema rappresentate dalle ipotesi - soprattutto dalle ipotesi fattuali che costituiscono vere e proprie anticipazioni immediate degli esiti cui si suppone di pervenire - costituiscono perciò la base dello schema logico-sequenziale cui tener conto nella messa a punto sia del Piano della ricerca, sia delle vie, dei tempi e degli strumenti da impiegare per assumere i dati valutativi iniziali, in itinere e finali della ricerca. Un piano - che può ben includere un PERT (*Project Evaluation Review Tecnique*) - col quale vengono perciò programmati, tra l'altro, gli itinerari da percorrere, i tempi e le tappe intermedie e finali da raggiungere; le operazioni da compiere e la specificazione di chi, quando e come deve compierle; ma anche delle tecniche e degli strumenti di rilevazione (trattamento e analisi) dei dati, che si vogliono e/o si devono impiegare; i criteri di verifica della congruenza delle attività via via svolte rispetto a quelle progettate e pianificate.

Quantunque la formulazione delle ipotesi fattuali derivi, come si è detto, dalle conoscenze di fondo che del problema si hanno, dagli esiti delle ricognizioni storiche e sincroniche dei tentativi riusciti e degli errori compiuti nella soluzione di questioni analoghe, essa costituisce pur sempre l'esito di una immaginazione che può e deve essere audace, anche perché se il problema è davvero nuovo, e ben posto, ci si trova a impegnarsi in un campo con pochi punti di appoggio. Quando il terreno della ricerca è stato poco dissodato, quando cioè ci si trova di fronte a situazioni problematiche inedite - non di rado prodotte dal ricercatore (anche per questo la ricerca non ha mai fine) dalle sue intuizioni circa le soluzioni probabili, corroborate soprattutto da procedimenti di tipo ipotetico-deduttivo - gli strumenti di assunzione dei dati indispensabili per corroborare l'affidabilità delle congetture fatte, devono essere anch'essi ideati, costruiti e provati dal ricercatore. Nelle ricerche di frontiera il ricercatore può diventare perciò egli stesso l'"ingegnere ideatore e costruttore" degli strumenti di rilevazione dei dati valutativi adeguati

alle esigenze dell'indagine.

Partendo, ora, dalla constatazione che la conoscenza delle caratteristiche essenziali dei più diffusi e provati strumenti di raccolta dei dati per un verso rappresenta la condizione necessaria per poterli usare in modo pertinente e adeguato, e che nel contempo possa meglio facilitare la ideazione e messa a punto di nuovi strumenti, e, ancora, considerando che fin qui si sono illustrati quelli particolarmente caratterizzanti la ricerca educativa - alcuni dei quali costituiscono nel contempo autonome procedure di indagine - in questo capitolo verranno illustrati brevemente i molti altri strumenti che trovano un impiego abbastanza diffuso nelle ricerche educative, a partire da quelli relativi all'osservazione sistematica, la cui trattazione generale è stata fatta nel Capitolo 7.

Concludiamo qui perciò la carrellata sugli strumenti della ricerca educativa, che tra i tanti possibili si è scelto di includere in questo lavoro perché assai diffusi e impiegati nella ricerca italiana e internazionale. Come più volte si è detto, altri ancora se ne possono ideare, tarare e impiegare in rapporto all'evoluzione del quadro conoscitivo e procedurale della ricerca empirico-sperimentale.

10.2 Gli strumenti della ricerca. L'osservazione sistematica

Gli strumenti e le tecniche di osservazione sistematica utilizzati in ambito educativo sono molteplici, e la loro scelta dipende - come si è visto - dalla situazione presa in esame, dal grado di accuratezza desiderato, dal tempo disponibile per effettuare le osservazioni e, soprattutto, dall'ipotesi di ricerca.

Alcuni dei più noti strumenti e tecniche di rilevazione e registrazione dei dati osservativi sono stati già esaminati. Occorre tuttavia ricordare che pur sviluppati per essere utilizzati in ambiti particolari di indagine (per esempio: l'attività in piccolo o grande gruppo, il lavoro cooperativo, la leadership, il gioco, il conflitto), risultano facilmente adattabili e utilizzabili in contesti diversi.

Occorre tenere presente che la scelta di una determinata modalità di osservazione sistematica, oltre che delimitare il campo di indagine, determina in parte l'elaborazione e la valutazione dei dati raccolti.

M. Postic e J.M. De Ketele (1993) hanno individuato otto tecniche di osservazione che ritengono particolarmente efficaci per condurre indagini in ambito educativo, e le distinguono in due raggruppamenti. Il primo comprende le *tecniche di osservazione sistematica*, volte a osservare attributi già prestabiliti, di cui fanno parte i sistemi di categorie, gli inventari o i sistemi di segni e le scale di valutazione; il secondo raggruppamento, definito delle *tecniche di osservazione esperienziale*, così definite in quanto si applicano all'indagine condotta a partire dall'esperienza di una situazione vissuta in un arco di tempo da se stessi o da altre persone (p. 31, 1993), include i diari, il metodo degli episodi critici, le tecniche di registrazione di una parte di vita e il diario di bordo. Qui di seguito presentiamo le caratteristiche formali e procedurali, le modalità di utilizzo, i punti di forza e di criticità di alcuni degli strumenti indicati da Postic e De Ketele (1993) con l'aggiunta del portfolio, strumento oggi assai diffuso in ambito educativo.

10.2.1 L'osservazione sistematica tramite il sistema di categorie

L'osservazione tramite il sistema di categorie consiste nella individuazione di un numero ristretto e molecolare di categorie che sono prese a riferimento dal ricercatore per classificare tutti gli eventi che si verificano nel contesto di studio considerato per l'intera durata temporale precedentemente stabilita. La messa a punto del sistema di categorie costituisce una aspetto fondamentale dello strumento, che caratterizza il piano di osservazione adottato. Occorre cioè stabilire, a priori, prima ancora di condurre l'osservazione sul campo, quali siano gli elementi da osservare, come si procederà alla loro registrazione e, successivamente, in che modo gli stessi verranno sottoposti alla analisi quantitativa e qualitativa. Il sistema di categorie deve essere:

- esaustivo e completo;
- formulato in modo che l'insieme delle categorie siano reciprocamente escludenti, ovvero tenendo conto del fatto che ogni elemento osservato debba essere classificato "uno alla volta" in una sola categoria.

Considerato che tutti gli eventi devono essere classificati, occorre prevedere nel sistema di categorie una categoria "ripostiglio", soluzione simile a quella presa da Flanders (1970) nella formulazione della nota griglia di rilevazione delle interazioni verbali, in cui l'autore ha previsto la categoria "silenzio-confusione" per classificare tutti quei comportamenti che non sarebbero rientrati nelle altre categorie.

La codifica può essere effettuata mediante un approccio di tipo unidimensionale, ma il più delle volte è consigliabile una codifica di tipo multidimensionale, volta a individuare più dimensioni da osservare, e a specificare per ciascuna di esse le categorie che le rappresentano. Le varie dimensioni, in base al principio che stabilisce che le categorie debbano essere reciprocamente autoescludenti solo se si trovano all'interno della medesima dimensione, consentono al ricercatore di classificare più volte quanto osservato, purché la registrazione sia effettuata una sola volta per ciascuna dimensione prevista.

L'operazionalizzazione delle categorie richiede un quadro teorico di riferimento robusto, in grado di garantire la pertinenza e l'affidabilità dello strumento. Inoltre è necessario avvalersi del confronto intersoggettivo tra osservatori affinché, a partire dalla familiarità con la situazione problematica oggetto di studio, possano pre-

disporre uno schema di categorie da sottoporre a pre-test con l'obiettivo di garantire la validità dello strumento. Per assicurare un alto grado di fedeltà di rilevazione, registrazione e codifica dei dati osservativi mediante il sistema di categorie è inoltre opportuno prevedere interventi specifici di formazione dei ricercatori coinvolti.

Un valido esempio di osservazione condotta tramite il sistema di categorie è individuabile nel modello classico elaborato da R.F. Bales (1950, 1970), volto a rilevare l'interazione sociale di gruppo. Tale proposta, che mantiene inalterata la sua efficacia, prevede l'utilizzo di dodici categorie per analizzare gli atti verbali delle interazioni. Bales distingue due aree fondamentali:

1. l'area delle comunicazioni di tipo socio-emotivo (categorie 1, 2, 3, 10, 11 e 12), che comprende un'area positiva (solidarizzare, rassicurare, concordare), e un'area negativa che indica la presenza nelle dinamiche di gruppo di tensioni, conflitti e contrapposizioni;
2. l'area delle comunicazioni riferite al compito (categorie 4, 5, 6, 7, 8 e 9), che comprende un'area della comunicazione diretta a guidare il lavoro di gruppo (offerta di suggerimenti, opinioni e orientamenti), e un'area della comunicazione diretta a informarsi (richiesta di orientamenti, opinioni e suggerimenti).

Nella Tabella 10.1 sono riportate le dodici classi di comportamento interattivo riguardanti l'Area socioemotiva (comportamenti affettivi positivi e negativi) e l'Area del compito (domande e tentativi di risposta).

Tabella 10.1 Categorie di osservazione del comportamento interattivo mediante il sistema di categorie di R.F. Bales (Bales,1950, 1970; Pierro & Livi, 2001)

Alunni / **Categorie**	A	B	C	D	E	F	G	H	I	L	...	**Tot.**
1) Si mostra amichevole												
2) Drammatizza												
3) Concorda												
4) Dà suggerimenti												
5) Esprime opinioni												
6) Dà informazione												
7) Chiede informazione												
8) Chiede opinioni												
9) Chiede suggerimenti												
10) Non concorda												
11) Manifesta tensione												
12) Si mostra ostile												
Totale												

Caso 10.1

IL CASO DELLA CONFERMA DELLE STRUTTURE TEORICHE SOTTESE AL SISTEMA DI CATEGORIE DI BALES

Pierro e Livi (2001) nel condurre una'indagine avvalendosi delle categorie di Bales hanno sottoposto lo strumento ad analisi fattoriale confirmatoria e hanno evidenziato che una struttura a 4 fattori latenti è quella che si adatta meglio ai dati di partenza. Nello specifico il modello presenta quattro categorie sovra-ordinate relative sia agli Atti Emessi sia agli Atti Ricevuti (Emozioni positive, Tentativi di risposta, Domande, Emozioni negative).

L'analisi fattoriale ha evidenziato che le strutture teoriche sottese al sistema di categorie ideato da Bales sono espressioni significative della teoria di riferimento e che le categorie individuate non sono elementi isolati tra loro, ma manifestano una loro dinamica che può essere compresa identificando alcuni pattern di risultati in relazione agli Atti emessi e a quelli ricevuti.

Il caso citato suggerisce che quando si utilizzano specifici sistemi di categorie, sia per condurre ricerche, sia per finalità didattiche e valutative, è sempre opportuno fare riferimento agli esiti scientifici consolidati. La conoscenza scientifica più aggiornata in-

fatti permette di avvalersi di uno strumento in modo consapevole e soprattutto consente di interpretare i dati osservativi non solo per analizzare la presenza o meno di taluni comportamenti, ma per comprenderne le dinamiche (riguardo all'esempio sopra riportato, la complementarità fra atti interattivi centrati sul compito e gli aspetti emozionali manifestati nelle discussioni di gruppo).

È sempre opportuno effettuare la triangolazione tra le osservazioni rilevate attraverso il sistema di categorie con quelle raccolte con altri strumenti, ad esempio il sociogramma o le scale di valutazione, che permettono di apprezzare l'intensità delle comunicazioni registrate. Come vedremo nel paragrafo successivo, i docenti, sulla base degli obiettivi esplicitati nel piano di osservazione, possono dotarsi di strumenti di rilevazione agili e flessibili, che permettono di scegliere autonomamente e di descrivere le categorie da sottoporre a osservazione.

Triangolazione

La triangolazione fa riferimento alla possibilità di osservare un dato fenomeno utilizzando più strumenti di rilevazione e confrontando tra di loro gli esiti rilevati.

Oltre alla triangolazione degli strumenti è possibile procedere alla triangolazione dei punti di vista dei diversi attori sociali coinvolti nella indagine osservativa (ad esempio: insegnanti, alunni, famiglie, dirigente scolastico).

L'interesse crescente volto a rilevare non solo la frequenza degli eventi, ma il loro significato in relazione al contesto e al punto di vista degli attori, suggerisce un uso integrato dei sistemi di categorie con altri strumenti (ad esempio: interviste e colloqui). In questa prospettiva, e anche nel caso di utilizzo di inventari o sistemi di segni, consigliamo di aggiungere in calce alla scheda di osservazione uno spazio libero in cui il ricercatore o il docente possono riportare le note sul campo per contestualizzare quanto registrato mediante la griglia o descrivere sinteticamente eventuali eventi imprevisti o di difficile classificazione. Ricordiamo che le note di campo sono anche una risorsa utile durante la fase di restituzione degli esiti della ricerca agli attori interessati.

10.2.2 Gli inventari o sistemi di segni

Gli inventari o sistemi di segni consistono nella individuazione di un numero assai ampio di categorie (indicativamente 80-100), che sono prese a riferimento per classificare esclusivamente quegli eventi che l'osservatore ha preventivamente selezionato perché sono considerati di interesse ai fini della ricerca e che ha descritto mediante una formulazione scritta (segno, descrittore o item).

La progettazione del piano di osservazione è dunque importante perché tramite di esso vanno indicati anticipatamente e in modo chiaro gli elementi che si vorrebbero osservare nel contesto scelto e nel tempo di osservazione stabilito.

A parere di Postic e De Ketele (1993), nel caso dei sistemi di segni, a differenza del sistema di categorie, non ci si riferisce più alla "codifica" ma alla "organizzazione" dei segni-azioni da osservare, e ci si può avvalere di una procedura sia unidimensionale sia multidimensionale.

L'organizzazione dell'inventario in varie dimensioni è consigliabile e dipende principalmente sia dalle caratteristiche del contesto analizzato, sia dagli obiettivi della ricerca o dell'azione didattica e valutativa intrapresa. Gli autori citati formulano alcune modalità di annotazione dei segni, alternative tra loro, nel seguente modo: «a) annotare un'azione una sola volta se compare durante il periodo; b) annotare la frequenza di apparizione di questa azione durante il periodo; c) registrare per mezzo di un cronografo la somma delle durate di apparizione di questo atto durante il periodo» (p. 32, 1993).

Poliandri, Muzzioli, Quadrelli e Romiti (2012) invitano a definire nell'osservazione la scelta dell'intervallo di tempo entro il quale rilevare le informazioni, aspetto che in letteratura propone diverse opzioni di campionatura del tempo:

- l'unità di tempo definita entro cui registrare quanto osservato (è possibile fare la regiatrazione ogni volta che accade un'azione, oppure si può annotare ciascuna categoria di azione una sola volta nel periodo di tempo prescelto, ad esempio ogni 15 minuti);
- il campionamento statico, che consiste nel rilevare una serie di istantanee in cui si codifica cosa avviene all'interno di intervalli regolari di tempo, non sequenziali;
- le unità naturali, che riflettono i periodi di tempo reali in cui hanno luogo alcune attività specifiche e/o interessanti.

L'operazionalizzazione dei segni-azioni richiede un solido quadro teorico di riferimento in grado di garantire la forza esplicativa, la pertinenza e l'affidabilità dello strumento.

La codifica, analogamente al sistema di categorie, alla check-list o al sistema di segni, può essere effettuata da uno o più osservatori e può essere supportata da un "manuale di codifica".

Nel Capitolo 4, a cui si rimanda, con riferimento al contesto italiano si illustra la proposta di Domenici di introdurre nei contesti educativi formali l'utilizzo sistematico dei sistemi di segni o "Descrittori dell'apprendimento" (1981) da parte dei docenti per la verifica dell'apprendimento, mediante la descrizione delle prestazioni (abilità-criterio), o per rilevare la presenza o meno di comportamenti ritenuti qualificanti l'intervento educativo svolto.

10.2.3 Le scale di valutazione e i questionari di autovalutazione

Le scale di valutazione consistono in una serie di elementi o item che sulla base di una teoria di riferimento consentono di declinare in termini operazionali alcuni indicatori specifici, o variabili, utili a rilevare il grado o livello con cui si manifestano interessi, atteggiamenti, opinioni o valori.

Le scale di valutazione presentano una serie di domande scritte accompagnate da risposte formulate con modalità differenti - numerica, grafica, descrittiva o ibrida - che permettono ai rispondenti di indicare, a seconda dei casi, il grado di intensità o la frequenza che caratterizzano il manifestarsi dell'elemento sottoposto a valutazione (Figura 10.1).

Scala numerica
– Il bambino presta attenzione in classe 1 2 3 4

Scala con avverbi
– Il bambino presta attenzione in classe mai qualche volta spesso sempre

Scala grafica
– Il bambino presta attenzione in classe 1 _____ × _____ 10

Scala descrittiva o di giudizio
– Il bambino presta attenzione in classe:
a) è attento in modo continuativo;
b) è attento se interessato;
c) a volte è distratto;
d) ha tempi di attenzione molto brevi;
e) è incapace di attenzione.

Figura 10.1 Tipi di scala. (Fonte: Coggi e Ricchiardi, Progettare la ricerca empirica in educazione, 2005, Carocci, Roma, p. 101.)

A seconda degli obiettivi e delle ipotesi di ricerca possiamo impiegare scale di valutazione con un numero di gradi differenti. La Tabella 10.2 riporta un esempio di scala a cinque gradi (mai, di rado, talvolta, spesso, sempre), che rappresenta una della soluzioni più diffuse.

Tabella 10.2 Esempio di scala di valutazione del comportamento a cinque gradi.

Comportamento osservato	Mai	Di rado	Talvolta	Spesso	Sempre
Gestisce positivamente i conflitti					X
Chiede aiuto agli altri					X
Mantiene gli impegni assegnati				X	
Interviene nelle discussioni				X	
Assume autonomamente decisioni			X		
Prima di decidere assume tutte le informazioni		X			
Riflette sugli errori compiuti			X		
.....					

In letteratura sono presenti alcune tipologie di scale assai note nell'ambito della ricerca sociale e di quella educativa proposte da Tursthon, Likert, Osgood e Guttman.

La *scala di Tursthon* ha l'obiettivo di apprezzare l'atteggiamento di un soggetto riguardo a un particolare argomento. La scala consiste in una serie numerata molto ampia di asserzioni (almeno 100), formulate in modo chiaro, semplice e univoco, alle quali il soggetto deve rispondere con sicurezza se si trova d'accordo o meno. Il problema di tale scala è la complessa e lunga procedura da seguire prima di arrivare alla somministrazione.

La *scala di Likert* presenta per iscritto una serie ampia di asserzioni e consente ai rispondenti di esprimere per ciascuna di esse il proprio grado di accordo scegliendo l'avverbio che meglio colloca la propria posizione nell'ambito della scala proposta. Spesso la scala Likert è articolata in numero dispari di livelli, tre (ad esempio: d'accordo, abbastanza d'accordo, in disaccordo), cinque o sette. Per evitare che i rispondenti possano collocarsi in una posizione intermedia, si preferisce di frequente adottare una scala con un numero pari di livelli (ad esempio quattro), per consentire una presa di posizione più decisa, in senso positivo o negativo, rispetto all'asserto proposto.

La scala Likert è molto utilizzata perché la sua costruzione è particolarmente agevole e il suo utilizzo non richiede necessariamente una formazione specifica. Suggeriamo di tenere conto dei seguenti passaggi nella fase di costruzione della scala:

- formulare una serie di asserti positivi e negativi riguardanti l'argomento di studio e di ricerca scelto;
- selezionare gli asserti migliori o più discriminanti mediante una prova sul campo su un gruppo ristretto di popolazione e avvalendosi del confronto intersoggettivo tra ricercatori ed esperti esterni;
- predisporre un elenco di asserzioni ritenute discriminanti e coerenti con gli obiettivi della ricerca (indicativamente non meno di 10);
- organizzare la successione degli asserti in modo da mescolare tra loro gli asserti positivi con quelli negativi;
- adottare per ogni asserto una scala a più livelli, pari o dispari, a seconda del piano di osservazione stabilito.

Le scale formulate con la tecnica proposta da Guttman sono definite "deterministiche" poiché la tipologia di risposta richiesta presuppone che la caratteristica o la qualità osservata siano individuabili in modo sicuro. Le risposte predefinite presenti in questo tipo di scala sono infatti dicotomiche (si/no, presente/assente).

La scala Guttman consiste in una serie limitata di asserti scritti, ordinabili in

modo univoco, ad esempio dal più positivo al più negativo o viceversa. Per ciascuno di tali asserti il rispondente deve scegliere tra sì/no o tra presente/assente. Il modo in cui sono ordinati gli asserti implica che nelle risposte si manifesti fino a un certo livello della scala la presenza cumulativa dei segni. Tale caratteristica della scala è utile al ricercatore per valutare e comprendere velocemente l'intensità dell'elemento studiato (Figura 10.3).

1. Ho avuto alcune insufficienze nella mia carriera scolastica	SI	NO
2. Ho avuto sistematicamente insufficienze in ogni livello scolastico che ho frequentato	SI	NO
3. Ho avuto insufficienze in tutti gli anni del mio curricolo scolastico	SI	NO
4. Sono stato bocciato almeno una volta	SI	NO
5. Sono stato bocciato più volte	SI	NO

Figura 10.3 Scala Guttman. (Fonte: Coggi e Ricchiardi, 2005. *Progettare la ricerca empirica in educazione,* Carocci, Roma, p. 101.)

La scala proposta da Osgood è nota in letteratura con l'espressione "differenziale semantico" e ha l'obiettivo di valutare il significato semantico attribuito da un soggetto a un concetto. La scala si presenta con una doppia serie di parole o di aggettivi, riportati in due distinte colonne, contrapposte tra loro (Figura 10.4). La colonna di destra presenta aggettivi o parole con polarità opposta a quella riportata nella colonna di sinistra. Tra le due colonne, per ciascuna coppia di aggettivi o parole, è riportata una scala, numerica o grafica, a più punti (ad esempio da 1 a 7), che permette al rispondente di indicare come si colloca nell'ambito dello spazio semantico del sistema dei significati rappresentato dai livelli della scala che definisce la distanza tra le polarità dei concetti presi a riferimento.

L'insegnante della classe è:

Lassista	1 2 3 4 5 6 7	Rigoroso
Distaccato	1 2 3 4 5 6 7	Empatico
Demotivato	1 2 3 4 5 6 7	Motivato
Neutrale	1 2 3 4 5 6 7	Appassionato
Indisponibile	1 2 3 4 5 6 7	Disponibile

Figura 10.4 Scala Osgood.

L'uso delle scale di valutazione ci consente di registrare la presenza sia dei comportamenti e dei tratti della personalità, sia degli elementi informativi che possono farne apprezzare il grado di intensità. Quest'ultima possibilità è molto importante

per leggere in modo dinamico il profilo del soggetto osservato e per ricostruirne l'evoluzione dei comportamenti manifestati. Risulta utile l'inserimento di alcune scale di valutazione all'interno dei questionari di autovalutazione dei soggetti. I questionari, detti anche "inventari", sono strutturati in maniera tale che l'allievo possa valutare se stesso rispondendo con sincerità alle domande proposte (Figura 10.5).

Quando non comprendi la spiegazione che ti è stata data per svolgere una attività:
a) chiedi chiarimenti all'insegnante
b) chiedi chiarimenti ai compagni
c) preferisci far finta di nulla
d) non sai che cosa fare

L'insegnante ti restituisce un lavoro che hai svolto e ti accorgi che hai commesso molti errori
a) chiedi spiegazioni all'insegnante per capire la causa degli errori
b) confronti il tuo quaderno con quello di altri compagni
c) cerchi di comprendere perché li hai commessi
d) chiudi immediatamente il quaderno e lo metti nello zaino

Figura 10.5 Esempio di item di Questionario di Autovalutazione volto a rilevare alcuni aspetti del comportamento dello studente nei confronti delle discipline e dell'esperienza scolastica.

Ricordiamo che l'uso delle scale deve tenere conto del quadro dinamico in cui sono svolte le valutazioni e le autovalutazioni dei soggetti interessati, sottoposto a modifiche in funzione delle esperienze personali e formative effettuate.

10.2.4 I diari

La scrittura "diaristica" consente di accompagnare in tempo reale gli avvenimenti osservati, e favorisce il contatto diretto con gli eventi e le esperienze, evitando una selezione preventiva dei materiali cognitivi ed emotivi da rilevare. Gli studi sono pressoché unanimi nel riconoscere alla scrittura autobiografica e ai diari in particolare la capacità di:

- rappresentare il corso dell'esperienza attraverso descrizioni che permettono di ritagliarne e fissarne gli eventi più significativi, facilitandone l'analisi e il controllo;
- definire, tra soggetti e contesti, quella opportuna distanza che permette a quanti sono implicati professionalmente in un determinato contesto e in specifiche situazioni, di esercitare pienamente la propria riflessività;

- assumere atteggiamenti aperti e flessibili nei confronti dell'esperienza vissuta;
- incoraggiare nelle persone l'adozione di comportamenti assertivi, che permettono di esprimere e sostenere rispetto al gruppo, senza timori e pregiudizi, le proprie idee e il proprio punto di vista, descrivendo anche i vissuti personali, le eventuali difficoltà e le contraddizioni riscontrate nel contesto preso in esame;
- valorizzare la funzione conoscitiva dei dettagli, dei piccoli cambiamenti apparentemente insignificanti, che attraverso la lettura sistemica e/o diacronica degli eventi e delle esperienze, permette spesso di effettuare osservazioni e documentazioni accurate e di svolgere riflessioni più articolate e sensibili riguardo agli aspetti contestuali e relazionali.

Possiamo distinguere diversi tipi di diari, sulla base del contenuto, della periodicità con la quale sono scritti, della funzione svolta, della modalità di narrazione scelta e del grado di coinvolgimento degli attori. Due sono i tipi generali di diario, il diario libero o spontaneo e il diario strutturato, che possono rispondere a diverse esigenze di ricerca, monitoraggio in itinere di processi e di documentazione. Le caratteristiche di ciascun diario sono:

- il *diario libero o spontaneo* è quella forma di diario in cui chi scrive è libero di esprimere il proprio punto di vista, di annotare in modo libero e tutte le volte che vuole le osservazioni personali, senza sottoporsi a vincoli esterni;
- il *diario strutturato* è quella forma di diario in cui chi scrive, pur parlando in prima persona, è invitato a ricostruire eventi, osservazioni, ambienti, o ad esprimere opinioni, convinzioni e sensazioni, nel rispetto dei vincoli prefissati da alcune voci prestrutturate indicate come traccia da seguire e delle aree tematiche di riferimento.

I diari, liberi o strutturati, sono strumenti narrativi assai importanti per la loro ricchezza informativa, per questo dovrebbero contenere sia elementi obiettivi e descrittivi sia elementi riflessivi, personali, emotivi e affettivi. I diari introspettivi, troppo centrati sulla vita personale, spesso sottovalutano l'importanza del contesto, al contrario i diari esclusivamente descrittivi non permettono di accedere alla dimensione personale degli attori implicati nelle attività di ricerca o coinvolti nell'intervento didattico.

La tecnica del diario può essere utilizzata flessibilmente con finalità differenti da parte del ricercatore, dell'insegnate o dell'alunno. In ogni caso la descrizione delle attività e dei comportamenti effettuata attraverso il diario è utile per tutti gli attori, rispettivamente, per la raccolta sul campo di evidenze nell'ambito della ricerca empirica di tipo esplorativo, per le finalità di sviluppo professionale, di qualificazione dei processi educativi e di apprendimento e acquisizione di maggiori livelli

di consapevolezza.

Il diario può essere sviluppato individualmente o in gruppo. La stesura o tenuta di un diario a livello di gruppo, soluzione spesso utilizzata nell'ambito dei modelli di ricerca partecipativa (per esempio, la Ricerca-Azione o la Ricerca-Formazione), è auspicabile quando i soggetti interessati sono messi a conoscenza del disegno della ricerca e sono invitati a partecipare al processo di individuazione del focus della ricerca.

Zabalza Beraza (p. 10, 2001), pur auspicando iniziative nelle quali i diari possano essere sviluppati "anche dagli alunni", focalizza l'attenzione sulle condizioni di utilizzo adeguato, etico e fattuale di un repertorio di diari di classe, tenuti dall'insegnante, individualmente o in gruppo. Tali strumenti sono utili sia per studiare qualitativamente le pratiche e i vissuti degli insegnanti, sia per favorirne lo sviluppo professionale, prendendo in considerazione quanto accadde in classe.

Nel paragrafo 10.2.5 presentiamo le caratteristiche del diario di bordo, tecnica che per le sue finalità e per i contenuti trattati è funzionale a un utilizzo collettivo da parte di un gruppo di ricerca o di un team di insegnanti. Anche gli alunni, oltre ai diari individuali, possono avvalersi dei diari di classe o di gruppo, per finalità sia didattiche sia di ricerca (si veda Approfondimento 10.2).

Approfondimento 10.2

DIARIO DI CLASSE SVILUPPATO DAGLI ALUNNI

Il diario di classe è uno strumento che può essere utilizzato dagli alunni appartenenti alla medesima classe o al medesimo gruppo di apprendimento. Benché si parli molto di dare voce agli studenti e di promuoverne il coinvolgimento attivo, l'uso collettivo del diario è poco diffuso in ambito scolastico.

Tale sottovalutazione dipende sia dalla scarsa attenzione dedicata agli aspetti cooperativi e collaborativi tra pari a livello di grande o di piccolo gruppo, sia dall'idea che gli studenti siano poco interessati a scrivere con modalità collaborative diari di classe o di gruppo.

Nella prospettiva di promuovere in forma attiva l'apprendimento e di coinvolgere come protagonisti gli studenti, si ritiene utile la tenuta dei diari di classe da parte degli stessi, senza la mediazione dei docenti. Tale possibilità è oggi facilitata dalle nuove forme di scrittura collettiva, tipiche dei processi di digitalizzazione e di comunicazione sociale, che possono essere agevolmente praticate in molti ambienti online, che a tale scopo dispongono di varie funzioni di supporto alla realizzazione di esperienze collaborative di scrittura tra e con i pari (ad esempio: blog, scrittura cooperativa Wiki).

I dati così raccolti, opportunamente trattati, possono ben essere impiegati in micro-indagini conoscitive.

Sono di seguito elencate le principali caratteristiche dei diari di classe o di gruppo in relazione ai soggetti, alle finalità e alla scelta dei focus tematici trattati.

- *Soggetti*: studenti, il gruppo classe nel suo insieme o nelle sue varie articolazioni (ad esempio: gruppi di apprendimento, di ricerca o di laboratorio). Alunni di classi diverse, della medesima scuola o di reti di scuole.
- *Finalità*: documentare, raccontare e descrivere i percorsi esperienziali, emotivi e cognitivi di apprendimento del gruppo classe e/o dei diversi gruppi indagati.
- *Focus tematici trattati*: le esperienze più significative sia di apprendimento sia affettive e relazionali; le fasi critiche di cambiamento, i passaggi critici affrontati dal gruppo, i modi di affrontare la conflittualità e la dissonanza cognitiva; il rapporto del gruppo di alunni con i docenti, con gli ambiti disciplinari e con la vita scolastica nel suo complesso. Tale dimensione partecipativa è molto importante, specie nei contesti formativi in cui la situazione, oltre che essere agita, è anche condivisa dai vari attori.

L'utilizzo del diario di classe da parte degli alunni dovrebbe rispettare in particolare tre regole deontologiche generali:

1. acquisizione del consenso esplicito manifestato dagli attori;
2. restituzione delle informazioni e delle riflessioni ai diretti interessati;
3. pubblicizzazione delle interpretazioni effettivamente condivise dagli attori.

10.2.5 Il diario di bordo

La tecnica del diario di bordo consiste nella narrazione individuale o di piccolo gruppo, sistematica e guidata, di azioni e intenzioni personali o riferite ad altri soggetti coinvolti direttamente o indirettamente nel contesto osservato.

M. Postic e J.M. De Ketele presentano il diario come tecnica di osservazione narrativa e retrospettiva "che consiste nel descrivere nel linguaggio abitualmente usato (cosa che non esclude l'utilizzazione di termini tecnici da parte dell'osservatore) le proprie attività o quelle degli altri" (p. 34, 1993). Gli autori citati inseriscono il diario di bordo tra le principali tecniche di osservazione sistematica e lo includono tra le tecniche di osservazione esperienziale, insieme agli episodi critici e alla registrazione di una parte rappresentativa di vita.

L'utilizzo del diario di bordo individuale, ma soprattutto di quello di gruppo, risulta efficace nelle iniziative di formazione e sviluppo professionale, nelle attività dei servizi educativi, sociali e sanitari, nel campo della educazione degli adulti, e più in generale negli ambienti di lavoro in cui più professionisti o diverse figure professionali sono chiamati a cooperare per raggiungere un medesimo risultato.

Consideriamo l'uso collettivo del diario di bordo particolarmente utile nell'ambito dei contesti educativi formali, in cui anche a seguito del processo di autonomia scolastica i ruoli assunti e le funzioni svolte dagli insegnanti non si limitano a quelli di aula e di insegnamento, ma si estendono alla gestione delle risorse e delle relazioni, al coordinamento dei gruppi di lavoro, alla partecipazione a progetti e iniziative in rete, sia tra scuole sia tra soggetti istituzionali. Sono sempre più diffuse le situazioni in cui l'insegnante è chiamato a far parte di gruppi di lavoro, di studio o di ricerca, che prevedono la partecipazione di professionalità differenti anche in rappresentanza di più istituzioni. Il diario di bordo collettivo utilizzato in tali contesti consente di annotare oltre alle azioni e alle intenzioni degli attori anche i personali punti di vista.

Nel corso di una indagine il singolo ricercatore, ma soprattutto i membri del gruppo di ricerca, attraverso la struttura condivisa del diario di bordo di fatto si impegnano a convergere tutti nella direzione di annotare esclusivamente il tipo di informazioni concordate in precedenza (ad esempio: indicazioni relative alle variazioni intervenute, ai processi decisionali, commenti, riflessioni e confronti tra intenzionalità ed effetti reali, e più in generale tutto ciò che a vario titolo accompagna le azioni e le relazioni sociali).

Prima di predisporre il diario di bordo a livello di gruppo è opportuno concordare con gli altri partecipanti il tipo di linguaggio da utilizzare e condividere il significato da assegnare ai principali concetti operativi che si intendono utilizzare per descrivere la situazione data. Tali accorgimenti possono sia facilitare la stesura partecipata del diario di bordo sia potenziarne la successiva lettura complessiva e consentirne una più approfondita e ricca analisi dei contenuti alla luce del quadro concettuale suggerito dalla ipotesi di ricerca.

La tenuta del diario di bordo - settimanale o anche giornaliera - consente a ciascuno degli osservatori-partecipanti di annotare, sia diacronicamente, sia sincronicamente mediante una forma di scrittura prevalentemente sintetica e narrativa, gli eventi critici rilevati, quelli ritenuti utili o strategici ai fini della ricerca o delle attività progettate.

Possiamo individuare due fasi distinte di utilizzo del diario di bordo: la prima, che coincide con la fase iniziale della ricerca, in cui il diario è funzionale alla possibilità di "ancorare" e fissare l'analisi sistemica che consente di inquadrare la situazione problematica; la seconda fase, in cui si possono mettere a fuoco alcuni sottoinsiemi da osservare, e il diario è lo strumento su cui si potranno rilevare gli elementi di riflessione per la definizione operazionale delle ipotesi di ricerca "da verificare ulteriormente con altre strategie e altri strumenti d'osservazione" (Postic, De Ketele, 1993, p. 38).

La scrittura del diario di bordo non è predeterminata, ma prevede una descrizione longitudinale delle realizzazioni effettuate; tuttavia è consigliata l'esplicitazione

e la predeterminazione del livello di esaustività delle annotazioni. Il diario di bordo è considerato una tecnica "più ricca e completa di quella del diario", in quanto prevede oltre all'annotazione delle attività svolte (come il diario) anche quella dei riferimenti all'insieme di elementi che possono consentire di ricostruire il contesto interno ed esterno della situazione osservata.

Chi utilizza il diario di bordo, dunque, può:

- annotare le azioni insieme con le proprie o altrui intenzioni;
- fare riferimento alle previsioni iniziali e confrontarle con gli effetti osservati;
- riportare informazioni e impressioni sui contesti di azione;
- descrivere gli altri attori eventualmente implicati e cercare di presentarne il punto di vista;
- riportare annotazioni sulle difficoltà incontrate e sulle modalità scelte per superarle.

Il diario di bordo risponde a diverse esigenze, di studio e di ricerca, ma il suo utilizzo è soprattutto efficace come tecnica di ricerca, di osservazione e di formazione degli adulti impegnati in specifici contesti professionali (comunità di pratiche o comunità di apprendimento). A tale riguardo Postic e De Ketele propongono un quadro concettuale, da prendere a riferimento per la stesura del diario, che prevede:

1. la descrizione contestualizzata delle attività della settimana;
2. l'analisi personale del vissuto;
3. la presentazione chiara dei contrasti, successivamente negoziati in termini di "Chi? Che cosa fa? Quando? Come? Perché? Entro quale termine?";
4. l'analisi che è stata fatta dai protagonisti e gli eventuali confronti intermedi, compresi gli esiti positivi o negativi di tali negoziazioni.

Il diario di bordo è uno strumento assai duttile e flessibile, che può essere strutturato in vario modo e può essere facilmente adattato per rispondere a molteplici esigenze informative e conoscitive comprese quelle condivise nel gruppo di ricerca o nel team dei docenti. La possibilità di condividere il diario in formato digitale (per esempio mediante Google Drive) ne rende più agevole la stesura collaborativa da parte di gruppi di ricerca, di team di insegnanti e gruppi di lavoro eterogenei per composizione professionale.

Le principali caratteristiche dei diari di bordo sviluppati da ricercatori e insegnanti in relazione ai soggetti, alle finalità e alla scelta dei focus tematici trattati sono le seguenti (Moretti, 2003):

- *Soggetti*: ricercatori, insegnanti, educatori, esperti e operatori che a vario titolo partecipano alla progettazione e realizzazione di attività di ricerca, di formazione e sviluppo personale e professionale, di sviluppo di attività educative e didattiche.
- *Finalità*: coordinare l'azione dei ricercatori, degli insegnanti e di altri operatori, facilitando la comunicazione orizzontale tra tutti gli adulti impegnati nella ricerca o nel progetto educativo e didattico. Favorire l'espressione di tutti i punti di vista, tra i membri del gruppo di ricerca, tra insegnanti, educatori e formatori, sia di quelli sistematicamente impegnati con la classe, sia di quelli che vi dedicano minore tempo perché impegnati su progetti specifici o circoscritti (esperto di laboratorio, educatore, colleghi di altre classi, ecc.), sia di coloro che si confrontano occasionalmente con gli alunni (esperti esterni, genitori-esperti, testimoni orali ecc.).
- *Nella strutturazione* del diario di bordo e nella focalizzazione delle aree tematiche si farà riferimento alle ipotesi di ricerca o alle finalità formative, educative e didattiche previste. Tra gli aspetti di rilievo da prendere in considerazione si segnalano: gli elementi sia cognitivi sia affettivo-motivazionali ed emozionali; l'esplicitazione delle decisioni assunte o eventualmente di quelle rimandate; i suggerimenti, le richieste di proseguire le osservazioni anche in altri contesti rispetto a quelli iniziali; i vissuti personali e professionali in relazione al gruppo e alle attività svolte con particolare riferimento a quelle che caratterizzano gli obiettivi della ricerca o le finalità progettuali e del Piano triennale dell'offerta formativa e del Piano di Miglioramento.

10.2.6 Il metodo degli episodi critici

Il metodo degli *episodi*, o *incidenti critici*, è stato impiegato con successo da Flanagan (1954) in ambito educativo al fine di rilevare e analizzare i tratti salienti, i comportamenti specifici, le caratteristiche dei processi mentali e i presupposti che sono alla base dei ragionamenti e dei giudizi formulati da attori che svolgono ruoli particolari (ad esempio: insegnanti o dirigenti scolastici), che per la loro ricchezza informativa osservabile "possono dare una spiegazione di un successo o di un insuccesso" (De Landsheere, 1973, p. 169).

De Frankrijker (1998, p. 105) suggerisce di considerare la tecnica dell'incidente critico come una raccolta di fatti importanti relativi al comportamento umano in determinate situazioni e nel renderne ragionevolmente predicibili le conseguenze.

La validità del metodo è testimoniata dall'applicazione effettuata in riferimento agli insegnanti e direttori didattici con la rilevazione e l'analisi di incidenti critici sia positivi sia negativi e per valutare il tipo di clima della classe. Tuttavia il metodo degli episodi critici ha riscontrato nel tempo una applicazione limitata che in

parte è dovuta ad alcune difficoltà connesse al suo utilizzo, già segnalate da De Landsheere (1973). Tali impedimenti consistono in particolare nel riferimento a comportamenti rari e quindi difficili da osservare in modo sistematico; alla tendenza delle persone interpellate a ricordare "l'incidente" unicamente per la sua eccezionalità, che per questo implica una difficoltà nel considerare tale eccezione come rappresentativa; al fatto che gli incidenti critici siano spesso legati tra loro, incoerenti e dunque difficili da utilizzare.

In Italia il metodo degli incidenti critici è stato applicato nell'ambito della formazione interculturale e della valutazione delle competenze interculturali (Damini e Surian, 2012) e in alcune pratiche di ricerca sulla riflessività (Nuzzaci, 2017). Nel corso di uno studio di caso, Petti e Triacca (2019) hanno utilizzato il diario e l'incidente critico in relazione all'uso responsabile dei dispositivi digitali a scuola e della promozione della cittadinanza attiva. Gli incidenti sia inappropriati (ad esempio: inserimento nelle cartelle condivise di file con contenuti inappropriati), sia virtuosi (ad esempio: aiuto dato ai compagni incapaci di utilizzare il tablet), sono stati scelti dagli insegnanti e proposti alla classe (2019, p. 94).

Nel complesso gli studi confermano la stretta relazione tra incidenti critici, pratica riflessiva sull'esperienza e apprendimento (Parker et al., 1995, p. 112), che li considerano come una risorsa per accompagnare quei processi di metariflessione ritenuti oggi indispensabili per lo sviluppo delle competenze personali e professionali degli insegnanti e degli educatori.

Tripp (1994), impegnato osservare la vita professionale degli insegnanti, ritiene interessanti non tanto l'analisi di biografie olistiche o di resoconti cronologici continui, ma l'esame di frammenti della vita professionale pratica degli insegnanti, attraverso il metodo degli incidenti critici. Tripp precisa che la scelta di individuare una determinata situazione come "incidente" e di considerare la sua interpretazione come "critica" implica di per sé un giudizio di valore sulla situazione stessa. Gli incidenti, infatti, non andrebbero considerati come fenomeni oggettivi, cioè esistenti indipendentemente dall'osservatore, ma dipendono dal modo in cui si osserva la propria e altrui esperienza (Mortari, 2003). Gli incidenti rimandano pertanto a una situazione o a un evento che può indurre un senso di potenziale dissonanza, offrendo l'occasione per sperimentare difficoltà elaborative fra rappresentazioni tra loro divergenti e per utilizzare in modo personale tale esperienza, fornendo l'opportunità di generare soluzioni a eventuali problemi, ma soprattutto dando la possibilità di vedere gli stessi eventi problematici da altri punti di vista, che vengono condivisi.

10.2.7 Il portfolio

Il portfolio è uno strumento sviluppato in campo artistico e professionale introdotto successivamente nei contesti educativi da chi non condivideva l'uso eccessivo

dei test nella valutazione scolastica e intendeva interpretare la valutazione come strumento di regolazione dei processi e di promozione della capacità autovalutativa e autoregolativa degli studenti.

Se Arter e Spandel (1992) definiscono il portfolio dello studente come "particolarmente utile per far riflettere ciascuno di essi sull'evolversi della qualità del loro lavoro" (p. 36, 1992), Castoldi (2005) si propone di aiutare gli insegnanti a superare le ansie di adempimento e di considerare il portfolio come risorsa per coltivare la riflessione nel contesto scolastico e per ripensare le modalità didattiche e valutative più diffuse.

Pellerey (2004) conferma l'idea del portfolio come dispositivo che si appoggia agli strumenti di osservazione sistematica e che per questo consente di valorizzare le pratiche formative, autovalutative, valutative e orientative sviluppate, in quanto «ne viene a costituire come una memoria, un patrimonio esperienziale, che consente riflessione critica su quanto già realizzato e progettazione più accurata per quanto riguarda il futuro» (p. 271).

Domenici e Moretti (2006) suggeriscono un'ampia gamma di strumenti e tecniche di supporto al portfolio inteso come dispositivo di regolazione dei processi formativi in grado di caratterizzare le scelte della scuola autonoma e di indirizzarla in favore della equità. Chiappetta Cajola (2006) riprende una analoga proposta, attenta ai processi di inclusione degli allievi disabili in una prospettiva ispirata a principi di democrazia e di uguaglianza.

In questo senso il portfolio è inserito in una prospettiva di personalizzazione dei processi di insegnamento-apprendimento.

Numerose sono le sperimentazioni di portfolio nell'istruzione superiore e nell'educazione degli adulti, sia in formato a stampa sia in formato digitale, il cui utilizzo è volto a favorire la continuità educativa e didattica e soprattutto l'occupabilità degli studenti nella fase di transizione dal sistema di istruzione al mondo del lavoro (Giovannini e Moretti, 2010). In questi casi il portfolio professionale è volto a documentare, formalizzare e favorire la certificazione delle competenze acquisite dal soggetto.

In Italia la ricerca valutativa e didattica nonché l'attività di istruzione e formazione fanno un modesto impiego del portfolio degli studenti. I motivi sono vari, in particolare si considera bassa l'affidabilità delle informazioni registrate nei portfolio per il fatto che in molti lo utilizzano in quanto obbligatorio ma lo considerano uno strumento amministrativo e burocratico piuttosto che una risorsa da utilizzare per qualificare i processi di istruzione.

In ambito internazionale e nazionale, accanto al portfolio dello studente è stato introdotto anche il portfolio del docente (*teaching portfolio*), che riscontra un utilizzo nel contesto universitario, specie in relazione allo svolgimento dei tirocini formativi e dei laboratori nella formazione iniziale dei docenti e degli educatori.

Rossi et al. (2015) approfondiscono la valenza formativa, valutativa e di ricerca del *teacher portfolio* nell'ambito della formazione dei neo-assunti. La C.M. 6768 del 27/2/2015 ha infatti modificato il modello di formazione rivolto ai docenti neoassunti per l'anno di prova ed è stata introdotta l'elaborazione di un portfolio digitale che prevede lo sviluppo di alcuni ambienti di lavoro interagenti (per esempio: Curricolo formativo, Attività didattica in relazione a progettazione, documentazione e riflessione, Bilancio delle competenze iniziale, finale e formazione in servizio futuro).

Del portfolio docenti viene apprezzata la sua capacità di predisporre uno spazio di "reflective thinking" che permette agli insegnanti di riflettere sulle proprie competenze personali e professionali in vista del miglioramento del proprio agire educativo (Seldin et al., 2010).

Riepilogo

In questo capitolo è stato affrontato il problema della coerenza e/o adeguatezza degli strumenti in relazione alle ipotesi formulate dal ricercatore (scopi, obiettivi) e ai contesti di ricerca (confermativa o esplorativa).

Sono stati forniti alcuni esempi di strumenti spesso utilizzati nell'ambito della ricerca educativa: dalle check-list ai questionari di autovalutazione, dai diari di bordo al portfolio, ecc., e ne sono state esaminate le principali caratteristiche, i punti di forza ed eventuali limiti.

Riferimenti bibliografici

Arter J.A., Spandel V. (1992). Using portfolios of student work in instruction and assessment. *Educational measurement: Issues and practice,* 11(1), 36-44.

Bales R.F. (1950). *Interaction Process Analysis, a method for the study of small groups.* Reading (Mass.): Addison-Wesley

Bales R.F. (1970). *Personality and interpersonal behavior.* New York: Holt, Rinehart & Winston.

Castoldi M. (2005). *Portfolio a scuola.* Brescia: La Scuola.

Coggi C., Ricchiardi P. (2005). *Progettare la ricerca empirica in educazione.* Roma: Carocci.

Chiappetta Cajola L. (2006). *Il portfolio dell'allievo disabile. Democrazia, uguaglianza e processo di integrazione.* Roma: Anicia.

Damini M., Surian A. (2012). L'uso degli incidenti critici nella valutazione dello sviluppo delle competenze interculturali. *Italian Journal of Educational Research,* pp. 291-302.

De Frankrijker H. (1998). Cross-cultural learning from incidents, the critical incident method: some applications concernig the practice of teacher education and parent support. *European Journal of Intercultural Studies,* 9, pp. 55-70.

De Landsheere G. (1973). *Introduzione alla ricerca in educazione.* Firenze: La Nuova Italia.

Domenici G. (1981). *Descrittori dell'apprendimento.* Teramo: Giunti & Lisciani.

Domenici G.,moretti G. (2006). *Il portfolio dell'allievo. Autonomia equità e regolazione dei processi formativi.* Roma: Anicia.

Flanagan J. (1954). The critical incident technique. *Psychological Bulletin,* 51, pp. 327-358.

Flanders N.A. (1970). *Analyzing teaching behavior.* Reading (Mass.): Addison-Wesley.

Giovannini M.L., Moretti M. (2010). L'e-portfolio degli studenti universitari a supporto del loro processo di sviluppo professionale. *Quaderni di Economia del Lavoro,* 92, pp. 141-163.

Moretti G. (2003). *Pratiche di qualità e ricerca-azione. Costruire la scuola dell'autonomia.* Roma: Anicia.

Mortari L. (2003). *Apprendere dall'esperienza. Il pensare riflessivo nella formazione.* Roma: Carocci.

Nuzzaci A. (2017). Applicazione e analisi della Tecnica dell'Incidente Critico in alcune pratiche di ricerca sulla riflessività. Possibilità e limiti. In L. Ghirotto, *Formare alla ricerca empirica in educazione* (pp. 72-85). Bologna: Alma Mater Studiorum.

Olmetti Peja D. (1998). *Teorie e tecniche dell'osservazione in classe.* Firenze: Giunti.

Parker D.L. et al. (1995). The value of critical incident analysis as an educational tool and its relationship to experiential learning. *Nurse Education Today,* 15, 111-116.

Pellerey M. (2004). *Le competenze individuali e il portfolio.* Milano: La Nuova Italia.

Petti L., Triacca S. (2019). Educare all'uso responsabile dei dispositivi digitali a scuola. L'uso del diario e dell'incidente critico. In *Le Società per la società*: ricerca, scenari, emergenze. Atti del Convegno Internazionale SIRD, Roma 26-27 settembe 2019.

Pierro A., Livi S. (2001). Tecniche di osservazione sistematica dei processi di gruppo: attendibilità e validità della Interaction Process Analysis (IPA) e della SYstematic Multiple Level Observation of Groups (SYMLOG). *Psychofenia: Ricerca e Analisi Psicologica,* (6), 79-100.

Poliandri D., Muzzioli P., Quadrelli I., Romiti S. (2012). La Scheda di osservazione in classe: uno strumento per esplorare le opportunità di apprendimento. *Italian Journal of Educational Research,* 173-187.

Popper, K.R. (1934). Logik der Forschung. Wien: Verlag von Julius Springer (trad. it. *Logica della scoperta scientifica.* Torino: Einaudi, 1970).

Popper K.R. (1983). *Postscript to The Logic of Scientific Discovery* . Hutchinson: W.W.

Bartley (trad. it. Postscritto alla Logica della scoperta scientifica. Milano: Il Saggiatore, 1984).

Postic M., De Ketele J.M. (1993). *Osservare le situazioni educative. I metodi osservativi nella ricerca e nella valutazione.* Torino: Sei.

Rossi P.G., Magnoler P., Giannandrea L., Mangione G.R., Pettenati M.C., Rosa A. (2015). Il TeacherPortfolio per la formazione dei neo-assunti. *Pedagogia Oggi,* 223-242.

Seldin P., Miller J.E.,Seldin C.A. (2010). *The teaching portfolio: A practical guide to improved performance and promotion/tenure decisions.* New York: John Wiley & Sons.

Tripp D. (2003). *Critical incidents in teaching: developing professional judgment.* London: Routledge.

Tripp D. (1994). Teachers' lives, critical incidents, and professional practice. *Qualitative Studies in Education,* 7(1), 65-76.

Zabalza Beraza A.M. (2001). *I diari di classe. Uno strumento per lo sviluppo professionale dell'insegnante.* Torino: Utet.

11

Docimologia e valutazione educativa

In questo capitolo parleremo della nascita e dello sviluppo della docimologia. Verranno descritte inoltre in modo analitico le caratteristiche peculiari degli strumenti di valutazione degli apprendimenti e il loro utilizzo nell'ambito della didattica.
Partendo dalla considerazione che la *strutturazione del giudizio valutativo,* avvalendosi di elementi conoscitivi e informativi sia di tipo qualitativo sia di tipo quantitativo è sempre, e non può non essere, un *giudizio squisitamente qualitativo,* analizzeremo il fatto che nel caso esso sia fondato solo su dati qualitativi o su dati quantitativi, può rappresentare l'espressione di una estrema soggettività, oppure l'espressione di un giudizio ad alto livello di affidabilità (*ovvero valido e attendibile*).
Particolare attenzione sarà rivolta alla misurazione, al ricorso a criteri assoluti o a criteri relativi per la costruzione delle prove di valutazione, e alla strutturazione dei giudizi valutativi a livello di classe, di scuola, di sistema complessivo.
Si porrà infine attenzione al problema della comunicazione del giudizio valutativo, alla produzione e all'uso di informazioni utili per l'assunzione di decisioni ai diversi livelli di responsabilità del processo educativo.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Nascita e sviluppo della docimologia.
- Strumenti di valutazione degli apprendimenti.
- Didattica e valutazione.
- Misurazione, verifica e giudizi valutativi.
- Dati valutativi. Criteri assoluti e criteri relativi.
- Comunicazione del giudizio valutativo e informazioni.
- Autonomia scolastica, organizzazione didattica e valutazione.
- Strutturazione di decisioni che abbiano alta probabilità di successo.

"Se il signor Palomar facesse uso d'un telescopio le cose sarebbero più complicate sotto certi aspetti e semplificate sotto altri; ma, ora come ora, l'esperienza del cielo che interessa a lui è quella a occhio nudo, come gli antichi navigatori e i pastori erranti. Occhio nudo per lui che è miope significa occhiali; e siccome per leggere la mappa gli occhiali deve toglierseli, le operazioni si complicano con questo alzare e abbassare degli occhiali sulla fronte e comportano l'attesa di alcuni secondi prima che il suo cristallino rimetta a fuoco le stelle vere o quelle scritte. Sulla carta i nomi delle stelle sono scritti in nero su sfondo blu e bisogna accostare la lampadina accesa proprio addosso al foglio per scorgerli. Quando si alza lo sguardo al cielo lo si vede nero, cosparso di vaghi chiarori; solo a poco a poco le stelle si fissano e dispongono in disegni precisi, e più si guarda più se ne vedono affiorare."

Italo Calvino, *Palomar*

11.1 Nascita e sviluppo della docimologia

Alle sue origini - inizi anni Venti dello scorso secolo - la *ricerca docimologica* si è occupata principalmente della *validità* e dell'*attendibilità* degli strumenti di rilevazione dell'apprendimento in ambito scolastico, cioè della loro capacità di far cogliere davvero le conoscenze oggetto della verifica e con esiti condivisibili da parte di più valutatori. Le prove oggettive di accertamento degli apprendimenti e il loro corretto impiego rappresentano, per molti versi, l'esito-proposta più ampiamente conosciuto di quelle prime ricerche sorte dalla critica sistematica dei limiti metrologici delle prove "tradizionali", prove che, purtroppo, risultano ancora oggi molto diffuse, anche nel nostro Paese, nonostante le vistose distorsioni informative, e di conseguenza valutative e formative che esse producono. Basti qui segnalarne i principali: *dall'effetto di alone a quello di contrasto*, ovvero, rispettivamente, dalla alterazione di un giudizio valutativo su una data prestazione o compito dovuta a una indebita estensione di un precedente giudizio valutativo, alla influenza del giudizio relativo alla prova di un soggetto dovuta alla qualità della prestazione di un altro: si pensi alle interrogazioni orali; *dall'effetto di stereotipia o pregiudizio*, cioè

dalla influenza negativa di tali caratteristiche di chi valuta, sui giudizi valutativi che egli stesso formula, all'effetto pigmalione o *della profezia* - fatta dal valutatore sulla base della propria equazione soggettiva - *che si autoverifica* poiché lui o chi per lui mette in atto un insieme di comportamenti, attesi e simili, che progressivamente producono l'effetto previsto e/o desiderato, o altro: tutti effetti prodotti non per una "cattiveria" fisiologica o "parzialità" consapevole dei docenti o degli esaminatori, ma a causa della mancata tendenziale oggettività della rilevazione dei dati sulla cui base viene espresso il giudizio valutativo.

Eppure, i primi strumenti di rilevazione delle conoscenze scolastiche volti a diminuire il peso delle componenti soggettive nella strutturazione del giudizio valutativo - veri e propri antenati dei test di profitto - vennero elaborati sul finire della prima metà dell'800 nelle scuole pubbliche di Boston. L'iniziativa fu assunta in seguito a una espansione della popolazione scolastica che fece temere un abbassamento della qualità della formazione scolastica. Come è inevitabile, tra le dinamiche socio-culturali e le strutture e i modelli educativi sussiste quasi sempre una stretta relazione. Furono così messe a punto serie di domande scritte finalizzate a sollecitare prestazioni che manifestassero in modo tendenzialmente univoco le reali competenze degli allievi.

Si deve tuttavia attendere l'inizio del Novecento perché compaiano i primi test standardizzati per misurare gli esiti dell'apprendimento scolastico. Sulla scia di quanto avveniva in campo psicologico con la produzione dei reattivi mentali, della scala di misura dell'intelligenza di Binet-Simon, e altro, negli Stati Uniti furono elaborate vere e proprie batterie di test di profitto tarati su campioni rappresentativi di studenti (capaci perciò di fornire parametri di confronto nella misurazione dell'apprendimento in molte materie d'insegnamento). Tra queste lo *Stanford Achievement Test* del 1923.

È proprio a partire dagli anni Venti del Novecento, infatti, che la riflessione circa l'affidabilità delle valutazioni scolastiche, in particolare di quelle condotte in sede d'esame con le tradizionali forme di accertamento delle conoscenze, comincia a farsi sistematica e a suscitare l'interesse di molti studiosi. Le ricerche nel settore, staccandosi dalla psicotecnica, vengono progressivamente a costituire *un corpo teorico e concettuale autonomo, con proprie procedure d'indagine in grado di accrescerlo significativamente*, che daranno origine a una nuova scienza dell'educazione, la *docimologia*, il cui campo privilegiato d'interesse sarà per lungo tempo il superamento dei limiti delle pratiche valutative diffuse, proprio attraverso la proposizione di strumenti oggettivi di controllo dell'apprendimento, in particolare di quello scolastico. Le prime ricerche scientifiche sulla validità delle prove d'esame, condotte da H. Piéron (da sua moglie M. Angenout e da H. Laugier), a ragione considerato il padre della docimologia - e certo non solo per averne coniato il nome unendo in un articolo apparso nel 1922, il termine greco *dokimazo* (prova, saggio) con la de-

sinenza *logia* (discorso su), quanto soprattutto perché a lui si devono i primi studi sistematici sulla valutazione.

11.2 Produzione e diffusione dei test oggettivi di conoscenza e prime indagini sul profitto scolastico

La pluralità d'impiego dei test di conoscenza, utilizzabili oltre che nelle scuole - per la verifica iniziale, processuale e finale degli apprendimenti - anche per la selezione dei candidati all'assunzione nell'amministrazione statale e nelle aziende pubbliche e private, ha decisamente favorito la nascita in molti Paesi industrialmente avanzati, in primis negli Stati Uniti, di veri e propri centri di servizio, come il CEEB (*College Entrance Examination Board*) e l'*Educational Testing Program* per la produzione e l'applicazione di test per conto di università, scuole e altre istituzioni. Inoltre il crescere dell'attenzione dell'opinione pubblica verso la qualità del prodotto scolastico col crescere delle dimensioni dei sistemi formativi e con l'internazionalizzazione della spendibilità delle competenze acquisite a scuola, ha favorito, nel secondo dopoguerra, la nascita di organismi e associazioni mondiali (come l'IEA, *International Association for the Evaluation of Educational Achievement*, con sede principale a Stoccolma e sede legale in Belgio), che proprio attraverso l'impiego di batterie di prove oggettive standardizzate hanno promosso indagini comparative internazionali sul rendimento e sul profitto scolastico. Tra queste, le indagini internazionali IEA, per l'appunto, e, per ultimo, le indagini conoscitive PISA (Programme for International Student Assessment) promosse dall'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico) che nella loro settima edizione (2018) hanno coinvolto ben 79 Paesi.

Anche in forza delle dinamiche socio-culturali e politiche a cavallo degli anni Sessanta e Settanta, e della partecipazione alla seconda indagine internazionale IEA, voluta principalmente da Aldo Visalberghi, in Italia viene sviluppandosi una particolare attenzione verso la formazione delle nuove generazioni. Affiora così, progressivamente, anche la necessità di far creare un apposito centro di elaborazione e impiego di prove oggettive di profitto, per la verifica e l'innalzamento del prodotto e della produttività della istruzione scolastica, presso l'Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione (INVALSI), nato nel 1999 dalla trasformazione del preesistente Centro Europeo dell'Educazione (CEDE) istituito nel 1974.

Proprio a partire dagli anni Settanta, i docimologi rivolgono la loro attenzione alle *relazioni formali* che sussistono tra la *qualità delle informazioni* (volta a volta raccolte o disponibili) e le caratteristiche dei processi valutativi, da una parte, e la strutturazione delle decisioni didattico-organizzative a *livello di micro-, meso- e di macrosistema formativi* dall'altra, cioè ai diversi piani di responsabilità dei processi di formazione. Di decisioni, dunque, rispettivamente interne a ciascuna classe scolastica, perciò prevalentemente utili per la regolazione della didattica; riferite alle singole unità scolastiche, impiegabili per favorirne il processo autovalutativo e permetterne il "pilotaggio" in presenza di una più accentuata autonomia didattico-organizzativa; coinvolgenti, infine, l'intero sistema di istruzione e formazione, decisioni, perciò, insostituibili per la progettazione e il controllo delle azioni di politica educativa a livello nazionale.

11.3 Orientamenti della ricerca docimologica

Sono venuti così sviluppandosi rilevanti settori di indagine, taluni tuttora in espansione, tra i quali - oltre quello appena indicato relativo ai *rapporti tra valutazione e processi di decisione* a diversi livelli di responsabilità (docenti, capi di istituto, uomini di governo, eccetera) - risultano degni di nota il filone rivolto ad approfondire la *conoscenza dei fattori o delle variabili qualificanti le attività di formazione*, ovvero la *qualità dell'istruzione*; le *relazioni implicite ed esplicite tra processi di apprendimento e processi valutativi*, e perciò dei mezzi più adatti per regolare con alta affidabilità le reciproche connessioni, di cui si sono da noi particolarmente occupati Benedetto Vertecchi e Mario Gattullo; il filone che mira alla elaborazione e alla messa in atto di procedure sempre più agevoli e funzionali di *autovalutazione delle singole unità scolastiche* non solo dal punto di vista dell'*assessment* (verifica e valutazione degli apprendimenti fatti conseguire agli allievi), ma anche, se non soprattutto, dell'*evaluation* (cioè e particolarmente la valutazione delle caratteristiche organizzative della scuola); il settore d'indagine, ancora, volto alla determinazione di *indicatori nazionali e internazionali della qualità dell'istruzione*, utili sia per *analisi valutative di sistema, interne* a ciascun Paese, sia per *analisi comparative internazionali*; infine - ma non per ultimo se si considera l'effetto di impatto sull'opinione pubblica che hanno prodotto i recenti lockdown dovuti al Covid-19 - la *valutazione nei e dei processi formativi a distanza* e l'impiego, a tal fine, delle più avanzate tecnologie dell'informazione e della comunicazione, di cui si tratterà in un apposito capitolo di questo volume.

Gli indirizzi di ricerca appena segnalati sono quelli che si sono progressivamente differenziati nello spazio geopolitico a livello mondiale, e lungo l'asse del tempo che va dagli anni Venti del secolo scorso agli attuali anni Venti, ma che hanno perseguito e perseguono un comune obiettivo: il miglioramento della proposta educativa e dei suoi risultati, impiegando a tal fine la leva dei processi di verifica e valutazione delle variabili educative più importanti, primariamente l'apprendimento e l'insegnamento, sia quando la soluzione didattica privilegiata risulti essere "in presenza", sia nelle formule didattiche "a distanza", "miste" o "blended". D'altro canto, la progressiva globalizzazione dei mercati, con la messa in atto di un processo

per molti versi paradossale, di competizione e nel contempo di cooperazione tra i diversi Paesi, e l'esplosione - in anni non molto lontani - di crisi economiche e finanziarie con i conseguenti problemi della disoccupazione e della bassa produttività del servizio scolastico hanno imposto all'attenzione pubblica e ai decisori politici la questione della razionalizzazione e al tempo stesso del miglioramento della spesa per l'istruzione e del contestuale incremento dell'efficacia culturale e formativa dei sistemi e degli interventi riformatori in campo educativo.

La docimologia ha così, via via, ampliato gli orizzonti dei propri interessi in stretto rapporto con il mutare dei problemi valutativi posti dal cambiamento dei contesti sociali e dei più o meno corrispondenti sistemi di istruzione e formazione. Quasi tutti gli indirizzi di ricerca prima indicati, contemplano, come si può facilmente intuire, aree tematico-concettuali tra loro intersecabili, non di rado sovrapponibili.

Tuttavia, l'*attenzione verso la struttura formale* e *le qualità metrologiche degli strumenti di raccolta dei dati valutativi*, ovvero, verso il problema dell'*adeguatezza strutturale delle prove di verifica alle tipologie storicamente mutevoli dei saperi e delle competenze da promuovere (e valutare) soprattutto in ambito scolastico* ha in qualche modo rappresentano il sottile filo rosso che, soprattutto in Italia, fino ad oggi ha collegato i diversi indirizzi di studio di questa disciplina. E ciò, certamente in relazione alle esigenze dello sviluppo tecnologico, ma anche in funzione delle richieste sempre più pressanti nei Paesi democratici di una partecipazione più consapevole e critica dei cittadini alla vita pubblica, quindi di una formazione mutevole, ma sempre congruente con tali apprezzabili finalità.

11.4 Nuovi strumenti per la verifica dell'apprendimento. Le prove semistrutturate

All'interno di quest'ultimo specifico filone d'indagine, in Italia, le proposte valutative che sono derivate dagli esiti delle ricerche docimologiche - avviate da Aldo Visalberghi e Luigi Calonghi - hanno via via trovato meritata fortuna soprattutto tra gli insegnanti più consapevoli e professionalmente attrezzati. Cioè tra quei docenti che proprio perché avevano cominciato a impiegare più o meno diffusamente la pratica del testing e ne apprezzavano l'apporto informativo nel processo di strutturazione delle decisioni didattiche e di espressione dei giudizi valutativi, più di altri avvertivano la necessità di integrare, se non addirittura superare, l'uso delle stesse prove oggettive. Superarlo, tuttavia, attraverso l'impiego di strumenti più flessibili e "intelligenti", in senso letterale, di rilevazione della molteplicità e complessità delle conoscenze e delle competenze che sempre più risultava indispensabile promuovere nelle nuove generazioni attraverso i processi formali di istruzione anche, se non soprattutto, attraverso il processo di valutazione, a causa del suo

forte impatto cognitivo ed emotivo.

È proprio partendo dal *bisogno empirico di rendere sempre coerente il mezzo di rilevazione e promozione dell'apprendimento con l'oggetto e l'obiettivo stesso della verifica* che molti docenti hanno usato o ogni giorno impiegano - spesso creandoli appositamente - gli strumenti valutativi più disparati, che poi essi stessi, e/o in collaborazione con esperti di settore, affinano sul piano strutturale e metrologico. Anche sulla base di tali spinte *bottom up*, sono state via via studiate e perfezionate sul piano formale, metrologico e strutturale, le cosiddette *prove semistrutturate di verifica degli apprendimenti.*

Prove che, sviluppatesi seguendo pressappoco tale logica, in tempi diversi e in modo differenziato e parcellizzato qua e là in differenti contesti geo-educativi (i saggi brevi, per esempio, soprattutto in area francofona, mentre le domande strutturate nei Paesi di cultura e lingua inglese), hanno paradossalmente trovato in Italia, paese fino a circa trent'anni fa sempre in ritardo rispetto agli orizzonti della ricerca didattico-valutativa internazionale, un terreno di coltura certo non molto esteso, ma sicuramente più fertile che altrove. Un fatto che ben si coglie dall'esame della pur limitata letteratura internazionale relativa alla tipologia di prove di verifica formalmente definibili come semistrutturate, manifestazione peraltro, della forza prorompente di quella corrente di pensiero, cioè, che se non soggiaceva supinamente al peso della consuetudine valutativa nazionale imperniata sulla soggettività degli strumenti e dei giudizi, certo non si attardava a proporre pedissequamente in Italia, per mera esterofilia, ciò che in altri paesi si andava facendo. Quelle prove nascono, non a caso, dalla convinzione teorico-culturale, nonché dall'idea e dalla constatazione empirica che la conoscenza, spesso assai ricca, complessa e articolata che si vuole promuovere, soprattutto in ambienti formali di insegnamento-apprendimento, possa, anzi, debba essere in buona parte sviluppata e poi verificata, facendo prevalentemente ricorso a contesti opportunamente simulati, poiché la conoscenza è per sua "natura" reticolare e modulabile sia nei processi che la incrementano e la potenziano, sul piano individuale come collettivo, sia nelle modalità di impiego richieste o "imposte" spesso a ognuno, dalle più disparate e inusuali situazioni sociali, culturali, di lavoro. Ragioni poste poi alla base di forme attuative di una cosiddetta *valutazione autentica.* Seppur in forma impropria, alcune di quelle prove sono state progressivamente introdotte negli esami di Stato al termine della scuola secondaria superiore italiana sul finire degli anni Ottanta, per la riconosciuta loro maggiore efficacia valutativa rispetto a buona parte dell'armamentario docimologico impiegabile per la rilevazione e la valutazione di apprendimenti complessi, quali certamente sono quelli da far conseguire nella scuola secondaria superiore, così da promuoverne l'uso di alcune in modo più sistematico negli esami riformati, a partire dal 1998.

Più o meno nello stesso periodo, quelle da noi definite *prove semistrutturate* (Do-

menici, 1981, 1991, 2001, 2020) hanno fatto il loro ingresso in modo parziale, anche se più diffuso che nel passato, nelle ricerche comparative internazionali promosse da differenti organismi (un primo tentativo si era registrato con l'indagine IEA del 1989-'92, in Italia contraddistinta con l'acronimo SAL - Studio Alfabetizzazione Lettura - seguito poi, più sistematicamente, dalle indagini OCSE-PISA sulle conoscenze dei quindicenni in matematica, scienze e "lettura").

Oggi, finalmente, va sempre più diffondendosi la consapevolezza della spiccata adeguatezza di queste prove rispetto a molte altre, di far rilevare competenze, ovvero modalità di impiego strategico e tattico delle conoscenze ovunque e comunque acquisite da ognuno, nei tentativi di risolvere un qualsiasi problema di natura pratica, come teorico-concettuale, in una situazione data; quindi la loro capacità di far cogliere taluni dei saperi posseduti dagli studenti e l'abilità di selezionare, all'occorrenza, quelli utili per risolvere un problema dato e per applicarli in situazioni nuove; di far conoscere le procedure valutative impiegate per selezionarli a tal fine, e le modalità di revisione e adattamento al contesto degli stessi saperi padroneggiati; di far affiorare gli schemi privilegiati nella strutturazione delle decisioni e l'abilità di ristrutturarli sulla base dei dati di fatto, di contesto e valutativi disponibili e/o appositamente acquisiti, in una parola, di far comprendere gli aspetti formali delle strategie di apprendimento.

Qualità, queste, dei processi di apprendimento e di impiego delle conoscenze individuali che né le prove tradizionali né quelle oggettive riescono a far cogliere e, al tempo stesso, contribuire a promuovere. D'altro canto, come si è più volte ribadito, gli strumenti impiegati per la rilevazione, la verifica e la valutazione degli apprendimenti rappresentano, pur sempre, dei veri e propri *costrutti teorici*. Ancor prima che «mezzi», «utensili», «oggetti materiali» in senso lato, rappresentano infatti una idea, un concetto, una teoria, una ipotesi interpretativa della realtà che cercano di cogliere e raffigurare: quella, nel nostro come in altri Paesi a democrazia liberale, della *qualità di quei saperi complessi, ritenuti di grande rilievo per la formazione dell'uomo nel mondo di oggi e di un domani più o meno prossimo, ovvero di chi deve poter apprendere, criticamente ma anche autonomamente, nei diversi periodi e contesti di vita*. Le caratteristiche formali degli strumenti valutativi impiegati diffusamente nelle strutture educative e nei sistemi di istruzione, non solo dunque contribuiscono significativamente a qualificare - in bene o in male ovviamente - la didattica, ma simboleggiano anche, se non soprattutto, le scelte di politica culturale di ogni Paese.

Se l'alta qualità dell'apprendimento diffusamente promosso in tutti gli allievi non può non essere considerata il *criterio aureo di confronto* con cui commisurare ogni altro elemento, qualità compresa, della struttura, della organizzazione e della funzionalità sociale di ogni sistema educativo al fine di stabilirne valore ed efficacia, si comprende allora quanta importanza assuma l'impiego di prove adatte allo scopo di istruire e formare cittadini, consapevoli e critici, del proprio Paese e del mondo. Peraltro, con il complicarsi del contesto socioculturale e produttivo entro cui soprattutto la scuola è chiamata a operare, il legame tra didattica e valutazione è venuto progressivamente intensificandosi fino a far assumere ai due elementi - sulla base di accorti processi di analisi e meta-analisi di evidenze fattuali e di ricerche empiriche - una interdipendenza reciproca così spinta da farli postulare e teorizzare, nelle strategie educative più avanzate, come processi per molti aspetti reciprocamente integrati. Nel senso almeno che le forme e le modalità attraverso cui vengono offerte agli alunni le opportunità di "apprendimento insegnato" - direttamente o indirettamente guidato - ormai derivano soprattutto, sebbene non solo, da quelle con le quali si dà corso al processo o percorso valutativo e autovalutativo, si assumono e si impiegano i dati valutativi soprattutto *durante*, oltre che *prima* e *dopo*, l'attività di apprendimento e di insegnamento da parte di allievi e docenti.

11.5 Modelli e strumenti valutativi come costrutti. Didattica e valutazione

A ragione, perciò, nel loro complesso tecnico-procedurale, le soluzioni date ai problemi della *valutazione del e nel sistema di istruzione e formazione* possono considerarsi come un vero e proprio costrutto politico-educativo oltre che teorico-culturale. Gli strumenti impiegati per la verifica e/o la misurazione del valore assunto da una o più variabili del processo di insegnamento-apprendimento; le procedure seguite per la rilevazione e il trattamento dei dati valutativi; i processi attivati per la costruzione e la comunicazione del giudizio e, soprattutto, le finalità delle *verifiche* e delle *valutazioni*, sono tutti elementi che complessivamente rappresentano un sistema teorico-pratico che, oltre a derivare dall'evoluzione della ricerca di settore, costituisce nella sua più estrema e positiva materialità, il concetto di formazione dell'uomo proprio di chi impiega quel sistema, talvolta persino oltre le sue stesse intenzioni manifeste.

L'attività valutativa, per quanto complessa e articolata, precede, accompagna e segue, soprattutto le azioni finalizzate, ma anche quelle più disinteressate, ed entrambe, come si sa, caratterizzano peculiarmente i processi formativi scolastici, in particolare la didattica. Gli esiti conoscitivi cui si perviene con il suo progressivo svolgimento orientano infatti l'agire indirizzandolo verso direzioni e traguardi desiderati. Ben lontana dal rappresentare solo l'espressione di un giudizio, peraltro considerato definitivo, come a livello di senso comune erroneamente si crede, l'attività valutativa si esplica attraverso una serie articolata di operazioni tra esse distinte ma reciprocamente connesse, tutte di grande rilievo, anche, se non soprattutto, sul piano formativo. Per cominciare, attraverso l'operazione di vera e propria *attribuzione di valore a fatti, dati, procedure, eventi (educativi) o a talune loro qualità, in rela-*

zione agli scopi che colui che "valuta" vuole perseguire: un fatto ritenuto positivo o negativo in un contesto e in relazione a certe finalità può infatti non essere considerato tale in ambienti differenti, con allievi diversi, in presenza di altri vincoli e in vista di ben altri obiettivi. Con un sistema esplicito di "classificazione" o discriminazione degli eventi e/o degli oggetti da valutare diventa perciò più facile promuovere volta a volta quelli ritenuti favorevoli ed evitare o inibire quelli considerati invece controproducenti. Disporre quindi di un tale sistema, per quanto empiricamente definito, aumenta di fatto la capacità modificativa del contesto, rende possibile sottoporre a vaglio critico le scelte (non solo didattico-valutative) compiute.

11.6 Misurazione, verifica e giudizi valutativi

Coerentemente con la precedente operazione, vi è poi quella che rimanda alla *verifica* attraverso la *rilevazione* - diretta o indiretta, durante lo svolgimento dell'azione formativa o in appositi momenti - delle informazioni e dei dati valutativi ritenuti di particolare rilievo, attraverso l'impiego di quegli strumenti, tra i tanti possibili - in primis prove strutturate e semistrutturate -, reputati più adatti alle necessità valutative. Gli strumenti di rilevazione da impiegarsi possono essere i più diversi e a differente grado di formalizzazione proprio perché, come si è visto, l'affidabilità dei dati raccolti deriva in gran parte dalla loro coerenza con le ragioni e gli "oggetti" della *verifica*; con le caratteristiche degli allievi; con la logica dei materiali e dei processi di apprendimento.

Le informazioni cui si perviene attraverso le rilevazioni e l'analisi critica dei dati - sia qualitativi che quantitativi - possono effettivamente abbassare l'ambiguità interpretativa dei problemi incontrati; chiarire meglio gli elementi costitutivi del contesto, talvolta persino contribuire a costruirlo in funzione delle esigenze di formazione; offrire indizi per far approfondire questioni complesse o mal poste, in definitiva possono far meglio strutturare il quadro conoscitivo a disposizione dei principali attori del processo di istruzione. Per tale ragione l'impiego di strumenti per la rilevazione delle conoscenze attraverso l'osservazione diretta e/o indiretta, sistematica o quantitativa, e/o non-sistematica, esperienziale o qualitativa (diari, diari di bordo, racconti di vita ecc.); oppure attraverso l'uso di prove assai differenziate (verbali, scritte, grafiche, operative, pratiche ecc.) a sollecitazione o *stimolo aperto e risposta aperta*, quindi a carattere soggettivo, oppure a *stimolo e risposta chiusi*, come le prove strutturate o oggettive, oppure, ancora, a *stimolo chiuso e risposta aperta* ma da formularsi rispettando vincoli formali assegnati, come le *prove semistrutturate* (questioni già accennate nei capitoli precedenti) e così elencando, e più in generale l'impiego di sistemi di *rilevazione* e *memorizzazione* dei dati valutativi, verranno decisi volta a volta in piena autonomia professionale dagli insegnanti o dai ricercatori, ma sempre in stretta relazione con le differenziate ragioni

della didattica e/o delle ipotesi di ricerca.

Quello che tuttavia occorre qui distinguere sono gli aspetti formali che caratterizzano i diversi momenti che portano all'espressione di un giudizio valutativo.

Le prove da somministrare vengono ovviamente scelte in ragione degli obiettivi della verifica. La *verifica* rappresenta perciò a sua volta l'atto di rilevazione di ben definiti aspetti qualitativi e/o quantitativi del grado di possesso di una certa qualità da parte dei soggetti cui la prova si somministra. Quando lo strumento di rilevazione lo permetta, assieme al dato qualitativo può determinarsi anche il valore da esso assunto sul piano quantitativo: nel primo caso l'apprezzamento valutativo potrà essere espresso con aggettivi, avverbi, voti e simili, che, come si è visto nel Capitolo 4, possono organizzarsi in forma di scala di misura nominale e/o ordinale, a prevalente *carattere qualitativo*; nell'altro caso, il giudizio potrà esprimersi anche attraverso un ben preciso riferimento a *dati quantitativi*, quindi a punteggi di una scala di misura a intervalli regolari. In questo caso alla rilevazione di certe qualità viene associata l'operazione di *calcolo della misura* in cui essa è posseduta da parte dei soggetti cui la prova è stata somministrata.

La *strutturazione del giudizio valutativo* poi, avvalendosi di elementi conoscitivi e informativi sia di tipo qualitativo sia di tipo quantitativo, è sempre, e non può non essere, un *giudizio squisitamente qualitativo*. E tuttavia, a seconda che sia fondato solo su dati qualitativi o su dati quantitativi (o su entrambi assieme), può rappresentare, nel primo caso, l'espressione di una soggettività al quadrato - data dall'interazione dell'equazione soggettiva dell'allievo e di quella di chi struttura il giudizio - oppure, nel secondo caso, l'espressione di un *giudizio ad alto livello di affidabilità (ovvero valido e attendibile) in quanto fondato su rilevazioni oggettive.*

11.7 Attribuzione di significato ai dati valutativi. Criteri assoluti e criteri relativi

Nelle macrodecisioni (per il governo dell'intero sistema educativo, delle unità scolastiche e dei processi di insegnamento-apprendimento) come nelle tante microdecisioni che docenti e allievi assumono più o meno consapevolmente ogni giorno, il livello di pertinenza e di efficacia delle azioni che in conseguenza di esse si intraprendono, quindi il grado della loro adeguatezza al contesto e alle necessità effettivamente emerse, la loro coerenza con gli obiettivi di fatto perseguiti, e in ultima analisi il loro successo educativo, dipendono non solo dall'affidabilità dei dati e delle informazioni raccolte, ma anche e soprattutto dalla *capacità di connetterli, interrogarli e farli "dialogare" tra loro.*

La messa in relazione dei dati così assunti può perciò far emergere finalmente, quella rete di significati, quel sistema di relazioni tra dati, fattori e variabili sulla cui base solamente diviene possibile *ri-adattare*, a seconda dei momenti, *tattiche e*

strategie formative, di apprendimento e auto-apprendimento adottate, contribuendo a rendere la proposta educativa sempre adeguata e significativa per ogni allievo e nel contempo coerente con gli obiettivi perseguiti, e a indirizzare verso il miglioramento della qualità dell'istruzione. La valutazione attua così pienamente la sua rilevante funzione formativa.

Quando si compie una simile operazione è bene sapere, tuttavia, se con gli strumenti di verifica degli apprendimenti si sono adottati *criteri valutativi normativi* (cosiddette *prove a norma*) o *criteri assoluti* (cosiddette *prove a criterio*) (cfr. anche Capitolo 12).

Approfondimento 11.1

COSA SIGNIFICA VALUTARE CON CRITERI RELATIVI E CON CRITERI ASSOLUTI

I primi di questi criteri mettono in particolare evidenza la posizione relativa del punteggio o del soggetto rispetto ad altri punteggi omologhi: il valore, ovvero l'apprezzamento del singolo risultato, dipende dall'andamento dei risultati complessivi del gruppo di soggetti o di prove prese a riferimento. Se, ad esempio, si accettano come positivi tutti i punteggi che si situano sopra alla media di tutti i punteggi, evidentemente i risultati individuali, di ognuno, dipenderanno molto dalla preparazione e perciò dai risultati dei componenti l'intero gruppo. Se il singolo fa parte di un gruppo di elementi appartenenti per la caratteristica rilevata a un particolare universo di soggetti, evidentemente il risultato individuale dipenderà dall'andamento della distribuzione della qualità esaminata, che è assai probabile che rispecchi la curva *normale,* da cui prende il nome questo tipo di valutazione. Una curva gaussiana, questa, in cui appunto media, mediana e moda tendono o risultano coincidenti - come ormai dovremmo sapere - così che le analisi statistiche volte eventualmente a meglio indagare i risultati della prova faranno riferimento alle dimensioni dello scostamento dei dati reali rispetto a quella curva, assunta perciò come parametro di riferimento. Le statistiche postuleranno la popolazione in esame come avente quelle caratteristiche. Le statistiche saranno anch'esse di tipo parametrico.

Se viceversa, anziché "valutare" una prestazione in rapporto a quelle degli altri la si apprezza in relazione a una *soglia,* a uno *standard prefissato* che assume per tutti un *valore assoluto di riferimento,* l'esito non dipenderà dall'andamento dei risultati complessivi, ma solo dal confronto del risultato di ciascuna prova rispetto al criterio valutativo adottato e fissato precedentemente.

Ovviamente, non può dirsi che una scelta valga più dell'altra. Sarebbe tuttavia buona norma impiegarle, quando sia possibile, entrambe, per evitare che i criteri assoluti siano troppo distanti dalle qualità possedute dalla popolazione esaminata. Tuttavia, si comprende bene che se si assumessero con prove parametriche dei cardiochirurghi, si correrebbe il rischio di scegliere i migliori, anche quando questi risultassero tanto ignoranti da non non sapere nemmeno cosa sia e come funzioni un cuore, ma meno ignoranti dei loro colleghi esaminandi. Una soglia minima, magari fissata anche sulla base di pregresse valutazioni normative, non farebbe incorrere in tale marchiano errore valutativo!

11.8 Restituzione e discussione dei dati valutativi

La "restituzione" dei dati valutativi agli allievi - ma anche ai soggetti coinvolti o interessati dal processo valutativo, ovvero agli *stakeholder* - costituisce senza dubbio l'operazione più importante, delicata e densa di conseguenze educative di tutta la complessa attività valutativa. In assenza di questo "momento" altamente formativo o con una eccessiva riduzione del tempo e delle energie a esso dedicati, la valutazione non ha senso se non quello della mera fiscalizzazione selettiva delle prestazioni degli studenti. A seconda dei casi, la "restituzione" dei dati valutativi può avvenire durante lo svolgimento (quando, in forma laboratoriale o in laboratorio prevale la dimensione dell'imparare facendo) o subito dopo il termine delle differenti attività di insegnamento e di apprendimento. La scelta sarà compiuta evitando che l'informazione di ritorno data all'allievo inibisca in qualche modo il processo di apprendimento o di affioramento della *consapevolezza delle proprie strategie cognitive e motivazionali.* La "restituzione" altro non è che riflessione, soprattutto "a freddo", sui punti forti e critici delle strategie individuali; discussione argomentata circa le più probabili concause degli esiti registrati; individuazione delle migliori vie da seguire per avvicinarsi ai traguardi desiderati. La valutazione, indicando e suggerendo positivamente a ciascuno direzioni e azioni di sviluppo adeguate; sostenendo l'impegno cognitivo e il coinvolgimento emotivo dell'allievo, si fa di fatto, in questa fase, "pro-attiva", costitutiva dello stesso processo di apprendimento.

11.9 Strutturazione (documentazione) e comunicazione del giudizio valutativo

Gran parte della ricostruzione, comunicazione e condivisione dell'esperienza educativa dei docenti, così come quella degli allievi, deriva dai processi di memorizzazione e di conservazione di particolari dati valutativi sintetici e/o analitici, di documenti, di strumenti impiegati magari come criteri valutativi di riferimento, e simili.

La documentazione sistematica delle attività formative e dei loro esiti facilita la riflessività individuale e di gruppo permettendo l'attivazione di procedure di analisi critica e di eventuale modifica delle finalità, dei criteri, delle metodologie e degli strumenti valutativi impiegati, della loro adeguatezza alle differenti funzioni della didattica e della valutazione nei particolari contesti operativi.

La registrazione e la documentazione relative allo sviluppo degli apprendimenti e delle competenze degli allievi può avvenire impiegando i più diversi strumenti, dal libretto al dossier personale dell'allievo fino al portfolio e all'e-portfolio.

La formulazione di un giudizio valutativo più degli altri momenti caratterizzanti il processo di valutazione etero e auto-condotto - *definizione degli obiettivi; costru-*

zione della prova di rilevazione e verifica degli apprendimenti; somministrazione dello strumento; rilevazione, trattamento, analisi e interpretazione dei dati valutativi - presuppone e comporta, come abbiamo detto già, la messa in atto di un attento e raffinato esame di tutti i dati e di tutte le informazioni utili e/o disponibili sull'oggetto specifico o su suoi elementi costitutivi per i quali si ritiene necessario esprimere apprezzamenti, ovvero una valutazione, un giudizio puntuale, articolato e argomentato.

Come si può ben comprendere, per il suo distacco dai singoli fatti procedurali dalla cui analisi tuttavia è venuto via via strutturandosi attraverso tanti giudizi parziali, interlocutori, ponderazioni di dati quali-quantitativi, il giudizio valutativo, specie se conclusivo di un consistente periodo di istruzione o formazione e/o di apprendimento, assume uno spiccato valore di analisi sintetica retrospettiva della qualità e del valore delle azioni compiute e e una forte valenza prospettica - di orientamento pro-attivo verso un futuro più o meno prossimo - se non addirittura prognostica.

11.10 Autonomia scolastica, organizzazione didattica e valutazione

L'eccessiva variabilità del prodotto culturale scolastico tra aree geografiche, tra centro e periferia dei grandi agglomerati urbani e persino tra gli allievi di un medesimo istituto, e di una stessa classe, permane ancora oggi, nonostante l'autonomia di ogni singola scuola rispetto ad alcuni dettati nazionali, come una delle costanti del nostro sistema, così come hanno fatto registrare tutte le indagini sui risultati formativi della scuola italiana svolte fino ad oggi dalla fine degli anni Sessanta, perpetuando forme inaccettabili di iniquità, perché dovute non al demerito o alla casualità delle dotazioni individuali, bensì alla casualità del luogo di nascita e della scuola frequentata.

L'istituto dell'autonomia era stato invece e giustamente ideato e pensato come il mezzo principe per far tendere la scuola alla promozione, nel breve termine, del pieno successo formativo di ciascuno e di tutti gli allievi di ognuna e di tutte le unità scolastiche del Paese. È con un'autonomia ben interpretata e governata che può, infatti, aver luogo la messa in atto di quel delicato, ma decisivo processo - per quanto apparentemente contraddittorio - di *valorizzazione delle diversità inter e intraindividuali; inter e intrascolastiche nonché inter e intraterritoriali* per promuovere in tutti i fruitori della formazione una più elevata base conoscitiva comune su cui solamente è possibile innestare una differenziazione non più legata alla casualità del luogo di nascita e della scuola frequentata, perciò meno discriminativa sul piano sociale e più equa.

L'autonomia andrebbe più correttamente intesa, perciò, come via per poter raccordare i processi formativi formali alle caratteristiche peculiari individuali e a quelle del contesto socio-culturale territoriale specifico di riferimento, poiché è tale operazione di individualizzazione della proposta didattica che può generare il successo educativo diffuso.

Se dunque l'istruzione promossa nella popolazione studentesca rappresenta il prodotto principale della scuola, è con essa che va parametrato ogni elemento costitutivo della sua struttura e delle sue parti, della sua organizzazione culturale e didattica, quando si voglia valutarne il contributo qualitativo che apporta al conseguimento intenzionale e condiviso dello scopo per cui la scuola esiste.

L'autonomia può rendere possibile, con le scelte di ogni scuola e pur sempre nel quadro dei vincoli nazionali, l'adeguamento di ogni proposta di istruzione tanto alle caratteristiche di ogni contesto socio-economico-culturale e di chi deve apprendere, quanto e contemporaneamente alle finalità sociali dell'istruzione, cioè al profilo culturale, definito e atteso, della popolazione in uscita dai diversi livelli e gradi di scuola. L'autonomia andrebbe perciò intesa come strumento di adeguamento di ogni proposta, in estrema sintesi, sia ai traguardi di istruzione e formazione nazionali, sia a quelli della quota parte locale dei curricoli scolastici.

Approfondimento 11.2
AUTONOMIA E RAPPORTO DI AUTOVALUTAZIONE

L'istituto dell'autonomia, di fatto, viene a rappresentare il tramite per innalzare processi e prodotti dell'istruzione, proprio attraverso la messa in atto di una didattica flessibile, capace di individualizzare l'insegnamento-apprendimento per consentire a tutti il raggiungimento di comuni, ma elevati, traguardi di istruzione.

Ma occorre che i dati valutativi siano affidabili, capaci cioè di rispecchiare di fatto la situazione che vogliono rappresentare, e che non mutino di significato complessivo a seconda di chi li abbia raccolti o li diffonda. Solo così diventa possibile, anzi abbastanza agevole far esercitare ad ognuno, ai pur differenti livelli di responsabilità, in primis ai dirigenti scolastici e ai docenti, una leadership effettiva, cioè le funzioni di indirizzo, attuazione, gestione, controllo e modifica del proprio e/o dell'altrui operato, ovvero di ognuna e delle differenti azioni e attività che hanno luogo nella e per la struttura educativa. L'alta qualità dei processi si ha tra l'altro quando la proposta culturale, didattica, risulti adeguata alle esigenze individuali, ma sempre in vista di ben precisi obiettivi educativi.

La coincidenza e/o lo scarto dell'esito delle varie scelte compiute, delle relative azioni e attività, così come, peraltro, dovrebbe emergere dal Rapporto di Autovalutazione (RAV), dai risultati attesi e persino, nei momenti delle macrodecisioni, dal profilo di allievo che si è delineato, offrono informazioni preziose sul grado di legame tra quelle scelte, i corrispettivi processi di attuazione, monitoraggio e controllo, da una parte, e gli effetti prodotti dall'altra. *Queste informazioni abbassano di molto il livello di ambiguità interpretativa dei problemi*, mettono tutti in gra-

do di porre in discussione criticamente idee e conoscenze che si credevano validi e affidabili – proprio così come si procede nel campo della ricerca attraverso *congetture* e *confutazioni* anche di teorie o assunti consolidati – di "suggerire" direzioni e interventi complessivi e/o specifici e articolati di miglioramento.

11.11 Promuovere la cultura della valutazione

Con le pur dovute differenziazioni di peso e di potere decisionale dei diversi attori coinvolti e delle qualità metrologiche degli strumenti utilizzabili, la metodologia procedurale di un affidabile processo valutativo rimane la stessa tanto che si operi a livello di classe e di istituto, quanto del sistema scolastico nel suo complesso o delle sue articolazioni orizzontali e verticali.

Grammatica e sintassi della valutazione scientificamente fondata e democraticamente accettabile sono regole formali che non cambiano significativamente col mutare dei livelli di riferimento indicati. Si evolvono come ogni sapere e ogni linguaggio, anche se l'uso che della valutazione si può fare dipende forse più di ogni altra pratica dalle scelte politiche. Una solida cultura valutativa critica, adatta quelle regole generali ai contesti e alle situazioni, progredisce non già inseguendo questa o quella moda, questa o quella proposta occasionale anche quando sia ministeriale, ma permette di "leggere" e interpretare con più dati informativi a propria disposizione i processi e gli eventi educativi scolastici, e offre maggiori strumenti di analisi e di proposte di intervento modificativo positivo.

Utilità, praticabilità, appropriatezza, accuratezza della valutazione; condivisione degli scopi e degli usi che di essa si faranno, così come la condivisione delle procedure, dei criteri e degli strumenti che verranno impiegati, ma anche degli oneri che in qualche modo essa comporta e conoscenza dei benefici che possono ricavarsi, sono elementi costitutivi di un sistema o modello valutativo affidabile, adeguato e moderno.

11.12 La valutazione come risorsa

La valutazione, come fin qui si è potuto constatare, rappresenta dunque, o dovrebbe rappresentare sempre, l'esito di una attenta ponderazione dei dati quantitativi – comprensiva delle misurazioni – e di quelli qualitativi, assunti con gli strumenti più disparati, ma affidabili e coerenti. Essa presuppone e comporta l'espressione di tanti giudizi progressivi che, soprattutto in taluni momenti, producono effetti di grande rilievo sulle azioni che a essa conseguono.

La valutazione, può non solo ergersi a sistema di promozione fattuale dell'equità, e quindi di alta formazione civile e democratica degli allievi, ma può persino diventare – come si è accennato – un volano della formazione, una vera e propria risorsa aggiuntiva quando venga impiegata come strumento principe per assumere il massimo di informazioni significative affidabili sugli elementi dei problemi formativi che strategicamente e tatticamente, ai differenti livelli di responsabilità e articolazione dell'intero sistema educativo, si affrontano al fine di ridurne l'ambiguità interpretativa: esattamente come avviene quando, inciampando in una difficoltà o in una situazione in cui un certo equilibrio culturale, per esempio tra due fatti o un fatto e una teoria, si è rotto o è venuto meno, il ricercatore con la ricognizione longitudinale e trasversale della letteratura cerca di disambiguarlo per giungere, successivamente alla formulazione delle ipotesi risolutive del problema. È infatti tale riduzione che può favorire l'identificazione delle vie didattiche risolutive e perciò le decisioni più appropriate per rendere qualitativamente alta la proposta di formazione.

È allora utile considerare alcuni dei possibili fattori di arricchimento dei contesti educativi formali che possono far diventare procedure ed esiti della valutazione in ambito formativo, non solo scolastico, vere e proprie risorse aggiuntive almeno rispetto a quelle di cui diffusamente o ordinariamente si dispone. Risorse strategiche e tattiche delle quali potrebbero finalmente disporre tanto i principali attori dei processi di insegnamento-apprendimento – in primis docenti, allievi e genitori – quanto i decisori politici e il pubblico direttamente o indirettamente interessato ai problemi educativi.

Affinché ciò sia possibile vi è una condizione necessaria anche se non sufficiente: la creazione di una rete entro cui far scorrere in tutte le direzioni provviste di senso, flussi di informazioni e dati valutativi affidabili, ritenuti utili per la messa in atto di processi di decisione a tutti i livelli di intervento (compreso, come abbiamo visto, quello che deve coinvolgere ogni allievo per renderlo progressivamente consapevole delle proprie *strategie cognitive e affettive*).

Messa in atto di macro- e microdecisioni che abbiano sì alta probabilità di successo, ma che siano, non ci si stancherà mai di ribadirlo, anche pertinenti, efficaci e, se possibile, efficienti.

Allo stato attuale – caratterizzato ancora da un uso prevalente dei modelli di valutazione altamente soggettivi – molti dei dati e delle informazioni disponibili nel nostro Paese risultano invece così poco validi e attendibili da non permettere opzioni di successo né nella pratica educativa quotidiana né nel campo delle scelte politiche proiettabili verso il futuro.

11.13 Informazioni, valutazione, decisioni in un'ottica sistemica

Il sistema scolastico preso nel suo insieme, nonché ciascuna delle sue singole unità operative – scuole, istituti ecc. – possono considerarsi, con buona approssimazione, veri e propri *sistemi complessi, aperti e finalizzati. Complessi,* giacché risultano

costituiti, come si sa, da un elevato numero di elementi organizzati e interdipendenti la cui interazione per un verso dipende dalle caratteristiche dei singoli elementi e dalle loro variazioni qualitative e quantitative, per altro determina struttura e processi del sistema stesso; *aperti* e non già chiusi, per quanto limitati, poiché, sia a livello macro che micro, sono distinti dall'ambiente esterno da confini - e non da barriere - che costituiscono una vera e propria sede elettiva di un continuo scambio di "energia" e di informazioni con l'esterno; *finalizzati*, perché, storicamente, ovvero intenzionalmente, elaborati per il perseguimento di ben precisi obiettivi programmatici. In tale *ottica sistemica*, assai proficua sia sul piano dell'analisi teorica dei problemi scolastici e dell'allestimento di accorte e opportune politiche, sia su quello più squisitamente organizzativo e didattico, la questione relativa alla produzione e all'impiego dei flussi informativi, soprattutto interni, ma anche esterni al sistema, si impone come cruciale per assicurarne l'efficacia di funzionamento e buona parte del suo continuo ammodernamento, cioè la rispondenza dei processi attivati e degli esiti perseguiti agli obiettivi volta a volta programmaticamente determinati. Come dire che l'informazione assume la qualità e la forza di un vero attivatore energetico del sistema oltre che di regolatore dei processi che in esso hanno luogo. Infatti, seppure a diversi livelli di intervento e responsabilità, tutti gli operatori del sistema, compresi i docenti e i capi di istituto ma anche e soprattutto gli allievi, per poter risolvere qualunque tipo di problema - primariamente connesso all'apprendimento - e poter decidere di conseguenza, devono non solo disporre tempestivamente delle informazioni di cui hanno bisogno, ma devono esser certi della loro *affidabilità* o, quanto meno, devono poterne verificare in ogni momento il grado di *attendibilità*.

Solo procedendo in tal modo, sviluppando cioè processi di autoapprendimento e autovalutazione, in una didattica bi- e multidirezionale (classe rovesciata, valutazione formante...) ogni allievo potrà acquisire progressivamente capacità autonome di apprendimento.

11.14 Quali informazioni

Tuttavia, a una analisi persino sommaria delle fonti, delle tecniche di rilevazione, dei modelli di analisi, e delle modalità di impiego delle informazioni scolastiche - non solo di tipo statistico di molti sistemi educativi, soprattutto del nostro - emerge, paradossalmente, un insieme che per alcuni versi produce una sovrabbondanza di informazioni ridondanti, spesso superflue, ma che risulta incapace di garantire flussi continui delle informazioni pertinenti e significative, di ampliare l'area di sovrapposizione tra *Esigenza, Domanda e Offerta* di informazioni così come si evince dalla Figura 11.1.

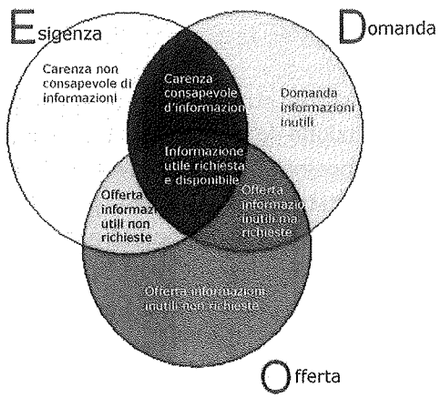


Figura 11.1 Il peso delle informazioni nei processi di decisione (da: A. Zuliani, *Informazione statistica e processi decisionali*, Roma, ISTAT, 1986).

Certo, come si può inferire dalla "lettura" del diagramma riportato - valido ancora oggi, cioè in presenza di un quadro informativo caratterizzato nel bene e nel male dai *Big Data* - lo spettro dei dati conoscitivi che si dovrebbero produrre e ricevere all'interno e all'esterno del sistema - macro, meso o micro che sia - è assai ampio, ma è anche vero che la produzione e la domanda di informazioni realmente offerte e richieste, generano e influenzano, in una sorta di circolarità virtuosa o viziosa a seconda dei casi, ulteriormente la nuova produzione e la nuova domanda di informazione.

Tuttavia, almeno per quanto concerne il pacchetto di informazioni che dovrebbero "fisiologicamente" derivare ed essere impiegate dallo e nello svolgimento dei processi previsti e attivati nel sistema stesso, la questione dovrebbe potersi risolvere facilmente.

Purtroppo però non vengono prodotti proprio quei blocchi e quei flussi informativi che, anche alla luce del cosiddetto modello CIPP (Context-Input-Process-Product) proposto da D.L. Stafflebeam negli anni Settanta (Alulli, 2000) dovrebbero rappresentare una sorta di sistema nervoso principale del più ampio sistema scuola, per esempio flussi e indicatori sulla consistenza e sull'uso dei laboratori, delle apparecchiature scientifiche e in generale sulle risorse formative e didattiche interne oltreché sulle caratteristiche processuali e sul prodotto qualitativo e quantitativo della formazione ai diversi livelli scolastici (ovvero, tipologia delle abilità e delle conoscenze fatte raggiungere dagli allievi al termine di un corso di studi e numero dei soggetti che le raggiungono). Accade invece che proprio in rapporto alla disponibilità di questi dati si avrebbe la possibilità reale di attivare processi intelligenti, in senso letterale, relativi alla pianificazione, alla organizzazione, alla

direzione, coordinamento e controllo del processo formativo a livello di micro-, meso- e macrosistema.

Molti dei dati e delle informazioni strettamente necessari per la pianificazione, la gestione, il controllo e la regolazione delle attività di formazione nella scuola dell'autonomia (cioè per la *governance*) possono essere rappresentati - di fatto secondo il modello CIPP - tra quelli appresso indicati, anche se ancora oggi non sempre quelle informazioni e quei dati vengono prodotti o resi disponibili e/o "consumati", pur se previsti e necessari per la compilazione e l'impiego del RAV (rapporto di autovalutazione):

- *informazioni e/o dati di contesto*: dal livello di istruzione della popolazione residente alla consistenza delle risorse sociocultuarli del territorio (biblioteche, associazioni culturali, teatri, cinema, ecc.);
- *informazioni e/o dati di input*: dalle caratteristiche strutturali della scuola (spazi, aule, laboratori, attrezzature ecc.) alle risorse umane (competenze professionali extraconcorsuali dei docenti; provenienza socioculturale, livelli di competenza e caratteristiche affettivo-motivazionali degli allievi);
- *informazioni e/o dati di processo*: organizzazione del tempo e dello spazio; strategie didattico-formative: frequenza delle verifiche e funzioni valutative per l'apprendimento; modalità di strutturazione, attuazione, controllo e modificazione delle decisioni; impiego delle risorse materiali: mezzi didattici: Pc, attrezzature biblioteche;
- *informazioni e/o dati di output*: prodotto qualiquantitativo della formazione; valore formativo aggiunto prodotto; motivazioni e atteggiamenti di allievi e docenti;
- *informazioni e/o dati di outcome*: riuscita scolastica e/o socio-lavorativa degli allievi, ecc.

11.15 Determinazione, uniformità, disponibilità e disaggregazione delle informazioni

Sulla base delle precedenti considerazioni, si possono trarre alcuni suggerimenti per la qualificazione delle informazioni scolastiche. Innanzitutto emerge l'esigenza di dover determinare lo spettro di quelle informazioni che, tra le infinite possibili, assumono rilevanza nel contesto dato; di garantire una vera e propria uniformità delle unità informative e dei criteri di aggregazione e che le informazioni circolino e siano disponibili, soprattutto per le attività ordinarie a tutti gli attori del processo e agli *stakeholder;* che le molteplici informazioni vengano opportunamente integrate - quando necessario - in blocchi sintetici fino ad assumere forme di specifici *indicatori*, e che si possano disaggregare in funzione della loro utilizzabilità fino a livello di singola classe, così come peraltro mira a fare il RAV.

Le informazioni prodotte dall'intero sistema assumerebbero così il carattere di veri e propri criteri rispetto ai quali poter e dover confrontare i dati omologhi relativi alle aggregazioni di scuole volta a volta considerate - fino al singolo plesso scolastico - in modo che dall'esito della comparazione si ottengano dati pertinenti per ricalibrare la pianificazione degli interventi formativi. Una sequenza di assunzione e attuazione di decisioni operative consentirebbe quindi di procedere progressivamente alla ottimizzazione dei processi e degli esiti dell'istruzione.

11.16 Valutazione e processi decisionali per l'apprendimento: come si struttura una decisione

Le variabili relative all'apprendimento costituiscono il gruppo più consistente e significativo di quelle da porre sotto un continuo controllo per regolare il micro-, il meso- e il macrosistema formativi e le loro reciproche relazioni, nonché per modificarle su rigorose basi conoscitive.

D'altro canto le finalità istituzionali della scuola e quelle perseguite nell'attuale contesto storico, sono principalmente la promozione di *conoscenze e abilità significative, sistematiche, stabili, spendibili e capitalizzabili* (Domenici, 1998, pp. 9.-14). Per queste ragioni il ventaglio aureo delle informazioni sulla efficacia della scuola è dato da quelle che si riferiscono agli apprendimenti prodotti e ai processi attivati per raggiungerli, in quanto da essi si può inferire l'adeguatezza dell'intero sistema verso i compiti socialmente assegnatigli.

Tuttavia sono proprio i dati sul processo e sul prodotto della formazione che, nonostante il gran numero dei momenti valutativi presenti nel fare scuola quotidiano e delle certificazioni formali dei risultati conseguiti - voti, pagelle, schede, scrutini, esami - lungo i percorsi scolastici, risultano i meno affidabili fra quelli pur limitati che vengono per così dire prodotti e da cui si dovrebbero poter ricavare informazioni sul funzionamento della scuola.

L'aggregazione e la disaggregazione dei dati faciliterebbero una lettura a più livelli dei fattori concomitanti che più di altri contribuiscono a determinare, come abbiamo visto, processi ed esiti dell'istruzione scolastica, condizione necessaria per l'attivazione di una vera e propria circolarità virtuosa tra attività e interventi compiuti nel macrosistema e quelli svolti dai singoli istituti scolastici, fino alla singola classe. Risulta pertanto utile considerare in questa sede il processo attraverso cui si attua una decisione. Strutturazione, analisi e controllo delle decisioni risultano essere, infatti, procedure che assumono un rilievo valutativo di primaria importanza per il miglioramento delle scelte di attività e azioni finalizzate.

Seppur rappresentato schematicamente il *percorso di una decisione* è dato da:

- affioramento, a livello di consapevolezza, dell'esistenza di un problema;
- formulazione del problema nella forma più chiara e precisa possibile;
- individuazione delle ipotesi risolutive, ovvero dalla determinazione, se non proprio dalla "generazione" creativa, delle possibili alternative di scelta delle azioni;
- assunzione delle informazioni relative al problema e alle scelte che si possono compiere nel contesto di riferimento;
- prefigurazione dei possibili (certi, incerti o probabili) esiti di ciascuna delle scelte praticabili;
- valutazione dei risultati previsti per ognuna delle possibili scelte;
- scelta vera e propria od opzione di scelta.

Come si può notare, senza la presenza di alternative di scelta - offerte dalla situazione o, quando sia possibile, appositamente ideate e "create" - non può strutturarsi né aversi una decisione. Occorre però considerare che, se è vero che quelli appena indicati sono i passi da compiersi per pervenire a un'opzione di scelta, è anche vero che l'itinerario di decisione illustrato può essere percorso in tempi persino brevissimi. I cosiddetti "decisori esperti" hanno sviluppato, e di norma posseggono, i requisiti necessari per poterlo fare e attuano le azioni indicate qui di seguito: colgono in un contesto informazioni che altri spesso non percepiscono (abilità percettivo-attentiva); utilizzano le più rilevanti informazioni disponibili (abilità discriminativa); tendono a rendere semplici anche i problemi complessi (capacità di andare al cuore delle questioni); selezionano le situazioni problematiche particolarmente importanti dal punto di vista della decisione (capacità di differenziare le situazioni problematiche); sanno creare nuove procedure di decisione (originalità e creatività); fanno spesso ricorso ad automatismi cognitivi nel processo di ideazione e modificazione delle strategie di decisione (capacità di ridurre lo sforzo cognitivo); sopportano, anzi affrontano bene situazioni stressanti legate alla delicatezza e all'importanza delle decisioni; accettano come "normale" eventi inattesi di disturbo della decisione; sanno valutare l'eventuale inadeguatezza, nel contesto specifico, di strategie di decisione collaudate, e generarne di nuove; si assumono la responsabilità, anche pubblica, degli esiti delle decisioni prese.

Quanto più si conoscono l'ambito generale nel quale i problemi cadono e le strategie decisionali mostratesi produttive in situazioni analoghe, tanto più diventa possibile mettere in atto quei processi mentali ricorsivi utili per rendere massimamente efficaci, entro i vincoli contestuali, le scelte da compiersi: in particolare nel processo di apprendimento e/o di *problem solving* sollecitato dal docente nell'allievo; nella progettazione e nel monitoraggio delle attività didattiche da parte del docente.

Ben oltre le decisioni da assumersi estemporaneamente giorno per giorno, momento per momento in relazione alle attività routinarie e agli imprevisti, le questioni più delicate del processo concernono: la chiarificazione e la formulazione dei problemi, giacché spesso si parte da situazioni non sempre chiare; l'analisi critica delle informazioni disponibili o di quelle necessarie; la scelta di "lavorare" sulle informazioni rilevanti escludendo quelle poco importanti se non addirittura fuorvianti; la selezione di situazioni (e decisioni) analoghe pescate nelle tracce della memoria individuale, dei gruppi o delle istituzioni che devono decidere (documentazioni, archivi di esperienze, ecc.) per confrontarle con quelle contingenti; l'esame e il riesame delle strategie di decisione adoperate in situazioni simili e di quelle che si vogliono adottare o si stanno adottando nella situazione data, da parte di allievi e docenti, ovvero la "metacognizione del processo decisionale", con lo scopo di renderla quanto più aderente alla specificità del problema.

Il riferimento alla teoria e alle pratiche di strutturazione delle decisioni diventa di enorme rilievo quando si vogliano evitare scelte e controlli delle stesse poco pertinenti e poco efficaci, quindi azioni che, oltre ad avere scarsa probabilità di successo, impediscano il ri-orientamento e il miglioramento continuo delle scelte e dei risultati procedurali e finali dell'apprendimento e dell'insegnamento.

In termini molto generali la valutazione assume dunque la funzione di regolazione (*ex ante, in itinere* ed *ex post*) tanto dei programmi e delle decisioni di politica scolastica a livello di *macrosistema*, (l'intero sistema educativo nazionale), quanto dei progetti e delle decisioni culturali e didattiche che i docenti rispettivamente elaborano e si trovano a dover assumere continuamente nel *microsistema educativo* (la classe e la singola scuola) nonché, infine, ma non per ultimo, delle decisioni in ambito operativo-culturale, conoscitive, degli allievi nello svolgimento della loro articolata esperienza di apprendimento.

All'interno del *microsistema* la *valutazione ex ante* ha lo scopo di far elaborare preventivamente quel progetto, tra i tanti possibili, a più alta rilevanza formativa nel contesto dato. Con essa si valutano preventivamente gli effetti d'impatto e risultato, attraverso un confronto sistematico di ipotesi complementari e/o alternative. Presuppone una conoscenza diffusa dell'ambiente in senso lato e stretto; una conoscenza delle più rilevanti strategie di intervento in rapporto agli obiettivi generali che si vogliono perseguire (il come apprendere, il come insegnare); la piena valorizzazione degli esiti valutativi (risultati di una precedente o contestuale valutazione ex post) di progetti analoghi già provati (quando la situazione non sia del tutto nuova) in quello stesso contesto. Da questo punto di vista, tanto la valutazione quanto la programmazione didattica garantiscono una circolarità virtuosa dell'operatività docente, fondata principalmente, ma non solo, sugli effetti di retroazione di informazioni ed esperienze acquisite e controllate.

La *valutazione in itinere* ha invece lo scopo di garantire una gestione oculata del progetto, un controllo della persistenza della sua adeguatezza al contesto, in modo che la proposta formativa offerta a ciascuno e a tutti gli alunni sia sempre adatta e

adattata, contemporaneamente, alle caratteristiche peculiari di ciascuno di loro e agli obiettivi formativi generali del progetto, in particolare all'obiettivo della promozione e sviluppo di capacità di apprendimento autonomo attraverso l'autovalutazione continua dei processi di apprendimento e autoapprendimento.

Infine, quella che abbiamo definito come *valutazione ex post* si propone di far controllare la *validità complessiva* del progetto educativo e delle strategie attuative in modo che sia reso possibile il miglioramento del primo e delle procedure seguite, nonché uno scambio intersoggettivo di esperienze tra scuole e tra docenti diversi, su basi di dati e di conoscenze affidabili e comparabili.

All'interno di tali funzioni generali vi sono ovviamente quelle specialistiche che, pur ormai acquisite dal lessico valutativo comune, stentano purtroppo a concretizzarsi nelle forme proprie che le contraddistinguono. Sono le funzioni diagnostico-iniziale, formativo-processuale, sommativo-intermedia e sommativo-finale, e per altri versi, orientativa, formante, pro-attiva ecc. cui rinviano, sul piano teorico-concettuale e pratico-operativo, i corrispettivi apparati didattici.

Occorre tuttavia fare riferimento a questioni generali per poter meglio analizzare e valutare anche questioni più circoscritte. Sarebbe perciò proficuo considerare che, in qualche modo, il processo valutativo assume sempre una doppia dimensione: quella *retrospettiva* e, nel contempo, *prospettica*: "opera nel presente come sintesi rappresentativa dei risultati di scelte passate, già compiute, ma è proiettata verso il futuro" attraverso il meccanismo di scelta, dell'alternativa decisionale ritenuta migliore tra quelle possibili, sulla base di uno dei più astratti processi mentali, consistente nella prefigurazione dei risultati di ognuna delle alternative praticabili, per poter scegliere quella ritenuta più opportuna.

Riepilogo

In questo capitolo è stato delineato il percorso di nascita e dello sviluppo della docimologia e sono state analizzate le principali caratteristiche degli strumenti di valutazione degli apprendimenti e il loro utilizzo nell'ambito della didattica.

Sono state esaminate le modalità di *strutturazione del giudizio valutativo*, avvalendosi sia di elementi conoscitivi e informativi di tipo qualitativo sia di tipo quantitativo, mettendo in luce il rischio che può rappresentare, nel primo caso, l'espressione di una estrema soggettività, e sottolinenando, nel secondo caso, il rilievo del ricorso a un *giudizio ad alto livello di affidabilità (ovvero valido e attendibile) in quanto fondato su rilevazioni oggettive.*

Particolare attenzione è stata rivolta alla misurazione, al ricorso a criteri assoluti o a criteri relativi per la costruzione delle prove di valutazione e all'allestimento dei giudizi valutativi e alla loro comunicazione, anche al fine di produrre una rete di informazioni utilizzabili per l'assunzione di decisioni efficaci.

Riferimenti bibliografici

Allulli G. (2000). *Le misure della qualità.* Roma: SEAM.

Ajello A.M. (a cura di) (2002). *La competenza.* Bologna: Il Mulino.

Anderson L.W., Krathwohl D.R. (eds.) (2001). *A taxonomy for learning, teaching, and assessing: a revision of bloom's taxonomy of educational objectives.* New York: Longman.

Beckers J., De Ball R., De Landsheere G. (1985.).. *Come formulare giudizi di valutazione didattica. La costruzione di scale descrittive.* Teramo: Lisciani & Giunti.

Benvenuto G. (2003). *Mettere i voti. Introduzione alla docimologia.* Roma: Carocci, Roma.

Bloom B.S., Engelhart M.D., Furst E.J., Hill W.H., Krathwohl D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.* New York: David McKay.

Broadfoot P.M., Daugherty R., Gardner J., Gipps C.V., Harlen W., James M., Stobart G. (1999). *Assessment for learning. Beyond the black box.* Cambridge: University of Cambridge.

Calonghi L. (1992). *Strumenti della valutazione. I saggi.* Teramp: Giunti e Lisciani.

Calonghi L. (1976). *Valutazione.* Brescia: La Scuola.

Calonghi L. (1978). *Valutazione delle composizioni scritte.* Roma: Armando Editore.

Chatterji M (2003). *Designing and using tools for educational assessment.* Boston: Allyn and Bacon.

Chiappetta Cajola L. (1998). *Handicap e Valutazione. La funzione di controllo e di regolazione nel processo formativo degli allievi disabili.* Roma: Anicia.

Chiappetta Cajola L., Domenici G. (2005). *Organizzazione didattica e valutazione.* Roma: Monolite.

Chiappetta Cajola L. (2006). *Il portfolio dell'allievo disabile. Democrazia, uguaglianza e processo di integrazione.* Roma: Anicia.

Clark A., Moss P. (2001). *Listening to young children: the Mosaic approach.* London: National Children's Bureau for the Joseph Rowntree Foundation.

De Landsheere G. (1998). *Il pilotaggio dei sistemi educativi.* Roma: Armando Editore.

Domenici G. (1991). *Gli strumenti della valutazione.* Napoli: Tecnodid.

Domenici G. (1998). *Manuale dell'orientamento e della didattica modulare.* Bari-Roma: Laterza (pp. 9-14).

Domenici G. (2000). *La valutazione come risorsa.* Napoli: Tecnodid.

Domenici G. (2020). *Manuale della valutazione scolastica.* Bari-Roma: Laterza.

Domenici G. (a cura di) (2005). *Le prove semistrutturate di verifica degli apprendimenti.* Torino: Utet.

Domenici G. (ed.) (2009). *PRIN 2006-2008. Valutazione a autovalutazione per la qualificazione dei processi formativi e-learning.* Lecce: Pensa MultiMedia.

Domenici G., Moretti G. (ed.) (2011). *Leadership educativa e autonomia scolastica.* Roma: Armando Editore.

Domenici G., Biasi V, Ciraci A.M. (2014). Evaluation of teaching and relational competencies for a flexible integrated didactic strategy: the CDVR Questionnaire. *Congreso Universidad,* 3(1), 1-15.

Domenici G., Ciraci A.M. (2016). Competencias y procesos de evaluación. Pruebas de verificación semiestructuradas. In: *Memorias, Universidad 2016, 10° Congreso Internacional de Educación Superior.* Havana (Cuba), 15-19 febrero 2016. CD-Rom (PED 291 - pp. 1826-1835).

Domenici G. (1981). *Descrittori dell'apprendimento.* Teramo: Giunti & Lisciani Editori.

Gattullo M. (1967). *Didattica e docimologia. Misurazione e valutazione nella scuola.* Roma: Armando Editore.

Gattullo M., Giovannini M.L. (1989). *Misurare e valutare l'apprendimento nella scuola media.* Milano: Edizioni Scolastiche Bruno Mondadori.

Gentile M. (2019). *Valutare per apprendere. Attività e strumenti per il lavoro in classe.* Torino: Utet.

Giovannini M.L. (1994). *Valutazione sotto esame.* Milano: Ethel Editoriale Giorgio Mondadori.

Grounlund N.E. (2003). *Assessment of student achievement.* 7th ed. Boston: Allyn and Bacon.

Harrison K., O'Hara J., McNamara G. (2015). Re-thinking assessment: Self- and peer-assessment as drivers of self-direction in learning. *Eurasian Journal of Educational Research,* 60, 75-88.

Hattie J., Timperley H. (2007). The power of feedback. *Review of Educational Research,* 77(1), 81-112.

Hudson B. (1987). *Introduzione alle tecniche di valutazione.* Bologina: Zanichelli.

Haladyna T.M. (1997). *Writing test item to evaluate higher order thinking.* Boston: Allyn and Bacon.

Mason L. (1996). *Valutare a scuola. Prodotti, processi, contesti dell'apprendimento.* Padova: CLEUP.

Notti A.M. (2000). *La valutazione nella scuola dell'autonomia.* Salerno: Edisud.

Pellerey M. (1996). *Valutare nella formazione professionale.* Provincia autonoma di Trento.

Pellerey M. (2010). *Ripensare le competenze.* Napoli:Tecnodid.

Topping K. (2017). Peer assessment: learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology,* 1(1), 1-17.

Torstein L., Rannveig G.S. (2019). Does the flipped classroom improve student learning and satisfaction? *A Systematic Review and Meta-Analysis.* AERA Open 5(3), 1-17.

Trinchero R. (2018). Valutazione formante per l'attivazione cognitiva. Spunti per un uso efficace delle tecnologie per apprendere in classe. *Italian Journal of Educational Technology* 26(3), 40-55.

Varisco B.M. (2000). *Metodi e pratiche della valutazione. Tradizione, attualità e nuove prospettive.* Milano: Guerini e Associati.

Vertecchi B (a cura di) (1989). *Valutazione e qualità degli studi,* Napoli:Tecnodid

Visalberghi A. (1955). Misurazione e valutazione nel processo educativo. Milano: Edizioni di Comunità.

Weaver M.E. (1995). Using peer response in the classroom: Students' perspectives. *Research and Teaching in Developmental Education,* 12, 31-37.

William D. (2011). What is assessment for learning? *Studies in Educational Evaluation,* 37, 3-14.

12

Pregi e limiti delle indagini internazionali su grandi campioni

In questo capitolo vengono presentate le principali indagini internazionali in ambito educativo, con particolare riferimento a quelle tuttora in corso.
A partire da un richiamo alle prime ricerche condotte dalla IEA (International Association for the Evaluation of Student Achievement, fondata nel 1958), saranno analizzati le finalità, gli obiettivi, i disegni di ricerca e gli sviluppi metodologici che hanno caratterizzato l'evoluzione degli strumenti impiegati dalla stessa IEA e, a partire dagli anni Novanta del secolo scorso, dall'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico).
L'attenzione si focalizza in particolare sul costrutto di "reading literacy", impiegato dalla IEA nell'indagine PIRLS e dall'OCSE per l'indagine PISA. Infine, verranno messi in evidenza i punti di forza e le criticità di tali indagini.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Scopi, struttura e funzionamento delle principali indagini internazionali in campo educativo.
- Fasi di definizione operativa e misurazione del costrutto di *reading literacy* impiegato nelle indagini condotte dalla IEA e dall'OCSE.
- Punti di forza delle principali indagini internazionali in campo educativo.
- Rischi legati a una lettura poco attenta dei dati forniti dalle indagini internazionali in campo educativo.

"Appuntare l'attenzione su un aspetto lo fa balzare in primo piano e invadere il quadro, come in certi disegni che basta chiudere gli occhi e al riaprirli la prospettiva è cambiata. Adesso in questo incrociarsi di creste variamente orientate il disegno complessivo risulta frammentato in riquadri che affiorano e svaniscono. S'aggiunga che il riflusso d'ogni onda ha anch'esso una sua forza che ostacola le onde che sopravvengono. E se si concentra l'attenzione su queste spinte all'indietro sembra che il vero movimento sia quello che parte dalla riva e va verso il largo."

Italo Calvino, *Palomar*

12.1 Le indagini internazionali in campo educativo

Nell'ultimo quarto di secolo con cadenza regolare si è assistito alla pubblicazione di rapporti di indagini internazionali sui sistemi d'istruzione. La ricerca che riceve maggiore attenzione presso l'opinione pubblica è senza dubbio PISA (Tabella 11.1) dell'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico), alla quale è stato assegnato sin dalla prima edizione (2000) un ruolo nella costruzione di benchmark per le politiche educative in sede nazionale ed europea (dagli Indicatori di Lisbona a quelli del 2030). Nel documento "Parametri di riferimento europei per l'istruzione e la formazione del 2002", la Commissione delle Comunità Europee afferma:

> «L'indicatore comparato più affidabile delle competenze fondamentali è rappresentato dall'indagine PISA dell'OCSE, che riguarda i livelli ottenuti nella capacità di lettura, in matematica e nelle scienze da parte dei quindicenni. Si tratta di statistiche che possono essere considerate uno specchio affidabile del complesso delle "competenze per la società della conoscenza", dal momento che individuano fasce della popolazione non sufficientemente preparate alle sfide contemporanee o all'apprendimento lungo tutto l'arco della vita.»

Tuttavia, PISA non è l'unica indagine internazionale in campo educativo come d'altro canto l'OCSE non è il primo organismo internazionale a realizzare ricerche di questo tipo. Se gettiamo uno sguardo alle principali indagini oggi attive (Tabella 12.1), possiamo notare come siano riconducibili a due organizzazioni internazionali: l'OCSE da un lato (PISA, PIAAC, TALIS) e la IEA (International Association for the Evaluation of Student Achievement) dall'altro (PIRLS, TIMSS, ICCS, ICILS).

Mentre l'impegno dell'OCSE in questo ambito è relativamente recente (risale all'ultimo decennio del XX secolo), la IEA è stata fondata nel 1958 allo scopo di fornire risposte a un'istanza che iniziava a essere avvertita nel mondo occidentale. Si trattava di colmare la mancanza di dati raffrontabili sul rendimento scolastico di Paesi che stavano esperendo una crescente fase di integrazione dal punto di vista economico e avevano intrapreso (o stavano per intraprendere) la strada di una progressiva democratizzazione dell'istruzione. In particolare, emergeva la necessità di verificare se l'allargamento dell'accesso all'istruzione secondaria comportasse un abbassamento del livello degli apprendimenti conseguiti nelle scuole secondarie, ovvero se un sistema più comprensivo e aperto perdesse di qualità. Fino ad allora i termini di raffronto tra sistemi scolastici si limitavano a iscrizioni e percentuali di titoli conseguiti ma, nonostante l'avvio di indagini nei decenni precedenti, non erano stati impiegati standard internazionali per misurare la qualità dell'istruzione. I passaggi intrapresi dalla IEA avrebbero condizionato tutte le indagini a venire (anche quelle OCSE) e meritano dunque di essere ricordati. A partire dalla sua fondazione, la IEA avvia uno studio di fattibilità con l'obiettivo di verificare la possibilità di realizzare un confronto fra diversi sistemi scolastici, considerando come variabile principale il profitto degli studenti (Asquini, 2007). Inizia ad affermarsi un principio fondamentale delle indagini successive: i risultati ottenuti da studentesse e studenti sono considerati variabili di output del sistema scolastico e vanno raffrontati con dati di input e di processo in modo tale da consentire una valutazione dei fattori che determinano, nei vari Paesi e a livello internazionale, il livello degli apprendimenti. Vengono così definite le strutture delle prime indagini IEA: la FIMS (First International Mathematics Study), la FIS (First International Science Study) e, a partire dal 1965, la fondamentale Six Subject. È in particolare con la Six Subject che l'IEA raggiunge lo scopo di fornire ai decisori politici dati raffrontabili sui sistemi educativi rispetto a un'ampia gamma di contenuti (le sei "materie" erano: scienze, comprensione della lettura, letteratura, inglese come lingua straniera, francese come lingua straniera, educazione civica).

La struttura che emerge nelle indagini IEA è affine al CIPP (Context, Input, Process, Product), il modello elaborato negli anni Sessanta da Daniel Stufflebeam e incentrato sulla necessità di mettere in relazione, nella valutazione dell'efficacia educativa, i risultati riscontrati con le informazioni su elementi contestuali, di ingresso e processuali che caratterizzano i diversi interventi oggetto di indagine. Si afferma dunque un modello di efficacia educativa di impronta economica (Scheerens e Bosker, 2007), che definisce l'apprendimento come prodotto finale (*output*) di una serie di interventi di lavorazione della materia grezza in ingresso (*input*). Parallelamente, le indagini IEA, allo scopo di individuare fattori di efficacia e di equità in campo educativo, correlano il rendimento ottenuto da studentesse e studenti ai test con variabili di sfondo come lo stato socio-economico e altre relative ai pro-

cessi educativi e organizzativi.

Sebbene le indagini IEA rendessero inevitabili meccanismi di raffronto sui risultati ottenuti nei diversi contesti, l'erogazione di classifiche non rientrava tra gli obiettivi di uno sforzo cooperativo che era finalizzato piuttosto a una migliore conoscenza dei punti di forza e di debolezza dei diversi sistemi scolastici al fine di individuare politiche di intervento in grado di migliorare la qualità e l'equità dell'istruzione (Husèn, 1967). A dispetto di queste finalità, va segnalato come sin dall'inizio nella comunità pedagogica internazionale fossero ben presenti preoccupazioni relative alla possibilità di un impiego inappropriato dal punto di vista scientifico ed educativo dei meccanismi di raffronto. In particolare, si temeva che tale raffronto si risolvesse in meccanismi competitivi, portando i diversi contesti a scegliere di concentrare i propri sforzi educativi verso obiettivi scelti non in forza della loro significatività, ma esclusivamente sulla base della loro misurabilità nell'agone internazionale.

Se è vero che tali preoccupazioni, come verrà evidenziato, non si sarebbero rivelate infondate, va riconosciuto alle prime indagini IEA il merito di aver affermato un principio fondamentale: le informazioni sui risultati, benché essenziali, non costituiscono un dato sufficiente per valutare i diversi sistemi educativi. Tale principio avrebbe caratterizzato tutte le indagini a venire, comprese quelle svolte dall'OCSE. E saranno proprio le indagini OCSE, a partire dagli anni Novanta, a introdurre due elementi che verranno poi inglobati nelle ricerche successive: la periodicità e la definizione dei risultati basata su livelli di prestazione. Le indagini OCSE prendono corpo a partire da esigenze poste dalla scelta di realizzare un sistema internazionale di indicatori sull'istruzione.

Nel 1992 viene pubblicata la prima edizione di *Education at a Glance* e nella parte relativa agli esiti si fa affidamento ai lavori IEA, che però sono occasionali e non consentono operazioni di raffronto periodico dei risultati. Inoltre, le indagini IEA forniscono stime sul rendimento a determinate prove, ma non consentono di ottenere informazioni analitiche utili a declinare positivamente i livelli di padronanza raggiunti dalle popolazioni coinvolte. L'OCSE introdurrà innovazioni fondamentali relativamente a questi due aspetti.

In primo luogo, l'OCSE rende periodiche le indagini, che vengono strutturate in cicli di diversa ampiezza a seconda del tipo di ricerca: tre anni per PISA, dieci per PIAAC. In secondo luogo, già a partire dalle prime indagini svolte sulla popolazione adulta (come la IALS, International Adult Literacy Survey del 1994, e la Second International Adult Literacy Survey del 1996; cfr. Gallina, 2005), l'OCSE fornisce informazioni sulla distribuzione della popolazione in fasce di rendimento declinandole esplicitamente in base alle caratteristiche dei compiti che i soggetti appartenenti a ciascun livello sono in grado di affrontare con una certa probabilità di successo. Si tratta di una scelta che, sebbene non comporti di per sé il passaggio da

prove a norma a prove a criterio (si veda Approfondimento 12.1), andando oltre la mera comparazione tra punteggi medi, orienta in senso criteriale la lettura dei risultati, fornendo indicazioni più analitiche sulla diffusione delle conoscenze e delle abilità rilevate nelle popolazioni coinvolte. La restituzione dei risultati per fasce di livello è un fattore che gioca un ruolo essenziale nell'assunzione delle misure delle indagini come benchmark politici per i diversi sistemi d'istruzione e nella loro conseguente trasformazione in veri e propri obiettivi educativi. Tale modalità di restituzione, che caratterizza i rapporti OCSE e IEA delle attuali ricerche (si veda Tabella 12.1), rappresenta uno dei contributi più rilevanti offerti dalle indagini internazionali allo sviluppo metodologico della ricerca educativa.

Approfondimento 12.1

PROVE A NORMA E PROVE A CRITERIO

Una *prova a norma* consente di confrontare i risultati delle studentesse e degli studenti con quelli ottenuti da un gruppo di riferimento (per esempio, la media di scuola, la media regionale o nazionale, quella IEA o OCSE). L'impiego di prove a norma standardizzate (ovvero tarate su campioni rappresentativi) consente di ottenere informazioni per valutare l'efficacia e l'equità degli interventi educativi. Inoltre esse sono utili perché consentono di ottenere informazioni sull'andamento rispetto a determinati gruppi di riferimento dello stesso soggetto in diversi ambiti disciplinari (per esempio, una studentessa che ottiene punteggi significativamente superiori rispetto alla media in matematica, ma non in lettura: rilevare un dato del genere può contribuire a tenere sotto controllo determinate distorsioni valutative, come l'effetto alone) e dello stesso soggetto nel tempo (per esempio, uno studente può ottenere punteggi inferiori rispetto alla media all'inizio dell'anno e superiori qualche mese dopo: rilevare un dato del genere può contribuire a tenere sotto controllo determinate distorsioni valutative, come l'effetto stereotipia).

Un *test o una prova a criterio* posiziona la prestazione delle studentesse e degli studenti rispondenti all'interno di una scala di padronanza stabilita a prescindere dal rendimento medio della popolazione. Le prove a criterio, oltre a indicare con chiarezza gli obiettivi di rilevazione, prevedono che i diversi livelli di padronanza vengano predeterminati senza ricorrere al punteggio medio di gruppi di riferimento ma, al contrario, sulla base di criteri analitici, definiti grazie a indicatori e descrittori, che segnalano chiaramente a quale livello della scala di padronanza sia possibile posizionare la prestazione di una studentessa o di uno studente. Un criterio è infatti definito in base alle dimensioni (indicatori) da considerare nella valutazione di una prestazione, alle indicazioni che consentono di riconoscere tali dimensioni (descrittori), e a una scala ordinale per la valutazione del raggiungimento di determinati livelli di padronanza. In forza del carattere analitico e progressivo del feedback che offrono, le prove a criterio - se supportate da strumenti validi e affidabili (griglie, rubriche) per il controllo della soggettività dell'accertamento - costituiscono uno degli strumenti più efficaci nell'ambito della valutazione educativa.

Sebbene le indagini internazionali prevedano, accanto alla procedura automatizzata di codifica delle risposte a quesiti strutturati, una valutazione delle risposte a domande aperte e complesse basata su processi di tipo criteriale (indicatori, descrittori e criteri di correttezza sono predeterminati), i punteggi e i livelli di abilità non sono predeterminati, ma dipendono anche dall'andamento medio della popolazione. Tuttavia la scelta di comunicare i risultati presentando fasce di punteggio e fornendo per ciascuna fascia una descrizione sintetica degli elementi che tendono a caratterizzare le prestazioni di chi vi rientra può orientare in senso criteriale la lettura dei dati.

Le attuali indagini IEA e OCSE tendono ad avere caratteristiche in comune quali la periodicità, la rilevazione di variabili di processo e contesto e la definizione dei risultati sulla base di livelli di rendimento. Inoltre, le ricerche attuali, non diversamente dalle prime indagini IEA, pur faticando a definire con chiarezza quali processi educativi e didattici incidano sui risultati di studentesse e studenti alle prove, attestano in maniera significativa il peso esercitato sugli output cognitivi da variabili di contesto come quelle socio-economiche.

Quanto alle differenze, le indagini IEA si focalizzano maggiormente sulla dimensione curricolare e disciplinare, mentre quelle OCSE si incentrano sull'impiego delle conoscenze in situazioni di vita quotidiana. La vocazione scolastica della IEA è testimoniata anche dalle scelte relative alla popolazione di riferimento. Infatti, mentre l'OCSE individua fasce d'età a prescindere dal livello scolastico, le indagini IEA sono condotte su popolazioni definite in base al grado di scolarità.

12.2 Le indagini in corso

In questo paragrafo si passano brevemente in rassegna le indagini internazionali attualmente svolte da IEA e OCSE, soffermandosi maggiormente su quelle che prevedono la somministrazione di test relativi a conoscenze e abilità che i diversi sistemi educativi si prefiggono di sviluppare. Nel paragrafo successivo l'attenzione è focalizzata sul processo di definizione del costrutto di *reading literacy* impiegato nelle indagini PIRLS e PISA.

Si precisa che si è preferito tradurre le diverse forme di "literacy" valutate dalle indagini con "abilità e conoscenze" e non con "competenza". Dal punto di vista pedagogico (Baldacci, 2010) le competenze hanno infatti una dimensione contestualizzata, attiva, sociale, dinamica, metacognitiva e affettiva e, sebbene i test impiegati dalle indagini internazionali possano offrire indicazioni preziose su alcune di queste dimensioni, essi non sono in grado di fornire una misura sufficientemente valida e affidabile di un costrutto tanto complesso.

Tabella 12.1 Sintesi delle principali indagini internazionali attualmente in corso *IEA-PIRLS*

Indagine	Principali caratteristiche
IEA-PIRLS Progress in International Reading Literacy Study	• Ambito: comprensione della lettura (reading literacy) • Popolazione: studentesse e studenti al IV anno di scolarità • Periodicità: cinque anni, dal 2001
IEA-TIMSS Trends in International Mathematics and Science Study	• Ambito: matematica e scienze • Popolazione: studentesse e studenti al IV e all'VIII anno di scolarità • Periodicità: quattro anni, dal 1995. Un ristretto gruppo di paesi partecipa a TIMSS Advanced, che prevede una somministrazione, in Italia, al V anno della scuola secondaria di secondo grado
IEA-ICCS International Civic and Citizenship Education Study	• Ambito: educazione civica e alla cittadinanza • Popolazione: VIII anno di scolarità
IEA-ICILS International Computer and Information Literacy Study	• Ambito: conoscenze e abilità digitali e informative • Popolazione: VIII anno di scolarità • Periodicità: 5 anni, dal 2013
OCSE-PISA Programme for International Student Assessment	• Ambito: comprensione della lettura, matematica, scienze, ed. finanziaria • Popolazione: quindicenni Periodicità: 3 anni, dal 2000
OCSE-PIAAC Programme for the International Assessment of Adult Competencies	• Ambito: comprensione della lettura, matematica, scienze • Popolazione: 16-65 anni
OCSE-TALIS Teaching and Learning International Survey	• Ambito: orientamenti pedagogici, pratiche didattiche e scelte educative di docenti e dirigenti

L'indagine PIRLS *(Progress in International Reading Literacy Study)*, avviata dal 2001 con cadenza quinquennale, ha come principale obiettivo la rilevazione, attraverso prove semistrutturate, dei livelli di comprensione del testo (*reading literacy*) di studentesse e studenti al quarto anno di scolarità (in Italia, quarta classe della scuola primaria). Per l'indagine è utilizzato un disegno di campionamento a due stadi. Nel primo stadio sono selezionate casualmente le scuole. Nel secondo stadio, all'interno di ciascuna scuola campionata sono estratte casualmente due classi. Partecipano all'indagine le studentesse e gli studenti delle classi campionate. Allo scopo di individuare i fattori che incidono maggiormente sulla comprensione della lettura, PIRLS raccoglie, per mezzo di questionari, informazioni su variabili quali i supporti disponibili a casa, il curricolo previsto e quello realizzato, le pratiche didattiche in classe e le risorse a disposizione. Nel 2016 PIRLS introduce anche prove computerizzate.

PIRLS fornisce i risultati facendo riferimento a una scala a livelli. Si tratta di quattro benchmark (basso, intermedio, alto, avanzato) che rendono conto del posizionamento del campione rispetto ai diversi elementi del costrutto di reading literacy (cfr. paragrafo 11.3). La scala a livelli, impiegata anche dalle altre indagini internazionali, è basata su un modello probabilistico, per cui ci si aspetta che i soggetti che rientrano in un determinato livello ottenendo un certo punteggio (standardizzato con media pari a 500 e deviazione standard pari a 100) abbiano una determinata probabilità di risolvere correttamente i quesiti relativi a quel determinato livello, una probabilità più elevata di rispondere correttamente a quesiti di livello inferiore e, infine, una probabilità minore di risolvere i quesiti del livello superiore.

L'indagine offre dunque informazioni preziose e confrontabili nel tempo sulla percentuale di studentesse e studenti che raggiungono i diversi livelli di comprensione del testo. La Tabella 11.2, riferita a quesiti posti su testi espositivi, evidenzia il legame tra i livelli (con le fasce di punteggio e la relativa distribuzione della popolazione), gli elementi del costrutto di reading literacy e gli elementi descrittivi delle prestazioni. In Italia l'indagine è condotta da INVALSI.

Tabella 12.2 Livelli, e relative descrizioni, dell'indagine IEA-PIRLS 2016 (riadattamento da INVALSI, Indagine IEA-PIRLS 2016, Rapporto nazionale)

Livello	Descrizione
BASSO Punteggio: 400- 474 Distribuzione: Italia 11%; UE 14% (<Liv basso: Italia 2%; UE 4%)	Quando leggono testi espositivi complessivamente più semplici, studentesse e studenti sono in grado di: *localizzare e riportare informazioni formulate esplicitamente in testi scritti e in altre modalità di rappresentazione delle informazioni (ad es. grafici, diagrammi ecc.); iniziare a fare inferenze dirette su fatti, descrizioni e spiegazioni.*
INTERMEDIO Punteggio: 475-549 Distribuzione: Italia 35%; UE 34%	Quando leggono testi espositivi più semplici o relativamente complessi, studentesse e studenti sono in grado di: *localizzare e riportare due o tre informazioni del testo; fare inferenze dirette per fornire spiegazioni di fatti; iniziare a interpretare e integrare informazioni per ordinare gli eventi.*
ALTO Punteggio: 550-624 Distribuzione: Italia: 41%; UE: 36%	Quando leggono testi espositivi relativamente complessi, studentesse e studenti sono in grado di: *localizzare e distinguere informazioni pertinenti in una porzione di testo densa o in una tabella complessa; fare inferenze su connessioni logiche per fornire spiegazioni e ragioni; integrare informazioni testuali e visive per interpretare relazioni tra idee; valutare e fare generalizzazioni su contenuto ed elementi testuali.*
AVANZATO Punteggio: >624 Distribuzione: Italia: 11%; UE: 12%	Quando leggono testi espositivi relativamente complessi, studentesse e studenti sono in grado di: *distinguere e interpretare informazioni complesse che si trovano in diverse parti del testo, tenendo conto di tutti gli elementi del testo che fondano/autorizzano una data comprensione; integrare informazioni che si trovano in diverse parti del testo per spiegare relazioni e ordinare una serie attività; iniziare a valutare elementi visivi e testuali per cogliere il punto di vista dell'autore.*

IEA-ICCS

Il progetto ICCS (International Civic and Citizenship Education Study) è il secondo ciclo d'indagine internazionale sull'educazione civica e alla cittadinanza e il quarto progetto promosso in questo ambito dalla IEA. L'ICCS indaga le conoscenze e la comprensione di concetti e problematiche nell'ambito dell'educazione civica e alla cittadinanza da parte di studentesse e studenti all'ottavo anno di scolarità (in Italia, terza classe della scuola secondaria di I grado), oltre che le loro opinioni, attitudini e comportamenti. Anche ICCS, come PIRLS, prevede un campionamento a due stadi (scuola/classe). Il quadro teorico di ICCS è strutturato in due parti. La prima è relativa al test cognitivo e al questionario sulle percezioni e sugli atteggiamenti di studentesse e studenti. In questa parte vengono presentate una dimensione relativa ai contenuti, ovvero agli argomenti di educazione civica e alla cittadinanza oggetto di rilevazione; una dimensione cognitiva, che descrive i processi di pensiero da valutare tramite il test; una dimensione affettivo-comportamentale, che descrive le percezioni e le attività degli studenti misurate attraverso il questionario studente. La seconda parte è relativa ai fattori contestuali che si prevede possano influire sugli esiti e spiegare la loro variazione. Attraverso la somministrazione di questionari vengono rilevate variabili relative ai contenuti e organizzazione dell'insegnamento dell'educazione civica alla cittadinanza, alle esperienze di docenti, alle pratiche didattiche, ai contesti ambientali, al clima scolastico, al contributo della famiglia e delle comunità locali.

Anche ICCS, come PIRLS, fornisce risultati per fasce di livello collegate a punteggi alle prove e descrizioni di prestazione (Schulz et al., 2017). Come evidenziato nel rapporto INVALSI anno 2016, le studentesse e gli studenti che raggiungono il livello più elevato svolgono collegamenti tra i processi di organizzazione e influenza sociale e politica e i meccanismi legali e istituzionali utilizzati per controllarli, generano ipotesi accurate sui benefici, le motivazioni e i probabili esiti delle scelte politiche; integrano, giustificano e valutano determinate posizioni, politiche o leggi basandosi sui principi che ne rappresentano il fondamento; mostrano inoltre di avere familiarità con i grandi processi economici a livello internazionale e sono consapevoli della natura strategica della partecipazione attiva.

IEA-TIMSS

L'indagine IEA-TIMSS (Trends in International Mathematics and Science Study), condotta ogni quattro anni con la partecipazione (nel 2019) di oltre 60 Paesi, misura la performance in matematica e scienze attraverso la somministrazione di quesiti strutturati e semistrutturati (a risposta aperta) a studentesse e studenti al quarto e ottavo anno di scolarità di scolarità (in Italia: IV classe della scuola primaria e III classe della scuola secondaria di primo grado). Obiettivo di TIMSS è esplorare la struttura e l'organizzazione dei sistemi d'istruzione dei Paesi partecipanti raccogliendo informazioni su curricolo, pratiche didattiche e risorse scolastiche impiegate in ambito scientifico. TIMSS, oltre a somministrare prove, raccoglie informazioni su contesti e processi (come le pratiche d'istruzione a casa e a scuola) attraverso questionari destinati a studentesse e studenti, famiglie, docenti, dirigenti scolastici. Anche in TIMSS, come per PIRLS, i risultati vengono forniti facendo riferimento a livelli: chi raggiunge il benchmark più avanzato in matematica (quarto anno di scolarità) è "in grado di applicare la comprensione e le conoscenze in una varietà di situazioni relativamente complesse e di spiegare il proprio ragionamento" (Emiletti, 2020). Un ulteriore punto di forza di TIMSS è rappresentato dalla capacità di fornire informazioni sul progresso di una coorte di studentesse e studenti dalla IV primaria alla terza secondaria di primo grado. In matematica (Figura 12.1), la coorte italiana (2011, quarta classe della scuola primaria) passa da 8 punti al di sopra della media internazionale TIMSS (pari a 500) a 6 punti al di sotto (2015, terza classe della scuola secondaria di II grado). In scienze il peggioramento relativo è ancora più evidente: rispetto alla media internazionale TIMSS, nel 2011 al quarto anno di scolarità il campione italiano fa registrare un punteggio di 24 punti superiore, mentre quattro anni dopo, nel 2015, il punteggio dell'ottavo anno è di un punto inferiore all'andamento medio internazionale. In Italia l'indagine è condotta da INVALSI.

2011 - Fourth Grade			2015 - Eighth Grade		
Country	Achievement Difference from TIMSS Scale Centerpoint (500)		Country	Achievement Difference from TIMSS Scale Centerpoint (500)	
Singapore	106 (3.2)	▲	Singapore	121 (3.2)	▲
Korea, Rep. of	105 (1.9)	▲	Korea, Rep. of	106 (2.6)	▲
Hong Kong SAR	102 (3.4)	▲	Chinese Taipei	99 (2.4)	▲
Chinese Taipei	91 (2.0)	▲	Hong Kong SAR	94 (4.6)	▲
Japan	85 (1.7)	▲	Japan	86 (2.3)	▲
England	42 (3.5)	▲	Russian Federation	38 (4.7)	▲
Russian Federation	42 (3.7)	▲	Kazakhstan	28 (5.3)	▲
United States	41 (1.9)	▲	United States	18 (3.1)	▲
Lithuania	34 (2.4)	▲	England	18 (4.2)	▲
Australia	16 (3.0)	▲	Slovenia	16 (2.1)	▲
Hungary	15 (3.4)	▲	Hungary	14 (3.8)	▲
Slovenia	13 (2.1)	▲	Lithuania	12 (2.9)	▲
Italy	8 (2.6)	▲	Australia	5 (3.1)	
Sweden	4 (2.1)		Sweden	1 (2.8)	
Kazakhstan	1 (4.5)		Italy	-6 (2.5)	▼
Norway (4)	-5 (2.8)		New Zealand	-7 (3.4)	▼
New Zealand	-14 (2.6)	▼	Norway (8)	-13 (2.0)	▼
Turkey	-31 (4.7)	▼	United Arab Emirates	-35 (2.0)	▼
Chile	-38 (2.3)	▼	Turkey	-42 (4.7)	▼
Georgia	-50 (3.7)	▼	Bahrain	-46 (1.4)	▼
Bahrain	-64 (3.2)	▼	Georgia	-47 (3.4)	▼
United Arab Emirates	-66 (2.0)	▼	Qatar	-63 (3.0)	▼
Iran, Islamic Rep. of	-69 (3.5)	▼	Iran, Islamic Rep. of	-64 (4.6)	▼
Qatar	-87 (3.4)	▼	Chile	-73 (3.2)	▼
Saudi Arabia	-90 (5.2)	▼	Oman	-97 (2.4)	▼
Oman	-115 (2.9)	▼	Morocco	-116 (2.3)	▼
Morocco	-165 (4.0)	▼	Saudi Arabia	-132 (4.6)	▼

Figura 12.1 Differenza dalla media TIMSS (pari a 500, con deviazione standard pari a 100) ed errore standard (tra parentesi) nei punteggi medi in matematica del 2011 e del 2015 delle coorti di studentesse e studenti al quarto anno di scolarità nel 2011 (rielaborazione da IEA's Trends in International Mathematics and Science Study – TIMSS 2015.. http://timss2015.org/timss-2015/mathematics/student-achievement/).

IEA-ICILS

L'indagine IEA-ICILS (International Computer and Information Literacy Study) prende avvio nel 2013, ha cadenza quinquennale e usa prove computerizzate per valutare abilità e conoscenze digitali e informatiche di studentesse (età media: 14 anni, in Italia terzo anno della secondaria di I grado). In particolare, la ricerca indaga due dimensioni. La prima è la CIL (Computer and Information Literacy), intesa come abilità di utilizzare il computer per ricercare, creare e comunicare al fine di partecipare attivamente nella vita domestica, scolastica, lavorativa e sociale. La seconda è il CT (Computational Thinking), ovvero la modalità di pensiero impiegata quando si programmano le operazioni da eseguire per risolvere un problema su un computer o un dispositivo digitale. Oltre alla somministrazione di quesiti a risposta chiusa e aperta complessa, ICILS prevede l'impiego di questionari (per studentesse e studenti, docenti, scuole, sistemi) per raccogliere informazioni sulle variabili di sfondo. La pubblicazione dei risultati si basa su livelli di rendimento.

OCSE-PISA

L'OCSE-PISA (Programme for International Student Assessment) è la più nota indagine internazionale in campo educativo. Nel 2018 ha coinvolto 79 Paesi e dal 2000 valuta la preparazione di studentesse e studenti quindicenni (a prescindere dalla classe e dalla scuola frequentate) per affrontare la vita adulta, rilevando il loro livello di conoscenze e abilità in matematica, scienze e comprensione della lettura. A questi tre ambiti, a partire dal 2012 si è aggiunta la financial literacy. PISA, per ogni ciclo triennale, pur somministrando test per ciascun ambito, approfondisce un dominio in particolare. Allo scopo di rilevare la capacità di impiegare quanto appreso in reading literacy, matematica, scienze ed educazione finanziaria per affrontare situazioni problematiche che possono presentarsi nella quotidianità, i diversi ambiti sono indagati attraverso prove incentrate su situazioni di vita reale (cfr. Paragrafo 12.3) piuttosto che su elementi curricolari. L'ampio ricorso a domande a risposta aperta e la distanza ravvicinata tra un'edizione e l'altra (tre anni) rappresentano due fondamentali punti di forza dell'indagine.

L'edizione del 2018 (INVALSI, 2019) – come quella del 2000 e quella del 2009 – ha avuto come ambito principale la reading literacy (cfr. Paragrafo 12.3). Oltre alle prove cognitive, l'indagine PISA somministra questionari (a studentesse e studenti, a dirigenti e famiglie) allo scopo di individuare fattori di contesto associati alle prestazioni. Il campionamento, diversamente dalle indagini OCSE, non è incentrato sull'estrazione di classi scolastiche, ma su quella di scuole e poi di studentesse e studenti.

I risultati di PISA, che ottengono una notevole eco presso l'opinione pubblica, vengono restituiti per livelli e, sin dalla delineazione degli obiettivi di Lisbona, rappresentano fondamentali indicatori di riferimento per le politiche educative dei

Paesi partecipanti. L'indagine restituisce preziose informazioni per valutare l'equità dei diversi sistemi educativi. In particolare PISA, oltre a testare, come TIMSS e PIRLS, la correlazione tra rendimento e stato socio-economico (elevata in Italia), genere (in Italia le ragazze ottengono mediamente punteggi migliori nella comprensione del testo, i ragazzi in matematica e scienze) e a fornire dati sulla diversa diffusione dei livelli di conoscenze e abilità nella popolazione di quindicenni per aree geografiche, consente all'Italia di ottenere dati sul tasso di segregazione scolastica, ovvero sulla tendenza ad aggregare negli stessi istituti studentesse e studenti con determinate caratteristiche dal punto di vista del livello di rendimento (Figura 12.2) e di quello sociale ed economico. Per l'Italia, l'indagine è curata dall'INVALSI.

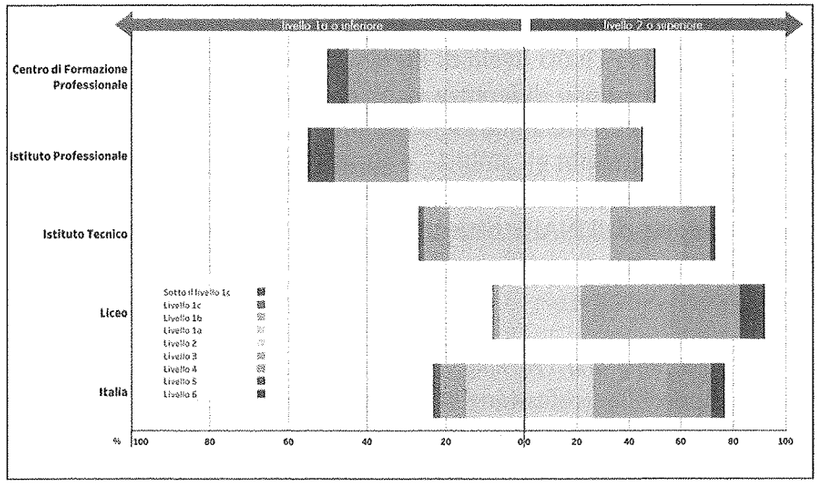


Figura 12.2 Diffusione dei livelli di reading literacy nei diversi indirizzi. Fonte: sintesi dei risultati italiani di PISA 2018 (INVALSI, 2019, p. 8).

OCSE-PIAAC

Erede della IALS-SIALS (Gallina, 2005), PIAAC (*Programme for the International Assessment of Adult Competencies*) ha l'obiettivo di ottenere informazioni sulla diffusione nella popolazione in età lavorativa (16-64 anni) di quelle che l'OCSE chiama "*foundations skills*", ovvero la comprensione della lettura (*Literacy*), le conoscenze e le abilità logico-matematiche (*Numeracy*) e quelle collegate alle tecnologie dell'informazione e comunicazione (ICT). Il primo ciclo (Di Francesco, 2014) si è svolto nel 2011/12. L'affinità degli ambiti consente significativi confronti tra i risultati PISA e PIAAC. Particolarmente innovativa è la scelta di PIAAC di indagare il tasso di impiego di tali conoscenze e l'abilità nell'attività lavorativa. Ai test, PIAAC af-

fianca infatti un questionario individuale sulle competenze utilizzate nel proprio lavoro (oltre alla condizione socio-anagrafica dell'individuo, i percorsi d'istruzione e la formazione svolta nel corso della vita lavorativa, la storia professionale e lo status occupazionale, oltre il retroterra sociale ed economico). In tal modo PIAAC rileva diverse dimensioni del mismatch, ovvero della mancata corrispondenza tra le conoscenze e le abilità acquisite e il loro effettivo impiego nella prassi lavorativa, con particolare riferimento alle diverse caratteristiche che sono correlate allo status di *overqualified*, ovvero di soggetto con competenze superiori al suo status professionale. L'indagine da questo punto di vista evidenzia, nel nostro Paese, la persistenza di discriminazioni di tipo sociale e di genere (le donne risultano particolarmente svantaggiate). L'ampiezza della gamma di età del campione (16-64 anni) consente a PIAAC di rilevare se le conoscenze e le abilità testate sono diversamente diffuse nella popolazione. Da questo punto di vista, i risultati del primo ciclo indicano in generale per l'Italia livelli di literacy e numeracy pari alla media internazionale, tuttavia emergono differenze significative per fasce di età: le fasce più giovani della popolazione ottengono infatti migliori risultati e questa differenza - di oltre venti punti in entrambi gli ambiti - è superiore alla media dei paesi partecipanti. Il campione è composto estraendo individui dalle liste anagrafiche. In Italia l'indagine è condotta dall'INAPP (Istituto Nazionale per l'Analisi delle Politiche Pubbliche).

OCSE-TALIS

L'OCSE-TALIS (*Teaching and Learning International Survey*) rileva informazioni su dirigenti (sviluppo della professionalità) e docenti (conoscenze e competenze). Nell'edizione 2018 (la terza), attraverso questionari sono stati esaminati temi come: le prassi di insegnamento, la leadership scolastica, la formazione e la preparazione iniziale del corpo docente, il clima scolastico. I risultati, rispetto alle altre indagini internazionali, ottengono una ridotta attenzione da parte dell'opinione pubblica, sebbene rappresentino una preziosa fonte informativa per la valutazione dei sistemi di formazione e selezione del corpo docente e dirigente. I dati dell'ultima edizione evidenziano come in Italia mediamente nella scuola secondaria di primo grado si ricorra in misura minore rispetto ad altri Paesi a pratiche che favoriscono l'attivazione cognitiva degli studenti. In Italia l'indagine è condotta dall'INVALSI

12.3 La definizione dei costrutti: la "reading literacy" tra IEA e OCSE

Abbiamo visto come le attuali indagini IEA e OCSE tendano a condividere scelte metodologiche come la rilevazione di variabili di processo e contesto e la definizione dei risultati sulla base di livelli di rendimento. A ciò si aggiunge una marcata accentuazione del ricorso a quesiti a risposta aperta complessa per la rilevazione degli apprendimenti, una scelta legata a una definizione degli obiettivi di indagine che chiama in causa la costruzione, oltre che il mero riconoscimento, delle risposte esatte. Allo scopo di illustrare come le indagini internazionali affrontino uno dei compiti più ardui della misurazione (intesa come collegamento di indicatori empirici a concetti astratti) dei propri obiettivi, in questo paragrafo vengono presentati i costrutti di reading literacy definiti all'avvio di IEA-PIRLS (2001) e dall'OCSE-PISA (2000). Tali costrutti iniziali resteranno un punto di riferimento fondamentale anche nelle edizioni successive (le modifiche apportate non ne modificano le dimensioni essenziali).

Negli anni Novanta, i lavori di elaborazione delle prove per l'ambito di reading literacy dell'indagine PISA (Nardi, 2002) si sono basatisulla seguente definizione di comprensione della lettura:

> «capacità necessaria per capire, utilizzare e riflettere a partire da testi scritti e elettronici, perché il soggetto possa raggiungere i suoi obiettivi, sviluppare le sue conoscenze e il suo potenziale, e svolgere un ruolo attivo nella società».

Negli stessi anni, anche sulla scorta di una pluridecennale esperienza - l'ultimo lavoro risale al 1994 ed è rappresentato dal Reading Literacy Study (Lucisano, 1994) - la IEA appronta il costrutto dell'indagine PIRLS, partendo da questa definizione di reading literacy:

> «capire e saper usare quelle forme di linguaggio scritto che la società ritiene indispensabili e che l'individuo apprezza come valori. I giovani lettori debbono saper costruire un significato da testi di vario tipo. Debbono inoltre leggere allo scopo di apprendere, far parte della comunità dei lettori e per godimento personale (Pavan de Gregorio, 2003).»

Queste due definizioni lasciano trasparire solo parzialmente differenze legate alla diversa età delle popolazioni testate e al difforme approccio di OCSE e IEA agli obiettivi delle indagini che abbiamo già evidenziato. PIRLS è interessata agli elementi curricolari comuni ai Paesi partecipanti e fa riferimento a un livello di scolarità che comporta che alla finalità di leggere per imparare a leggere inizi ad affiancarsi quella di leggere per acquisire informazioni e per fruire di esperienze esteticamente ed emotivamente appaganti. PISA non rivolge lo sguardo alla verifica di conoscenze e abilità direttamente connesse col curricolo dei diversi Paesi, ma mira a indagare il loro impiego in situazioni quotidiane. Tuttavia, le similarità delle due definizioni sono più evidenti delle loro differenze: gli aspetti costitutivi fanno in entrambi i casi riferimento al ricorso a processi di comprensione orientati dalla partecipazione al contesto sociale e culturale e da motivazioni individuali. Come

verrà evidenziato, tali similarità si traducono nella definizione operativa di costrutti che, nelle loro dimensioni, risultano quasi sovrapponibili.

PIRLS (Mullis et al., 2002) individua tre ordini di fattori allo scopo di definire la comprensione del testo: (1) i processi di comprensione; (2) gli scopi della lettura e i tipi di testo che vi si collegano; (3) gli atteggiamenti e le pratiche di lettura. Mentre il terzo ordine di fattori informa la costruzione dei questionari, il primo e il secondo guidano l'elaborazione delle prove. Rispetto al secondo, PIRLS individua lo scopo di leggere per acquisire e usare informazioni e quello di leggere per usufruire di un'esperienza estetica ed emotiva. Vengono così selezionati testi narrativi e testi informativi.

Il primo ordine di fattori, relativo ai processi di reading literacy, è determinato facendo riferimento a un modello di comprensione del testo che prevede l'integrazione tra processi bottom-up e top-down. I primi si incentrano su attività induttive di acquisizione di informazioni dal testo, mentre i secondi sono meccanismi deduttivi che prevedono l'impiego di aspettative e capacità inferenziali da parte di chi legge. In tal modo, PIRLS delinea i processi di comprensione del testo di seguito riportati:

1. ricavare informazioni e concetti esplicitamente espressi nel testo;
2. fare inferenze;
3. interpretare e integrare informazioni e concetti;
4. analizzare e valutare il contenuto, la lingua e gli elementi testuali.

Sul fronte OCSE, concentrandosi sui soli processi di comprensione del testo, riscontriamo come il lavoro che impronta la costruzione di prove per l'indagine PISA sia praticamente sovrapponibile a quello svolto dalla IEA. La prima edizione di PISA individua cinque aspetti della comprensione della lettura di seguito riportati.

1. Individuare informazioni.
2. Comprendere il significato globale del testo.
3. Sviluppare un'interpretazione.
4. Riflettere sul contenuto e valutarlo.
5. Riflettere sulla forma e valutarla.

Tuttavia, come indicato chiaramente nei rapporti PISA, la difficoltà di garantire per ciascun aspetto un numero sufficiente di quesiti determina la loro aggregazione nei tre processi di seguito riportati.

1. Individuare informazioni.
2. Interpretare il testo (che aggrega l'aspetto 2 e il 3).

3. Riflettere e valutare (che aggrega l'aspetto 4 e 5).

La similarità nelle definizioni di reading literacy garantisce un buon livello di confrontabilità tra i risultati delle indagini IEA e OCSE, consentendo a quei Paesi che partecipano a entrambe di acquisire informazioni utili a valutare lo sviluppo della comprensione del testo tra la scuola primaria e quella secondaria. È inoltre comune a entrambe le definizioni la scelta di rendere conto di un processo di comprensione inteso come costruzione attiva di significato. Dal punto di vista della definizione operativa, questa costruzione attiva si riflette nei processi di interpretazione, riflessione e valutazione.

La scelta di dare spazio a processi di costruzione attiva del significato del testo comporta una notevole sfida, rappresentata dalla necessità di porre non solo quesiti a risposta multipla o a corrispondenza, ma anche a risposta aperta complessa, item che non richiedono agli studenti il mero riconoscimento della risposta esatta, ma la sua produzione attiva. Le domande a risposta aperta complessa garantiscono a PIRLS e PISA una adeguata validità di contenuto, dato che consentono di ottenere informazioni anche su processi che richiedono di interpretare il testo o riflettere su di esso per valutarne forma o contenuto. In tal modo, PISA e PIRLS possono restituire i risultati indicando il posizionamento delle popolazioni rispetto a livelli di padronanza per ciascun processo rilevato (Asquini e Corsini, 2010). La scelta di porre domande simili fa però lievitare non poco i costi delle indagini, dato che comporta sia la formazione di gruppi di figure esperte che hanno il compito di valutare in maniera valida e affidabile le risposte secondo linee guida negoziate a livello internazionale, sia l'impiego di rigorose procedure di accertamento della coerenza delle misure ottenute. Anche in questo caso, le indagini offrono un prezioso contributo allo sviluppo metodologico della ricerca educativa. L'Approfondimento 12.2 riporta un esempio di quesito a risposta aperta complessa rilasciato dalla prova di reading literacy PISA 2000 e relativo alla capacità di riflettere sul contenuto del testo.

Approfondimento 12.2

ESEMPIO DI DOMANDA RELATIVA AL PROCESSO "RIFLETTERE SUL CONTENUTO DEL TESTO"*.

GRAFFITI

Sono furibonda: è la quarta volta che il muro della scuola viene pulito e ridipinto per cancellare i graffiti. La creatività è da ammirare, ma bisognerebbe trovare canali di espressione che non causino ulteriori costi alla società. Perché rovinare la reputazione dei giovani dipingendo graffiti dove è proibito? Gli artisti di professione non appendono i loro dipinti lungo le strade! Al contrario, cercano fondi e diventano famosi allestendo mostre legalmente autorizzate. Secondo me gli edifici, le recinzioni e le panchine nei parchi sono opere d'arte in sé. È davvero assurdo rovinare l'architettura con i graffiti e, peggio ancora, il metodo con cui vengono realizzati distrugge lo strato di ozono. Davvero non riesco a capire perché questi artisti criminali si diano tanto da fare, visto che le loro "opere d'arte" vengono cancellate sistematicamente.

Olga

Sui gusti non si discute. La società è invasa dalla comunicazione e dai messaggi pubblicitari. Simboli di società, nomi di negozi. Grandi poster che invadono i lati delle strade. Sono tollerabili? Sì, per la maggior parte. E i graffiti, sono tollerabili? Alcuni dicono di sì, altri no. Chi paga il prezzo dei graffiti? In fin dei conti, chi paga il prezzo degli annunci pubblicitari? Giusto. Il consumatore. Chi ha affisso i tabelloni, ha forse chiesto il vostro permesso? No. Allora perché chi dipinge graffiti dovrebbe farlo? Il proprio nome, i nomi delle bande e delle grandi opere pubbliche: non è solo una questione di comunicazione? Pensiamo ai vestiti a strisce e quadri apparsi nei negozi qualche anno fa. E all'abbigliamento da sci. I motivi e i colori sono stati presi in prestito direttamente dai variopinti muri di cemento. È piuttosto curioso che questi motivi e colori vengano accettati ed ammirati, mentre i graffiti dello stesso stile sono considerati orrendi. Tempi duri per l'arte.

Sofia

DOMANDA. Con quale delle due autrici delle lettere sei d'accordo? Spiega la risposta con parole tue, facendo riferimento a quanto affermato in una o in entrambe le lettere.

_____

_____

*Tratta dalle prove di reading literacy OCSE-PISA 2000.

12.4 Potenzialità e cautele: un bilancio

Come rilevato nei paragrafi precedenti, le indagini internazionali rappresentano un potente fattore di sviluppo metodologico in campo educativo (in particolare, nei processi di accertamento delle conoscenze e delle abilità oggetto di studio) e forniscono, se lette tenendo conto delle specificità dei diversi contesti (Vertecchi, 2004), dati utili per informare riflessioni sull'efficacia e sull'equità dei sistemi d'istruzione. In particolare, come sin qui evidenziato, la partecipazione alle indagini consente ai diversi Paesi di ottenere informazioni relative all'andamento alle prove delle diverse popolazioni coinvolte e tenere sotto controllo la correlazione di tale rendimento con variabili come il retroterra socioeconomico e culturale. A partire dagli anni Settanta, il riferimento agli indicatori impiegati nelle indagini IEA invita il discorso pubblico sulla scuola e sulle politiche educative a fare i conti con evidenze empiriche che sfidano alcuni dei luoghi comuni che lo caratterizzano. Un esempio è offerto dall'impiego dei dati delle indagini IEA (TIMSS e PISA) per confutare empiricamente uno degli stereotipi più retrivi e radicati in ambito educativo come quello che vuole le ragazze "naturalmente" meno competenti dei ragazzi in matematica e scienze. Il rapporto Eurydice del 2013 indica che le differenze di rendimento in matematica e scienze tra ragazzi e ragazze alle prove IEA e OCSE non risultano significative in alcuni paesi, e già lo studio di Guiso et al. (2008) evidenziava come il gender gap in ambito matematico a svantaggio delle quindicenni – calcolato sui dati di PISA 2003 – fosse negativamente correlato all'indice di emancipazione femminile e dunque imputabile al contesto culturale piuttosto che a presunti fattori biologici.

Tuttavia, se la partecipazione a indagini internazionali comporta importanti benefici, gli aspetti problematici non mancano. Per illustrarli è utile partire da quanto scritto decenni fa da Aldo Visalberghi, uno dei massimi sostenitori della partecipazione dell'Italia alle indagini IEA, e confrontarlo con quanto rilevato, sullo stesso argomento, quaranta anni dopo, da Paul Morris, eminente comparatista britannico. Se Visalberghi nel 1977 consiglia di concepire le rilevazioni internazionali come occasione per ottenere informazioni sui fattori che incidono sugli apprendimenti e non come "Olimpiadi del sapere", nel 2016 Morris rileva che in Corea del Sud le studentesse e gli studenti si preparano ad affrontare le prove TIMSS marciando verso il testing center cantando l'inno nazionale, ben consapevoli di rappresentare il Paese nell'equivalente educativo dei giochi olimpici.

Il caso riportato da Morris è indicativo della propensione dell'opinione pubblica dei diversi Paesi partecipanti a focalizzare la propria attenzione sulla posizione occupata nelle classifiche erogate da IEA e OCSE. Questa tendenza ci consente di segnalare due fattori di rischio collegati alla partecipazione alle indagini. Il primo è connesso a errori di valutazione della loro portata informativa, il secondo alla

semplificazione degli oggetti delle loro misure. Questi due fattori assumono rilevanza grazie al cambio di passo nelle indagini internazionali avvenuto nell'ultimo decennio del XX secolo (Corsini, 2019). Le indagini internazionali negli anni Settanta erano orientate da uno sforzo cooperativo incentrato sulla necessità di sostenere i processi di democratizzazione in atto nei diversi Paesi attraverso l'accesso alla scuola secondaria di strati di popolazione sin lì esclusi. Dalla fine degli anni Novanta il focus delle indagini si orienta verso la necessità di quantificare il capitale umano, una variabile chiave nella competizione economica a livello globale che viene declinata nei termini di rendimento alle prove somministrate da OCSE e IEA.

In primo luogo, la portata informativa delle indagini internazionali rischia di essere sminuita dalla sopraccitata scelta, operata con frequenza da decisori politici e opinione pubblica, di concentrarsi sul posizionamento dei diversi paesi nelle graduatorie internazionali (Agrusti, 2009). Sono tuttavia operanti anche meccanismi di sopravvalutazione, legati alla tendenza, affermatasi con forza nell'ultimo ventennio, a considerare i dati forniti da IEA e OCSE misure valide e affidabili della qualità dei processi formativi. Basandosi su disegni correlazionali - non sperimentali - e non operando alcuna osservazione dei processi, le indagini internazionali non sono in grado di misurare la qualità dei processi organizzativi e didattici. Esse infatti non descrivono compiutamente alcun processo educativo - limitandosi a ottenere informazioni attraverso questionari - e, non lavorando su gruppi che prevedono un'assegnazione di studentesse e studenti casuale (disegni sperimentali) o controllata (disegni quasi sperimentali), non sono in grado di produrre stime valide e affidabili dell'impatto che sul rendimento in lettura, matematica o scienze ha la scuola rispetto a quello che, più in generale, ha la famiglia o il contesto sociale, culturale ed economico. Tali ricerche non hanno sin qui contribuito, a differenza di quelle operanti attraverso disegni sperimentali o quasi sperimentali, a stimare l'efficacia di determinate scelte educative, organizzative e didattiche. Benché, come visto, ricerche come IEA TIMSS e OCSE PISA offrano elementi utili per ragionare sulle correlazioni tra il rendimento alle prove e alcune variabili di sfondo, l'erronea tendenza, affermatasi nell'ultimo ventennio, a considerarle in grado di stimare l'effetto netto di determinati processi educativi si associa alla propensione a trattare l'educazione alla stregua di variabile indipendente rispetto a fattori sociali ed economici. Questo comporta un'attribuzione eccessiva di responsabilità al sistema scolastico, che viene costantemente messo sotto pressione e si fa carico anche di iniquità che, al contrario, sono proprie del sistema sociale ed economico, mai messo in discussione (Komats e Rappleye, 2017).

Il secondo fattore di rischio legato alla partecipazione alle indagini internazionali è legato al processo di misurazione di obiettivi educativi. Benché le ricerche, agendo con piena trasparenza, rendano noti i processi di traduzione operativa di

concetti complessi, inevitabilmente il processo comporta l'eliminazione di elementi che sono ritenuti non misurabili o non rilevanti in un'ottica internazionale. I problemi legati a questo inevitabile processo di semplificazione sarebbero trascurabili se non venissero notevolmente amplificati dalla tendenza a introiettare come obiettivi educativi le misure rilevate dalle indagini internazionali. Inoltre, per alcuni Paesi la partecipazione a tali ricerche rappresenta l'unica fonte informativa strutturata sugli apprendimenti della propria popolazione studentesca, e questo può comportare una definizione di tratti essenziali del curricolo incentrata sull'importazione di obiettivi dall'esterno e non già su meccanismi decisionali legati a dinamiche interne ai diversi contesti. Questa tendenza si registra in misura crescente nei Paesi in via di sviluppo e rischia di tradursi in una forma di neocolonialismo educativo (Addey, 2020). Inoltre, se è vero che, come abbiamo visto, la partecipazione alle ricerche internazionali comporta indubbi vantaggi per lo sviluppo metodologico, la rilevanza politica assunta dalle misure tende a produrre rischiose ipersemplificazioni. Un primo esempio è quello rappresentato dalla scelta di definire "indagini comparative" ricerche che, pur procurando informazioni utili per raffrontare i diversi sistemi educativi, propongono in realtà una visione impoverita della comparazione educativa (Gaudio, 2020), basata su un'algebra sociale offerta dai ranking piuttosto che sul puntuale confronto tra analitiche descrizioni finalizzate alla comprensione di dinamiche complesse (Cowen, 2014; Auld e Morris, 2016). Un secondo esempio, in Italia, è fornito dal processo di ipersemplificazione implicito nella scelta, effettuata nel dibattito pubblico e, non di rado, nella documentazione ufficiale, di tradurre con "competenza" le diverse forme di "literacy" che le indagini internazionali indagano (Giancola e Viteritti, 2018). Se quello dell'effettiva traducibilità di concetti e processi rimane un nodo cruciale nelle indagini internazionali (Lê Thành, 1981; Nardi, 2000), in particolare questa scelta conduce alla riduzione di un costrutto complesso, pedagogicamente e didatticamente rilevante come quello di competenza, ai suoi soli tratti misurabili attraverso le prove delle indagini. Risulta così tradita la dimensione attiva, sociale, metacognitiva, contestuale, dinamica e plurale (mai riconducibile a un solo accertamento) delle competenze. L'assunzione dei risultati delle ricerche internazionali come *benchmark* di politica educativa, conferendo rilevanza a una versione impoverita delle competenze, rischia così di tradursi in un impoverimento delle prassi didattiche volte allo sviluppo delle competenze. Un fenomeno che comporterebbe anche una compromissione della validità delle stesse misure. Rispetto a queste dinamiche la ricerca educativa aveva ammonito da decenni, come evidenziato dalla "legge di Campbell": *quando i punteggi dei test diventano l'obiettivo del processo di insegnamento, perdono il loro valore come indicatori dello stato di istruzione e distorcono il processo educativo in modi indesiderabili* (Campbell, 1979).

Il bilanciamento tra benefici e rischi legati alla partecipazione alle ricerche in-

ternazionali rappresenta una delle sfide più complesse per la politica e la ricerca in campo educativo, e per affrontarla è necessario tenere conto del rapporto tra i preziosi dati forniti da tali indagini e gli strumenti, i disegni, le ipotesi, gli assunti e i valori che li hanno prodotti.

Riepilogo

Nel capitolo sono state presentate le caratteristiche principali (finalità, obiettivi, disegni di ricerca) delle indagini internazionali in campo educativo.

Sono stati descritti i passaggi che hanno portato dalle prime indagini IEA alle ricerche che attualmente vengono svolte dalla stessa IEA e dall'OCSE, focalizzando l'attenzione sull'evoluzione metodologica e sui cambiamenti che hanno fatto sì che, nel corso degli anni e in particolare nell'ultimo ventennio, i risultati di alcune indagini abbiano assunto una notevole rilevanza politica ed educativa.

Infine, sono stati evidenziati i punti di forza e le criticità delle indagini internazionali.

Riferimenti bibliografici

Addey C. (2020). La global data infrastructure di PISA e la politica OCSE nascosta in piena vista: data infrastructure alternative per valutare la qualità dell'educazione. In C. Corsini, G.C. Pillera, C. Tienken, M. Tomarchio, *Evaluating Educational Quality* (pp. 29-37). Milano: Franco Angeli.

Agrusti G. (2009). Idola theatri e indagini comparative internazionali. *Cadmo*, XIX, 1, 35-46.

Asquini G. (2007). Che cosa ci dicono le indagini internazionali sull'istruzione. *Economia e Lavoro*, XVI, 57-75.

Asquini G., Corsini C. (2010). L'evoluzione dei risultati di lettura nelle diverse edizioni di PISA. In INVALSI, *Pisa 2006 Approfondimenti tematici e metodologici* (pp. 177-201). Roma: Armando.

Auld E., Morris P. (2016). PISA, Policy and persuasion: translating complex conditions into education 'best practice'. *Comparative Education*, 52,(2)) 202-229.

Baldacci M. (2010). *Curricolo e competenze*. Milano: Mondadori.

Campbell D.T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning* 2,67-90.

Corsini C. (2019). Dalle comparazioni alle classifiche. *Cadmo*, XXVII, 2, 40-47.

Cowen R. (2014). Ways of knowing, outcomes and 'comparative education': be careful what you pray for. *Comparative Education*, 50(3)) 282-301.

Di Francesco G. (2014). *PIAAC-OCSE. Rapporto nazionale sulle competenze degli adulti*. Roma: ISFOL.

Emiletti M. (2020). Risultati in matematica della quarta primaria. In L. Palmerio, E. Caponera (a cura di), *IEA-TIMSS 2019. I risultati degli studenti italiani in matematica e scienze*. Roma: INVALSI.

Fraillon J., Ainley J., Schulz W. et al. (2020). *Preparing for life in a digital world. IEA International Computer and Information Literacy Study 2018. International Report*. Amsterdam: IEA-Springer.

Gallina V. (2005, a cura di). *Letteratismo e abilità per la vita*. Roma: Armando Editore.

Gaudio A. (2020). La valutazione nel gioco degli specchi comparativi."*Scholé*", LVIII:1, 11-26.

Giancola O., Viteritti A. (2019). Le competenze nello spazio globale dell'educazione. Discorsi, modelli e misure. *Scuola Democratica*, 1, 11-40.

Guiso L., Monte F., Sapienza P., Zingales L. (2008). Culture, gender, and math. *Science*, 320, (5880), 1164-1165.

Husèn T. (1967), *International Study of Achievement in Mathematics. A Comparison of twelve Countries*. Stockholm: Almqvist & Wiksell.

Komatsu H., Rappleye J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education* 53(2), 166-191.

INVALSI (2019). *OCSE Pisa 2018. I risultati italiani in lettura, matematica, scienze*. Roma.

Lê Thành K. (1981). *L'éducation comparée*. Paris: Colin.

Lucisano P. (a cura di, 1994). *Alfabetizzazione e lettura in Italia e nel mondo. I risultati dell'indagine internazionale IEA SAL*. Napoli: Tecnodid.

Morris P. (2016). *Education policy, cross-national tests of pupil achievement, and the pursuit of world-class schooling: A critical analysis*. London: UCL.

Mullis I., Martin M., Kennedy A., Flaherty C.M. (2002). *PIRLS 2001. Encyclopedia: A Reference guide to reading education in the countries participating in IEA's Progress in Reading Literacy Study (PIRLS)*. Boston College: Chestnut Hill.

Nardi E. (2000). Le traduzioni nelle indagini internazionali. Un problema di metodo, *Cadmo*, VII(23), 41-49.

Nardi E. (2002), *Come leggono i quindicenni. Riflessioni sulla ricerca OCSE-PISA*. Milano: Franco Angeli.

Pavan De Gregorio G. (2003). *Quadro di riferimento del progetto*. In: INVALSI, *Studio IEA-ICONA, Rapporto di ricerca*. INVALSI, Frascati.

Scheerens J., Bosker R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.

Schulz W., Ainley J., Fraillon J., Losito B., Agrusti G., Friedman T. (2017), *Becoming Citizens in a Changing World. IEA International Civic and Citizenship Education Study 2016 International Report*, IEA, Amsterdam.

VERTECCHI B. (2004), Competizione imperfetta, in *Cadmo*, XII (2), 2-5.
VISALBERGHI A. (1977), *Rapporto generale sulle ricerche IEA e sui risultati connessi. Valutazione complessiva dei risultati*". In *Misurazione del rendimento scolastico, Indagini IEA e situazione italiana, Quaderni degli annali della Pubblica Istruzione, 5, 125-150.*

Sitografia

Eurydice. Differenze di genere nei risultati educativi: studio sulle misure attuate e sulla situazione attuale in Europa.
https://eurydice.indire.it/wp-content/uploads/2017/06/Gender_IT.pdf *(ultimo accesso: 01/12/2020)*
IEA, IEA's Trends in International Mathematics and Science Study - TIMSS 2015.
http://timss2015.org/timss-2015/mathematics/student-achievement/ *(ultimo accesso: 01/12/2020)*
INVALSI, Indagine IEA-PIRLS 2016. *https://www.invalsi.it/invalsi/ri/pirls2016/documenti/risnaz/Rapporto_Nazionale_Pirls_2016.pdf (ultimo accesso: 01/12/2020)*
INVALSI, Indagine IEA-ICCS 2016: i risultati degli studenti italiani in Educazione civica e alla cittadinanza: https://www.invalsi.it/invalsi/ri/iea-doc/iccs2016/Rapporto_ICCS_2016_definitivo.pdf *(ultimo accesso: 01/12/2020)*
INVALSI, Sintesi dei risultati italiani di OCSE-PISA 2018:
https://www.invalsi.it/invalsi/ri/pisa2018/docris/2019/Sintesi%20dei%20risultati%20italiani.pdf *(ultimo accesso: 01/12/2020)*
INVALSI, Risultati OCSE-TALIS 2018:
https://www.invalsi.it/invalsi/ri/talis.php?page=talis_it_01 (ultimo accesso: 01/12/2020)
OECD, 2018 TALIS Results: Teachers and School Leaders as Valued Professionals: http://www.oecd.org/education/talis/ *(ultimo accesso: 01/12/2020)*

13

Ricerca e processi valutativi nella formazione e-learning

Con il termine e-learning ci si riferisce a una vasta gamma di modalità di costruzione di conoscenze, abilità e competenze assistita da tecnologie dell'informazione e della comunicazione. Gli interrogativi di ricerca riguardano il come utilizzare gli strumenti messi a disposizione dall'e-learning per promuovere apprendimenti efficaci degli studenti, in presenza e a distanza.
Spunti interessanti provengono dall'approccio "Evidence Informed Education", che considera prioritaria la consultazione delle evidenze prodotte dalla ricerca del passato prima di prendere decisioni in merito alle strategie didattiche e valutative da utilizzare.
Alcuni fattori di efficacia sembrano essere ormai consolidati: le strategie adottate dal docente, la qualità delle interazioni docente-studente e studente-studente, l'uso della valutazione per promuovere gli apprendimenti. Per questo è importante che i docenti siano formati all'uso didattico delle tecnologie (che non coincide con l'uso strumentale). Questo significa promuovere il ruolo cognitivamente attivo dei discenti e renderli progressivamente più autonomi nel controllo del proprio processo di apprendimento, tutti elementi che l'e-learning può potenziare, a patto di avvalersene in modo opportuno.
Una formazione specifica è necessaria anche per ribaltare la prospettiva con cui vedere i momenti valutativi: si valuta anzitutto per far apprendere e, contemporaneamente ma non primariamente, per ottenere dati sugli apprendimenti acquisiti. Anche su questo aspetto l'e-learning offre potenzialità rilevanti, che vanno dal consentire allo studente di mettersi alla prova in situazioni difficilmente esperibili altrimenti, al tenere traccia di prodotti e processi dell'apprendimento, al rendere possibili interazioni cognitive in telepresenza ad alto valore didattico.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Principi e metodi di didattica digitale efficace, in presenza e a distanza.
- Criteri per la valutazione dell'uso delle tecnologie nell'insegnamento e nell'apprendimento.
- Uso della valutazione per promuovere gli apprendimenti degli studenti.
- Attività formative efficaci in e-learning, in presenza e a distanza.
- Osservazione dello studente nella sua interazione con i pari e con il docente mediata dalle tecnologie, allo scopo di ottenere dati valutativi.

"Ma come si fa a guardare qualcosa lasciando da parte l'io? Di chi sono gli occhi che guardano? Di solito si pensa che l'io sia uno che sta affacciato ai propri occhi come al davanzale d'una finestra e guarda il mondo che si distende in tutta la sua vastità lì davanti a lui. Dunque: c'è una finestra che s'affaccia sul mondo. Di là c'è il mondo; e di qua? Sempre il mondo: cos'altro volete che ci sia? Con un piccolo sforzo di concentrazione Palomar riesce a spostare il mondo da lì davanti e a sistemarlo affacciato al davanzale. Allora, fuori della finestra, cosa rimane? Il mondo anche lì, che per l'occasione s'è sdoppiato in mondo che guarda e mondo che è guardato.
E lui, detto anche "io", cioè il signor Palomar? Non è anche lui un pezzo di mondo che sta guardando un altro pezzo di mondo? Oppure, dato che c'è mondo di qua e mondo di là della finestra, forse l'io non è altro che la finestra attraverso la quale il mondo guarda il mondo. Per guardare se stesso il mondo ha bisogno degli occhi (e degli occhiali) del signor Palomar"

Italo Calvino, *Palomar*

13.1 Tecnologie per apprendere, insegnare, valutare

Sotto l'etichetta "e-learning" viene ricompresa una vasta gamma di modalità di costruzione di conoscenze, abilità e competenze assistita da tecnologie dell'informazione e della comunicazione: apprendimento/insegnamento attraverso strumenti multimediali (*multimedia learning*; Mayer, 2010), apprendimento arricchito da sistemi tecnologici (*technology-enhanced learning*; Morgan e Johansson, 2016), istruzione assistita dal computer (*computer-based instruction, computer-based training, computer-assisted instruction, computer-aided instruction*; Lowe, 2001), istruzione assistita dalla Rete (*internet-based training, web-based training, online education, virtual education*; U.S. DoE, 2010), ambienti virtuali di apprendimento (*virtual learning environments, digital educational collaboration*; Chou e Liu, 2005), sistemi mobili per l'apprendimento (*mobile learning*; Al Zahrani e Laxman, 2016).

La tecnologia mette a disposizione contenuti (testi, audio, immagini, animazioni, video, app interattive) che possono stimolare l'esercizio di numerosi processi cognitivi da parte dei discenti e supportare molteplici strategie didattico-valutative adottate dal docente.

L'emergenza sanitaria del 2020 ha (ri)portato alla ribalta una domanda cruciale: sappiamo davvero come utilizzare le tecniche e gli strumenti messi a disposizione dall'e-learning per promuovere apprendimenti efficaci degli studenti, in presenza e a distanza? Un'importante ricerca nazionale sulle pratiche didattiche degli insegnanti italiani nei mesi del lockdown (marzo-maggio 2020), a cui hanno partecipato 16.000 insegnanti italiani (si veda Lucisano, 2020; Trinchero, 2020), ha messo in luce le seguenti tre macro-strategie didattico-valutative che hanno guidato la formazione a distanza (*forzata*, perché imposta dal lockdown) assistita dalle tecnologie info-telematiche.

1. *Tecnologie come supporto trasmissivo*: il docente usa le tecnologie come veicolo di trasmissione di materiali di studio, accompagnati da interventi di spiegazione scritta dei materiali stessi, i quali vengono visti come integrazione dei contenuti presenti sui libri di testo e supporto per compiti a casa e studio individuale. L'interazione telematica, sincrona e asincrona, viene utilizzata per accompagnare la fruizione dei testi trasmessi con interventi successivi di chiarimento e restituzione. Gli apprendimenti ottenuti vengono valutati con compiti scritti e test a distanza.
2. *Tecnologie come mezzo di telepresenza*: il docente "entra" nelle case degli studenti con lezioni espositive in videoconferenza (sincrona), a cui possono seguire interventi successivi di chiarimento, e valuta gli apprendimenti ottenuti con colloqui orali e interrogazioni da remoto.
3. *Tecnologie come mezzo di costruzione collaborativa di saperi*: il docente chiede agli studenti di svolgere lavori individuali e di gruppo supportati dalla Rete (ad esempio, ricerche a tema, esperimenti su laboratori virtuali, uso di app e programmi di simulazione) e di presentarli in videoconferenza al gruppo-classe, valutando gli apprendimenti conseguiti sulla base dei lavori svolti e della loro presentazione e discussione.

Le tre macrostrategie non sono tra di loro esclusive (gli stessi docenti possono utilizzarle in momenti differenti e in misura diversa) e non sono utili solo nella didattica a distanza: potrebbero tranquillamente essere applicate in approcci *blended* (mix distanza-presenza; Vo, Zhu e Diep, 2017) o di *flipped classroom* (lavoro degli studenti a casa complementato da discussione in classe di quanto svolto; Torstein e Rannveig, 2019).

La ricerca educativa sull'e-learning parte proprio da queste considerazioni per chiedersi: com'è possibile valutare l'efficacia di questi interventi? come progettare materiali didattici e attività che abbiano maggior probabilità di essere efficaci? come bilanciare momenti didattici e momenti valutativi? come integrarli e renderli coerenti tra di loro? Per poter produrre risposte attendibili a queste domande è necessario partire da alcune considerazioni chiave, ormai consolidate nel panorama di ricerca internazionale.

Anzitutto le tecnologie in sé *non* sono il fattore che produce apprendimento. La ricerca dimostra che i fattori essenziali sono la *qualità delle interazioni didattiche e di*

apprendimento e le *strategie adottate dal docente* (Calvani e Vivanet, 2016; U.S. DoE, 2010; Higgins, Xiao e Katsipataki, 2012). Rispetto a questi elementi le tecnologie possono svolgere una funzione supportiva ma non sostitutiva. In particolare, fattori che giocano a favore dell'efficacia delle tecnologie nella scuola sembrano essere: (a) la capacità dell'insegnante di utilizzare *più strategie didattiche*, adattandole alle diverse situazioni di apprendimento che si trova a gestire; (b) la formazione del docente all'uso *didattico* delle tecnologie, che non coincide con l'uso *strumentale* delle stesse; (c) il *controllo* da parte dello studente del proprio processo di apprendimento; (d) la *strutturazione* dei momenti di apprendimento tra pari e del feedback bidirezionale che si instaura tra studenti e insegnante (l'insegnante ottiene informazioni sul processo di apprendimento degli studenti e le usa per intervenire sul processo stesso) (Calvani e Vivanet, 2016; Hattie, 2009, 2017; Higgins, Xiao e Katsipataki, 2012). Approfondiremo questo aspetto nel paragrafo 13.2.

In secondo luogo, la ricerca dimostra che *la separazione tra momenti didattici e momenti valutativi è in larga parte artificiosa*. Se non limitati alla semplice trasmissione di informazioni dal docente al discente, tutti i momenti didattici fanno emergere dati utilizzabili a fini valutativi e tutti i momenti valutativi contribuiscono in modo determinante all'apprendimento (Dann, 2014). In particolare, la valutazione può essere vista secondo tre prospettive. La prima è quella classica della *valutazione sommativa* (*assessment of learning*), ovvero il momento valutativo segue il momento formativo e viene utilizzato per fare un bilancio su di esso. La seconda è quella della *valutazione formativa* (*assessment for learning*), in cui i momenti valutativi vengono inframmezzati ai momenti formativi allo scopo di ottenere dati sulle acquisizioni degli studenti e operare "correzioni di rotta" nei momenti formativi seguenti. La terza prospettiva (che sarà approfondita nel paragrafo 13.3) è quella della *valutazione formante* (*assessment as learning*; Earl, 2014): il momento valutativo è esso stesso momento formativo e l'allievo impara svolgendo, da solo o con altri, un compito che deriva da una consegna valutativa.

13.2 Cosa ci insegna la ricerca sull'efficacia delle tecnologie per insegnare e apprendere

Quali sono le interazioni didattiche e di apprendimento e le strategie che la ricerca considera essere efficaci quando supportate dalla tecnologia? Esamineremo in questo paragrafo alcune sintesi di risultati di ricerca tratte da meta-analisi, secondo l'approccio Evidence-Informed Education (Hattie, 2016; Calvani, 2012). I risultati presentati saranno accompagnati dal relativo indice di *effect size* (ES).

Approfondimento 13.1

L'APPROCCIO EVIDENCE-INFORMED EDUCATION: META-ANALISI ED EFFECT SIZE

L'*Evidence-informed education* è un approccio all'istruzione che promuove un processo decisionale basato sull'integrazione delle capacità di giudizio professionali dell'insegnante con la miglior evidenza empirica disponibile, tratta dalla ricerca sul campo nella scuola, in psicologia dell'apprendimento, nelle neuroscienze cognitive. In questo contesto viene considerata *evidenza* un'affermazione derivante dall'esperienza sensibile su cui vi sia convergenza di più rilevazioni rigorosamente condotte, coerente con un quadro teorico esplicito e traducibile in una decisione o pratica professionale.

Per poter tener conto nel suo processo decisionale dei principali risultati di ricerca, l'insegnante necessita di avere a disposizione delle sintesi di tali risultati, che possono venire da rassegne sistematiche di ricerche empiriche (*systematic review*) e da meta-analisi, ossia analisi secondarie di studi già condotti che mirano a sintetizzare i risultati emersi e a leggerli secondo un quadro di riferimento unitario.

L'indice di *effect size* (ES) medio di un fattore, calcolabile attraverso una meta-analisi di studi già condotti, dà una misura dell'incidenza di tale fattore sui risultati di apprendimento. Viene calcolato rapportando la differenza tra gli esiti di una rilevazione strutturata delle conoscenze, abilità, competenze del gruppo sperimentale *prima* e *dopo* l'intervento con il fattore considerato, alla variabilità congiunta degli esiti stessi.

Dire ad esempio che "dalla meta-analisi di più studi, la costruzione di mappe concettuali da parte degli allievi ha un ES di 0,57, mentre la loro esposizione ad audiovisivi ha un ES di 0,22", significa che mediamente, per l'insieme degli studi considerati, far costruire mappe concettuali agli allievi migliora i risultati di apprendimento circa tre volte in più rispetto alla visione di audiovisivi.

In genere si considera rilevante per l'apprendimento un fattore che ha un ES superiore al valore soglia di 0,4 (Hattie, 2009).

Le tecnologie digitali per la formazione possono essere classificate in cinque categorie (Li e Ma, 2010):

1. *programmi tutoriali*, che forniscono informazioni, dimostrazioni e verifica delle conoscenze acquisite (*drill & practice*);
2. *programmi di comunicazione e accesso all'informazione*, ad esempio software per gestire posta elettronica, navigare in Rete, comunicare in audio- e videoconferenza uno a uno e uno a molti;
3. *ambienti di apprendimento*, ad esempio laboratori virtuali, programmi di simulazione e ambienti ipermediali;

4. *strumenti di produzione*, ad esempio word processor, fogli elettronici, programmi di presentazione, software per l'editing foto-audio-video;
5. *ambienti di programmazione*, che permettono di scrivere software con linguaggi generali (per esempio, Java) o per usi specifici (per esempio, Logo).

Preso nel suo complesso, l'uso delle tecnologie in presenza nella formazione scolastica ha un effetto positivo, ma moderato, sugli apprendimenti ottenuti (ES = 0,35; Tamim et al., 2011). L'effetto è maggiore per gli studenti del primo ciclo (studenti "K-12", ES = 0,40) rispetto a quelli dei cicli successivi (ES = 0,29, Tamim et al., 2011; ES = 0,28, Schmid et al., 2009). Come accennato, tutto dipende da *come* la tecnologia viene usata: è più efficace se supporta l'istruzione (ES = 0,42), piuttosto che come semplice veicolo per distribuire materiali didattici (ES = 0,31), e se supporta l'elaborazione cognitiva (ES = 0,41; Schmid et al., 2009) piuttosto che come mezzo di presentazione delle informazioni (ES = 0,10).

Rosen e Salomon (2007) evidenziano che l'uso delle tecnologie per l'insegnamento della matematica porta, in media, un ES di 0,46, che sale a 0,60 in presenza di *approcci costruttivisti* all'insegnamento. Gli esiti positivi di questi approcci si vedono soprattutto sulla creatività e la capacità di collaborazione degli allievi, più che su aspetti puramente automatici ed esecutivi (per esempio, la rapidità di calcolo).

Per l'insegnamento della lettura l'ES medio è 0,16 (Cheung e Slavin, 2012), che sale a 0,38 se la tecnologia non viene usata come semplice appendice all'istruzione tradizionale, ma viene integrata all'interno delle attività di classe. Per l'insegnamento della scrittura l'ES è pari a 0,54 sulla quantità della scrittura e a 0,40 sulla qualità, che si accompagna però con un incremento significativo della capacità di collaborare con gli altri (Goldberg et al., 2003). Per l'apprendimento della seconda lingua si ha un effetto molto più marcato (ES = 1,12) (Zhao, 2003).

Tutti questi risultati evidenziano che il "*come* si usa la tecnologia" è molto più importante del "*se* si usa la tecnologia" (Mayer, 2013). La ricerca mette in luce quattro elementi particolarmente importanti per stabilirne l'efficacia d'uso: (a) la guida istruttiva che il docente offre agli allievi; (b) l'adozione di strategie che promuovano un'elaborazione profonda e significativa dei contenuti da apprendere; (c) l'utilizzo del feedback per promuovere la capacità metacognitiva degli studenti; (d) la strutturazione dei momenti di apprendimento in gruppo.

Approcci costruttivisti

Sono approcci che prevedono la costruzione attiva di conoscenze da parte dei discenti, ad esempio attraverso il coinvolgimento attivo in situazioni di problem-solving di gruppo guidato dal docente. Sono contrapposti ad approcci centrati sulla trasmissione di informazioni non accompagnata da momenti di coinvolgimento attivo (Rosen e Salomon, 2007).

13.2.1 L'importanza di una buona guida istruttiva nell'uso delle tecnologie

Due sono i rischi principali legati all'uso di tecnologie per la didattica: il *disorientamento cognitivo* e il *sovraccarico cognitivo*. Un eccesso di attività aperte, autodirette dallo studente (per esempio, ricerche non guidate su Internet, preparazione di presentazioni su temi specifici senza precisi criteri di riferimento, apprendimento per scoperta su laboratori virtuali), genera facilmente disorientamento cognitivo: è normale che gli studenti abbiano difficoltà nell'orientare l'attenzione su ciò che conta, perché si stanno formando e quindi non possono già sapere cosa conta veramente nei materiali che stanno studiando; è quindi molto più probabile che si perdano in dettagli scarsamente rilevanti, disperdendo gli sforzi in molteplici rivoli improduttivi (Hattie, 2016, p. 165). Un caso paradigmatico è l'esplorazione autonoma di contenuti in Rete: non guidati da criteri precisi, gli studenti rischiano di concentrare l'attenzione su materiali di scarsa qualità, di non cogliere i concetti chiave e di fermarsi a un'elaborazione superficiale dei contenuti (Hattie Yates, 2014, pp. 201-204).

L'esplorazione autonoma di materiali in Rete è consigliabile quindi solo agli studenti che abbiano già acquisito una prima rappresentazione di base dell'argomento in questione, mentre è del tutto sconsigliabile agli studenti che devono costruirsi in modo autonomo una rappresentazione. Il docente può guidare l'attività di esplorazione fornendo materiali introduttivi, *organizzatori anticipati, carte di studio*.

Organizzatori anticipati e carte di studio

Gli organizzatori anticipati (*advance organizers*; Ausubel, 1978) sono schemi forniti prima dell'esplorazione dei materiali di studio che facilitano la corretta collocazione delle informazioni via via esperite. Le *carte di studio* (*cartes d'étude*, Nunzati, 1990) sono guide strutturate che, fornite in anticipo, orientano l'azione dello studente nello svolgere un compito e forniscono i criteri di qualità necessari per autovalutare il proprio operato.

Un secondo pericolo è quello del *sovraccarico cognitivo*. Dare più informazioni agli studenti non significa ottenere più apprendimento, ma rischiare di generare confusione e rifiuto. La teoria del *carico cognitivo* (Sweller e Chandler, 1991; Sweller et al., 1998; Van Merriënboer e Ayres 2005), supportata da numerose evidenze empiriche (si vedano ad esempio Clark et al., 2006; Clark, 2010), sottolinea l'importanza di un'adeguata mediazione del docente nel proporre all'allievo sequenze istruttive che ottimizzino l'uso della memoria di lavoro. Secondo tale teoria, il soggetto che apprende si trova sottoposto a un carico cognitivo *intrinseco*, ossia legato alla naturale complessità del compito e al livello di expertise attuale dello studente (e quindi non riducibile), e a un carico cognitivo *estraneo*, ossia che non contribuisce agli

apprendimenti desiderati e che può essere ridotto con un'adeguata progettazione degli interventi e dei materiali didattici e con un'adeguata preselezione dei materiali di studio da parte del docente. A questi si affianca il carico cognitivo *pertinente*, ossia non strettamente legato alla naturale complessità del compito, ma comunque positivo perché utile per costruire saperi stabili e sviluppare capacità trasversali (per esempio, imparare a gerarchizzare i concetti principali presenti in un testo e costruire un diagramma ad albero rafforza la conoscenza dei contenuti oggetto di apprendimento e sviluppa anche la capacità di costruire rappresentazioni schematiche della conoscenza).

È possibile ridurre al minimo il carico cognitivo estraneo e ottimizzare il carico cognitivo intrinseco curando: (a) la *segmentazione* del compito in elementi più semplici (*chunking*), che significa anche ridurre obiettivi complessi in insiemi di obiettivi più semplici; (b) la sua *sequenzializzazione* in fasi (*sequencing*), che significa dare agli obiettivi una sequenza opportuna che ne rispetti la propedeuticità; (c) il *controllo dello studente sul ritmo di presentazione* delle informazioni (*pacing*), ossia la sua possibilità di non passare all'obiettivo successivo se non padroneggia pienamente quelli precedenti.

13.2.2 L'importanza di un'elaborazione profonda e significativa dei contenuti da apprendere

Un'elaborazione superficiale delle informazioni acquisite non è funzionale al conseguimento di apprendimenti *significativi* (Ausubel, 1978), che invece richiedono un'elaborazione "profonda" (*deep processing*; Anderson, 2009, pp. 151-152 e 166-167), ossia orientata al recupero dei significati più importanti (e funzionali agli obiettivi di apprendimento) insiti in un materiale di studio, alla loro riorganizzazione e al loro collegamento stabile ai saperi già posseduti dal soggetto (*elaboration and organization*, ES = 0,75; *strategy to integrate with prior knowledge*, ES = 0,93; Hattie, 2017).

La possibilità di operare una corretta assegnazione di significato dipende dalle preconoscenze che hanno gli studenti in relazione ai temi trattati (*prior ability*, ES = 0,94; *prior achievement*, ES = 0,55; Hattie, 2017; Clark et al., 2006; Clark, 2010), ma anche da come vengono organizzati i materiali didattici e di apprendimento. In particolare la ricerca descrive come efficaci:

1. l'uso da parte del docente di esposizioni che fanno uso coordinato di testo e grafica, in modo da indurre gli studenti a utilizzare il doppio canale (verbale e visuale) per elaborare e assegnare significato alle informazioni a disposizione (*implicit instructional methods*, ossia il testo spiega l'immagine e l'immagine spiega il testo, senza dover ricorrere ad attività istruttive esplicite; Clark, 2010, p. 36);

2. il proporre domande di comprensione, al termine di una esposizione (o di un testo) o all'interno di essa, che inducano lo studente a riflettere sui significati incontrati (Mayer et al., 2009; Anderson, 2009, pp. 167-169) e a controllare di averli colti in modo corretto (*strategy monitoring*, ES = 0,58; Hattie, 2017);
3. il proporre momenti di pratica dei saperi acquisiti su problemi di difficoltà calibrata (*deliberate practice*, ES = 0,79; Hattie, 2017; *practice exercises*; Clark, 2010, p. 166), che prevedano anche il ricorso ad aiuti esterni predefiniti e controllati (*help seeking*, ES = 0,72; Hattie, 2017);
4. la costruzione da parte degli studenti di autospiegazioni (*self-explanations*, ES = 0,61; Fiorella e Mayer, 2015 e Clark et al., 2006, p. 230), che chiedano per esempio allo studente di giustificare il processo adottato per adempiere a una consegna, sia esso costruito dallo studente stesso o proposto dal docente allo studente.

In questo scenario, i prodotti ipermediali offrono la possibilità, se correttamente progettati, di avere rappresentazioni multiple dell'informazione e di instaurare numerose forme di interattività tra informazione e utente. Ad esempio, prodotti quali i video interattivi (filmati inframmezzati da azioni che il fruitore deve svolgere, ad esempio rispondere a domande) ottengono alti indici di efficacia nelle meta-analisi (ES = 0,54; Hattie, 2017) soprattutto se vengono utilizzati come integrazione dell'insegnamento in presenza piuttosto che come sua sostituzione. Laddove manca l'interattività, ad esempio nei metodi istruttivi che utilizzano audiovisivi tradizionali, gli indici di efficacia sono molto più bassi (ES = 0,22; Hattie, 2017).

13.2.3 Il ruolo del feedback e della metacognizione

Per poter mettere in atto una efficace guida istruttiva, il docente deve prevedere frequenti momenti di sperimentazione e di "messa in pratica" delle rappresentazioni mentali costruite dagli studenti, su compiti in grado di promuovere il loro coinvolgimento (*appropriately challenging goals*, ES = 0,59; *concentration/persistence/engagement*, ES = 0,56; Hattie, 2017). Questi momenti possono essere ottime occasioni di feedback studenti -> docente (il docente vede come gli studenti applicano ciò che hanno appreso e si rende conto se è stato compreso o meno) e docente -> studenti (il docente interviene per correggere eventuali miscomprensioni). Hattie (2017) assegna una forte efficacia al feedback (ES = 0,70) e agli interventi che prevedono l'uso della valutazione formativa (ES = 0,48), così come Marzano et al. (2001) sottolineano l'importanza del fornire feedback agli studenti su processi e strategie messe da loro in atto (ES = 0,74).

Nei percorsi di formazione assistita dalle tecnologie, il feedback può essere fornito dalla macchina. L'interazione con un tutor umano offre indubbiamente molti

vantaggi, ma anche il feedback della macchina ne ha di suoi peculiari: ha meno probabilità di essere visto come minaccioso per l'immagine di competenza che l'allievo può voler dare all'esterno, può essere maggiormente strutturato, personalizzato e adattivo in relazione ai bisogni dello studente, può essere fornito in tempo reale appena commesso l'errore (Blok et al., 2002). I *sistemi di tutoraggio intelligenti* (ES = 0,48; Hattie, 2017), che sfruttano le possibilità offerte dall'intelligenza artificiale, possono aprire interessanti opportunità in tal senso.

I Sistemi di tutoraggio intelligenti

I sistemi di tutoraggio intelligenti (ITS, *Intelligent Tutoring Systems*; Graesser, Hu e Sottilare, 2018) sono programmi per calcolatore pensati per offrire un'assistenza personalizzata all'apprendimento, sulla base di una conoscenza dello studente che il sistema stesso sviluppa e affina man mano che interagisce con l'utente. Il sistema propone contenuti e consegne con cui cimentarsi e, nel far questo, costruisce e contemporaneamente monitora conoscenze, abilità e modi di approcciare i problemi, individuandone e colmandone le lacune. L'azione tutoriale si serve di tecniche di intelligenza artificiale, che supportano sia l'interazione con l'utente, allo scopo di personalizzarla, sia quella con l'esperto di dominio, allo scopo di elicitarne l'expertise e integrarla nella base di conoscenza del sistema.

Scopo del feedback, umano o computerizzato, deve essere quello di rendere lo studente progressivamente sempre più autonomo nel regolare il proprio apprendimento (secondo un processo detto *scaffolding*, ES = 0,82; Hattie, 2017), portandolo a essere in grado di scoprire da solo quando le cose "vanno bene" o "vanno male", senza dover sempre dipendere da un valutatore esterno, e a intraprendere correttivi opportuni. Questo processo di autonomia progressiva è legato anche alla capacità dello studente di trasferire i saperi acquisiti a nuove situazioni, che può essere formata dal docente mediante strategie opportune (*transfer strategies*, ES = 0,86; Hattie, 2017). Un buon feedback è quello che promuove l'uso di strategie metacognitive di controllo della propria comprensione e dell'efficacia delle strategie adottate da parte dello studente (*metacognitive strategies*, ES = 0,60; Hattie, 2017). Sono importanti: (a) i percorsi di riflessione guidata sui propri errori (*evaluation and reflection*, ES = 0,75; Hattie, 2017; Clark, 2010, p. 229); (b) i processi di autoverbalizzazione di strategie messe in atto e gli esiti ottenuti, anche supportati dal porsi domande opportune (*self-verbalization and self-questioning*, ES = 0,55; Hattie, 2017).

13.2.4 Il ruolo dell'interazione con i propri pari

Con le tecnologie o senza, non basta far "lavorare in gruppo" gli studenti per migliorare l'apprendimento: è necessario che i membri del gruppo seguano un copione preciso, dettato dal docente. Strategie quali il *reciprocal teaching* (Rosenshine e Meister, 1994), il *peer tutoring* (Rohrback et al., 2003), lo *small group learning* (Springer et al., 1999) si dimostrano particolarmente efficaci nelle interazioni didattiche tra studenti, anche supportate dalle tecnologie. Nel *reciprocal teaching* (ES = 0,74; Hattie, 2017), ogni studente a turno assume il ruolo di "insegnante" e illustra i significati associati a un materiale di studio (testo, video, multimedia ecc.), stimolando la discussione del gruppo. Questo favorisce l'elaborazione "profonda" dei materiali studiati, dato che gli studenti devono mettere in atto operazioni cognitive quali riassumere, chiarificare, formulare domande e prevedere il seguito di un brano. I feedback che giungono dal gruppo aiutano nel controllo della propria comprensione. Nel *peer tutoring* (ES = 0,53; Hattie, 2017) uno studente assume il ruolo di "docente" di uno o più compagni, non sostituendo ma affiancando l'insegnante. Tale cambiamento di ruolo, oltre a stimolare un'elaborazione "profonda" dei materiali di studio, favorisce anche lo sviluppo di capacità di autoregolazione dello studente-tutor sul proprio apprendimento e porta numerosi benefici agli studenti "tutorati", soprattutto in relazione alla comprensibilità dell'esposizione che lo studente-tutor è in grado di produrre. Nello *small group learning* (ES = 0,47; Hattie, 2017), il docente forma coppie o piccoli gruppi mirati di allievi e assegna loro una consegna di difficoltà calibrata, da svolgere utilizzando materiali didattici e tecnologie di supporto. L'interazione a coppie o piccoli gruppi mirati, anche supportata dalle tecnologie (Hattie 2009, pp. 225-226), espone gli studenti a prospettive multiple, offre continue possibilità di revisione del proprio pensiero, fa conoscere ipotesi e spiegazioni diverse per risolvere problemi, fornisce molteplici sorgenti di feedback, occasioni di correzione degli errori, modi alternativi di costruire conoscenza. Tutti questi meccanismi operano bene nel piccolo gruppo, mentre nel gruppo ampio le opportunità sono ridotte.

13.3 Integrare formazione e valutazione con l'ausilio delle tecnologie: un modello per la didattica in presenza e a distanza

Che la valutazione fosse un momento di per sé formativo è noto da anni. Già Aristotele sottolineava che l'esercizio di recuperare ripetutamente qualcosa dalla memoria ne consolida il ricordo. In tempi successivi Gates (1917) ha definito l'*effetto testing*, che ci ricorda che le prove di valutazione non servono solo a valutare l'apprendimento, ma modificano la situazione stessa di apprendimento, facilitando il ricordo dei materiali testati e indirizzando gli sforzi degli allievi verso determinati obiettivi, dedotti sulla base delle domande presenti nella prova. A partire dagli anni Settanta in Francia è stata definita la *valutazione formatrice* (Nunziati, 1990), ossia una forma di valutazione che punta, più che alla valutazione degli apprendimenti, a far assimilare agli allievi i criteri valutativi che gli insegnanti applicano alle loro

performance e a sviluppare la capacità di gestione autonoma degli errori e la padronanza degli strumenti di anticipazione e di pianificazione dell'azione. Tutte queste intuizioni sul ruolo formante della valutazione sono confermate da una messe di interessanti risultati di ricerca (Della Sala, 2016):

- svolgere un test in cui si chiede di richiamare alla memoria concetti precedentemente studiati migliora il ricordo dei concetti stessi (*effetto fortificante delle prove*) (Izawa, 1970);
- questo effetto si verifica anche quando il test viene sottoposto allo studente *prima* che il materiale venga studiato (Kornell et al., 2009);
- sottoporsi a una prova di valutazione iniziale o *in itinere* in un percorso formativo produce prestazioni migliori all'esame finale e prolunga il ricordo di quanto appreso (Roediger e Karpicke, 2006);
- il livello di prestazione iniziale (rilevato attraverso una prova iniziale) è un fattore critico per spiegare i risultati ottenuti dagli allievi in una prova finale, ma quando viene fornito un feedback agli allievi stessi a partire dagli errori nella prova iniziale, l'effetto fortificante dovuto alle prove iniziali si ottiene anche nei soggetti per i quali la prestazione nella prova iniziale è stata scarsa (Kang et al., 2007);
- i riscontri differiti (ossia il ritornare sulle domande a distanza di tempo) costituiscono una forma di pratica distribuita (*spaced practice*) che agevola maggiormente il ricordo di quanto appreso rispetto alla pratica concentrata in un tempo ristretto (Butler et al., 2007);
- le prove a risposta aperta favoriscono un'elaborazione maggiormente profonda e significativa e quindi danno un effetto testing ancora maggiore (Kang et al., 2011);
- gli studenti ricordano e organizzano meglio l'informazione in loro possesso quando svolgono delle attività in cui devono richiamare tante informazioni a partire da pochi indizi (come accade nelle prove di valutazione), rispetto a quando gli si chiede di studiare ripetutamente le informazioni (Zaromb e Roediger, 2010);
- il richiamo di quanto appreso attraverso prove di valutazione facilita il transfer degli apprendimenti a situazioni nuove (Butler, 2010);
- gli studenti che ottengono risultati peggiori a scuola coincidono spesso con quelli che mostrano una fiducia eccessiva nelle strategie di apprendimento che usano, quindi la prova di valutazione può consentire a questi studenti di rendersi conto della reale efficacia delle proprie strategie (Hacker et al., 2008).

Questi risultati sono poi coerenti con i principi derivanti dalle metanalisi di ricerche enunciati nel paragrafo precedente. Infatti, proporre momenti frequenti di valutazione:

1. consente al docente di monitorare e regolare in modo costante processi e prodotti di apprendimento degli studenti, ottenendo le informazioni necessarie per poter esercitare una guida istruttiva più puntuale e direttiva all'inizio, e man mano lasciando che essi interiorizzano la capacità di autovalutarsi;
2. promuove l'elaborazione profonda e significativa dei materiali da apprendere da parte degli allievi, perché il doverli utilizzare per rispondere a domande precise induce ad esplorarli in maniera non superficiale;
3. fornisce agli allievi occasioni per mettere alla prova i propri saperi e ottenere feedback sui propri punti di forza e di debolezza, aiutandoli ad avviare un processo di riflessione metacognitiva;
4. fornisce agli allievi - se le consegne valutative sono strutturate in modo opportuno - occasioni per interagire con altri al raggiungimento di un obiettivo e quindi confrontare le proprie rappresentazioni e azioni con quelle degli altri membri del gruppo di lavoro e apprendere osservando e dialogando con gli altri.

Tutto ciò rende più chiaro il concetto di *valutazione formante*: una valutazione che è essa stessa momento di apprendimento. Secondo Earl (2014), l'*assessment as learning* indica un processo regolatorio, attivato dalla metacognizione, che ha luogo quando lo studente monitora personalmente e sistematicamente cosa sta imparando attraverso frequenti e sistematiche prove di valutazione, e usa il feedback di tale monitoraggio per operare aggiustamenti, adattamenti e cambiamenti, anche sostanziali, nella propria comprensione. Oltre a fornire occasioni di elaborazione e confronto, il momento valutativo porta a: (a) acquisire nuove informazioni (se la prova consente di utilizzare materiali informativi di supporto); (b) chiarire con precisione gli obiettivi dell'apprendere (ossia "cosa il docente si aspetta dall'allievo"); (c) indirizzare gli sforzi dell'allievo nelle direzioni volute (riducendo il carico cognitivo estraneo, dato che specifica quali sono gli elementi considerati importanti); (d) facilitare il transfer degli apprendimenti (soprattutto se le prove chiedono all'allievo di mettere in gioco quanto appreso in più contesti operativi).

Accanto a questi aspetti formativi, vi sono gli aspetti valutativi. Osservare gli allievi mentre cooperano per adempiere a una consegna, ascoltarne le verbalizzazioni dei processi messi in atto, le argomentazioni in difesa delle proprie scelte, le modalità utilizzate per trasferire i propri saperi e adattarli a nuove situazioni, le modalità con cui recepiscono e utilizzano i feedback del docente, la crescita in relazione all'istruzione ricevuta e alle esperienze compiute, forniscono una pluralità di informazioni che possono essere utilizzate dal docente per esprimere giudizi valutativi. La prospettiva è quella di una valutazione *dinamica* (Symons, 2011): non si valutano i saperi qui ed ora, ma come questi evolvono nel corso di un processo che è allo stesso tempo formativo e valutativo.

Il processo di valutazione formante è quindi un processo ciclico e continuo di

Esperienza (derivante dal mettersi alla prova nell'affrontare specifiche consegne valutative) e di *Riflessione* (operata mediante verbalizzazione, autovalutazione, feedback da parte del docente e dei propri pari), come schematizzato in Figura 13.1. Entrambe le fasi *non sono spontanee ma guidate dal docente*: il docente propone stimoli per avviare le giuste Esperienze da parte del discente e guida il processo di Riflessione, orientando in entrambi i momenti gli sforzi degli studenti nelle direzioni opportune. Questo viene fatto: (a) costruendo materiali e consegne che puntano alla profondità di elaborazione; (b) dando tempi e ritmi di apprendimento; (c) stimolando l'autoriflessione e l'interazione costruttiva tra pari; (d) fornendo feedback mirati su tutti gli elementi che compongono il processo, sulla base dei dati raccolti con l'osservazione e l'ascolto. Le tecnologie supportano tutto il processo e ne facilitano la gestione mettendo a disposizione supporti informativi, simulativi, comunicativi, elaborativi.

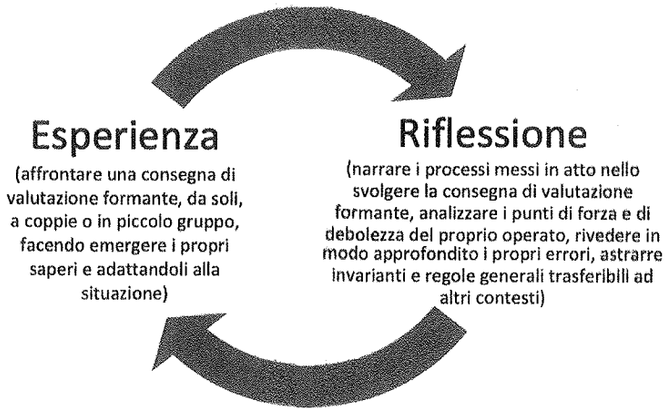


Figura 13.1 Un possibile modello generale per strutturare momenti di valutazione formante in classe e a distanza.

Organizzare le attività didattico/valutative con un momento di Esperienza (svolgere una consegna mirata a coppie o in piccolo gruppo) e un momento di Riflessione (esposizione dei prodotti e narrazione dei processi da parte degli allievi, loro analisi da parte dell'insegnante e del gruppo classe, definizione ed esposizione di regole generali di azione da parte dell'insegnante), forma gli allievi e consente al docente di farne emergere i saperi, i modi con cui interpretano e affrontano le consegne proposte, i modi in cui usano il feedback per riflettere sulle proprie interpretazioni e azioni. L'acquisizione di questi elementi informativi permette di valutarne le acquisizioni, formulare giudizi *in itinere* e riorientare il momento formativo.

L'Approfondimento 13.2 riporta esempi di "strutture" di consegne di valutazione formante atte a far partire un processo di questo tipo con attività supportate dalle tecnologie. Le strutture presentate, se riempite di contenuti opportuni, possono generare molteplici attività adatte a più ordini di scuola, dalla primaria alla secondaria di secondo grado, e queste attività possono essere svolte nell'ambito di percorsi di didattica in presenza, a distanza o mista (Trinchero, 2018).

Approfondimento 13.2

ESEMPI DI CONSEGNE VALUTATIVE ATTE A STIMOLARE MOMENTI DI ESPERIENZA

- Leggete questi due testi (o visionate questi due diagrammi, immagini, audio/video, app) ... e trovate tutte le similarità e le differenze tra di loro.
- Leggete questo testo (o visionate questo diagramma, immagine, audio/video, app) ... e indicate: a) il concetto principale che viene trattato; b) i possibili concetti di secondo livello; c) i possibili concetti di terzo livello. Evidenziateli con colori diversi.
- Leggete questo testo (o visionate questo diagramma, immagine, audio/video, app) ... e trovate tutte le incongruenze interne che presenta.
- Leggete questo testo (o visionate questo diagramma, immagine, audio/video, app) ... e trovate tutte le incongruenze che presenta con ciò che avete studiato precedentemente e/o che sono presenti sul libro di testo.
- Leggete (o ascoltate) questo testo ... e trovate tutti gli errori ortografici (*se il testo è scritto*) e sintattici.
- Analizzate queste quattro soluzioni al problema proposto ... e ordinatele dalla migliore alla peggiore, spiegando anche perché avete messo ciascuna soluzione in quella posizione. Potete aiutarvi cercanto informazioni in rete.
- Leggete questo testo (o visionate questo diagramma, immagine, audio/video, app...), scegliete un concetto tra quelli presentati (quello che più è vicino ai vostri interessi) e descrivetelo approfonditamente in massimo 10 righe.
- Leggete questo testo (o visionate questo diagramma, immagine, audio/video, app...) e trasformatelo in mappa concettuale servendovi del software fornito.
- Leggete questi tre testi (o visionate questi diagrammi, immagini, audio/video, app...) e trovate tutti gli elementi che hanno in comune.
- Leggete questi tre testi (o visionate questi diagrammi, immagini, audio/video, app...) e trovate tutte le differenze.
- Ascoltate questo testo in lingua (da audio o video) e trascrivetelo.
- Ascoltate questo audio/video e scrivete su una tabella a due colonne le cose che già conoscevate e quelle che avete sentito per la prima volta.
- Ascoltate questa descrizione audio/video e componete un disegno, schema, diagramma o

mappa concettuale che lo sintetizzi, servendovi del software fornito.

- Ascoltate questa descrizione audio/video e collegate quanto detto al disegno, schema, diagramma o mappa concettuale che vi è stato fornito, scrivendo su di esso con l'app apposita.
- Leggete questo problema che non avete mai visto prima ... e dite cosa bisognerebbe fare secondo voi per risolverlo, utilizzando le vostre conoscenze attuali e i materiali a vostra disposizione. Se non vi sembrano sufficienti cercate altre informazioni utili in Rete.
- Leggete questo problema ... e descrivete almeno tre modi possibili per risolverlo. Potete aiutarvi cercando informazioni in Rete.
- Cosa vuol dire secondo voi questa parola ...? Descrivetene il significato in massimo 20 righe servendovi degli strumenti che avete a disposizione e cercando in Rete le informazioni che pensate di non avere.
- Leggete questo testo (o visionate questo diagramma, immagine, audio/video, app) ..., elencate i punti che non vi sono chiari o di cui non conoscete il significato. Fatene un elenco e cercate materiali, servendovi degli strumenti che avete a disposizione (Rete inclusa), che li chiariscano, scrivendo, a fianco di ciascun punto non chiaro, il significato che voi avete assegnato sulla base del materiale trovato.
- Data questa definizione del termine ... rintracciate in questo testo (o diagramma, immagine, audio/video, app) ... esempi che possono essere legati a quel termine.
- Collocate i fatti descritti in questo testo (o diagramma, immagine, audio/video, app) ... nella linea del tempo fornita.
- Collocate i fatti/oggetti descritti in questo testo (o diagramma, immagine, audio/video, app) ... nella tabella bidimensionale fornita.
- Costruite un testo, diagramma, immagine sinottica, audio, video mappa concettuale, che riassuma tutto ciò che sapete (o che è stato spiegato nella lezione precedente) sul tema ...
- Costruite un articolo di quotidiano (o un audio/video) che faccia la cronaca di quanto successo nell'ultima lezione che avete seguito. L'articolo (o la clip audio) deve essere accompagnato da un'immagine significativa.
- Costruite tutte le possibili domande che l'insegnante potrebbe farvi sul testo (o diagramma, immagine, audio/video, app o ultima lezione) ... e poi scrivete a fianco le relative risposte.
- Come potrebbe evolvere il seguente sistema (da testo, diagramma, immagine, audio/video, app di simulazione) ... ? Formulate uno scenario possibile e descrivetelo in un testo o diagramma.
- Definite quali criteri dovrebbe rispettare una buona soluzione al seguente problema .... Potete aiutarvi mediante ricerca di informazioni in Rete.

L'Approfondimento 13.3 riporta alcune domande guida di stimolo per la fase di Riflessione, che ovviamente deve essere guidata dal docente e può essere supportata dal confronto con i propri pari.

Approfondimento 13.3

ESEMPI DI STIMOLI PER AVVIARE MOMENTI DI RIFLESSIONE GUIDATA

- Descrivi i passi che hai compiuto per svolgere il compito e i ragionamenti che hai fatto per prendere le decisioni relative.
- Quali sono secondo te gli elementi importanti presenti nei materiali a tua disposizione che ti hanno consentito di svolgere il compito assegnato?
- A quali altri compiti che hai svolto in passato assomigliava il compito che hai appena svolto?
- Quali difficoltà hai incontrato nello svolgere il compito?
- Come sei riuscito a superarle?
- Quali sono secondo te i punti di forza della tua soluzione?
- Quali sono secondo te le cose che si potrebbero migliorare nella tua soluzione?
- Se dovessi risolvere lo stesso problema una seconda volta, cosa cambieresti e cosa rifaresti allo stesso modo?
- Quale contributo hai fornito alla coppia/gruppo con cui hai lavorato per costruire una soluzione al compito proposto?
- Cosa sapevi degli argomenti trattati prima di svolgere questa attività?
- Cosa pensi di aver imparato nello svolgere questa attività?
- Senti di poter affrontare da solo compiti analoghi a quello appena svolto? Quali difficoltà potresti avere?
- Sulla base dei seguenti criteri di qualità: ... ... come potresti valutare la tua soluzione?
- Proponi tre criteri di qualità per valutare le soluzioni proposte al compito assegnato.

Vedere come cresce la preparazione dello studente man mano che si cimenta con consegne sempre più impegnative consente di costruirne un profilo ricco e dinamico dei saperi, anche avvalendosi delle opportunità di rilevazione, tracciamento, conservazione e analisi dei dati offerte dalle tecnologie. Lo studente stesso può percepire la crescita della propria preparazione e incrementare il proprio senso di auto-efficacia (*self-efficacy*, ES=0,92; Hattie, 2017) man mano che procede nel percorso di Attività.

13.4 Considerazioni conclusive

Sia che le tecnologie vengano utilizzate come supporto trasmissivo per le informazioni, sia che vengano utilizzate come strumenti di telepresenza, sia che vengano utilizzate come strumenti di costruzione collaborativa, l'efficacia della didattica messa in atto dipende da come l'insegnante è in grado di organizzare un buon ambiente di apprendimento, al fine di combinare sinergicamente didattica e valutazione.

Le considerazioni esposte in questo capitolo possono essere utilizzate sia in un'ottica di valutazione degli apprendimenti conseguiti dagli studenti, sia in un'ottica di valutazione degli ambienti formativi in cui tali apprendimenti hanno luogo. Esse, infatti, delineano un insieme di fattori di efficacia che, se presenti in un intervento concreto, aumentano la probabilità di costruire apprendimenti efficaci e danno un quadro di cosa significa nella pratica un *uso didattico* della tecnologia.

Riepilogo

Un percorso formativo efficace in e-learning, in presenza o a distanza, deve tenere conto di alcune istanze chiave. Anzitutto deve limitare le attività "aperte", auto-organizzate dallo studente, e privilegiare attività strutturate dal docente, che facciano uso di organizzatori anticipati e carte di studio per prevenire il *disorientamento cognitivo*. Il docente deve poi mettere a disposizione materiali che prevengano il *sovraccarico cognitivo*, tenendo presente che una quantità eccessiva di informazioni, soprattutto se non correttamente organizzate, non gioca a favore di un apprendimento significativo.

Le attività proposte devono favorire un'elaborazione profonda dei materiali oggetto di studio, facendo emergere i saperi pregressi dello studente e utilizzandoli in modo funzionale agli obiettivi di apprendimento. A tal proposito è necessario mettere frequentemente alla prova i saperi che egli ha costruito, in modo che il docente possa avere dati che gli consentano il controllo sul processo di apprendimento e l'intervento con feedback opportuni. Gli sforzi del docente devono essere orientati allo sviluppo delle capacità autovalutative dello studente, anche attraverso stimoli di autoriflessione, e alla promozione delle interazioni cognitive con i propri pari, a coppie e in piccolo gruppo.

Tutte queste istanze possono essere realizzate implementando attività che sono contemporaneamente formative e valutative e consentono quindi di sfruttare al meglio il tempo-scuola a disposizione. Nel Capitolo vengono indicati alcuni esempi di strutture di attività che rispettano questi requisiti, adattabili a numerosi contenuti, e un insieme di domande di autoriflessione. L'obiettivo finale è quello di aiutare i docenti ad acquisire consapevolezza di quale può essere un utilizzo didatticamente accorto delle tecnologie.

Riferimenti bibliografici

Al Zahrani H., Laxman K. (2016). A critical review of meta-analysis studies on mobile learning. *Technology, instruction, cognition & learning*, 10(3), 245-258.

Anderson J (2009). *Cognitive psychology and its implications* (7th ed.). New York: Worth.

Ausubel D.P. (1978; ed. or. 1968). *Educazione e processi cognitivi. Guida psicologica per gli insegnanti*. Milano: Franco Angeli.

Blok H., Oostdam R., Otter M., Overmaat M. (2002). Computer-assisted instruction in support of beginning reading instruction: a review. *Review of Educational Research*, 72, 101-130. doi.org/10.3102/00346543072001101

Butler A.C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118-33. doi: 10.1037/a0019902

Butler, A.C., Karpicke, J.D., Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology*, 13(14), 273-281. doi:10.1037/1076-898X.13.4.273

Calvani A. (2012). *Per un'istruzione evidence based. Analisi teorico-metodologica internazionale sulle didattiche efficaci e inclusive*. Trento: Erickson.

Calvani A., Vivanet G. (2016). Le tecnologie per apprendere nella scuola. Oltre il fallimento. *Pedagogia oggi*, 2, 155-178.

Cheung A., Slavin R. E. (2012). *The effectiveness of educational technology applications for enhancing reading achievement in K-12 classrooms: a meta-analysis*. Baltimore (Md): Johns Hopkins University, Center for Research and Reform in Education.

Chou S.W., Liu C.H. (2005). Learning effectiveness in a web-based virtual learning environment: A learner control perspective. *J. Comp. Assisted Learning*, 21, 65-76.

Clark R.C., Nguyen F., Sweller J. (2006). *Efficiency in learning. Evidence-based guidelines to manage cognitive load*. San Francisco: Pfeiffer Wiley.

Dann R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice, Assessment. *Education: Principles, Policy & Practice*, 21(2), 149-66.

Della Sala S. (2016). *Le neuroscienze a scuola. Il buono, il brutto, il cattivo*. Firenze: Giunti Scuola.

Earl L.M. (2014). *Assessment as learning. using classroom assessment to maximize student learning*. Cheltenham (Victoria): Hawker Brownlow.

Fiorella L., Mayer R. (2015). *Learning as a generative activity. Eight learning strategies that promote understanding*. Cambridge: Cambridge University Press.

Gates A.I. (1917). Recitation as a factor in memorizing. *Archives of Psychology*, 6 (40).

Goldberg A., Russell M., Cook A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, 2(1), 1-52.

Graesser A.C., Hu X., Sottilare R. (2018). *Intelligent tutoring systems*. In F. Fisher, C.E. Hmelo-Silver, S.R. Goldman, P. Reimann (eds.), *International handbook of the learning sciences*. New York: Routledge.

Hacker D.J., Bol L., Bahbahani K. (2008). Explaining calibration accuracy in class-

room contexts: the effects of incentives, reflection, and explanatory style. *Metacognition and Learning* 3(2), 101-121. https://doi.org/10.1007/s11409-008-9021-5

Hattie J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hattie J. (2016). *Apprendimento visibile, insegnamento efficace*. Trento: Erickson.

Hattie J. (2017). *Visible learning plus: 250+ Influences on Student Achievement*. https://visible-learning.org/wp-content/uploads/2018/03/VLPLUS-252-Influences-Hattie-ranking-DEC-2017.pdf

Hattie J., Yates G. (2014). *Visible learning and the science of how we learn*. London: Routledge.

Higgins S., Xiao Z., Katsipataki M. (2012). *The impact of digital technology on learning: a summary for the Education Endowment Foundation*. Retrieved from http://educationendowmentfoundation.org.uk/uploads/pdf/The_Impact_of_Digital_Technologies_on_Learning_(2012).pdf

Izawa C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, 83(2, Pt.1), 340-344. http://dx.doi.org/10.1037/h0028541

Kang S.H.K., McDermott K.B., Roediger H.L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European journal of cognitive psychology*, 19(4/5), 528-558. doi: 10.1080/09541440601056620

Kang S.H.K., Pashler H., Cepeda N.J., Rohrer D., Carpenter S.K. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, 103(1), 48-59. doi: 10.1037/a0021977

Kornell N., Hays M.J., Bjork R.A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(4) 989-998. doi:10.1037/a0015729

Lowe J. (2001). Computer-based education. *Journal of Research on Technology in Education* 34(2), 163-171.

Lucisano P. (2020). Fare ricerca con gli insegnanti. I primi risultati dell'indagine nazionale SIRD "Per un confronto sulle modalità di didattica a distanza adottate nelle scuole, italiane nel periodo di emergenza COVID-19". *Lifelong, Lifewide Learning* 17(36), 3-25.

Marzano R.J., Pickering, D.J., Pollock J.E. (2001). *Classroom instruction that works: research-based strategies for increasing student achievement*. Alexandria (Virginia): ASCD.

Mayer R.E. (2009). *Multimedia learning*. 2nd ed. Cambridge: Cambridge University Press.

Mayer R.E. (2010). *Research-based principles for designing multimedia instruction*. In: Benassi V.A., Overson C.E., Hakalà C.M. (eds.). *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 59-70). http://teachpsych.org/ebooks/asle2014/index.php

Mayer R.E. (2013). *Multimedia learning*. In: Hattie J., Anderman E.M. (eds.). *International guide to student achievement*. London: Routledge.

Morgan M, Johansson L. (2016). *A systematic mapping of the effects of ict on learning outcomes*. Oslo: Kunnskapssenter for utdanning, Norges Forskningsråd.

Nunziati G. (1990). Pour construire un dispositif d'évaluation formatrice. *Cahiers Pedagogiques* 280, 47-64.

Roediger H.L., Karpicke J.D. (2006). Test-enhanced learning: taking memory tests improves long-term retention. *Psychol Sci.* 17(3), 249-255. doi.org/10.1111/j.1467-9280.2006.01693.x

Rohrback C.A., Ginsburg-Block M.D., Fantuzzo J., Miller T.R. (2003). Peer assisted learning interventions with elementary school studies: A meta-analytic Review. *Journal of Educational Psychology* 95, 240-257.

Rosen Y., Salomon G. (2007). The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: a meta-analysis. *Journal of Educational Computing Research* 36(1), 1-14. doi.org/10.2190/R8M4-7762-282U-554J.

Rosenshine B., Meister C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research* 64(49), 479-530.

Schmid R.F., Bernard R.M., Borokhovski E., Tamim R., Abrami P. C., Wade C.A. et al. (2009). Technology's effect on achievement in higher education: a stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education* 21(2), 95-109.

Springer L., Stanne M.E., Donovan S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta analysis. *Review of Educational Research*, 69(1), 50-80.

Sweller J., Chandler P. (1991). Evidence for cognitive load theory. *Cognition and Instruction* 8, 351-362.

Sweller J., van Merriënboer J.J.G., Paas,F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251-296.

Symons D.K. (2011). *Dynamic assessment*. In S. Goldstein, J.A. Naglieri (eds), *Encyclopedia of child behavior and development*. Boston (MA): Springer.

Tamim R.M., Bernard R.M., Borokhovski E., Abrami P.C., Schmid R.F. (2011). What forty years of research says about the impact of technology on learning: a second-order meta-analysis and validation study. *Review of Educational Research* 81(1), 4-28. doi: 10.3102/0034654310393361

Timmerman C.E., Kruepke K.A. (2006). Computer-assisted instruction, media richness, and college student performance. *Communication Education* 55, 73-104. doi.org/10.1080/03634520500489666

Torstein L., Rannveig G.S. (2019). Does the flipped classroom improve student learn-

ing and satisfaction? A systematic review and meta-analysis. *AERA Open* 5(3), 1-17.

Trinchero R. (2018). Valutazione formante per l'attivazione cognitiva. Spunti per un uso efficace delle tecnologie per apprendere in classe. *Italian Journal of Educational Technology* 26(3), 40-55.

Trinchero R. (2020). Insegnare ai tempi del lockdown. *Rivista dell'Istruzione*, 6, novembre-dicembre 2020.

U.S. DoE (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington(DC) 2010. Retrieved from https://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf

Van Merriënboer J.J.G., Ayres P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development* 53(3), 5-13.

Vo, M.H., Zhu, C. & Diep, A.N. (2017). The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation* 53, 17-28.

Zaromb F.M., Roediger H.L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition*. 38(8), 995-1008.

Zhang M., Trussell R.P., Gallegos B., Asam R.R. (2015). Using math apps for improving student learning: an exploratory study in an inclusive fourth grade classroom. *TechTrends* 59(2), 2-39. doi.org/10.1007/s11528-015-0837-y

Zhao Y. (2003). Recent development in technology and language learning: A literature review and meta-analysis. *CALICO Journal* 21(1), 7-27. doi: 10.1558/cj.v21i1.7-27

14

Comunicazione dei risultati della ricerca scientifica e problematiche connesse

In questo capitolo presenteremo le principali modalità condivise dalla comunità scientifica per la stesura di un rapporto di ricerca.
Saranno affrontate questioni relative all'articolazione delle parti del rapporto al fine di inquadrare lo studio di cui si riferisce nell'ambito delle ricerche di settore e di assicurare al lettore la comprensione delle varie fasi dell'indagine condotta in modo da poter rendere possibile una replicazione della stessa e una verifica dei risultati raggiunti.
Verranno precisate inoltre le implicazioni deontologiche legate alle varie fasi della pubblicazione di scritti scientifici: dalla dichiarazione di autenticità dei dati e della paternità del testo, alla regolarità della procedura peer review in particolare per gli articoli in riviste.
Verranno specificati alcuni dei principali errori che si possono compiere nella redazione di un rapporto di ricerca e puntualizzati i criteri di qualità (originalità, rilevanza, rigore metodologico ecc.) da perseguire anche in vista della scelta delle sedi editoriali per la diffusione internazionale delle conoscenze scientifiche.

OBIETTIVI DI APPRENDIMENTO

Con lo studio di questo capitolo impareremo a riflettere e/o ad operare sui seguenti temi:

- Stesura di un rapporto di ricerca.
- Tipologie di scritti scientifici.
- Fasi per la pubblicazione di lavori scientifici: procedura peer review.
- Implicazioni deontologiche: paternità dello scritto, autenticità, ecc.
- Criteri per la valutazione della qualità della ricerca (originalità, rilevanza, rigore metodologico).
- Criteri per la scelta della sede editoriale.

"In tempi di generale silenzio, il conformarsi al tacere dei più è certo colpevole. In tempi in cui tutti dicono troppo, l'importante non è tanto il dire la cosa giusta, che comunque si perderebbe nell'inondazione di parole, quanto il dirla partendo da premesse e implicando conseguenze che diano alla cosa detta il massimo valore. Ma allora, se il valore d'una singola affermazione sta nella continuità e coerenza del discorso in cui trova osto, la scelta possibile è solo quella tra il parlare in continuazione e il non parlare mai. Nel primo caso il signor Palomar rivelerebbe che il suo pensiero non procede in linea retta ma a zigzag, attraverso oscillazioni, smentite, correzioni, in mezzo alle quali la giustezza di quella sua affermazione si perderebbe. Quanto alla seconda alternativa, essa implica un'arte del tacere più difficile ancora dell'arte del dire."

Italo Calvino, *Palomar*

14.1 La comunicazione scritta dei risultati scientifici attraverso la stesura di un rapporto di ricerca

Nel corso degli anni 2005-2009 da parte dell'American Educational Research Association (AERA) è stato sviluppato un progetto ambizioso: elaborare e approvare alcuni standard fondamentali di riferimento per preparare, valutare e pubblicare rapporti di ricerca in ambito educativo, distinguendo tra ricerche empiriche e ricerche orientate in senso umanistico.

Precisiamo che il termine "ricerca umanistica in ambito educativo" include - ma non è limitato ai seguenti ambiti:

1. studi educativi in cui i problemi identificati e i metodi impiegati rientrano nella competenza delle tradizionali discipline umanistiche come la linguistica, la letteratura, la storia, la giurisprudenza, la filosofia e la religione;
2. studi educativi che hanno un livello interpretativo teorico importante e che rientrano nella competenza generale delle scienze sociali a partire da discipline quali l'antropologia culturale, la sociologia, l'economia e le scienze politiche;
3. una serie di altri approcci agli studi in materia di istruzione, come studi di critica letteraria, artistica, narrativa, che non sono esclusivamente identificabili con una particolare disciplina, ma che più da vicino rientrano nei metodi generali delle scienze umane.

Successivamente sono stati definiti altri modi di fare ricerca in ambito educativo basandosi su metodologie empiriche ed evidenze (per esempio secondo il modello "Evidence Based Research"): ci si riferisce a dati o materiali empirici raccolti o identificati per affrontare i problemi allo studio.

Esaminando le Linee Guida proposte dall'APA (American Psychological Association (5ª ed.) si possono facilmente identificare le parti di un manoscritto scientifico.

La maggior parte degli articoli di riviste pubblicati in psicologia - compresi i settori educativi - è costituita peraltro da resoconti di studi empirici. È pertanto scopo di questo capitolo dare indicazioni per la loro preparazione.

La comunicazione scritta dei risultati di una indagine scientifica si articola in diverse sezioni: *introduzione e definizione del problema allo studio, rassegna della letteratura specialistica sull'argomento, ipotesi, metodo e procedura, strumenti e partecipanti, analisi dei dati, risultati, conclusioni e discussione, bibliografia.* Si tratta quindi di allestire un Rapporto di Ricerca dettagliato in modo tale da rendere sempre possibile la ripetizione delle osservazioni svolte.

Riportiamo di seguito le sezioni necessarie alla realizzazione di una buona comunicazione degli esiti di una ricerca scientifica tipo empirico:

- Abstract o Riassunto (con la sintesi dei metodi, dei risultati e delle conclusioni raggiunte);
- Introduzione e definizione del problema allo studio;
- Rassegna della letteratura specialistica sull'argomento;
- Ipotesi (scopi della ricerca e ipotesi operativa);
- Metodo e procedura;
- Strumenti e partecipanti;
- Analisi dei dati;
- Risultati;
- Conclusioni;
- Discussione (analisi dei risultati, interpretazione dei dati, confronti con altri studi, spiegazione di limiti ed eventuali errori);
- Bibliografia;
- Ringraziamenti e/o finanziamenti ricevuti;
- Appendici.

Nel dettaglio per ogni sezione si prevede la trattazione di specifici punti come riassunto di seguito (si veda anche Domenici, Lucisano e Biasi, 2017).

Nell'ambito dell'*Abstract* occorre tener conto della lunghezza richiesta e specificata nelle norme di una rivista scientifica: si tratta di allestire una breve descrizione del lavoro di ricerca (per esempio, da un minimo di 100/150 parole a un

massimo di 250). Tale descrizione va strutturata in modo da presentare chiaramente l'oggetto di studio, l'ipotesi che si è inteso testare, la metodologia utilizzata e i risultati principali ottenuti.

L'Abstract può prevedere brevi paragrafi ben strutturati e concisi, deve naturalmente contenere le stesse informazioni dell'articolo esposte in modo essenziale, e deve in genere essere comprensibile per una audience allargata.

Approfondimento 14.1
SCELTA DEL TITOLO DELL'ARTICOLO SCIENTIFICO E STRUTTURA DELL'ABSTRACT

Titolo. Un titolo deve riassumere l'idea principale dell'elaborato in modo semplice e chiaro

Nome/i dell'Autore/i e appartenenza istituzionale. Vengono spesso indicati anche i cosiddetti "running heads": si tratta di titoli abbreviati posti in cima alle pagine di un articolo per facilitare l'identificazione dell'articolo da parte del lettore.

Abstract. Un abstract è un riassunto, breve ed esauriente, dei contenuti dell'articolo; permette al lettore di stimare i contenuti di un articolo velocemente e, al pari del titolo, svolge funzioni informative e riassuntive per indicizzare e recuperare gli articoli.
In particolare:
Un **abstract di un rapporto di *uno studio empirico*** deve descrivere:
- il problema sotto indagine, in un singolo periodo se possibile;
- i partecipanti o soggetti, specificando caratteristiche pertinenti, come il numero, l'età, il genere ecc.;
- il metodo empirico, includendo l'impianto, le procedure di raccolta dei dati, i nomi completi dei test ecc.;
- i risultati, includendo i livelli di significatività statistica;
- le conclusioni e implicazioni o applicazioni.

Un **abstract per un articolo teorico o di rassegna critica** deve descrivere:
- l'argomento, in una frase;
- lo scopo, la tesi o il costrutto che guida e la portata (allargata o selettiva) dell'articolo;
- i riferimenti usati (esempio, osservazioni personali, letteratura pubblicata);
- le conclusioni.

Un **abstract per un paper metodologico** deve descrivere:
- la classe di metodi generali proposti o discussi;
- le caratteristiche essenziali del metodo proposto;
- il campo di applicazione del metodo proposto;
- il comportamento della metodologia, includendo la sua efficacia e resistenza (robustezza) alle violazioni degli assunti.

Un **abstract per lo studio del caso singolo** deve descrivere:
- il soggetto e le caratteristiche rilevanti dell'individuo o dell'organizzazione presentata;
- la natura della soluzione del problema illustrata dall'esempio del caso;
- le questioni sollevate per una ulteriore ricerca o una teoria.

Nell'ambito dell'*Introduzione* si illustra al lettore la rilevanza dello studio condotto inquadrandolo nel contesto nazionale e internazionale (*background*) costituito da studi precedenti e da questioni ancora non risolte. Occorre definire con precisione il problema oggetto della ricerca e gli scopi dello studio cercando di essere concisi senza però semplificare eccessivamente il discorso.

Si passa quindi all'esame della ricerca svolta precisando la *Metodologia* utilizzata e la *Procedura* seguita (per esempio, le fasi della sperimentazione allestita). Ciò allo scopo di permettere ad altri ricercatori di valutare l'appropriatezza dei metodi usati e di poter ripetere l'esperimento o l'osservazione per ulteriore verifica.

Occorre inoltre precisare gli *Strumenti* applicati per la raccolta dei dati.

È necessario definire la popolazione oggetto dello studio (i *Partecipanti*) e le variabili oggetto dell'analisi statistica. Non è pertinente invece introdurre discussione o conclusioni in questa sezione.

Nell'ambito del paragrafo che illustra i *Risultati* occorre procedere a una presentazione oggettiva dei dati evitando interpretazioni e confronti. Si descrivono nel testo le *tecniche di analisi dei dati* adottate (le quali possono costituire anche un paragrafo indipendente che precede immediatamente l'esposizione dei risultati conseguiti), i risultati principali e secondari, anche illustrati con diagrammi e figure (nelle quali si possono inserire vari dettagli), indicando eventuali conversioni di dati se si sono rese necessarie.

Per chiarezza è utile presentare i dati secondo un ordine logico evitando duplicazioni o ripetizioni ridondanti che potrebbero appesantire l'esposizione.

Nella parte dedicata alle *Conclusioni* occorre evidenziare i risultati più significativi e le conclusioni raggiunte dalla ricerca: si debbono riassumere quindi i punti chiave dello studio senza ancora introdurre nuove analisi o interpretazioni.

Tali riflessioni potranno invece venire sviluppate nella specifica sezione *Discussione* dedicata all'interpretazione dei risultati, supportando le conclusioni e confrontando i dati dello studio con quelli già presenti in letteratura. È sempre utile discutere eventuali errori e limiti dello studio, offrire spiegazioni alternative dei propri dati - se ne esistono - e discutere i quesiti irrisolti e le prospettive di sviluppo futuro dello studio.

Seguono naturalmente i *Riferimenti bibliografici* i quali devono essere completi e

aggiornati, ma non ridondanti. Si aggiungono eventuali *Ringraziamenti* e/o riferimenti a *finanziamenti* ricevuti. In calce si possono inserire delle *Appendici* in cui mostrare altri dettagli dello studio ed il cui inserimento nel testo avrebbe potuto appesantire gli aspetti narrativi.

L'indicazione dei nomi degli Autori concerne naturalmente coloro che hanno contribuito alla ricerca. Il primo autore è in genere colui che ha dato il contributo maggiore anche in termini di ideazione della ricerca; l'ordine dei nomi degli autori in genere riflette il contributo di ognuno, altre volte è invece alfabetico. L'ultimo autore è solitamente colui che ha dato il secondo importante contributo e spesso è il membro senior del gruppo di ricerca.

Approfondimento 14.2
CHECK-UP SUL RAPPORTO DI RICERCA

In sintesi possiamo considerare che le varie Sezioni di una *buona relazione* scientifica debbono ottemperare i seguenti quesiti:

- La ricerca risponde a un interrogativo significativo e il lavoro è originale ed importante (rilevante)?
- Gli strumenti si sono rivelati affidabili e validi?
- Le misure degli esiti sono chiaramente connessi alle variabili concernenti la ricerca?
- Il disegno sperimentale esamina pienamente e senza ambiguità l'ipotesi?
- I partecipanti sono rappresentativi della popolazione a cui mirano le generalizzazioni?
- La ricerca osserva gli standard etici nel trattamento dei partecipanti?
- La ricerca è a una tappa abbastanza avanzata da rendere significativa la pubblicazione dei risultati?

Si vuole qui sottolineare come sia la presentazione scritta sia quella orale di comunicazioni scientifiche - ossia *i resoconti o rapporti di ricerca - rappresentino in definitiva un passo critico del metodo scientifico e siano una delle chiavi del progresso del progresso all'interno della ricerca scientifica stessa.*

La costruzione di un manoscritto scientifico deve in sostanza tener strettamente conto di tre distinti elementi principali: (1) l'idea generale; (2) l'esecuzione della ricerca, e (3) la presentazione della sintesi di quella ricerca.

Si può peraltro facilmente convenire sul fatto che un'idea di scarso valore scientifico o un'indagine mal progettata non possa essere salvata da una presentazione, seppur eccellente. Ciò significa, in sintesi, che la prima condizione necessaria per allestire un buon resoconto scientifico è certamente quella di aver condotto una ricerca di qualità, che risponda cioè a criteri condivisi dalla comunità scientifica quali: *originalità, innovatività e rilevanza dello studio nel quadro nazionale e internazionale, rigore metodologico, coerenza formale tra ipotesi e impianto procedurale adottato, verificabilità delle asserzioni e/o degli obiettivi e chiarezza espositiva.* Sarà poi importante per il confronto scientifico anche il rispetto del *criterio di pertinenza dell'indagine rispetto ai temi trattati nella sede editoriale prescelta.*

14.2 Tipologie di scritti scientifici

Gli scritti scientifici si articolano generalmente in *rapporti di studi empirici, articoli metodologici, articoli di rassegne critiche, articoli teorici* e *studi di casi.*

Gli articoli originali di ricerca possono essere costituiti sia da *rapporti di studi empirici,* sia da *note brevi di ricerca,* sia da *articoli metodologici.*

- **Rapporti di studi empirici.** Sono comunicazioni di ricerche originali. Sono composti tipicamente da sezioni distinte che riflettono le tappe del processo di ricerca così come sopra descritto. In linea di massima devono comparire almeno le seguenti quattro sezioni:
  1. Introduzione: sviluppo del problema investigato e dichiarazione del fine della indagine.
  2. Metodo: descrizione del metodo usato per condurre l'indagine.
  3. Risultati: comunicazione dei risultati individuati.
  4. Discussione: interpretazione e discussione delle implicazioni dei risultati).

- **Comunicazioni brevi (note di ricerca).** hanno le stesse caratteristiche degli articoli originali di ricerca salvo il fatto che si tratta di ricerche ancora in corso di cui si vuole dare notizia alla comunità scientifica rispetto ai primi dati ottenuti. Si tratta quindi di resoconti scientifici più brevi, progetti pilota, ricerche in progress.

- **Articoli metodologici.** sono elaborati in cui sono presentati alla comunità dei ricercatori nuovi approcci metodologici, modifiche ai metodi esistenti e discussioni di approcci quantitativi e tecniche di analisi dei dati. Questi elaborati devono focalizzarsi su approcci metodologici e analisi dei dati accessibili e dovrebbero introdurre dati empirici soltanto come illustrazione dell'approccio. Gli articoli metodologici devono essere presentati a un livello che li renda accessibili al ricercatore preparato e presentare sufficienti dettagli affinché i ricercatori possano valutare l'applicabilità della metodologia al loro problema di ricerca. Inoltre l'articolo deve permettere al lettore di paragonare ragionevolmente l'approccio proposto con approcci alternativi usati abitualmente. Negli articoli metodologici i documenti altamente tecnici (per esempio, fonti, prove, dettagli di simulazioni) devono essere presentati in appendice per ampliare la comprensione complessiva dell'articolo.

- **Articoli teorici.** Sono elaborati nei quali l'autore riflette sulla ricerca esistente in letteratura per proporre una teoria in qualsiasi area dell'ambito di studio. Gli articoli teorici e quelli riguardanti rassegne sono spesso simili nella struttura, ma gli articoli teorici presentano informazioni empiriche solo quando riguardano questioni teoriche. L'autore traccia lo sviluppo di una teoria per ampliare o perfezionare i costrutti teorici. Comunemente l'autore può presentare una nuova teoria; in alternativa potrebbe analizzare una teoria esistente sottolineando i limiti o dimostrando la superiorità di una teoria sull'altra. In questo tipo di analisi teorica l'autore esamina abitualmente la coerenza interna ed esterna della teoria, ovvero se una teoria si auto-contraddice o se teoria e osservazione empirica si contraddicono tra loro. Le sezioni di un articolo teorico, come quelli della rassegna critica, sono comunemente ordinate secondo un nesso di contenuto piuttosto che secondo la cronologia.

- **Rassegne critiche (o review) e meta-analisi.** Gli articoli costituiti da rassegne critiche *(review)*, incluse le meta-analisi, sono valutazioni critiche di materiale già pubblicato. Attraverso l'organizzazione, l'integrazione e la valutazione di materiale pubblicato precedentemente, l'autore considera lo stato dell'arte della ricerca attraverso una chiarificazione del problema. In un certo senso un articolo di analisi critica ha un carattere didattico quando l'autore:
  1. definisce e chiarifica il problema;
  2. riassume le indagini precedenti al fine di informare il lettore sullo stato dell'arte della ricerca;
  3. identifica relazioni, contraddizioni, vuoti e incoerenze nella letteratura;
  4. suggerisce il prossimo o i prossimi passi nella soluzione del problema.

Le parti di una rassegna, a differenza delle sezioni di un rapporto di studi empirici, sono disposte secondo un nesso logico più che cronologico.

- **Studi di caso (*case study*) o casi clinici o case report.** Gli studi di caso sono elaborati in cui l'autore descrive il materiale relativo a un caso ottenuto lavorando con un individuo o una organizzazione per illustrare un problema, indicare un mezzo per risolverlo o far luce su ricerche necessarie o questioni teoriche. Nello scrivere gli studi di un singolo caso, gli autori considerano attentamente l'equilibrio tra fornire documenti illustrativi importanti e usare documentazione confidenziale sul caso in modo responsabile. La riservatezza è generalmente affrontata in uno dei due modi seguenti. Un'opzione è preparare la documentazione descrittiva del caso e presentarla al soggetto dello studio per ottenere un consenso scritto. L'altra opzione è mascherare alcuni aspetti della documentazione così che né il soggetto, né coloro che conoscono il soggetto possano identificarlo. Tale mascheramento del caso è una questione delicata perché è essenziale non modificare le variabili connesse con il fenomeno che si sta descrivendo. A tal fine, sono possibili tre strategie principali : (1) alterare specifiche caratteristiche, (2) limitare la descrizione di specifiche caratteristiche, (3) offuscare dettagli del caso aggiungendo della documentazione estranea. Per informazioni aggiuntive sulla presentazione della documentazione sul caso, si veda Vanden Bos (2001).

- **Editoriali.** Possono rappresentare un *punto di riflessione* sullo stato dell'arte in merito a un argomento di ricerca ed essere occasione di riflessioni per ricerche future.

Altri tipi meno frequenti di articoli pubblicati nelle riviste scientifiche includono *brevi rapporti, commenti e repliche su articoli precedenti e monografie.*

Sebbene i contenuti di questi articoli siano dissimili, i manoscritti devono essere logicamente e coerentemente organizzati in sostanziale accordo con le linee guida sopra descritte.

Approfondimento 14.3

PRINCIPALI ERRORI DA EVITARE NELLA REDAZIONE DI UN RAPPORTO DI RICERCA

In genere occorre evitare i seguenti frequenti errori:

1. stesura di una pubblicazione ridondante (per esempio frutto della divisione di un singolo rapporto in una serie di elaborati che si sovrappongono);
2. presentazione di dati insufficienti per interpretare il fenomeno allo studio;
3. presentazione di risultati negativi senza tener conto della "potenza" dell'analisi statistica sottostante;
4. mancanza di concordanza tra specifiche operazioni di uno studio (incluse quelle relative al disegno e all'analisi statistica) e l'interpretazione e la discussione degli esiti dello stesso;
5. omissioni sulle dimensioni dell'effetto;
6. omissioni dei controlli necessari relativamente a un aspetto particolare ma importante dello studio;
7. livello irrilevante di dettaglio della ricerca nell'affrontare il problema: c'è una differenza tra esplorare i limiti della generalizzabilità delle conclusioni di una ricerca e la produzione infinita di elaborati che comunicano irrilevanti cambiamenti di ricerche precedenti.

Si annoverano, inoltre, tra gli *errori maggiori*, i seguenti

- non identificare chiaramente i concetti da trasmettere al lettore;
- scrivere una traccia senza ripartirla in sezioni e paragrafi;

- non distinguere chiaramente le sezioni dei Metodi e dei Risultati;
- esporre le varie informazioni in modo generico e non dettagliato;
- scrivere usando un linguaggio circonvoluto, ambiguo, criptico o troppo complesso;
- usare frasi troppo lunghe e/o confuse;
- dare informazioni ridondanti;
- fare affermazioni del tutto soggettive;
- non quantificare le informazioni;
- non illustrare i dati con figure e diagrammi.

14.3 Questioni deontologiche nella conduzione della ricerca e nella rendicontazione dei risultati

14.3.1 Principi etici legati alla conduzione della ricerca

Come sappiamo in letteratura sono noti i limiti fisici ed etici alla ricerca: tra i primi si trova per esempio l'impossibilità di tornare a un periodo storico antecedente per poter osservare in condizioni naturali i fenomeni cui siamo interessati. Ci si può avvalere allora, talvolta, di ricostruzioni e modelli ipotetici per avanzare e testare ipotesi.

Più delicato ancora è, in qualche modo, il problema dei limiti etici alla conduzione della ricerca i quali sono stati via via sempre più evidenziati dalla varie Società Scientifiche.

Si tratta di rispettare la volontà e la dignità dei partecipanti alla ricerca, che devono essere informati e reclutati in modo consapevole e volontario. Naturalmente la partecipazione all'indagine proposta non deve mettere in pericolo la loro salute e quindi non deve arrecare danni e tantomeno effetti irreversibili.

Si deve dare inoltre garanzia dell'anonimato dei partecipanti e del trattamento dei dati solo per fini di conoscenza scientifica.

14.3.2 Principi etici legati al processo di valutazione e pubblicazione dell'articolo

Tali principi possono essere legati: all'autore che deve essere garante dell'autenticità dei dati che presenta; all'Editor o Direttore di una Rivista che deve assicurare la serietà e l'originalità dei lavori che vengono pubblicati; alla procedura di *peer review* che deve essere condotta in modo ortodosso; al conflitto di interesse che deve essere declinato da tutte le parti contraenti il contratto di edizione; al diritto di privacy e confidenzialità da rispettare garantendo l'anonimato dei partecipanti allo studio; al processo di pubblicazione nel suo insieme che non deve ledere la libertà di opinione e deve garantire la trasparenza per la riproducibilità della ricerca medesima.

Un altro aspetto etico riguarda i contenuti da pubblicare: si devono pubblicare tutti gli studi ben condotti su un argomento importante, anche se i risultati sono negativi.

14.3.3 Considerazioni etiche legate all'autore

Tali considerazioni sono legate alla garanzia di autenticità e originalità dei dati e all'approvazione finale della versione da pubblicare. Se vi sono più autori, essi devono assumersi la responsabilità diretta del manoscritto nelle sue varie parti.

Occorre scogliere eventuali problemi per la paternità dell'articolo: alcune riviste (Lancet, BMJ, Annals of Internal Medicine, Journal of Clinical Oncology) attualmente richiedono agli autori di dichiarare il loro ruolo nella ricerca presentata.

Precisiamo che vi è un "conflitto di interesse" quando un reviewer o un editor hanno interessi finanziari, personali, commerciali, politici o accademici che potrebbero influenzarne le azioni. Quindi per evitare questi problemi è buona norma che tutti i partecipanti al processo di pubblicazione dichiarino eventuali relazioni che potrebbero presentare o essere viste come un potenziale conflitto di interesse.

14.4 La scelta delle Sedi Editoriali

14.4.1 Criteri per la proposta di un articolo a una rivista scientifica che adotta la procedura peer-review

Com'è noto, nell'ambito delle scienze che si annoverano nel sistema bibliometrico, la scelta della rivista dovrebbe orientarsi su quella con massimo *impact factor* (IF) e massimo *citation index* (CI).

Tuttavia, va tenuto in considerazione il fatto che l'articolo sia "appropriato" per lo stile della specifica rivista e quale sia il "pubblico" (o target) a cui è destinata la sede editoriale.

È utile quindi esplorare qual è il focus della rivista; che tipo di articoli pubblica (per esempio, *clinically based* o *basic science based*); se il format scelto per l'articolo (per esempio, review, case report ecc.) viene pubblicato da quella rivista; oppure se l'articolo che intendiamo pubblicare risulta "troppo specialistico" per quella sede editoriale.

È naturalmente utile scegliere la rivista prima di iniziare a scrivere l'articolo, leggerne attentamente le "linee guida per gli autori" e consultare i numeri recenti per capire qual è il tipo di articoli che viene normalmente pubblicato.

Una scelta inappropriata della rivista può ritardare la pubblicazione di un articolo.

Un fattore importante per la scelta della sede editoriale è il tipo di "indicizzazione" seguito. La possibilità di trovare un articolo rilevante nelle banche dati elettroniche dipende infatti in larga parte da come è stato indicizzato: ciò è pertanto determinante per la sua diffusione.

Le Banche Dati più utili ai fini dell'indicizzazione sono Medline, PsychInfo, WOS e SCOPUS.

Fondamentale risulta a questo punto la scelta delle cosiddette "parole chiave" o "keywords". In genere ne vengono chieste 3-5 da inserire dopo l'abstract, e devono cogliere l'essenza dell'articolo.

Parole chiave o keywords

Si tratta di scegliere 3-5 parole che identificano in modo pregnante gli argomenti trattati nell'articolo scientifico.

Tali parole devono cogliere l'essenza dell'indagine svolta per facilitarne la catalogazione nei vari database: ciò al fine di permettere al lettore l'acquisizione e la diffusione delle conoscenze proposte.

Possono essere utilizzati dei Thesaurus noti per individuare parole chiave già condivise dalla comunità scientifica.

Lo stile della rivista espresso nelle "Norme redazionali" deve essere rispettato, con particolare attenzione allo stile della Bibliografia.

Le riviste richiedono che i contenuti dello studio non siano stati già oggetto di pubblicazione, che siano quindi originali, e chiedono altresì espressamente che eventuali conflitti di interesse vengano segnalati.

Nell'ambito del processo editoriale, l'editore può decidere di rifiutare un articolo senza passarlo ai reviewer perché palesemente non adatto nei contenuti o nella metodologia allo standard della rivista.

Se giudicato pertinente un articolo è inviato al vaglio di due o più revisori che si devono esprimere in termini di accettazione, richiesta di revisioni minori, richiesta di revisioni maggiori, o di rifiuto.

La velocità del processo editoriale è direttamente proporzionale al prestigio della rivista.

Con le nuove procedure elettroniche il processo editoriale può essere completato in tempi abbastanza brevi (alcuni mesi o talvolta meno di un mese).

La scelta dei reviewer è fondamentale poiché sul loro lavoro si basa la credibilità della rivista: un reviewer deve essere un esperto del settore e deve emettere giudizi obiettivi, motivati, formulando critiche costruttive per gli autori

Ricevute le proposte di revisione all'articolo, l'Editor si esprime in via definitiva sull'accettazione o il rifiuto del medesimo. Nel caso di giudizi contrastanti dei Referee, l'Editor può prendere la decisione finale sulla base della qualità dei giudizi ricevuti oppure inviare a un terzo reviewer l'articolo.

I referee che chiedono revisioni importanti (cosiddette "maggiori") sono invitati a riesaminare il manoscritto per verificare i cambiamenti introdotti dagli autori.

Alcune riviste stilano un elenco di Reviewer cosiddetti inefficienti (in quanto sono risultati troppo lenti o superficiali o ipercritici).

A questo punto vediamo come affrontare le critiche dei Reviewer.

Se un reviewer richiede revisioni maggiori, la pubblicazione dell'articolo non è scontata: i punti sollevati dai reviewer vanno affrontati meticolosamente e alcune riviste scompongono in punti le critiche dei reviewer, che devono essere affrontati uno per uno.

L'autore può comunque controbattere le critiche di un reviewer se ha argomentazioni forti e può anche opporsi alla decisione di rifiuto dell'Editor chiedendo di far riesaminare l'articolo da altri reviewer, ciò nel caso in cui le critiche ricevute risultino immotivate, speciose o superficiali.

Nel caso di alcune Riviste che lo prevedano espressamente, l'autore può esprimere all'Editor la preferenza di far esaminare l'articolo da uno o più revisori di sua scelta. L'autore ha anche il diritto di segnalare uno o più reviewer con cui ha conflitti di interesse. L'Editor ha la facoltà di accettare o meno la richiesta dell'autore.

14.5 La diffusione delle conoscenze scientifiche in ambito educativo: potenzialità e aspetti problematici

La diffusione internazionale delle conoscenze è affidata in particolare ai database internazionali disponibili sui principali motori di ricerca che permettono di calcolare una *misura di impatto* della produzione scientifica.

Per le cosiddette scienze dell'uomo - ad eccezione delle scienze psicologiche che recentemente in Italia hanno adottato il sistema bibliometrico - di cui fanno parte i settori di ambito storico, umanistico, sociale, filosofico, pedagogico, si utilizzano i *sistemi non bibliometrici.*

A tale riguardo la diffusione delle conoscenze scientifiche viene assicurata attraverso la pubblicazione di "Monografie" curate da Case Editrici preferibilmente a circolazione nazionale o internazionale. Occorre avere la garanzia di un Comitato Scientifico che assicuri la valutazione o referaggio delle opere da pubblicare. Sono di particolare valore inoltre in questi settori le cosiddette *Riviste di Fascia A*, cioè classificate secondo i criteri proposti dall'ANVUR, come riviste di alto livello nei rispettivi settori scientifici. Tali sedi editoriali garantiscono la pubblicazione di articoli scientifici caratterizzati quindi da originalità, correttezza metodologica, ecc. e applicano la procedura di peer review (ossia la valutazione in condizioni di doppio cieco) di ogni testo da parte di esperti del settore.

Si annoverano anche altre sedi editoriali come Riviste Scientifiche che ancora

non rientrano nella cosiddetta fascia A ma rispettano comunque le norme condivise per la stesura di rapporti scientifici.

I database per la circolazione degli articoli pubblicati su Riviste di Fascia A corrispondono ai più noti come Google Scholar e, in particolare, Scopus.

A livello internazionale però spesso - sia nel mondo anglosassone sia in alcuni contesti latini - anche le Scienze dell'Educazione (ossia il settore "Education") utilizza il *sistema bibliometrico* per la diffusione delle conoscenze scientifiche. In questo sistema, al fine di ottenere una maggiore diffusione possibile delle ricerche scientifiche condotte e quindi, in ultima istanza, per valutare il peso scientifico di una pubblicazione, i principali database permettono di calcolare il valore dell'*Impact Factor della sede editoriale* (Rivista) e il numero di citazioni ottenute per articolo e autore (*Citation Index*) fino a generare l'*indice di Hirsch* (uno scienziato ha un indice *n* se ha pubblicato almeno *n* lavori, ciascuno dei quali sia stato citato almeno *n* volte: il che permette di escludere i lavori che non hanno avuto alcuna citazione).

Il Citation Index è l'indice delle citazioni tra le varie pubblicazioni e permette al lettore di stabilire quali articoli citano quali altri articoli. Nel 1960, l'*Institute for Scientific Information* (ISI) di Eugene Garfield ha introdotto il primo Citation Index per gli articoli pubblicati nel Science Citation Index (SCI), oggi noto in internet come *Web of Science* (WOS).

Il valore dell'Impact factor diviene così una misura dell'importanza scientifica di una rivista, pubblicata sul Journal Citation Reports. Viene calcolato annualmente dall'Institute for Scientific Information (ISI) per le riviste da esso indicizzate con riferimento all'anno precedente sulla base di:

- A = numero di volte in cui un articolo pubblicato per esempio nel 2007 e nel 2008 viene citato da riviste del 2009;
- B = numero totale di "citable items" pubblicati per esempio nel 2007 e 2008 (tra cui articoli, reviews; sono esclusi scritti minori come Editorials o Letters-to-the-Editor).

Il coefficiente IF (Impact Factor) si ottiene dal rapporto tra questi due valori (IF= A/B).

L'IF di una rivista è calcolato su base triennale e può essere considerato come la media del numero di volte in cui gli articoli pubblicati sono citati nei 2 anni dopo la pubblicazione. Ne consegue che eventuali nuove riviste appena indicizzate non vedranno pubblicato il loro IF prima di tre anni

Attraverso il sistema citazionale bibliometrico, WOS indicizza più di 9000 riviste provenienti da 60 Paesi e quindi garantisce una larga diffusione internazionale

della ricerca scientifica.

A tale importante database si aggiunge, in ordine di diffusione, il famoso *Scopus* che costituisce la più grande banca dati volta a raccogliere abstract e citazioni di letteratura peer-reviewed e fonti web di qualità.

In sintesi, conoscere l'impatto che il nostro lavoro sta avendo sulla comunità dei ricercatori è utile per misurare l'influenza del lavoro dei nostri colleghi e concorrenti; seguire la scia delle idee e dei concetti più importanti e interessanti; determinare se una teoria è stata confermata, cambiata o implementata; vedere come è stato applicato un concetto di base; valutare negli anni lo sviluppo di una argomento di ricerca nella letteratura; verificare l'accuratezza della bibliografia.

Vanno tuttavia sempre tenute presenti alcune importanti criticità che possono rendere un circuito virtuoso affetto da pratiche viziose e di abuso degli indici citazionali. Sono noti infatti i rischi di pratiche scorrette (una versione degenerativa rispetto alle famose "buone pratiche") per alzare l'indice di citazione di una Rivista, e/o abusi di citazioni per contrattazioni "amichevoli" tra colleghi.

Inoltre, conoscere l'impatto che il nostro lavoro sta avendo sulla comunità dei ricercatori, visto che il numero di citazioni orienterà la carriera del ricercatore stesso, rischia di spostare gli interessi di ricerca in modo strumentale. Per questo motivo l'IF può correre il rischio di rappresentare un *indicatore di conformismo* penalizzando ricerche originali che non hanno ancora un fondo di citazioni rilevante (in quanto pionieristiche o in quanto appartenenti, per esempio, a settori altamente specialistici, di nicchia, che vedono pochi studiosi coinvolti).

Occorrerà quindi fare un uso virtuoso degli indici citazionali attribuendo loro un peso equilibrato, evitando quindi sia una dipendenza passiva sia una esaltazione magnificante.

Riepilogo

In questo capitolo sono state presentate le principali modalità condivise dalla comunità scientifica per la stesura di un rapporto di ricerca. È stata poi descritta analiticamente l'articolazione delle parti del rapporto che permettono di inquadrare lo studio nell'ambito delle ricerche di settore e di assicurare al lettore la comprensione delle varie fasi dello studio condotto in modo da poter rendere possibile una replicazione dello stesso e una verifica dei risultati raggiunti.

Una parte del capitolo è stata dedicata all'illustrazione delle implicazioni deontologiche legate alle varie fasi della pubblicazione di scritti scientifici: dalla dichiarazione di autenticità dei dati e della paternità del testo, alla regolarità della procedura peer review.

Sono stati commentati criticamente alcuni dei principali errori che si possono

compiere nella redazione di un rapporto di ricerca e puntualizzati i criteri di qualità per la realizzazione di una buona ricerca (originalità, rilevanza, rigore metodologico ecc.) da perseguire anche nella scelta delle sedi editoriali per la diffusione internazionale delle conoscenze.

Riferimenti bibliografici

AERA (2006). Standards for reporting on empirical social science research in AERA Publications: American Educational Research Association. *Educational Researcher* 35(6), 33-40.
https://journals.sagepub.com/doi/abs/10.3102/0013189x035006033?journalCode=edra

APA (2005). *American Psychological Association Statement. Policy statement on evidence-based practice in psychology.*
http://www.apa.org/practice/guidelines/evidence-based-statement.aspx

Domenici G., Lucisano P., Biasi V. (2017). *La ricerca empirica in educazione. Elementi introduttivi.* Roma: Armando Editore.

Questo volume, sprovvisto del talloncino a fronte, è da considerarsi copia saggio-campione gratuito fuori commercio. Fuori campo applicazione IVA ed esente da bolla di accompagnamento (art. 22 L.67/1987, art. 2, lett. I D.P.R. 633/1972 e art. 4 n. 6 D.P.R. 627/1978).

Gaetano Domenici · Pietro Lucisano · Valeria Biasi
RICERCA SPERIMENTALE E PROCESSI VALUTATIVI IN EDUCAZIONE
Experimental Research and Evaluation Processes in Education
McGraw-Hill Education (Italy) S.r.l.
9788838655463