

*A Babbo e Mamma, che mi hanno sempre permesso di inseguire i miei sogni
con incondizionato amore e fiducia.
A Pino, Francesca e Samuele che riempiono la mia quotidianità di amore,
condividendo sempre con pazienza le mie esperienze.
Ai miei figli, ai miei nipoti e ai miei studenti, che mi insegnano, nel loro
cammino verso la conoscenza, che il limite della linea dell'orizzonte si sposta
ogni volta che compiamo un passo avanti.
Ai miei preziosi angeli custodi, nonna Dina e Marisa, che mi hanno
insegnato a vivere la vita con il sorriso sul volto e l'amore nel cuore.*

*In copertina "L'uomo dei palloncini" (2002, G. Burchianti)
Artista di grande genialità, estro e creatività,
uomo di grande intelletto e profonda sensibilità e umanità,
zio amorevole e maestro di grandi insegnamenti.
Grazie, zio bello!*

*"Tutti gli uomini per natura tendono al sapere."
(Aristotele)*

Indice

Prefazione	11
Introduzione	15
1 Il nuovo scenario di istruzione e formazione nell'economia della conoscenza	19
1.1 Conoscenza e globalizzazione	20
1.2 Economie basate sulla conoscenza e sistemi di istruzione	26
1.3 Conoscenza e formazione	34
1.4 Conoscenza e valutazione	35
2 Origini e sviluppi della ricerca valutativa in campo formativo	43
2.1 Misurazione della conoscenza e valutazione	43
2.1.1 Origini e sviluppi della ricerca valutativa nell'istruzione... un po' di storia	45
2.1.2 International Association for Evaluation of Educational Achievement (IEA)	55
2.1.3 Dagli anni '60 ai giorni nostri: gli albori e le nuove frontiere della IRT	59
3 Il quadro europeo della ricerca valutativa	63
3.1 Europa2020... il post Lisbona 2010: gli obiettivi dell'istruzione in Europa	68
3.2 Gli indicatori internazionali della valutazione	77
3.3 Sistemi di Istruzione e di Valutazione in Europa	82
3.3.1 Il sistema di valutazione in Francia	82
3.3.2 Il sistema di valutazione in Inghilterra	85
3.3.3 Il sistema di valutazione in Germania	89
3.3.4 Il sistema di valutazione in Spagna	92

3.4	Autonomia scolastica e valutazione di sistema in Italia	94
3.4.1	Autonomia scolastica ed autovalutazione d'Istituto	101
3.4.2	La valutazione degli insegnanti	111
4	Approccio olistico alla persona: valutare le competenze	119
4.1	Le Indicazioni per il curricolo: un lungo cammino lungo per il primo ciclo della scuola italiana	119
4.2	Valutare per competenze	126
4.3	Le competenze e il curricolo scolastico	133
4.4	Le scale di misura nella valutazione scolastica	137
4.5	Metodi qualitativi della valutazione	143
4.5.1	L'osservazione	145
4.5.2	La valutazione autentica	153
4.5.3	Motivare alla valutazione: l'autovalutazione	163
4.6	I metodi della valutazione quantitativa	164
4.6.1	Le prove oggettive di valutazione	172
4.6.2	Tipologie di prove strutturate	178
4.6.3	La costruzione delle prove di verifica strutturate.	188
4.7	La valutazione scolastica degli apprendimenti: tradizione e innovazione nello scenario italiano	193
5	Classical Test Theory (CTT) vs Item Response Theory (IRT)	205
5.1	Classical Test Theory (CTT): principi e assunti di base	212
5.1.1	Analisi delle risposte ad un questionario	217
5.1.2	Analisi dell'affidabilità di un test	222
5.1.3	Analisi della validità di un test	230
5.1.4	La misurazione dei risultati dell'apprendimento e la correzione per guessing	231
5.1.5	Analisi di un test secondo la CTT: studio di caso	238
5.2	Item Response Theory (IRT)	247
5.2.1	Assunti del modello	254
5.2.2	I postulati della Item Response Theory	256
5.2.3	Il modello di Rasch e il modello logistico ad un parametro (1PL)	257
5.2.4	Il metodo di massima verosimiglianza (ML) per le stime del modello	264
5.2.5	La funzione di informazione	271

5.2.6	Il modello logistico a due parametri (2PL) . . .	273
5.2.7	Il modello logistico a tre e a quattro parametri .	277
5.2.8	Esempio di caso: le Olimpiadi Nazionali di Statistica 2013)	279
5.3	Classical Test Theory (CTT) e Item Response Theory (IRT): modelli a confronto	288
6	La valutazione nella Didattica Speciale	291
6.1	Gli speciali bisogni educativi	294
6.2	Valutare tutti, valutare ciascuno	301
6.3	la valorizzazione della persona per mezzo del Computer Adaptive Testing (CAT)	307
	Conclusioni	311
	Bibliografia	327

Prefazione

"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be." (Sir William Thomson Lord Kelvin, 1824-1907)

Nella società attuale, che vede la Conoscenza quale protagonista della crescita economica e sociale di un Paese, le esperienze multidisciplinari, la passione per la ricerca e la scoperta, la motivazione alla crescita continua sono aspetti preponderanti ed imprescindibili.

Il presente manuale nasce quale risultato di un cammino di ricerca, portato avanti con determinazione e coraggio, che ha guidato l'autrice a scoprire nuovi campi del sapere, epistemologicamente molto diversi dalla sua formazione universitaria.

L'autrice è una maestra di scuola primaria, esperta nel settore educativo e didattico, che dopo un'esperienza ventennale nel mondo della scuola, decide di conoscere ed approfondire i metodi della ricerca sperimentale e di comprendere le opportunità offerte dalla disciplina statistica nella conoscenza dei fenomeni reali.

Dopo aver vissuto personalmente, come docente, le fasi più caratteristiche nella storia della valutazione scolastica dell'ultimo ventennio; dopo aver vissuto il passaggio dai giudizi descrittivi generali, ai giudizi sintetici e al voto scolastico in decimi; dopo essersi dovuta adeguare ai continui cambiamenti della scuola italiana, alle svariate riforme e pseudo-riforme, alle politiche dei risparmi nel settore di istruzione e formazione... non perde la pazienza, non si demoralizza, non si sente frustrata nella sua professione ma reagisce con curiosità e motivazione alla ricerca di nuovi saperi e alla scoperta di nuovi strumenti, che consentano di osservare e conoscere i processi di insegnamento-apprendimento, non solo dal punto di vista psico-pedagogico, ma

con un approccio statistico al problema della misurazione in campo educativo.

L'interesse per l'analisi di variabili latenti, quali l'apprendimento, l'autostima, la motivazione, la competenza e la sensibilità nei confronti del problema della valutazione del capitale umano in istruzione, la inducono a frequentare un dottorato di ricerca in Economics & Statistics, durante il quale studia ed approfondisce la Item Response Theory in campo educativo.

La consapevolezza del problema della valutazione nella quotidianità della funzione docente e l'esigenza di pervenire a forme oggettive di valutazione degli apprendimenti e delle competenze, hanno maturato un desiderio di condivisione con i docenti del percorso individuale di formazione al fine di trasmettere l'idea della fattibilità di forme di misurazione e valutazione caratterizzate da rigore metodologico e scientifico, che consentano comunque di salvaguardare l'attenzione alla persona e alla sua valorizzazione.

Il presente manuale è il risultato di questo processo di integrazione tra approccio pedagogico e approccio quantitativo alla valutazione scolastica. Il volume presenta un'analisi dettagliata degli aspetti rilevanti del settore dell'istruzione nell'ambito dell'Economia della Conoscenza, evidenziando punti di forza e criticità che devono essere superate. I temi dell'abbandono scolastico, della dispersione, della valorizzazione di tutti e di ciascuno nelle proprie peculiari specificità, sono affrontati con competenza e sensibilità, bilanciando ed integrando gli aspetti scientifici ed umanistici.

L'approccio quantitativo alla valutazione è trattato con rigore scientifico e metodologico, tuttavia l'uso di un linguaggio fluido e chiaro e la capacità di esplicitare e semplificare i contenuti consentendo una facile lettura anche ai meno esperti nella disciplina statistica. Questa impostazione consente, quindi, una fruizione agevole dei contenuti statistici anche ai docenti delle scuole di ogni ordine e grado e agli studenti universitari che, pur frequentando percorsi di studio di tipo umanistico, necessitano di analizzare con strumenti rigorosi fenomeni reali in campo educativo.

L'approccio qualitativo alla valutazione consente di cogliere alcune sfumature nell'analisi dei fenomeni reali che, integrate con la lettura quantitativa delle caratteristiche dei più svariati aspetti dei processi di

insegnamento-apprendimento, consentono una lettura più organica e dettagliata della complessità dei fenomeni educazionali oggetto di indagine.

Nel manuale la valutazione dello studente in istruzione si connota non come semplice attribuzione di valore ad una prestazione, ma come valorizzazione delle specifiche abilità e competenze di ognuno, al fine di favorire nei giovani lo sviluppo di forme di pensiero critico, creativo e divergente, che consentiranno loro di divenire protagonisti attivi nella Società della Conoscenza.

Tonio Di Battista

Professore Ordinario di Statistica
Dipartimento di Scienze Filosofiche, Pedagogiche ed
Economico-Quantitative
Università agli studi "G. d'Annunzio"
Chieti-Pescara

Introduzione

*“Scopo basilare della valutazione è stimolare la crescita e il miglioramento.
Tutte le altre finalità, pur rispettabili, sono solo sfaccettature dello sforzo generale che consiste nel
valutare le condizioni presenti come base per migliorare.
Una valutazione che non porti a un perfezionamento delle pratiche è sterile.”*
(Kempfer H. H., 1955)

In Italia il concetto di misura in campo educativo si trova a vivere oggi, almeno per ciò che concerne ricerche a livello nazionale, un periodo di transizione, che apre la strada all'introduzione di modelli statistici atti a cogliere e misurare tratti latenti, ossia non direttamente osservabili, quali sono appunto le abilità, le conoscenze, le competenze e più in generale l'apprendimento. Tuttavia, la valutazione degli apprendimenti nell'esperienza quotidiana di vita scolastica, ad esclusione di alcuni rari movimenti di sperimentazione, risulta ancora troppo ancorata ad aspetti soggettivi e demandata all'esperienza professionale maturata dal docente nel corso della sua carriera lavorativa.

Prima di procedere alla descrizione dello svolgersi di questo lavoro, ritengo opportuno spiegare le motivazioni che hanno fatto nascere l'interesse per questo specifico campo della ricerca. Il presente lavoro è impostato in modo da integrare aspetti del curriculum pedagogico, economico e statistico, nel rispetto dell'ottica peculiare di implementare attività di ricerca che vedano coinvolte più discipline che concorrano all'acquisizione di forme di conoscenza olistica di un fenomeno osservato, evitando settorialismi e frammentazioni. L'interesse per il capitale umano in istruzione, che accompagna la mia attività lavorativa di docente di scuola primaria da vent'anni, e che si inquadra nel curriculum economico e statistico, mi ha condotto ad approfondire i temi specifici del Capitale Umano, nell'ottica dell'Economia della Conoscenza, e mi ha avvicinato allo studio di metodi statistici che potessero risultare maggiormente utili per la misurazione e lo studio delle relazioni tra variabili latenti, quali appunto l'intelligenza, l'apprendimento, l'autostima, ecc., che concorrono alla formazione del

capitale umano. Cipollone *et al.* [2010] mettono in evidenza quanto segue: *“Una delle alternative per misurare la qualità del capitale umano è costituita dalla rilevazione diretta di capacità e competenze. Mentre la valutazione del titolo di studio non tiene conto delle capacità e delle competenze acquisite con il completamento di un percorso formativo, né del deterioramento delle abilità causato da un mancato esercizio delle stesse, la valutazione diretta delle competenze può fornire evidenza di varie caratteristiche in un certo momento. Anche se i test sulla preparazione degli studenti non riescono a cogliere completamente attitudini e motivazioni, che pure rientrano nell’ampia definizione di capitale umano, i loro risultati possono fornire una robusta evidenza della qualità del capitale umano stesso e arricchire la ricerca sulla relazione causale tra il livello di istruzione e il reddito oppure tra iniziative di prolicy e risultati scolastici”*. In questi anni la mia produzione scientifica si è basata su esperienze di ricerca empirica, condotte in ambito scolastico, applicando il Modello di Equazioni Strutturali, per la misurazione e l’analisi delle relazioni tra variabili, e la Classical Test Theory (CTT) e la Item Response Theory (IRT), per la calibrazione dei test e la classificazione dei livelli di abilità possedute dagli studenti. La letteratura scientifica nel campo di valutazione scolastica si sta indirizzando, nel corso degli ultimi anni, all’ampliamento del tema della valutazione e della certificazione delle competenze, tanto che si è verificata una proliferazione di definizioni di tale costrutto. A fronte di una definizione di competenza non univoca, è condivisa invece l’idea che la competenza sia un costrutto latente inferibile dall’apprendimento, quale insieme dei processi che, provocando una modificazione di uno stato posto in essere, consente all’individuo di far propri una migliore conoscenza ed un migliore adattamento all’ambiente. L’apprendimento è, quindi, conseguente all’esperienza, all’esercizio, all’osservazione e all’adattamento e può essere valutato dalla comparazione di più prestazioni degli individui, antecedenti e successive ad una data sollecitazione esperenziale. L’assunto teorico di base, in relazione alla misurazione degli apprendimenti, ritiene che i comportamenti osservabili siano il risultato dell’utilizzo di capacità cognitive e che per mezzo di essi sia possibile misurare e quantificare l’apprendimento. L’interesse verso la misurazione psicometrica di attitudini, abilità, conoscenze e competenze ha radici storiche molto antiche. Addirittura uno studioso cinese riconduce al III secolo d.C. i

primi tentativi di costruire misure di abilità cognitive al fine di selezionare il personale per la Corte Imperiale della Cina in base alla rapidità nel parlare e nello scrivere e più tardi usando analogie verbali e frasi da completare. Tuttavia, come vedremo, si deve attendere il 1905 per vedere la nascita della prima scala metrica ufficiale dell'intelligenza ad opera di Binet e Simon e, bisognerà arrivare alle soglie del nostro secolo per il perfezionamento di metodi di stima di variabili latenti in campo educativo. All'interno di questa cornice si inserisce il presente contributo, che si pone l'obiettivo di studiare e misurare il costrutto della competenza, nella sua accezione multidimensionale e di individuare uno strumento informatizzato che consenta ai docenti di ogni ordine e grado di strutturare questionari di verifica degli apprendimenti che siano oggettivamente calibrati e standardizzati.

Il lavoro si dipana, quindi, a partire dall'introduzione del problema, attraverso un excursus storico della valutazione in campo formativo. Si sviluppa in un'analisi dettagliata delle criticità che emergono nell'atto valutativo del docente, per giungere alla proposta di una possibile soluzione del problema della calibrazione dei questionari per la misurazione degli apprendimenti, alla luce dei modelli statistici più adeguati. Il manuale si compone in sei capitoli. Il primo capitolo introduce il quadro del nuovo scenario del settore di istruzione e formazione nell'ottica dell'economia della conoscenza. Dopo una breve definizione dei costrutti di intelligenza e di attitudine, si inquadra il nuovo ruolo del settore Istruzione e Formazione nell'era dell'Economia della Conoscenza, prendendo forma a partire da una riflessione sui risultati emersi da indagini in campo nazionale ed internazionale sulle competenze degli studenti italiani. Il secondo capitolo illustra le principali teorie della ricerca in campo valutativo rintracciate in letteratura, attraverso un excursus storico delle origini e dello sviluppo della ricerca valutativa in campo educativo. Al termine del capitolo vengono presentate le nuove prospettive della ricerca scientifica in campo valutativo con riferimento alla Item Response Theory (IRT), quale metodo statistico applicabile alla valutazione scolastica per la misurazione delle performance degli studenti e per la calibrazione dei questionari, utilizzati nella misurazione delle competenze stesse. Il terzo capitolo illustra il quadro europeo dei Sistemi di Istruzione, inquadrando il tema della ricerca valutativa in riferimento agli obietti-

vi che gli stati membri hanno programmato per la realizzazione del piano operativo “Europa 2020”. La riflessione, prendendo le mosse dall’attuale situazione dell’Italia, in riferimento ai livelli di istruzione e formazione, pone in evidenza l’importanza della condivisione di indicatori internazionali della valutazione per favorire la crescita del capitale umano. Nel capitolo viene analizzata l’organizzazione dei Sistemi di Istruzione e di Valutazione di Francia, Inghilterra, Germania, Spagna e Italia per condurre una riflessione sull’importanza di trovare punti di condivisione per la crescita dell’Europa. Il quarto capitolo affronta il tema della valutazione delle competenze quale approccio olistico alla persona ed illustra le metodologie più diffuse per la valutazione scolastica degli apprendimenti tra tradizione e innovazione nello scenario italiano. Il quinto capitolo illustra la Classical Test Theory (CTT), che per decenni è stata protagonista indiscussa nella docimologia per lo studio dei questionari di rilevazione degli apprendimenti e per la valutazione dei livelli di competenza degli studenti. Si introduce, quindi, il confronto con l’Item Response Theory che viene qui esplicitata e approfondita. Viene quindi presentato un contributo empirico: si tratta della somministrazione di una batteria di test per la rilevazione dei livelli di apprendimento ad un gruppo di studenti di età compresa tra i 9 e i 14 anni; l’analisi dei dati, trattata sia con la CTT che con la IRT messe a confronto, evidenzia le criticità emerse nella costruzione dei questionari, avvenuta ad opera dei docenti dell’istituto di appartenenza degli studenti, ponendo particolare rilievo alla diffusa pratica attuata nella scuola italiana della costruzione di strumenti di rilevazione che non fondano su rigorose procedure scientifiche. Nel sesto capitolo è affrontato il tema della valutazione degli alunni diversamente abili e vengono analizzati gli strumenti utilizzati dalle equipe multidisciplinari e dai docenti per il monitoraggio del processo formativo degli studenti con bisogni educativi speciali.

Cristiana Ceccatelli

Docente di Scuola Primaria

Ph.D. in Economics & Statistics

Presidente provinciale UNIPED (Unione Italiana Pedagogisti)

1 Il nuovo scenario di istruzione e formazione nell'economia della conoscenza

*“Desidererei ricordare a tutti, soprattutto ai governanti impegnati a dare un profilo rinnovato agli assetti economici e sociali del mondo, che il primo capitale da salvaguardare e valorizzare è l'uomo, la persona, nella sua integrità:
L'uomo infatti è l'autore, il centro e il fine di tutta la vita economico-sociale”*
(Papa Benedetto XVI, Caritas in Veritate)

Il *capitale umano* è rappresentato dal bagaglio di conoscenze e di capacità produttive, acquisite da un individuo attraverso le esperienze vissute e le competenze maturate in campo di istruzione, di formazione e di attività lavorativa. Tali capacità e conoscenze influenzano non solo la personale realizzazione economica e sociale del singolo individuo, ma hanno un impatto ed una ricaduta sul contesto nel quale l'individuo vive. Lo studio del capitale umano viene solitamente diviso in due macroaree:

- il capitale umano acquisito attraverso l'istruzione;
- il capitale umano acquisito attraverso la formazione sul posto di lavoro.

Tale divisione è necessaria in quanto i problemi economici e le implicazioni coinvolte nei due diversi processi di accumulazione di conoscenze e capacità produttive (skill) da parte degli individui, sono concettualmente diversi e diversa è la loro rilevanza in termini di politiche e welfare.

Questo lavoro si sviluppa all'interno della prima macroarea, in quanto ci occuperemo dell'importanza che assume l'istruzione quale strumento di crescita del capitale umano ed, in particolar modo, accentreremo il discorso sulla valutazione scolastica quale mezzo di valorizzazione delle risorse umane in istruzione.

Le scuole e le università sono le istituzioni preposte alla promozione

del capitale umano e a ciò adempiono dotando i giovani di un adeguato bagaglio culturale , trasmettendo conoscenze e saperi, stimolando la motivazione ad apprendere ed educandone i comportamenti verso obiettivi socialmente condivisi.

Il primo ciclo di istruzione, quello che va dagli anni della scuola dell'infanzia al termine della scuola secondaria di secondo grado, rappresenta il terreno più fertile nel quale seminare se si vogliono ottenere risultati futuri di crescita economica e sociale.

Ciò che l'individuo apprende ed interiorizza nei primi quindici anni di vita in termini di atteggiamenti e comportamenti rispetto all'apprendimento, alla gestione delle relazioni sociali, alla capacità di collaborazione, all'assunzione di responsabilità, va a costituire la sua personalità e la sua predisposizione nei confronti di ogni evento della vita.

1.1 Conoscenza e globalizzazione

Il forte processo di globalizzazione, che ha avuto avvio nel secolo precedente e che caratterizza l'epoca storica, sociale ed economica attuale, ha condotto alla strutturazione e all'organizzazione di quella che può essere definita oggi una società-mondo, caratterizzata da impellenti e specifici bisogni, soprattutto in campo formativo e sociale. Il turbinio dei cambiamenti in campo politico, sociale ed economico, l'impetuoso dilagare delle tecnologie dell'informazione e della comunicazione, la smoderata, ma talvolta infruttuosa ed inadeguata, ricerca del ben-essere dell'individuo, sono infatti solo alcuni dei cogenti e delicati aspetti che devono essere presi in considerazione e criticamente analizzati per cercare soluzioni che consentano al settore istruzione e formazione di operarsi al fine comprendere e gestire il cambiamento stesso.

L'interesse manifestato dalle diverse scienze su aspetti specifici dell'essere umano ha generato, nel tempo, una parcellizzazione del sapere e la promozione di specializzazioni settoriali che hanno determinato, a lungo andare, una frammentazione della conoscenza della persona a discapito di una visione più olistica dell'individuo. Se è vero e indiscutibile che in ogni campo della ricerca scientifica si è assistito al raggiungimento di importanti traguardi che hanno

garantito un'esponenziale trasformazione tecnologica, una più diffusa crescita economica, una migliore condizione di salute, una più ampia diffusione dell'istruzione, è altrettanto vero e indiscutibile che si sono diffuse forme nuove di mal-essere dell'individuo dovute ad una diffusa incapacità di trarre beneficio dal cambiamento, a causa di una difficoltà nella gestione delle trasformazioni in atto, che ha fatto sì che una gran fetta della popolazione mondiale si sia trovata travolta dal cambiamento stesso senza essere in grado di gestirlo.

Si è verificato una sorta di tsunami, di un anomalo moto ondoso generato dal progresso, originato da eventi storici, economici, politici e sociali a livello globale. Questo fenomeno ha comportato la crescita improvvisa di una grande massa di conoscenza in alcune aree del globo che, non potendo essere contenuta e relegata in un luogo fisico, per le sue specifiche caratteristiche di immaterialità e dinamicità, ha rotto i confini dilagando nell'intero pianeta. Purtroppo, però, a causa di forme sempre più diffuse di analfabetismo funzionale nella popolazione mondiale, non sempre il singolo individuo è stato in grado di comprendere, affrontare e gestire il cambiamento e da esso si è lasciato travolgere.

A livello mondiale i sistemi di istruzione e formazione si trovano oggi a compiere profonde riflessioni sul loro ruolo in questo processo di globalizzazione, poiché essi stessi sono stati travolti da questa massa di conoscenza, sia perché la maggioranza della classe docente non era di fatto preparata alla gestione del cambiamento, sia perché le radicate convinzioni metodologiche e didattiche della tradizione scolastica non hanno consentito di analizzare correttamente il fenomeno e di prendere le adeguate decisioni formative in merito allo stesso.

Nel campo di istruzione e formazione diventa, oggi più che mai, sempre più pressante la necessità che *ognuno e ciascuno* sia in grado di comprendere e gestire il cambiamento. Con l'espressione *ognuno e ciascuno* ci riferiamo, al tempo stesso, alla necessità di una formazione che promuova in ogni individuo, e quindi in *tutti*, un pensiero capace di gestire il cambiamento nell'era dell'Economia della Conoscenza, ossia di un pensiero caratterizzato da spirito critico, flessibilità e creatività, e alla convinzione che *ciascuno* abbia il diritto di essere valorizzato nelle sue specificità e singolarità, che lo rendono un individuo unico ed irripetibile. Educatori e formatori hanno, infatti, l'obbligo professionale

e deontologico di valorizzare il soggetto in formazione, rendendolo protagonista attivo della sua crescita e favorendo la formazione, al tempo stesso, dell'uomo e del cittadino.

La trasmissione della conoscenza, per lunghi anni, si è centrata sull'idea di un processo unidirezionale che vede da una parte il docente che trasmette sapere e dall'altra il discente che apprende. Il processo di insegnamento-apprendimento era, quindi, ingabbiato in strategie e metodologie statiche di trasmissione basate sull'ascolto, sull'imitazione e sulla riproduzione mnemonica degli apprendimenti. Molteplici studi in campo pedagogico e psicologico hanno evidenziato come un processo di insegnamento-apprendimento così rigidamente strutturato non dia, però, garanzia di crescita della conoscenza poiché non favorisce la nascita di un pensiero divergente e creativo.

La conoscenza, infatti, non è mai un sapere compiuto, ma è sempre in fieri; è un processo di ancoraggio continuo di nuove scoperte a vecchi saperi. Il sapere, per essere valido e affidabile, deve passare, inoltre, anche attraverso l'esperienza: per imparare qualcosa bisogna sperimentarla, viverla nella sua complessità, decontestualizzarla e scoprirne i significati meno evidenti. Risulta più che mai attuale l'affermazione di un antico proverbio cinese che recita: "Dimmelo e lo dimenticherò. Mostramelo e lo ricorderò. Fammelo fare e sarà parte di me". La diffusione di conoscenza per essere efficace deve passare attraverso un'approccio olistico all'essere in formazione in quanto l'individuo apprende nella sua interezza.

La conoscenza è, quindi, un'entità astratta e complessa che troppo spesso, anche in campo formativo, viene confusa con la pura trasmissione di informazioni. L'informazione, infatti, può essere facilmente trasmessa tra individui, anche se utilizzano a codici comunicativi diversi, e può essere acquisita anche passivamente ed inconsciamente, sia in contesti di apprendimento formale che in contesti informali. Al contrario, l'acquisizione di conoscenza necessita di un processo più complesso che richiede una motivazione attiva del soggetto che apprende ed un coinvolgimento olistico della persona.

Negli anni la ricerca, in campo psicologico, pedagogico e didattico, ha messo in evidenza come non sia sufficiente trasmettere informazioni, saperi e nozioni, per promuovere la conoscenza. Quest'ultima, infatti, non è insita nei contenuti del sapere, come insieme di dati

strutturati e formali, ma è necessario che il soggetto che apprende sia motivato e desideroso di allargare i propri orizzonti del sapere. Se l'apprendimento è attivo, partecipato e consapevole, struttura solide fondamenta sulle quali costruire conoscenze sempre nuove.

Il processo di insegnamento-apprendimento deve esplicitarsi in un clima sociale fertile che assicuri uno scambio ricco e dinamico di conoscenza. L'insegnante che riesce a promuovere un clima sociale fertile, infatti, stimola nei propri alunni il desiderio della scoperta e della cooperazione e, al tempo stesso, consente a se stesso di acquisire sempre nuove conoscenze poiché in un'interazione educativa valida anche colui che insegna è messo in condizione di apprendere. La ricchezza della diversità consente, dunque, al singolo individuo, sia esso bambino o adulto, di interagire con una pluralità di informazioni ed esperienze, ampliando l'orizzonte conoscitivo.

Nel nuovo panorama società-mondo, inoltre, l'apparato tecnologico di cui si dispone, grazie alle tecnologie dell'informazione e della comunicazione, diventa un fattore centrale nella diffusione della conoscenza e nell'interazione con la diversità. Già il sociologo Collins ha elaborato il concetto di *"organizzazioni senza confini"* [Collins, 1992] per rappresentare l'ampliamento degli ambiti di interazione delle organizzazioni moderne, grazie all'interrelazione con l'insieme di tutti gli stakeholder che possono afferire ad un'organizzazione.

La complessità, soprattutto in campo formativo, si pone, quindi, come la caratteristica più evidente di questo nostro secolo, ma essa non deve essere interpretata come una situazione che determina confusione e difficoltà nell'osservare, nel comprendere e nel fare previsioni su fenomeni reali. Non si deve, infatti, incorrere nell'errore di identificare la complessità con la complicazione ma, al contrario, occorre inquadrarla come linea di indagine che introduce approcci innovativi in molteplici settori di ricerca e che si qualifica come più articolata, in quanto tenta di riuscire a cogliere nel loro significato gli elementi apparsi come accidentali; un tipo di approccio che si articoli su più punti di vista in modo da essere più esaustivo. Considerare la complessità come elemento concreto e decifrabile diventa, quindi, una sfida che richiede appunto una riforma della conoscenza, del pensiero e dell'insegnamento.

Il compito di istruzione e formazione, oggi più che nel passato,

è articolato e complesso poiché, come abbiamo anticipato, riteniamo che sia necessario mettere tutti e ciascuno in condizioni di acquisire le competenze strumentali di base, di maturare un pensiero critico, di possedere un buon livello di alfabetizzazione funzionale, di saper utilizzare le nuove tecnologie dell'informazione e della comunicazione e di essere in grado di gestire la complessità e il cambiamento grazie a forme di pensiero creativo e flessibile. Occorre promuovere nel discente l'attitudine a comprendere e inferire per consentirgli di fronteggiare il cambiamento, che ha un ruolo strategico e caratterizzante nell'economia della conoscenza in quanto promuove l'innovazione permanente. Possiamo, quindi, parlare di economia della conoscenza in termini di economia del cambiamento permanente, che richiede maggiori livelli di formazione, competenze specifiche che privilegiano l'adattabilità, la mobilità e la flessibilità [Foray, 2006], nonché forme reticolari di condivisione e cooperazione.

La formazione continua ed il lavoro di équipe risultano essere elementi estremamente vantaggiosi nella produzione di conoscenza poiché consentono di superare i limiti della autoreferenzialità, potendosi avvalere di un patrimonio più cospicuo di competenze. Il flusso e la condivisione di informazioni, necessario alla produzione della conoscenza può avvenire in modo informale e spontaneo e, di conseguenza risultare occasionale, sporadico e solitamente finalizzato ad un utilizzo individuale della conoscenza (richiesta di consigli, scambio di opinioni,...) o attraverso processi formali di cooperazione che consentono di organizzare forme di strutturazione del lavoro e di condivisione dei risultati.

Non è superfluo soffermarsi a puntualizzare e a riflettere su quali differenze intercorrono tra i concetti di *informazione*, *conoscenza* e *sapere*, che troppo spesso, anche in ambito formativo e nel campo della ricerca, vengono utilizzati pressoché come sinonimi.

L'*informazione* è costituita da una serie di dati strutturati ed organizzati secondo un codice specifico ai quali l'individuo può accedere, per interpretarli ed elaborarli, solo se conosce la chiave di decodifica opportuna. Tuttavia nella comunicazione interagiscono tra loro una pluralità di codici che integrano la trasmissione dell'informazione e che possono consentire la decodifica e l'interpretazione anche nel caso in cui gli interlocutori non possiedano lo stesso codice comunicativo.

Si tratta di tutte quelle forme di comunicazione non verbale, iconografica e cinestesica. Grazie all'innovazione tecnologica e all'avvento delle reti globali si è assistito ad un'amplia diffusione dell'informazione, che può essere oggi fruita con facilità ed immediatezza. Questa diffusione ha comportato, però, in una larga parte della popolazione, forme di analfabetismo funzionale legate alla mancanza di competenza nell'utilizzo corretto, consapevole e responsabile dell'informazione stessa. È necessario, quindi promuovere negli studenti competenze che gli consentano di fare scelte pertinenti, attraverso un pensiero critico e consapevole, in modo da metterli in condizione di saper integrare le diverse informazioni tra loro, ancorandole alle conoscenze pregresse, e di saper gestire l'imprevisto e il cambiamento grazie a forme di pensiero flessibile e creativo. La *conoscenza* è appunto la capacità cognitiva che consente di gestire l'integrazione delle informazioni grazie agli attributi di razionalità, criticità e creatività. Le istituzioni e le agenzie formative hanno la necessità di certificare il possesso di conoscenza al fine di fare della conoscenza stessa un bene "investibile" in altre forme di istruzione e formazione o nel modo del lavoro.

Numerose ricerche hanno evidenziato come il livello di istruzione rivesta un peso determinante nello spiegare i processi di crescita economica di un Paese, sia perché l'accrescimento di capitale umano alimenta l'efficienza produttiva, sia perché favorisce la formazione di capitale sociale, che rappresenta una risorsa per la creazione di benessere, in quanto favorisce reti di relazioni internazionali che risultano fondamentali per lo sviluppo di una nazione. Diversi studi hanno, inoltre, evidenziato come l'istruzione influenzi positivamente il salario e aumenti la probabilità di trovare un'occupazione [Card, 2003; Checchi, 2003]. I sistemi di Istruzione e Formazione si trovano, quindi, nell'ottica dell'Economia della Conoscenza a dover strutturare nuove forme di valutazione delle abilità, delle conoscenze e delle competenze maturate degli studenti nel corso dei loro studi, poiché la competitività economica di ogni paese risulta essere oggi una competitività soprattutto di tipo intellettuale. Quando una conoscenza può essere esplicitata e dichiarata per mezzo di una certificazione formale possiamo esprimerla in termini di *sapere*. La certificazione di conoscenza deve essere in grado di dare un'idea il più precisa e dettagliata possibile delle abilità e delle competenze possedute da un

individuo, anche se è necessario ricordare che, come afferma Polanyi [1979] *sappiamo più di quanto non possiamo dire* e che, di conseguenza, individui che possiedono una medesima certificazione non necessariamente possiedono la stessa quantità e qualità di conoscenza, che per ciascuno è ricca di aspetti taciti che la certificazione da sola non è grado di documentare. La conoscenza tacita, non articolata né codificata, risiede nelle persone e la sua trasferibilità è favorita dalla condivisione con un numero sufficientemente elevato di individui.

Affrontando il discorso sulla conoscenza e sulla sua certificazione, in campo formativo, dobbiamo porci il problema della misurazione della conoscenza stessa. A tale scopo può risultare utile evidenziare alcune caratteristiche salienti della conoscenza, che sono da individuare nell'*eterogeneità*, nella *non completa osservabilità* e nell'*intangibilità*. Queste caratteristiche fanno comprendere che, seppure sia possibile utilizzare forme di certificazione della conoscenza, fortemente necessarie in un'economia globale, esse non possono ritenersi mai completamente esaustive nella misurazione della conoscenza stessa se, oltre agli aspetti direttamente osservabili e misurabili, non si prendono in considerazione le numerose variabili latenti che influenzano la conoscenza stessa. La conoscenza non è infatti un'entità materiale, non è un bene tangibile, ma è un bene prodotto dall'uomo in tutte le sue attività, non solo nel campo dell'istruzione formale ma anche in situazioni contestualizzate quali quelle di *learning by doing* e *learning by using*.

Il focus in questo lavoro di ricerca si incentra, nello specifico, sul problema della misurazione della conoscenza, che si ritiene possa essere affrontato per mezzo dell'utilizzo di specifiche metodologie statistiche atte alla misurazione di variabili latenti e allo studio delle relazioni tra le diverse variabili che interagiscono tra loro nei processi di apprendimento.

1.2 Economie basate sulla conoscenza e sistemi di istruzione

L'affermarsi di economie basate sulla conoscenza è dovuto non solo all'avvento delle tecnologie dell'informazione e della comunicazione, ma anche ad un iniziale consolidarsi di politiche pubbliche destinate alla crescita del capitale intellettuale (istruzione, formazione continua,

ecc.) che hanno determinato nel tempo la diminuzione dei costi di codificazione, di trasmissione e di acquisizione delle conoscenze. L'aumento degli investimenti nella conoscenza è un elemento considerato cruciale per l'avvenire e la crescita dell'Europa. In questo contesto l'istruzione rappresenta una delle chiavi per garantire la crescita di una nazione¹.

Le teorie economiche tradizionali ritengono che le differenze di crescita dei diversi paesi siano dovute all'incremento degli input, del capitale e del lavoro, di contro le teorie economiche più moderne enfatizzano l'importanza del Capitale Umano come fattore guida per la crescita a lungo termine di una nazione. La nuova visione economica, proposta dall'Economia della Conoscenza, è basata infatti sulla promozione e sulla diffusione della formazione lungo tutto l'arco della vita e sulla diffusione ed implementazione delle tecnologie dell'informazione e della comunicazione [Becker, G. S., 1964].

Come ha evidenziato il Rapporto sullo sviluppo umano dell'UNDP (Programma delle Nazioni Unite per lo Sviluppo) del 2010, però, non tutti gli abitanti del nostro pianeta appartengono al *villaggio globale* in quanto ancora una larga percentuale della popolazione rimane in condizioni di sottosviluppo e, anche laddove le infrastrutture tecnologiche lo consentono, si rileva tuttavia un'elevata percentuale di cittadini che non possiedono neanche le competenze strumentali di base per il loro utilizzo². In relazione a ciò, Foray [2006] mette in evidenza che anche in Europa, nonostante la qualità dei programmi di istruzione e formazione, alcuni strati della popolazione rimangono sostanzialmente esclusi dalle competenze legate all'economia della conoscenza.

Maggiori preoccupazioni nascono, per quel che riguarda il nostro Paese, se ci proiettiamo al futuro poiché dai dati delle ultime indagini Ocse-Pisa, che riguardano i quindicenni scolarizzati, emerge un basso livello di istruzione degli studenti italiani, frequentanti la scuola secon-

¹Comunicazione della Commissione al Consiglio Europeo: "Un'Europa moderna e favorevole all'innovazione"

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2006:0589:FIN:IT:HTML>

²Rapporto sullo sviluppo umano del 2010

http://hdr.undp.org/en/media/HDR10%20IT%20summary_without%20table.pdf

daria di primo e secondo grado, rispetto alla media dei paesi OCSE. In Italia ben un adolescente su cinque non va a scuola e non lavora. PISA 2009 ha organizzato i livelli di prestazione degli studenti in sette livelli. Se si considera la distribuzione degli studenti italiani in questi livelli di competenza, si rileva come circa il 21% di essi si collochi al di sotto del livello 2, considerato il livello al di sotto del quale gli studenti presentano forti rischi di insuccesso, con alto rischio di abbandono e di non realizzazione del proprio potenziale di apprendimento³. Un altro dato estremamente preoccupante è che in Italia ben un adolescente su cinque non va a scuola e non lavora, tanto che appena il 22,2 % dei giovani di età compresa tra i 15 e i 19 anni hanno una scolarizzazione secondaria. Un quadro allarmante se si considera l'alta correlazione tra il possesso di istruzione e i titoli di studio e l'opportunità di trovare un lavoro più qualificato. Ai primi posti per livello di istruzione, Corea (il 97% dei giovani ha un diploma di scuola secondaria), Norvegia (il 95%), Giappone e Slovacchia (il 94%). Agli ultimi posti, Italia, Turchia, Messico, Portogallo, Usa e Gran Bretagna⁴. Il basso livello di istruzione dei giovani adulti è una delle principali criticità evidenziata in tutte le indagini internazionali. Nel nostro Paese la percentuale dei giovani sprovvisti di un titolo di studio secondario è molto alta. Si rende necessario, pertanto, predisporre azioni organiche in grado di favorire, in coerenza con gli obiettivi europei, il rientro in formazione del maggior numero possibile di giovani in modo da far conseguire loro un titolo di studio secondario di secondo grado, nonché competenze di base funzionali al pieno inserimento nel mondo del lavoro. Questi dati fanno cogliere una stretta correlazione tra istruzione e crescita economica, in quanto proprio la Corea e il Giappone hanno avuto un'esponentiale crescita economica dovuta al grande investimento in Capitale Umano e alla scelta di basare l'organizzazione economica su un'economia fondata sulla conoscenza. Ma la formazione scolastica da sola non basta per rispondere efficacemente alle sfide lanciate dal mondo del lavoro: nel vecchio continente si spende ancora poco per i programmi di aggiornamento professionale. In Usa, Finlandia, Svezia

³Per approfondimenti si guardi

http://www.cidi.it/documenti/PISA2009_cidi_01209.pdf

⁴Dati rilevati dall'indagine del 2010 condotta dal SISTAN - Ministero della Pubblica Istruzione

TITOLI DI STUDIO	Totale (1.074.000)	
	15-24 anni	25-34 anni
Licenza elementare	9.000	22.000
Licenza media	162.000	186.000
Diploma 2-3 anni	39.000	39.000
Diploma 4-5 anni	217.000	235.000
Laurea breve, laurea, dottorato	22.000	143.000
Totale	449.000	625.000

Figura 1.1: Giovani in cerca di occupazione classi di età e titolo di studio (Dati ISTAT2009)

e Svizzera più del 40% dei lavoratori ogni anno partecipa a programmi di aggiornamento, mentre la percentuale scende al 10% nel caso di Italia, Spagna, Grecia, Portogallo e Ungheria. L'Italia, però, si mantiene in testa alla classifica per quanto concerne le ore di insegnamento agli studenti di età compresa tra i 7 e i 14 anni, circa 8.000 ore, contro una media globale di 6.852 ore. Resta basso, però, l'investimento annuo sul Pil dell'istruzione: meno del 5% in Italia, contro il 6,1% degli Usa e il 7,98% della Danimarca⁵.

Mentre nel mondo si registra una tendenza al miglioramento del tasso di scolarizzazione (il 75% delle persone nate dopo gli anni settanta ha terminato la scuola secondaria, mentre negli anni quaranta la percentuale si fermava ad appena il 50%) in Italia assume ancora dati rilevanti il problema della dispersione scolastica. Dai dati trasmessi nel 2007 dal Ministero della Pubblica Istruzione risulta che gli "early school leavers", coloro che si sono congedati un po' troppo presto dalla scuola, in Italia, nel 2006, risultavano ben 890 mila: ragazzi di età compresa fra i 18 e i 24 anni - pari al 20,6% del totale di quella fascia - in possesso della sola licenza media e che non partecipano a nessuna forma di istruzione o formazione, insomma, giovani usciti definitivamente dai circuiti formativi. Nonostante l'incidenza ancora elevata di abbandoni scolastici, pari al 19,2% nel 2009, negli ultimi quattro anni il valore dell'indicatore si è ridotto di quasi tre punti percentuali. Tradotto nel sistema d'istruzione italiano, l'indicatore

⁵<http://www.oecd.org/edu/skills-beyond-school/37393537.pdf>

equivale alla percentuale di popolazione appartenente alla fascia d'età 18-24 anni che, dopo aver conseguito la licenza media ("scuola secondaria di primo grado" nella riforma varata con la legge n. 53 del 2003), non ha terminato un corso di formazione professionale riconosciuto dalla Regione di durata superiore ai due anni e non frequenta corsi scolastici o altre attività formative. Ebbene, i giovani con esperienza di abbandono scolastico precoce sono ancora oltre 800 mila. Su dieci giovani tra i 18 e i 24 anni che hanno abbandonato gli studi, sei sono maschi. Siamo preceduti solo da Spagna e Portogallo. La media UE cala al 14,9%, mentre Finlandia e Danimarca sono già al di sotto del 10% e Francia, Germania, Olanda, Regno Unito e Grecia sono molto vicini l'obiettivo del 10%, fissato dalla Strategia di Lisbona⁶. Enorme anche il numero dei ragazzi promossi con debito alle scuole secondarie di secondo grado che risulta essere quasi un milione⁷. Nonostante gli

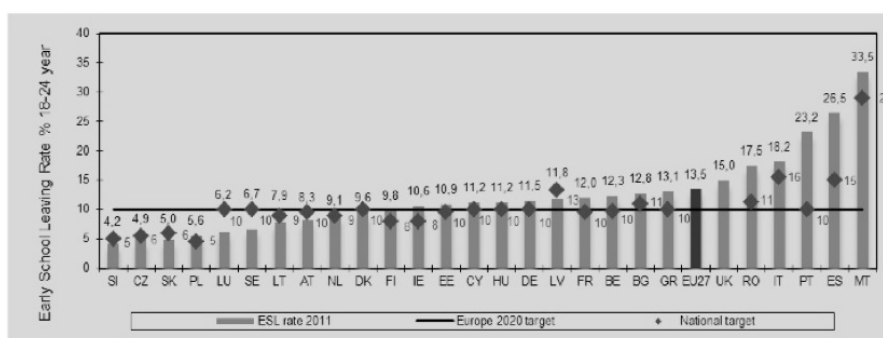


Figura 1.2: ESL rate 2011, Europe 2020 target and national targets

sforzi compiuti non si riesce, inoltre, a contenere il fenomeno della dispersione scolastica. Sono, infatti, quasi mezzo milione gli alunni italiani che, ogni anno, interrompono gli studi o vanno incontro ad una bocciatura. I dati sul preoccupante bilancio della dispersione scolastica in Italia sono stati pubblicati dal Ministero stesso con un dossier dal titolo "*La dispersione scolastica: indicatori di base per l'analisi del fenomeno*". Un fenomeno che al Paese incide per circa 3 miliardi

⁶http://ec.europa.eu/europe2020/pdf/themes/21_early_school_leaving.pdf

⁷<http://www3.istat.it/dati/catalogo/2011052300/rapporto2011.pdf>

di euro l'anno di costi diretti e una cifra non quantificabile di costi indiretti che possono essere molto pesanti. Il Commissario europeo per l'istruzione, la formazione, la cultura e il multilinguismo, J. Figel, avverte: *"Sistemi d'istruzione e di formazione efficienti possono avere un notevole impatto positivo sulla nostra economia e società ma le disuguaglianze nell'istruzione e nella formazione hanno consistenti costi occulti che raramente appaiono nei sistemi di contabilità pubblica. Se dimentichiamo la dimensione sociale dell'istruzione e della formazione, rischiamo di incorrere in seguito in notevoli spese riparatorie"*.

Negli ultimi anni si è dato largo spazio al dibattito sull'importanza di avvicinare maggiormente la scuola al mondo del lavoro poiché la scuola è risultata spesso inadeguata nel formare cittadini-lavoratori del domani. Questo non significa, tuttavia, che la scuola non debba essere "fagocitata" dal mondo delle imprese, perché la funzione primaria a cui deve assolvere resta, sempre e comunque, quella educativa, ovvero quella di assicurare l'apprendimento culturale, la capacità di analisi critica e di trasferibilità ed interconnessione delle conoscenze, la flessibilità e la creatività, indipendentemente dall'inserimento in uno specifico settore lavorativo. Bastano pochi numeri per collocare l'Italia ai primi posti in Europa per la dispersione scolastica. Una nota del Ministero dell'Istruzione, dell'Università e della Ricerca, del giugno 2010, riporta le proiezioni elaborate con i primi dati parziali sulle scuole secondarie di primo e secondo grado dove mette in evidenza che aumentano le bocciature nella scuola italiana: il 5,7% degli studenti della quinta superiore non è stato ammesso agli esami di maturità mentre i cinque in condotta dovrebbero essere quest'anno circa 10 mila. Secondo i dati del Miur, aumenta dal 10,9% al 11,4% il numero complessivo di studenti non ammessi alla classe successiva nelle scuole secondarie di secondo grado, con la percentuale di bocciati maggiore al primo anno (15,3%) e poi decrescente negli anni seguenti. Cresce di circa un punto percentuale il numero di maturandi obbligati a ripetere l'anno, con il maggior aumento relativo al liceo classico (dal 3,5% al 4,2%) e con la punta massima negli istituti tecnici (14,4% nel 2009/2010). Più contenuto, invece, il numero di non ammessi alle classi successive nella scuola secondaria di primo grado: pari al 4,3% nel primo e nel secondo anno. Il 3,5% degli studenti non è stato invece ammesso agli esami di stato a conclusione del primo ciclo di

istruzione. In più, 100 mila tra ragazzini della scuola media e studenti delle superiori dopo essersi iscritti a scuola si sono ritirati, in molti casi, senza comunicare niente a nessuno. La dispersione scolastica è più alta nei primi anni della scuola secondaria di primo e secondo grado. In particolar modo nelle prime classi degli istituti professionali, dove i tassi di bocciatura raggiungono livelli a dir poco stratosferici: 42%. Ma è soprattutto in tutte le regioni meridionali che si annida la dispersione con valori superiori alle medie nazionali anche del 30%. Segno che il fenomeno è legato anche ad aspetti economici, culturali e sociali. Quindi, se ai bocciati e ritirati si sommano gli evasori totali, che superano il mezzo milione di alunni e se, nella scuola secondaria di secondo grado, si aggiungono i promossi con debito, l'insuccesso colpisce più di metà degli studenti italiani.

Dall'ultima indagine Eurostat pubblicata a gennaio del 2010 (Key Data on Education in Europe 2009) risulta che nell'UE la spesa pubblica per l'istruzione è stata pari al 5,1% del PIL; risulta peraltro che in Italia la spesa pubblica per l'istruzione primaria, secondaria e universitaria si colloca al 18° posto nell'elenco dei Paesi dell'Unione Europea. Si tratta dei dati comparabili più recenti basati sugli indicatori messi a punto, con una metodologia comune, da Eurostat assieme a UNESCO e OCSE. La natura complementare delle informazioni qualitative e quantitative è stata potenziata con l'inserimento di due nuove fonti di informazione, in quanto il rapporto è stato integrato dai risultati sugli apprendimenti conseguiti dagli studenti nelle prove PISA (2006) e PIRLS (2006), indagini empiriche sugli apprendimenti effettuate dall'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE) e l'Associazione internazionale per la valutazione del rendimento scolastico (IEA). Dell'Italia ne è emerso un quadro piuttosto infelice, che non fa altro che confermare i dati diffusi dall'OCSE nel rapporto annuale *Uno sguardo sull'educazione 2009*. Secondo l'indagine Eurostat, in media, i paesi dell'UE spendono per l'istruzione il 5,05% del PIL. La percentuale maggiore è in Danimarca (7,98%) e a Cipro (7,02%), mentre la più bassa in Slovacchia (3,79%). L'Italia si colloca al 18° posto con il 4,73%. La maggior parte degli Stati membri hanno dichiarato che la spesa pubblica per l'istruzione rappresentano tra il 4% e il 6% del loro PIL, anche se questa percentuale è scesa al di sotto del 4% del PIL in Slovacchia e Romania. Va inoltre rilevato che la cre-

scita del PIL può mascherare aumenti significativi che sono stati fatti in termini di spesa per l'istruzione negli ultimi dieci anni all'interno di alcuni Stati membri. Per l'istruzione universitaria, addirittura, solo la Bulgaria spende meno dell'Italia. Infatti, i 27 Paesi Ue spendono per l'università in media l'1,1% del PIL, l'Italia lo 0,8%, mentre la Danimarca spende ben il 2,3% del PIL. La spesa media per l'istruzione di uno studente ammonta a 5.970 euro l'anno. In particolare, 4.921 euro si spendono per un bambino della scuola elementare, 6.038 per un allievo delle scuole superiori e 8.591 per uno studente universitario. Per uno studente universitario in Italia lo Stato spende in media 6.984 euro l'anno, circa la metà di quanto si spende in Svezia, Danimarca e Olanda, tre volte meno di quanto si spende a Cipro. Esigui sono anche i fondi privati a disposizione della scuola italiana. Infatti, mentre in media i fondi privati rappresentano il 12,5% delle risorse complessive a disposizione degli istituti di istruzione, in Italia la percentuale è del 7,7%. In media, il 6% della spesa complessiva finisce direttamente nelle tasche degli studenti e delle famiglie sotto forma di aiuti, quali borse di studio e buoni libro. In Italia, inoltre, gli aiuti coprono solo il 4% della spesa pubblica totale, contro il 16% della Danimarca. Non se la cava bene neanche la scuola privata: la media dei Paesi dell'Unione Europea è del 12,5%; la spesa in Italia è soltanto del 7,7%.

Un'altra indagine molto interessante condotta da Eurostat (database tsii040 e tin00073), inoltre, mostra il basso grado di penetrazione di Internet in Italia e come l'uso di Internet sia legato al livello d'istruzione delle persone. L'indagine è particolarmente interessante perché mostra come il cosiddetto "*digital divide*", cioè la frattura di competenze e potenzialità che esiste tra chi è in grado di usare i computer (e Internet) e chi invece no, abbia una stretta correlazione con il livello di istruzione. L'Italia, anche in questo settore, è indietro rispetto agli altri membri dell'UE anche in questa classifica, molto probabilmente a causa del livello medio di istruzione della sua popolazione, che è appunto più basso della media europea. Se si mettono in correlazione il livello di studio, l'investimento nell'istruzione e nella formazione e la diffusione delle nuove tecnologie si comprende quali siano alcune delle grandi problematiche di fondo che condizionano la crescita economica italiana nel mondo.

Oggi, più che mai, ognuno ha bisogno di comprendere la com-

pietà degli eventi ma, ciononostante, i sistemi di insegnamento tradizionali adottati in grande maggioranza in tutti i Paesi continuano a separare, a disgiungere le conoscenze che dovrebbero invece essere interconnesse, e continuano, troppo spesso, a formare menti unidimensionali ed esperti riduzionisti che privilegiano una sola dimensione dei problemi umani, occultando tutte le altre. Ciò non significa che nell'istituzione scuola non si stia cercando di procedere in questi ultimi anni ad un processo di rinnovamento, ma che sicuramente questo stesso procede troppo lentamente rispetto alla repentinà innovativa che caratterizza l'economia della conoscenza. Affinché tutti siano in grado di essere cittadini attivi in questo momento storico e nel futuro, si rende necessaria una riforma del nostro modo di conoscere, una riforma del nostro modo di pensare, una riforma dell'insegnamento: tre riforme strettamente interdipendenti tra loro.

1.3 Conoscenza e formazione

Formazione, capitale umano ed e-learning organization sono capitoli importanti della strada che oggi tutti i capitalismi nazionali stanno intraprendendo, affidandosi non solo alle innovazioni di singole imprese ma anche a forme di intelligenza collettiva [Levy, P., 1996]. L'Italia, al pari degli altri stati, deve prevedere grandi investimenti in questi settori. Il continuo divenire che caratterizza questa era richiede, quindi, che gli individui maturino una *professionalità polivalente e flessibile*. Per rendere trasferibili le proprie conoscenze e competenze bisogna imparare a lavorare in collaborazione con altri, cominciando ad adottare una visione generale e condivisa delle cose, che consenta di inquadrare i problemi da punti di vista differenti, non ripetitivi e standardizzati. Il processo di cambiamento delle professionalità che si sta mettendo in moto ha, quindi, proporzioni enormi e richiede tempi lunghi.

La complessità, come abbiamo detto, implica che, quando ci si trova a dover prendere una decisione, sia necessario compiere un'insieme di azioni mirate ad osservare, analizzare e conoscere tutte le scelte possibili, al fine di selezionare quella che risulta più opportuna. Decidere è insomma tutt'uno con lo sperimentare le soluzioni che vengono di volta in volta prescelte. Del resto, solo quando ci si riferisce ad un'esperienza concreta diventa possibile precisare il contesto in cui si pone un

problema reale. Mondo del lavoro e apprendimento devono, quindi, interfacciarsi attraverso la mediazione di un sistema formativo che garantisca esperienze concrete. Tutte le istituzioni scolastiche percepiscono, infatti, che il mercato non gradisce più i prodotti standard. Ciò presuppone un cambiamento dell'obiettivo formativo da perseguire. Ossia occorre formare capacità generali di ragionamento, di iniziativa, di relazione perché ciascun utente della formazione possa egli stesso divenire capace di personalizzare la definizione del suo specifico problema e di trovare il percorso di sperimentazione e ricerca utile per risolverlo. Non si tratta solo di un metodo di apprendimento più efficace, in condizioni di complessità, si tratta anche di un metodo che distribuisce l'onere dell'investimento formativo nella filiera, rendendo il lavoratore parte attiva, e vigile, di questo processo. D'altra parte, su questo terreno, non va bene nemmeno la fuga verso le competenze generaliste, negando del tutto le specializzazioni. Lo specialista deve sapere tutto del suo settore ma deve sapere anche collegarsi con altri specialisti e formare di volta in volta dei sistemi virtuali efficienti che mettono insieme molte competenze diverse. Dunque la formazione non può insegnare solo a relazionarsi attraverso saperi di tipo generale, ma deve anche aiutare a fare alcune esperienze specialistiche, che siano però molto focalizzate e reversibili. Non si tratta infatti di rinunciare agli specialisti, ma di creare un meccanismo che sia in grado di generare specializzazioni sempre nuove e di relazionare in modo rapido ed efficiente quelle che già esistono. L'obiettivo che si presenta ora per l'università è, in definitiva, quello di identificare le modalità per affrontare concretamente la creazione di strutture permanenti in grado di fornire un'efficace e moderna risposta al potenziamento del ruolo propulsore delle strutture universitarie nel territorio.

1.4 Conoscenza e valutazione

Abbiamo già accennato alla necessità delle istituzioni formative di certificazione delle competenze al termine di un percorso di formazione, al fine di esplicitare formalmente i risultati raggiunti dal singolo individuo e consentire a quest'ultimo di poter acquisire un titolo spendibile per proseguire gli studi o per immettersi nel mondo del lavoro. L'atto valutativo in campo formativo è stato per lungo tempo conside-

rato esclusivamente il momento conclusivo di un percorso; il momento in cui terminato un processo di insegnamento-apprendimento il docente è chiamato a compiere una valutazione dei risultati e ad attribuire un giudizio. La motivazione al fatto che per secoli la valutazione nel settore istruzione e formazione sia stata vissuta come atto conclusivo di un percorso ha radici lontane, in quanto con la nascita stessa dell'istruzione insorge la necessità di attribuire un valore a quanto il discente apprende. La questione non si modifica molto neanche nel momento in cui, agli inizi del novecento, nasce la ricerca in campo valutativo che, come vedremo più nel dettaglio nel prossimo capitolo, ha origini come ricerca docimologica. Solo negli ultimi decenni, grazie all'interessamento della ricerca statistica ad un approccio scientifico più articolato e complesso al problema della misurazione in campo formativo, la valutazione assume un ruolo centrale e diviene il perno attorno al quale costruire l'intero processo di insegnamento-apprendimento. Numerose ricerche mettono, inoltre, in evidenza che il docente che possiede buone competenze nella valutazione formativa favorisce forme di apprendimento più efficace negli studenti. Tuttavia nel capitolo seguente seguirà un dettagliato excursus storico della ricerca in campo valutativo, nel quale sarà esplicitato ed ampliato quanto qui accennato ed anticipato.

Valutare significa attribuire un valore e, affinché tale attribuzione risulti obiettiva e affidabile, è necessario che colui che valuta sia in grado di attribuire tale valore in modo univoco, privo di condizionamenti e interferenze dovute a fattori soggettivi che possono alterarne l'attendibilità. Il momento della valutazione deve essere progettato e strutturato secondo modalità e tempistiche che necessitano di competenze specifiche e professionalità.

La valutazione all'interno delle agenzie formative riveste, quindi, un ruolo fondamentale e deve attenersi a canoni di qualità che devono essere fissati e stabiliti ancor prima di procedere alla realizzazione di un percorso formativo. Prima di programmare un intervento formativo è fondamentale, infatti, aver un quadro ben preciso del contesto nel quale si va ad operare e ciò comporta già una prima valutazione fondata sull'analisi di contesto socio-culturale, sulla valutazione delle risorse, in termini di infrastrutture e capitale umano disponibile, e una profonda conoscenza della letteratura psico-pedagogica che consenta

di individuare le strategie e le metodologie di intervento educativo più idonee. L'apparato valutativo di una istituzione formativa risulta estremamente complesso poiché interessa, al contempo, una valutazione di sistema e una valutazione degli apprendimenti, pertanto tutti coloro che sono coinvolti in questo processo valutativo devono avere le necessarie competenze pedagogiche, didattiche e statistiche necessarie.

In campo educativo la progettazione di percorsi formativi ha un carattere dinamico che richiede flessibilità e capacità di gestire le interrelazioni, il cambiamento e l'imprevisto. La valutazione assume, di conseguenza, un valore centrale nel processo di insegnamento-apprendimento, in quanto deve accompagnare tutto l'intero processo formativo garantendo fattibilità e consentendo di prendere le decisioni più idonee a far sì che la progettazione stessa risulti valida.

Ogni studente, per il fatto di essere una persona unica ed irripetibile, presenta una sua specificità che la valutazione deve valorizzare; proprio in questo è il cardine della valutazione: valutare non solo come attribuzione di un valore, ma soprattutto come valorizzazione delle specificità del singolo individuo. Occorre, infatti, dare valore a tutto il processo di insegnamento-apprendimento per valorizzare l'intera formazione della persona. L'attività valutativa dovrà considerarsi, per tanto, come un'*attività di pensiero produttivo*, poiché deve essere in grado, appunto, di espletare la sua funzione regolativa del processo formativo che consente di gestire il cambiamento e l'imprevisto; come un'*attività di pensiero comparativo* poiché nasce sempre da un confronto tra ciò che si programma e ciò che si realizza, tra aspettative e risultati, tra un livello di competenze iniziale ed uno finale, tra diversi stili di apprendimento; come un'*attività di pensiero critico* perché comporta un'accurata analisi, riflessione, progettazione e continua verifica da parte dell'insegnante che deve contestualizzare l'esperienza formativa e da parte dello studente che deve costruire in prima persona la propria crescita culturale e sociale; e ancora come *attività di pensiero ermeneutico* poiché l'educazione e la formazione non sono avulse dal contesto storico, politico, economico, culturale e sociale. I sistemi formativi, infatti, sono chiamati ad interpretare la realtà e a progettare interventi finalizzati alla formazione di uomini che siano in grado di vivere la loro esperienza di vita con razionalità, criticità, flessibilità e creatività.

Un docente, chiamato a compiere una valutazione, deve tener presente che valutare non significa mai emettere un giudizio sulla persona ma misurare il suo rendimento, la sua performance, i progressi compiuti e le difficoltà che devono essere superate. L'insegnante, infatti, è chiamato a valutare prima di tutto l'efficacia del suo stesso intervento educativo attraverso la verifica costante del processo. Questa verifica si basa sulla misurazione degli apprendimenti, sull'osservazione delle modificazioni dei comportamenti, sull'analisi delle competenze, coniugando processi e prodotti, ma sarebbe deontologicamente scorretto se comportasse, al contempo, un giudicare la persona. Se poi, come tutt'oggi talvolta accade, venisse usata come "arma" per esercitare la propria autorità sugli studenti, denoterebbe, di fatto, un'incapacità dell'educatore di affermazione delle proprie competenze professionali. L'autorità non va confusa con l'autoritarismo: contenere, limitare, dare delle regole, definire dei confini non significa reprimere ma, al contrario, fungere da guida rispettando il desiderio e il bisogno di individualità e di auto-affermazione dell'altro. In questa ottica potremmo affermare che lo studente, per maturare un concetto corretto di libertà e di conquista di individualità, necessita sicuramente di una forma di contenimento, in quanto ha il bisogno di trovare nell'educatore una guida autorevole che insegni a pensare, a fare scelte, ad agire nel rispetto degli altri e di se stessi. Ecco, quindi, che la valutazione diventa uno strumento utile allo studente per indirizzarlo nel suo stesso percorso formativo. Le agenzie formative, attraverso la valutazione, la co-valutazione e l'autovalutazione, promuovono gradualmente i processi di identificazione e di differenziazione, che consentono la maturazione della persona nella sua visione olistica e la rendono un soggetto autonomo e libero. Il docente non può favorire questa crescita se impone la sua autorità, ma deve essere invece responsabile e autorevole, poiché solo in questo modo consente allo studente di separarsi, individuarsi e differenziarsi. Gli stessi conflitti circa i ruoli e le responsabilità possono avere un valore costruttivo e formativo, poiché consentono ai ragazzi di confrontarsi e di diventare consapevoli delle differenze, nonché di assumere posizioni attive di responsabilità. Il docente deve saper essere autorevole, non autoritario, e la valutazione è uno degli strumenti che può consentirgli di esprimere la propria autorevolezza, più che l'autorità in sé. I docenti che fanno leva sulla sola

autorità, senza dare sostegno alle proprie motivazioni, senza guidare l'alunno a comprendere il significato di un giudizio espresso in una prova di verifica, o che tendono ad imporsi, escludono il confronto, facendo sentire gli studenti giudicati. La valutazione, per diventare un momento esperienziale fondante per la formazione della propria personalità, deve essere condivisa e deve guidare gradualmente lo studente ad auto valutarsi, a conoscersi, a comprendere i propri punti di forza e di debolezza, a trovare le strategie per superare le difficoltà e quelle per migliorarsi, in una presa di coscienza del fatto che ognuno di noi è artefice della costruzione del proprio sapere e del proprio essere. Il ruolo dell'educatore, dell'insegnante, del formatore è primariamente quello di mettere ogni discente in condizione di crescere, nel suo processo di formazione, in relazione alle proprie attitudini e, soprattutto, che si coltivi in lui lo stupore, la meraviglia, quell'attimo di sospensione del pensiero che ti lascia rapito di fronte a qualcosa di nuovo e che genera la scintilla, il desiderio di scoprire, di percorrere nuove strade della conoscenza. Il segreto di un apprendimento duraturo e significativo è, quindi, da ricercare nella capacità del docente di rendere il sapere un oggetto di desiderio, di saper motivare l'alunno alla scoperta di nuove conoscenze. La valutazione degli apprendimenti, pertanto, deve essere vissuta dall'alunno come un momento positivo del processo di insegnamento-apprendimento, evitando sia che si generino ansie e preoccupazioni, sia che questo momento, cruciale ed indispensabile, venga vissuto come una situazione nella quale si ritiene di dover dimostrare la propria efficienza solo rispetto ad un apprendimento specifico. È necessario, in quest'ottica, che il docente metta l'alunno in condizione di saper auto-valutare il proprio processo di apprendimento. La valutazione non si deve, quindi, limitare ad una verifica di prestazione ma, al contrario, deve essere il momento in cui l'alunno stesso prende coscienza del proprio livello di abilità, conoscenze e competenze maturate. Chi apprende, in questo modo, è reso protagonista del suo stesso processo di apprendimento, poiché è messo in condizione di comprendere il livello di padronanza raggiunto, di conoscere i propri punti di debolezza e, soprattutto, i propri punti di forza e le potenzialità su cui poter far leva per superare le difficoltà incontrate per gestire, affrontare e superare i propri limiti. La valutazione si connota, di conseguenza, non più come semplice

attribuzione di valore ma come *valorizzazione* dell'individuo nell'ottica di una visione olistica della persona.

La valutazione è, quindi, un'operazione complessa che mette in gioco diverse variabili: la qualità e la quantità delle conoscenze acquisite, la capacità di trasferire le competenze acquisite ad altri contesti per la soluzione di problemi diversi, la corrispondenza tra le aspettative degli insegnanti e le risposte degli allievi, gli aspetti emotivi, affettivi e relazionali, i fattori ambientali, le condizioni socio-culturali e molto altro ancora. Ne consegue che la qualità del giudizio di valutazione spesso non risulta, sia in contesti diversi, sia in uno stesso contesto ma con docenti diversi, uniforme e costante. Questa difformità nelle procedure valutative non è da inquadrare solo in una problematica di ordine didattico, legata al fatto che determina valutazioni molto eterogenee tra loro, e in una non conformità della certificazione di competenze, ma esplicita un problema ancora più complesso, che è quello di una disomogeneità nell'interpretazione complessiva dell'azione formativa. La valutazione non può limitarsi alla sola funzione certificativa poiché è necessario promuovere la sua funzione regolativa, la sola in grado di consentire, sulla base delle informazioni raccolte in itinere nel processo di insegnamento-apprendimento, un continuo adeguamento delle proposte di formazione alle reali esigenze degli alunni ed ai traguardi programmati. In tal senso deve essere privilegiata una valutazione formativa che prenda in considerazione non solo i risultati raggiunti dal singolo alunno ma l'intero percorso formativo dello stesso, i passi realizzati, la motivazione, l'impegno, le aree di sviluppo prossimale. L'importanza di una valutazione in itinere sta nel consentire di procedere per passi, correggendo o modificando strumenti e modalità di apprendimento sia da parte dei docenti che da parte degli alunni e, al tempo stesso nell'avere una valenza diagnostica in quanto consente di orientare docenti ed alunni verso gli sviluppi successivi del processo di apprendimento. Una valutazione di tipo *ideografico*, che consenta un confronto diacronico fra la situazione iniziale e la situazione finale di un percorso di apprendimento, mirata al recupero dell'originalità del singolo e delle singole situazioni e che sia al tempo stesso in grado di dare una certificazione oggettiva e confrontabile delle competenze acquisite, è quella che si ritiene più auspicabile. Questo tipo di valutazione, infatti, consente di effettuare

una riflessione profonda dei risultati conseguiti progressivamente dal singolo studente, di rilevare carenze da superare ma soprattutto le potenzialità sulle quali ipotizzare gli sviluppi successivi del processo di insegnamento- apprendimento. La condivisione, con le famiglie e con gli alunni stessi, delle metodologie di valutazione utilizzate dal team docente e delle osservazioni rilevate consente di promuovere nell'alunno una maggior motivazione ad apprendere e una miglior capacità di autovalutazione e, al tempo stesso, promuove un senso di co-protagonismo e co-partecipazione di insegnanti-genitori-alunni al processo di insegnamento-apprendimento. Gli studenti, dunque, se si sentiranno coinvolti in prima persona, vivranno esperienze di *co-valutazione* valide, che consentiranno loro di giungere ad una capacità di autovalutazione, poiché gradualmente comprenderanno il proprio stile di apprendimento e le proprie capacità e scopriranno i propri limiti. In tal modo saranno maggiormente motivati a migliorarsi e affronteranno meglio i possibili insuccessi, in quanto loro stessi hanno partecipato a fissare i parametri della valutazione. Con il coinvolgimento e la partecipazione degli alunni si assiste all'emersione delle diversità e alla loro valorizzazione, poiché ognuno diventa consapevole del proprio ruolo e del ruolo degli altri. È utile, inoltre, che la valutazione sia condivisa e discussa con i genitori in tutte le sue fasi, perché in questo modo si costruirà un protocollo di lavoro condiviso scuola-famiglia. Infatti, attraverso la condivisione, sia l'alunno, sia la famiglia sono messi a conoscenza del percorso formativo, dello step in cui si colloca il proprio figlio nei diversi ambiti di apprendimento e negli aspetti educativi e sociali, dei passi successivi che egli deve compiere e degli interventi mirati di supporto ed aiuto che possono essere necessari per superare difficoltà o per potenziare eccellenze.

La valutazione, in quest'ottica, diventa quindi un processo di ricerca sociale applicata che ha il suo focus nel confronto con le complessità generate da un processo di interazioni e relazioni che si influenzano reciprocamente. In questo senso diventa fonte di apprendimento sia per il docente che per lo studente, poiché determina un continuo rinvio dalla dimensione progettuale a quella più specificatamente valutativa e viceversa, che, come abbiamo visto, devono essere fortemente condivise tra tutti i protagonisti del processo formativo, affinché l'apprendimento sia significativo. Lo sfondo concettuale nel quale si

muove la valutazione oggi è, infatti, quello del *paradigma della complessità*, nel quale, come afferma Morin [2005], la valutazione stessa diventa un costrutto complesso in interazione con il fenomeno valutato, finalizzato a rinforzare le dinamiche più che a stabilire un controllo. Per questo motivo appare necessario, in primis, che l'educatore maturi competenze specifiche in campo valutativo, acquisendo tutte le possibili metodologie e strategie che la ricerca pedagogica, didattica e statistica ha, nel correre del tempo, sperimentato; così da riuscire, in un proficuo lavoro di team, a trovare le interconnessioni tra le diverse teorie in modo da strutturare le modalità più idonee a fare della valutazione una concreta esperienza di interpretazione e valorizzazione della complessità. Un formatore non deve mai perdere di vista il principio della complessità, che è necessario porre alla base di qualsiasi valido discorso in campo educativo, poiché complesso è l'essere umano, complessa è la società in cui esplica se stesso, complessa è la sua formazione e complesso il compito di valutarla.

2 Origini e sviluppi della ricerca valutativa in campo formativo

“La docimologia è ricerca sulla valutazione. Sulla quale noi insegnanti agiamo spesso in maniera incauta. Ad esempio effettuando medie aritmetiche come se esistesse un’unità di misura costante. Utilizzando voci valutative come cinque meno, sei e mezzo, tra il sei ed il sette elaborate come se si fosse in possesso di chissà quali strumenti capaci di compiere discriminazioni molto sottili e precise. In sostanza pensando che i voti siano espressi in base ad un’unità di misura: in realtà essi esprimono piuttosto un giudizio qualitativo.(Peppe E. su “Orizzonte scuola” del 1 maggio 2005)”

2.1 Misurazione della conoscenza e valutazione

Il problema più complesso in campo formativo è quello della misurazione della conoscenza, ossia di assegnare una misura ad un fenomeno reale che non è di per sé direttamente misurabile. È quello, quindi, di attribuire un valore ad una grandezza non direttamente osservabile, a quella che in statistica viene definita come *variabile latente*. Un esempio tipico è quello del concetto di quoziente intellettivo che, non essendo appunto misurabile direttamente, viene misurato attraverso la somministrazione di una serie di test. Nel contesto educativo esistono molteplici fenomeni che non possono essere misurati direttamente quali le abilità, le conoscenze, le attitudini, il livello di autostima, la capacità di relazione, le competenze e molti altri ancora che concorrono a definire ed a specificare quella splendida e complessa creatura che è l’essere umano.

Il problema della misurazione nelle scienze educazionali, psicologiche e sociali è abbastanza complesso e variegato proprio perché queste scienze si trovano ad indagare fenomeni che possono essere descritti da una serie di variabili molte delle quali di tipo latente. Nella storia della misurazione, tuttavia, qualunque strumento costruito convenzionalmente per garantire una misura univoca e condivisibile, non ha avuto una nascita semplice, né tanto meno una semplice divulgazione.

In campo educativo viviamo, quindi, oggi un problema che in passato hanno affrontato e superato molte scienze.

La necessità di scegliere opportune unità di misura e i relativi campioni risale a tempi antichissimi, in quanto era già avvertita dai primi studiosi della geometria e dell'astronomia, ma l'esigenza di disporre di adeguate unità di misura universali diventò cogente con l'avvento del metodo sperimentale, introdotto da Galilei intorno al 1600. La diffusione dei dati sperimentali raccolti da ricercatori di diversa nazionalità evidenziava la necessità, per potersi intendere e confrontare, di sostituire l'enorme numero di unità di misura locali in uso a quel tempo, con unità di misura univoche e condivise. Pensiamo, ad esempio, alla misura del tempo; de Lalande [1790] scrive : Gli italiani contano le ventiquattro ore di seguito da una sera all'altra. La ventiquattresima ora che si chiama l'Ave Maria suona una mezz'ora o tre quarti d'ora dopo il calar del sole e cioè a notte calata. Secondo questo sistema di misurazione del tempo, Mezzogiorno a Venezia il 1° gennaio cadeva dopo le diciannove poiché, essendo calato il sole il 31 dicembre alle sedici, di conseguenza il suono dell'Ave Maria corrispondeva alle sedici e quarantacinque e la Mezzanotte alle sette e quarantacinque. In questa maniera il Mezzogiorno non cadeva sempre alla stessa ora, poiché la sua misurazione era basata su un metodo che possiamo definire naturale, in quanto fino a quando non si sono diffusero le macchine per misurare il tempo non si poteva partire che dal punto più sensibile agli occhi, quello della fine del giorno. L'introduzione delle varie grandezze fisiche avvenne in tempi e con modalità differenti. Dalla geometria, una delle scienze più antiche, già nota al tempo degli antichi Egiziani e dei Babilonesi, emerse il concetto di lunghezza al quale l'astronomia associò quello di tempo (più esattamente intervallo di tempo o durata). Con la definizione di lunghezza e di tempo è stato possibile costruire quell'importante branca della fisica che si chiama cinematica (dal greco *kínema* = movimento). Quando in seguito si decise di indagare sulle cause legate al movimento dei corpi si presentò l'esigenza dell'impiego di una terza grandezza fisica, la massa. All'inizio del 1800, lo studio dei fenomeni termodinamici impose l'introduzione di una quarta grandezza fondamentale, la temperatura, la quale, per la verità, più che una grandezza è un indice di stato fisico. Successivamente lo studio dei

fenomeni elettrici rese necessaria l'adozione di una quinta grandezza fondamentale, che venne individuata nella intensità di corrente elettrica, alla quale si aggiunse l'intensità luminosa, quando prese avvio lo studio dei fenomeni di ottica. Il quadro si andò infine completando nel 1971 con l'adozione di una settima grandezza fondamentale che fu riconosciuta nella quantità di sostanza indispensabile per descrivere i processi chimici. Da queste sette grandezze fondamentali fu possibile ricavare tutte le altre, necessarie per la descrizione dei diversi fenomeni naturali. E chi sa che in un futuro, riferendosi agli attuali sistemi di valutazione, i posteri racconteranno delle valutazioni scolastiche odierne come di stravaganti procedure che si sono risolte con l'avvento dell'ottava grandezza fondamentale.

2.1.1 Origini e sviluppi della ricerca valutativa nell'istruzione... un po' di storia

In campo educativo il problema della misurazione è quello di disporre di un metodo standardizzato, riconosciuto e condiviso che consenta anche di effettuare confronti a livello territoriale e/o temporale. Illustriamo in questo paragrafo un breve excursus storico degli sviluppi della ricerca nel campo della valutazione degli apprendimenti, precisando che questa trattazione non vuole pretendere di essere esaustiva, considerata la vastità dell'argomento, ma si pone solo l'obiettivo di fornire uno sguardo d'insieme sulle principali teorie che si sono sviluppate nel tempo in questo settore.

Dal periodo ellenistico (323 - 31 a.C.) inizia nell'antica Grecia la diffusione sistematica della scuola pubblica, promossa dai cambiamenti politici determinati dall'espansione del regno macedone con Alessandro Magno e dal fiorire dei regni ellenistici. Questa diffusione, che accompagna di pari passo il crescere della cultura, portò al sorgere di scuole frequentate da giovani nobili, destinati a formare la classe dirigente. La scuola si caratterizzava per il suo carattere elitario e selettivo che, attraverso l'uso di test, riconosciuti come strumenti istituzionalizzati del processo educativo, valutava le capacità fisiche ed intellettuali degli individui. I test somministrati non avevano validità scientifica ed erano strutturati esclusivamente dai "maestri" per i loro discenti. Oggi nelle scuole di ogni ordine e grado i metodi di

valutazione utilizzati quotidianamente vengono progettati e strutturati dai docenti e non presentano carattere di scientificità. Ma in fin dei conti sono passati solo poco più di duemila anni!

Il cammino della misurazione degli apprendimenti, finalizzato alla valutazione degli alunni, scientificamente strutturato e testato è, però relativamente breve. Fino alla seconda metà dell'ottocento, infatti, l'educazione faceva riferimento, quasi esclusivamente, alla morale. La pedagogia, d'altro canto, non era ancora una scienza a sé, poiché risultava strettamente interconnessa con la filosofia. Le riflessioni sull'educazione erano, di conseguenza, affrontata sul piano puramente teorico e chi si preoccupava di analizzare concretamente gli aspetti delle pratiche educative era accusato con accezione negativa di eccessivo empirismo. L'introduzione del concetto di misura, conseguente all'avvio della ricerca sperimentale sull'educazione, ha rappresentato, quindi, un'autentica svolta culturale [Vertecchi, 1993].

A Londra Bain [1859] in *The emotions and the will* comincia ad ipotizzare la possibilità di costruire test attitudinali quali strumenti scientifici per la misurazione di variabili non direttamente osservabili. Tuttavia la ricerca docimologica, che si occupa appunto dello studio della misurazione e della valutazione in ambito scolastico mediante l'ausilio della psicotecnica e della statistica, affonda le sue radici alla fine del 1800, quando l'inglese Edgeworth [1888] scrisse *The statistics of examinations*, ritenuto il primo studio scientifico sulla distribuzione dei voti e sulla loro variabilità. Questi studi rimasero, tuttavia, per lungo tempo lasciati all'approfondimento teorico poiché mancò un'applicazione pratica delle teorie stesse. Alfred Binet, psicologo, conduce studi sulla misurazione dell'intelligenza e realizza i primi test di intelligenza. Nel 1904 il ministero della Pubblica Istruzione francese incarica un gruppo di lavoro, del quale faceva parte Binet, di studiare nuovi metodi per l'educazione dei bambini con uno scarso sviluppo intellettuale delle scuole di Parigi. Il primo problema che coglie Binet è quello di individuare uno strumento valido per la classificazione degli studenti in relazione al livello di abilità mentale posseduto. Nasce così la prima scala metrica di misurazione del livello intellettuale. La scala, composta da una serie di 30 test, ha lo scopo di offrire una valutazione di alcuni aspetti dell'intelligenza, come la capacità di comprensione, di ragionamento logico e di giudizio. A Binet va il merito di aver colto

l'importanza di una valutazione multidimensionale dell'intelligenza, di aver compreso la complessità dei processi cognitivi, di aver studiato la correlazione tra livello socioculturale e successo scolastico e di aver individuato uno strumento di misurazione che valutasse l'eterogeneità che caratterizza l'intelligenza stessa.

È sempre nel 1904 che lo psicologo statunitense Edward Lee Thorndike pubblica *An introduction to the Theory of Mental and Social Measurements* nel quale troviamo la sua celebre affermazione "Tutto ciò che esiste, esiste in una determinata quantità e può essere misurato" [Sears, 2007]. Nella sua opera Thorndike affronta il problema della misurazione mettendo in evidenza l'importanza della *validità*, ossia il grado di precisione con cui un test riesce a misurare ciò che si propone di misurare, e ponendo l'attenzione sulla considerazione dell'errore nella misurazione. Thorndike ha il merito di dare un grande contributo alla ricerca valutativa in campo formativo in quanto getta le basi della *valutazione formativa*, ossia di una valutazione che miri ad individuare attitudini ed abilità di uno studente al fine della costruzione del curriculum di studi; di una valutazione, quindi, che assume un ruolo centrale nel processo di insegnamento-apprendimento in quanto non vista come momento finale ma come valutazione ex-ante ai fini della progettazione formativa, educativa e didattica. Sempre Thorndike, nel 1907, conduce la prima interessante ricerca sull'abbandono scolastico, gettando le basi di un vasto filone di ricerca nel settore istruzione e formazione che si concentra maggiormente su indagini mirate all'analisi della distribuzione territoriale dei livelli di istruzione, dell'offerta formativa e delle organizzazioni scolastiche, per quest'ultime soprattutto in relazione allo studio delle risorse in termine di strutture, strumenti e personale docente.

La prima ricerca valutativa applicata si svolge negli Stati Uniti ad opera di Starch ed Elliot, che nel 1910 condussero un'indagine in sull'esattezza dell'attribuzioni di voti nella scuola secondaria [Starch and Elliott, 1912]. L'indagine è stata condotta su 142 insegnanti di inglese di scuola secondaria, ognuno dei quali disponeva di una scala numerica da 0 a 100 per correggere due compiti di inglese. Dall'analisi dei risultati ottenuti dagli studenti emerge un notevole scarto nelle valutazioni ottenute da correttori diversi che hanno esaminato elaborati svolti dagli estesi studenti: nella prima prova compiti svolti in modo

sostanzialmente uguale ottennero un punteggio diverso, oscillante da 64 a 98; nella seconda prova da 50 a 98. I risultati evidenziano, quindi, la scarsa oggettività nella valutazione dovuta ad un uso non adeguato degli strumenti da parte degli insegnanti correttori. Negli Stati Uniti si assiste dal primo decennio del Novecento ad un'ampia diffusione dell'uso dei test nella scuola per la valutazione dei livelli di apprendimento e delle attitudini degli studenti, tanto che dal 1911 la *National Education Association* approva l'uso dei test di valutazione per la selezione degli studenti nell'ammissione scolastica e per gli esami conclusivi.

Daniel Starch continua il suo lavoro nel settore della ricerca valutativa applicata all'istruzione e nel 1916 pubblica *Educational Measurement* nel quale espone la sua teoria in base alla quale tutte le caratteristiche dell'individuo, siano esse di ordine fisico o psicologico, si distribuiscono secondo una *normale gaussiana* e che tale principio possa essere applicato anche in tema di valutazione degli apprendimenti. Sul versante europeo comincia a nascere anche l'interesse alla ricerca valutativa con uno scopo di studio comparativo dei sistemi di istruzione; così nel 1917 Marc Antoine Julien a Parigi conduce una ricerca internazionale sull'analisi comparata dei diversi sistemi di istruzione al fine di individuarne punti di forza e debolezza. Nel 1922 McCall, professore della *Teachers College*, scrive *How to measure in Education* e l'anno successivo *How to experiment in Education*, opere considerate da Campbell e Stanley quali migliori opere del suo tempo nel campo della metodologia statistica applicata all'istruzione [Campbell *et al.*, 1963].

Nel 1918, intanto, viene pubblicato "The Seventeenth Yearbook of The National Society for the Study of Education - Part II: The Measurement of Educational Products" ad opera di E. J. Ashbaugh, W. Averill, L. P. Atees, F. W. Ballou, E. Beyner, B. E. Buckingham, S. A. Courtis, M. E. Haggeety, C. H. Judd, Geo. Melcher, W. S. Moneoe, E. A. Nifeneckee ed E. L. Thoendikb, che rappresenta una preziosa testimonianza dello stato dell'arte della ricerca valutativa agli inizi del Novecento a livello mondiale e che per lungo tempo ha rappresentato lo spunto per l'avvio di nuove ricerche in campo valutativo, nonché per la produzione negli Stati Uniti di un gran numero di batterie di test standardizzati per la valutazione quantitativa degli apprendimenti

[Thorndike, 1918].

Le ricerche docimologiche si diffusero contemporaneamente in tutta Europa grazie agli studi di Piéron [1922] in Francia e Decroly [1929] in Belgio. In un primo tempo esse misero in luce il limite della soggettività della misurazione educativa e successivamente proposero strumenti più obiettivi di misurazione del profitto. Nel 1923 fu sviluppata da Kelley, Ruch e Terman la prima batteria di test standardizzati, Stanford Achievement Test, per valutare gli apprendimenti degli studenti nelle diverse discipline [Ruch, 1923].

Il termine *docimologia* (dal greco *dokimázo*, esaminare, e *logo*, discorso, riflessione scientifica), per indicare la disciplina che ha per oggetto lo studio sistematico degli esami e della valutazione scolastica in genere, è stato introdotto dal francese H. Piéron nel 1922. Piéron conduce interessanti studi sulle correlazioni che possono esistere tra successi e fallimenti nei diversi settori della conoscenza

La prima ricerca di docimologia, in Francia, fu condotta proprio da Piéron che esaminò 17 scolari, che sostenevano l'esame di licenza elementare, con una batteria di 6 test concernenti diverse capacità. Piéron correlò i risultati ottenuti dai suoi test con gli esiti delle votazioni dello scrutinio finale della licenza elementare. Dai risultati di questa indagine, emerse che i voti riportati dagli allievi di scuola elementare negli apprendimenti erano, di fatto, poco correlati con i risultati rilevati con le prove oggettive di conoscenza. L'intento di Piéron era appunto quello di dimostrare che i risultati degli esami di verifica delle conoscenze, ottenuti dalle prove da lui somministrate, verificavano abilità diverse da quelle dichiarate dagli insegnanti in quanto, questi ultimi, non utilizzavano nella valutazione un metodo scientifico e attendibile.

Oltremania, nel 1927, lo psicologo e statistico Spearman conduce ricerche nel campo psicometrico relative alla misurazione degli apprendimenti, che sono identificati come costrutti latenti derivati dalla correlazione di fattori diversi. Spearman elabora, quindi, una teoria gerarchica dell'intelligenza la cui formulazione è frutto dell'applicazione di una nuova tecnica statistica di analisi dei dati creata proprio dallo stesso Spearman. Tale teoria, nota come *analisi fattoriale* mette in evidenza che, sottoponendo ad un gruppo di individui una serie di test, che consentano di misurare variabili osservate, e calcolate le

correlazioni tra i diversi test, si possa vedere come i coefficienti di correlazione, una volta posti in ordine decrescente, conservino tra loro rapporti costanti. Su questa base, Spearman enunciò la nota *teoria dei due fattori*, secondo la quale esiste un fattore generale, detto *fattore G*, che esprime l'intelligenza generale, che è ereditaria e compare in maggiore o minor peso nella risoluzione di tutti i test di apprendimento, e diversi fattori specifici, *fattori s*, che rappresentano proprio abilità specifiche, che emergono in ogni singolo test in modo differenziato. L'analisi fattoriale, quindi, è un metodo statistico che consente di individuare uno o più fattori che riflettono la similarità tra una serie di variabili osservate. Il punto di partenza di Spearman si dipana, quindi, dall'osservazione delle correlazioni positive fra diversi test di abilità. Sostanzialmente l'assunto di base muove dalla constatazione che soggetti che ottengono risultati sopra la media in un'abilità specifica tendono ad ottenere punteggi sopra la media anche in altre abilità. Il notevole vantaggio di questo metodo è che i pochi fattori latenti individuati riassumono la stessa informazione contenuta nelle variabili osservate, che sono invece molto più numerose. Ancora oggi, in campo psicometrico, uno dei test più utilizzati è il Reattivo di Raven (o Matrici di Raven) che si fonda proprio sulla teoria di Spearman [Misiti, 1954]. Oltre a Spearman, anche Thurstone, come vedremo tra breve, applicò la tecnica statistica dell'analisi fattoriale per studiare la struttura dell'apprendimento. Tuttavia sebbene entrambi gli autori utilizzassero fondamentalmente lo stesso approccio di analisi, i risultati a cui giunsero furono molto diversi.

Negli anni trenta del novecento, nasce in molti Paesi, in relazione all'istruzione, l'esigenza di comprendere i grandi cambiamenti di inizio secolo: la crescita della domanda di istruzione in tutti i ceti della popolazione e il conseguente avvento della scuola di massa. Nel 1931, infatti, la Carnegie Corporation decide di finanziare un'inchiesta internazionale sugli esami finali della scuola secondaria. A tale indagine comparativa partecipano gli Stati Uniti, la Francia, la Germania, l'Inghilterra, la Scozia, la Finlandia e la Svizzera. La ricerca riguarda le sole prove scritte ed è condotta da sei gruppi esaminatori, a ciascuno dei quali vengono assegnati 100 elaborati da esaminare: composizione in lingua francese, versione latina, composizione in lingua inglese, prova di matematica, dissertazione filosofica, prova di fisica. I risultati

dell'indagine evidenziano che gli strumenti utilizzati dagli insegnanti per la valutazione mancano di oggettività poiché risultano influenzati dal docente. In particolar modo spicca le notevoli divergenze tra i docenti nel valutare gli elaborati, nel non applicare uno stesso "metro di misura" ma nell'essere o troppo "larghi" o troppo "stretti" nell'attribuzione dei voti [Vertecchi, 2003].

In seguito di questa indagine, il diffuso interesse in merito all'argomento, portò un ulteriore sviluppo delle ricerche sulla misurazione degli apprendimenti soprattutto nell'area geografica europea, dove si assiste all'implementazione dell'uso di test standardizzati per la misurazione degli apprendimenti degli studenti. Grande apporto alla Pedagogia viene nuovamente dato, proprio negli anni trenta del Novecento, dalla ricerca psicometrica. L'uso dei test per la misurazione delle abilità mentali e delle attitudini viene largamente utilizzato nelle scuole, ma non tanto a scopo educativo e formativo quanto piuttosto per stabilire livelli di padronanza in relazione ai quali costituire classi differenziate per omogeneità di abilità possedute. In tal modo si amplifica l'aspetto elitario della scuola e vengono istituite classi differenziate soprattutto per quelli alunni che presentano uno basso livello intellettuale, generando fenomeni di emarginazione scolastica e sociale. Un esempio importante si trova nell'uso del "*California Test of Mental Maturity*" di Sullivan, Clark e Tiegs [1937], diffuso in campo internazionale, per individuare i livelli intellettivi degli studenti e le specifiche attitudini in modo da operare una valutazione orientativa che consentisse la distribuzione degli studenti in classi omogenee in relazione al livello intellettuale. Nel 1935 nasce, in collaborazione con l'Università di Chicago, ad opera di L.L. Thurstone, la *Psychometric Society* che compie grandi passi avanti nell'applicazione del metodo statistico alla valutazione in quanto sperimenta l'uso dell'analisi fattoriale multipla.

Nel 1938 Thurstone introduce, infatti, un modello di misura multidimensionale dell'intelligenza, evidenziando una struttura gerarchica dei fattori multipli che non possono essere ordinati, in quanto ogni dimensione riveste la stessa importanza e ha lo stesso peso. L'intuizione della multidimensionalità delle caratteristiche psicologiche porta, quindi, Thurstone alla contestazione dei sistemi psicometrici allora utilizzati in quanto ritenere inappropriata la pratica diffusa, soprattutto

nell'uso dei test scolastici, di rappresentare su un continuum lineare le abilità e le attitudini dell'individuo. La contestazione prende le mosse dall'osservazione del fatto che, con questa prassi, vengono poste su una scala di misura lineare caratteristiche che di fatto non possono essere direttamente confrontabili: prendiamo in considerazione due studenti e mettiamoli a confronto per valutare quale dei due sia più preparato in una certa disciplina; sicuramente i loro percorsi formativi e i vissuti personali risulteranno almeno in qualcosa differenti, pertanto non possono essere realisticamente confrontati e posti in punti ben precisi su una linea, neanche in termini di più abile e meno abile, per cui, se anche costruiamo un metodo per farlo, il continuum lineare implicato nel confronto è essenzialmente solo concettuale. Per qualunque punto posto su di una retta nello spazio euclideo sappiamo, infatti, che esiste un *intorno*. L'intorno di un punto è un concetto fondamentale, che è alla base delle nozioni di funzione continua e limite. Un intorno di un punto è intuitivamente un insieme di punti vicini al punto considerato. Ogni intorno, quindi, individua un insieme differente di punti vicini. Di conseguenza se ponessimo l'abilità di ognuno dei due studenti su una linea retta, collocando entrambe in un preciso punto, ognuno di essi avrebbe il suo intorno e poiché, nello spazio euclideo, un intorno di un punto è sempre un insieme aperto contenente il punto stesso, accadrà che l'intorno centrato in ognuno dei due punti avrà un intervallo aperto sia a sinistra che a destra di lunghezza arbitraria con centro nel punto. Questo significa che l'intorno dell'abilità del primo studente potrebbe sovrapporsi, anche solo parzialmente, all'intorno dell'abilità del secondo studente e che, pertanto, le due abilità così misurate non possono essere considerate propriamente precise e distinte. Nondimeno, non siamo in grado di distanziare le due abilità sul continuum in base ad una unità di misura, ed è questo, vedremo, il compito principale dello scaling. Se le posizioni occupate su una retta da differenti studenti rappresentano possibili livelli di abilità in una disciplina, dalla disposizione sul continuum possiamo esprimere un giudizio di abilità in relazione all'ordinamento dei punti sulla retta stessa. Pertanto diremo ad esempio che, in uno spazio che varia da $-\infty$ a $+\infty$, lo studente con abilità a si colloca ad un livello di abilità superiore o inferiore rispetto allo studente b in relazione alla specifica posizione occupata e che, ancora, uno studente

c si collocherà in una posizione che possiamo confrontare con a e con b. Occorre, inoltre, stabilire tutta la gamma di abilità che uno studente può possedere. Una unità di misura dovrebbe consentirci di stabilire quanto sono distanti le tra loro le diverse abilità, ma poichè, come abbiamo premesso, i percorsi formativi e i vissuti personali degli studenti risulteranno almeno in qualcosa differenti, non possiamo escludere che nell'intorno di ogni abilità possa trovarsi la caratteristica propria di un'abilità differente o di un componente psicologica dell'individuo. Pertanto un approccio unidimensionale potrebbe comportare distorsioni evidenti nella misurazione. Thurstone risolve il problema introducendo un modello di misura multidimensionale, caratterizzato da una struttura gerarchica di fattori multipli che non possono essere ordinati, in quanto ogni dimensione riveste la stessa importanza e ha lo stesso peso. L'analisi fattoriale, introdotta da Thurstone, assume, da quello momento, una valenza notevole nella storia della ricerca in campo educativo e, ancora oggi, trova largo impiego soprattutto per valutare la capacità dei fattori estratti di riprodurre la correlazione delle variabili originarie.

Occorre evidenziare come, a distanza di molti anni, i problemi riscontrati dalle prime ricerche sulla valutazione, primo fra tutti, appunto, il grado di soggettività che influenza le misurazioni educative, risulta assolutamente attuale. Purtroppo, infatti, ancora oggi la mancanza di una specifica formazione degli insegnanti in materia di valutazione, fatte poche e rare eccezioni, determina l'utilizzo di prove che seguono criteri non scientifici, arcaici, selettivi e, ancor più grave, finalizzati alla sola valutazione sommativa e non all'analisi dell'intero processo formativo.

Tornando alle principali ricerche svolte nei primi decenni del XX secolo in Francia, Inghilterra e USA sono emerse le seguenti acquisizioni, per le quali si ribadisce il carattere di attualità:

- *Il disaccordo tra i correttori non è determinato dalla materia a cui si riferisce il compito.* L'opinione che la distribuzione dei voti sia dovuta alla materia a cui si riferisce la prova è, di fatto, da considerarsi priva di fondamento. Se nelle prime ricerche di Starch e di Elliot, che riguardavano due composizioni nella lingua madre, si poteva ipotizzare che lo scarto dei voti dei

correttori dipendesse proprio da due compiti diversi e dalla comune opinione di ritenere le prove di composizione linguistica altamente soggettive. In successive ricerche uno stesso compito di geometria, revisionato da 114 insegnanti di questa materia, ottenne una gamma di voti da 28 a 92. Una valutazione analoga è stata attribuita ad un compito di storia corretto da 70 insegnanti di storia. Nel risultato della valutazione si ottennero voti variabili da 43 a 90.

- Laugier and Weinberg [1935] confrontarono le votazioni espresse, a distanza di tempo, dallo stesso correttore su 37 compiti di scienze, somministrate, ovviamente, a gruppi di alunni diversi. Si rilevò che le valutazioni formulate dallo stesso esaminatore, a tre anni di distanza, avevano scarsa correlazione con le valutazioni precedenti di compiti svolti nello stesso modo. Questo risultato evidenziò la variabilità *di voti di uno stesso correttore nel tempo, quindi un sostanziale disaccordo di un correttore con se stesso*. Le ricerche di Starch, Elliot, Lauger e Weinberg mettono, quindi, in evidenza che vi è disaccordo tra i correttori sia se si tratta di persone diverse che correggono uno stesso compito, sia se ci si riferisce allo stesso correttore a distanza di tempo. Naturalmente questo disaccordo non è spiegabile ritenendolo imputabile alla sola soggettività personale del correttore.
- Laugier and Weinberg [1935], ipotizzando che il disaccordo nelle correzioni dipendesse dalla competenza dei correttori, affidarono i 37 compiti di scienze, di cui sopra, ad una studentessa liceale, priva di competenze epistemologiche e metodologiche specifiche della disciplina. I ricercatori correlarono i voti da lei assegnati con quelli ottenuti dai professori universitari della materia e si ottennero analoghi coefficienti di correlazione, scoprendo, quindi, che *la competenza specifica nella disciplina non assicura l'obiettività delle valutazioni*.

Gli studiosi di docimologia, nel corso degli anni, hanno effettuato ricerche e sperimentazioni atte a risolvere e superare i problemi legati alla corretta costruzione delle prove di verifica, in modo da renderle più affidabili, nonché a fissare norme procedurali mirate a garanti-

re oggettività e generalizzabilità nella somministrazione. In ambito scolastico la docimologia si è interessata soprattutto alla ricerca di strumenti di verifica degli apprendimenti, incrementando un tipo di valutazione esclusivamente quantitativa.

2.1.2 International Association for Evaluation of Educational Achievement (IEA)

Un pietra miliare nella ricerca valutativa nel settore istruzione e formazione viene posata nel 1958, quando un gruppo di studiosi, psicologi, sociologi, pedagogisti e psicometrici si incontrano presso l'Istituto per l'Istruzione ad Amburgo, in Germania, per partecipare ad una tavola rotonda sui problemi che interessano il mondo della scuola, con particolare attenzione al problema della scientificità dei metodi di valutazione utilizzati per la misurazione degli apprendimenti negli studenti. Nasce, quindi, una riflessione sulle caratteristiche che deve possedere la valutazione in campo di istruzione e formazione a garanzia di efficacia ed emerge l'importanza di forme di valutazione che non siano limitate esclusivamente alla misurazione degli apprendimenti ma ad un approccio più complesso che consenta di prendere in esame sistemi di istruzione, i curricula, gli apprendimenti, i docenti. I lavori di questo gruppo di studiosi si protraggono e nel 1967 nasce ed acquisisce personalità giuridica l'IEA (International Association for Evaluation of Educational Achievement). Gli studiosi che aderiscono all'IEA vedono il mondo come un maxi laboratorio di istruzione e formazione nel quale i diversi sistemi di istruzione adottano specifici modelli organizzativi e differenti metodologie e strategie di insegnamento al fine di ottenere risultati ottimali. Il confronto e l'analisi comparata consentono di individuare punti di forza e debolezza di ogni esperienza educativo-formativa, condotta nei diversi paesi membri, in modo da poter beneficiare delle migliori pratiche e di superare le problematiche laddove rilevate. Sin dalla sua fondazione l'IEA ha condotto più di 30 studi di ricerca comparata, che si concentrano su argomenti di particolare interesse per i membri dell'IEA. Questi includono studi distinti per settori disciplinari (matematica, scienze, lingua ...), per educazione alla cittadinanza, alfabetizzazione informatica, formazione degli insegnanti e valutazione degli apprendimenti e delle

competenze. L'IEA, con sede giuridica in Belgio, è un'organizzazione non governativa, senza scopo di lucro che si è evoluta nel corso del ultimi 50 anni, possiede centri di ricerca internazionali situati in ogni continente, rappresentati da 68 membri in tutto il mondo. La maggior parte di questi rappresentano sistemi nazionali di istruzione, anche se alcuni paesi hanno una rappresentanza per più di un sistema di istruzione. Tutti i membri sono istituti o agenzie, associate ad una forte tradizione di ricerca in campo formativo. L'IEA vanta una stretta collaborazione con i responsabili delle politiche nazionali e le agenzie di finanziamento formativo. Ogni due o tre anni, l'IEA organizza una conferenza internazionale che raduna ricercatori che lavorano con i dati IEA per favorire lo scambio di informazioni sulle criticità educative. L'IEA offre anche due premi annuali quali riconoscimento per esperienze di alta qualità nel campo della ricerca educativa empirica basata sull'analisi dei dati in possesso della stessa IEA. Lo scopo dell'IEA è quello di effettuare ricerche e studi comparativi in materia di istruzione per fornire parametri di riferimento internazionali, in modo da aiutare i responsabili delle politiche statali ad individuare i punti di forza e di debolezza dei loro sistemi di istruzione e guidarli all'individuazione delle aree di interesse e di azione per la preparazione e valutazione delle riforme scolastiche. Uno degli obiettivi principali è da ricercare nella promozione di azioni che mirano allo sviluppo di strategie nazionali per il monitoraggio e il miglioramento educativo dei diversi sistemi di istruzione, contribuendo allo sviluppo di una comunità mondiale di ricercatori in valutazione educativa. I paesi membri attuali afferenti all'IEA sono: Argentina, Armenia, Australia, Austria, Autorità Filippine, Belgio (Fiammingo e Francese), Bosnia Erzegovina, Botswana, Brasile, Bulgaria, Canada, Chile, Cipro, Corea, Croazia, Danimarca, Egitto, Emirati Arabi, Estonia, Federazione Russa, Finlandia, Francia, Georgia, Germania, Giappone, Giordania, Grecia, Hong Kong SAR, Indonesia, Inghilterra, Iran, Irlanda, Islanda, Israele, Italia, Kazakistan, Kenia, Kuwait, Lettonia, Lituania, Lussemburgo, Macedonia, Malaysia, Messico, Marocco, Nigeria, Norvegia, Nazionale Palestinese, Nuova Zelanda, Paesi Bassi, Porcellana, Portogallo, Qatar, Repubblica Ceca, Repubblica Slovacca, Romania, Scozia, Singapore, Slovenia, Spagna, Stati Uniti, Sudafrica, Svezia, Taipei cinese, Thailandia, Turchia, Ungheria. Altri sistemi educativi che attualmen-

te partecipano a studi dell'IEA sono: Algeria, Azerbaijan, Bahrain, Colombia, Repubblica Dominicana, El Salvador, Ghana, Guatemala, Honduras, Libano, Libia, Liechtenstein, Malta, Mongolia, Oman, Paraguay, Polonia, Arabia Saudita, Serbia, Svizzera, Siria, Tunisia, Ucraina, Yemen. Essendo un'organizzazione non-profit, l'IEA è sostenuta dai suoi membri e da finanziamenti di agenzie governative, fondazioni e altre organizzazioni. Recenti co-finanziatori dei progetti dell'IEA sono: la Commissione europea, la Ford Foundation, la Banca Interamericana di Sviluppo, i Ministeri della Pubblica Istruzione dei paesi membri, il Ministero della cultura e della scienza dei Paesi Bassi, la Banca mondiale, il Programma di sviluppo delle Nazioni Unite, il Dipartimento della Pubblica Istruzione degli Stati Uniti, il Centro nazionale per le statistiche dell'istruzione degli Stati Uniti, diverse agenzie di ricerca e di istruzione in Taipei, Giappone e Norvegia. Altri partenariati includono: Educational Testing Service (ETS), International Institute for Educational Planning (IIEP UNESCO), Organization of Ibero-American States (OEI), Educational Revitalization in the Americas (PREAL), Programme d'Analyse des Systèmes Éducatifs des États de la ConfEmEn (PASEC), Statistics Canada, The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), The West African Examinations Council (WAEC). Gli studi dell'IEA si concentrano sul rendimento scolastico degli studenti e su fattori ad esso correlati. L'IEA promuove valutazioni internazionali sistematiche rispetto a discipline di insegnamento (ad esempio, PIRLS e TIMSS) e a temi legati all'educazione alla cittadinanza (ad esempio, ICCS). La sistematicità con la quale vengono condotte le indagini internazionali permette ai paesi di monitorare i cambiamenti in materia di istruzione e risultati scolastici nel corso del tempo. Gli studi dell'IEA prendono in considerazione il processo di insegnamento-apprendimento, ponendo particolare attenzione agli effetti della formazione sull'individuo e sulla società, utilizzando il concetto di "*opportunità di apprendimento*", al fine di comprendere i legami tra *curriculum programmato*, ossia il curriculum previsto dalle politiche ministeriali, *curricolo effettivamente svolto*, ossia i programmi svolti nelle singole istituzioni scolastiche, e *curriculum ottenuto*, ossia i risultati effettivamente raggiunti dagli studenti. Per studiare queste relazioni, IEA raccoglie i dati di rendimento degli studenti, le informazioni di sistema relative ai dirigenti

scolastici, agli insegnanti, agli studenti (in alcuni studi intervistando anche i genitori), e analizza le politiche di istruzione e formazione dei diversi paesi. Gli studi dell'IEA hanno una forte base empirica: si basano principalmente su dati trasversali e longitudinali, non basati su disegni sperimentali, effettuati per mezzo di raccolta dei dati attraverso metodi di indagine campionaria. L'IEA, oltre ad utilizzare metodi quantitativi di analisi, integra anche metodi qualitativi, come casi di studio e tecniche di osservazione diretta. Al centro della maggior parte studi IEA si pone la valutazione del rendimento degli studenti in una o più materie scolastiche o in uno spazio interdisciplinare di apprendimento. I risultati degli studenti sono misurati attraverso la somministrazione di test oggettivi su un campione di studenti, che sono stati selezionati come rappresentativi della popolazione nazionale. Specifici questionari di contesto vengono utilizzati per raccogliere informazioni sulle caratteristiche delle scuole selezionate, sulle pratiche didattiche e sui finanziamenti erogati. L'IEA ha condotto più di 30 studi comparativi d'istruzione che coprono una vasta gamma di argomenti. Importanti indagini internazionali che è opportuno ricordare riguardano lo studio delle competenze possedute dagli studenti in matematica e nelle scienze, si tratta degli studi TIMSS 1995 , TIMSS 1999 , TIMSS 2003 , TIMSS 2007 , TIMSS 2011, e nelle competenze di lettura in lingua madre, PIRLS 2001 , PIRLS 2006 , PIRLS 2011. L'IEA conduce anche studi su argomenti di particolare interesse per i paesi membri dell'associazione, come la matematica e la fisica avanzate (TIMSS Advanced 1995 , TIMSS Advanced 2008), l'educazione civica e alla cittadinanza (CIVED , ICCS 2009), la tecnologia informatica e le informazioni in materia di istruzione (SITI-M1, SITI-M2, SITI 2006 e, in fase di preparazione, ICILS 2013), e ancora alcune indagini condotte a livello pre-scolare, come la PPP. Nel 2005, l'IEA ha avviato il suo primo studio di istruzione terziaria (TEDS-M). Gli studi dell'IEA sono riconosciute a livello mondiale per l'eccellenza tecnica e metodologica e per il rispetto di standard elevati di innovazione. La forza degli studi dell'IEA è da ricercare prima di tutto nella cooperazione internazionale; ogni studio, infatti, dispone di un centro internazionale di studio, responsabile del coordinamento internazionale, al quale afferiscono tutti i ricercatori in materia degli stati membri. All'interno di ogni paese, un centro nazionale coordina e svolge le attività nazionali le-

gate ad ogni specifica indagine, sotto la direzione di un coordinatore nazionale di ricerca. Ogni indagine dell'IEA è coordinata, quindi, dal centro di studio internazionale, che è il responsabile ultimo per l'esecuzione dell'indagine stessa e da centri di studio internazionali che si trovano in diverse sedi nel mondo. I maggior referenti per il coordinamento delle ricerche internazionali dell'IEA sono: il Consiglio Australiano per la Ricerca Educativa (ACER), il Boston College (USA), il Michigan State University (USA), la Fondazione Nazionale per la Ricerca Educativa (il NFER, in Inghilterra), l'Università di Roma Tre (Italia), l'Università di Twente (Paesi Bassi) e l'Università di Hong Kong (Hong Kong).

2.1.3 Dagli anni '60 ai giorni nostri: gli albori e le nuove frontiere della IRT

Da excursus storico si possono delineare tre filoni che si sono poi sviluppati nella moderna ricerca nella rilevazione delle abilità scolastiche. Il primo fa capo a Thurstone e si è sviluppato a metà anni venti; all'inizio degli anni trenta si sviluppa il filone di Likert; il terzo si afferma nei primi anni cinquanta e fa capo a Guttman. Thurstone [1959] cerca di classificare una batteria di item sul continuo in base alla difficoltà stimata degli stessi, rendendosi però conto di dover posizionare anche i soggetti sullo stesso continuo degli item non riesce a formalizzare un modello che inglobi questa idea. Lord [2005] utilizza una batteria di item, tentando di ordinare i soggetti sul continuo, dividendo le risposte in cinque livelli del tipo: "molto d'accordo"; "d'accordo"; "incerto"; "in disaccordo"; "molto in disaccordo", assegnando un peso a tutti i livelli, 1 al più basso, quindi in disaccordo, e 5 al più alto. Likert definisce una distanza a livello concettuale tra le varie categorie che difficilmente, come abbiamo visto, può essere accertata. Guttman [1949] propone un metodo per trasformare in scale non metriche, ma ordinali ed unidimensionali, un largo numero di items. Obiettivo principale di Guttman è, cioè, di fornire una soluzione al problema dell'unidimensionalità della scala, uno dei punti deboli della tecnica di Likert. La scala Guttman è composta da un insieme di items i cui contenuti sono tali da rappresentare un crescendo (o un diminuendo) di un determinato atteggiamento; in altre parole, gli item

rappresentano una sequenza di gradini, una successione di elementi aventi difficoltà crescente. In questo modo, gli elementi di una scala (item) sono cumulativi (scalabili) ed è possibile immaginare che essi si collochino su un continuum sottostante del quale gli item sono gli indicatori. Per questo motivo la scala Guttman è una scala cumulativa e viene, generalmente, chiamata "scalogramma". La tecnica dello scalogramma prevede come modalità di risposta esclusivamente elementi dicotomici, in questo modo ogni item può avere solo due risposte opposte e distinte (Sino, di solito contrassegnate con i numeri 0 e 1). La scala parte dall'assunto che se un individuo X è, ad esempio, più abile di un individuo Y in una disciplina, possederà tutte le stesse abilità di Y più qualche altra. Il risultato finale è un ordinamento interdipendente di risposte e individui, dove gli individui sono ordinati in base alle risposte date e le risposte in base agli individui. Il punteggio complessivo di ogni individuo è ottenuto sommando i punteggi ottenuti su ogni elemento (item). Guttman pone, quindi, sullo stesso continuo soggetti ed item, ma adotta modelli con scarso rigore statistico ed eccessiva rigidità che portano ad una ridotta capacità di analisi di fenomeni complessi. Il suo modello è di tipo "deterministico" in quanto se un soggetto ha una capacità superiore alla difficoltà dell'item non può fallire, di conseguenza due individui con capacità diverse ma comunque superiori a quella di uno stesso item non possono essere distinte perché la risposta attesa è la medesima per entrambe. Il modello sviluppato da Rasch [1960], come vedremo, fornisce la soluzione ai problemi riscontrati nei tre filoni di ricerca poiché supera il problema della dipendenza tra individuo e test. Il modello di Rasch è un modello di tipo probabilistico, che stima la probabilità di riuscita ad un item in funzione dell'abilità della persona. Ciò che lega e accomuna i tre filoni è come il problema della valutazione degli apprendimenti risulti strettamente connesso all'analisi della validità e dell'attendibilità degli strumenti che vengono utilizzati nella misurazione della performance di uno studente. La nascita della ricerca in campo valutativo, nel settore dell'istruzione e della formazione, muove infatti i primi passi per trovare una risposta al problema della verifica dello strumento di misurazione utilizzato nelle prove di accertamento dei livelli di apprendimento. Abbiamo visto, infatti, come il primo problema che coglie Binet nelle sue ricerche sia stato

proprio quello di individuare uno strumento valido per la classificazione degli studenti in relazione al livello di abilità mentale posseduto. È proprio lo stesso Binet ad improntare le basi teoriche su cui si fonda la *Classical Test Theory* (CTT) con l'intento di studiare l'affidabilità e la validità dei metodi psicometrici fondati sui risultati ottenuti dalla somministrazione di questionari per la verifica degli apprendimenti (*test scores*). Il problema al quale Binet e i suoi collaboratori cercano di dare una risposta è quello di trovare rigorose modalità scientifiche per la calibrazione dei questionari costruiti per la misura di caratteristiche psico-attitudinali, non direttamente osservabili in natura, delle persone esaminate.

La formalizzazione dell'equazione fondamentale su cui si basa la CTT viene proposta originariamente da Spearman [1904]; tale equazione ipotizza una relazione lineare e additiva tra il punteggio osservato di un test (X), la misura della variabile latente oggetto di studio (Θ) e la componente casuale dell'errore (ϵ). Proprio a causa della componente erratica, sempre presente in qualsiasi tipo di misurazione, nasce l'esigenza dello studio di metodi che fossero in grado di determinare il livello di confidenza della misurazione stessa attraverso una stima della correlazione delle singole misure di uno stesso fenomeno ottenute con due strumenti differenti, per esempio con la somministrazione di due questionari di verifica distinti, e per mezzo di una misurazione aggregata, ricavata dalla media di più misurazioni della stessa variabile. Vedremo più nel dettaglio in seguito come la CTT affronta l'analisi della calibrazione dei test e della misurazione degli apprendimenti. L'idea della necessità di calibrazione dei test deriva direttamente dal problema della misurazione nell'ambito delle scienze fisiche: infatti come osserva Taylor and Beach [1984] "Nessuna quantità fisica (una lunghezza, un tempo, una temperatura, ecc.) può essere misurata con assoluta precisione. Operando con cura, possiamo essere capaci di ridurre le incertezze fisiche finché esse sono estremamente piccole, ma eliminarle del tutto è impossibile". Proprio perché la misura sempre affetta da errore, nel caso in cui si debba misurare una variabile latente insorge la necessità di tecniche che ne valutino l'attendibilità, che ne determinino il livello di confidenza attraverso una stima della correlazione delle misurazioni di uno stesso oggetto, ottenute con due strumenti differenti (per esempio due test diversi) e per mezzo di

una misurazione aggregata ricavata dalla media di più misurazioni della stessa variabile. La Teoria Classica dei Test mantiene il primato metodologico nel campo della ricerca valutativa fino agli anni sessanta del Novecento, quando nasce, proprio in sua contrapposizione per superarne le limitazioni, la Teoria della Risposta all'Item (TRI), più nota al mondo scientifico come Item Response Theory (IRT). La IRT, come vedremo, presenta infatti una serie di vantaggi di misurazione che la rendono sicuramente più attendibile a livello applicativo rispetto alla CTT. Le principali differenze tra le due teorie, che saranno affrontate in seguito, riguardano soprattutto i concetti relativi all'errore standard di misurazione, all'attendibilità della misurazione, alla difficoltà dell'item, alla stima del livello di abilità e interpretazione dei punteggi, alla relazione tra variabile dipendente e indipendente e alla bontà dell'item.

Negli anni '60 il matematico danese George Rasch formalizza il modello base della IRT e teorizza il concetto di oggettività specifica che risulterà essere una proprietà fondamentale dei modelli logistici ad un parametro. Il modello di Rasch si è largamente diffuso nelle scienze umane ed è stato utilizzato nelle più svariate modalità. Oggi conosce un successo crescente e l'Istituto per la Misura Oggettiva, che ha sede a Chicago (Institute for Objective Measurement, IOM), è incaricato di promuovere la teoria e le applicazioni del modello. Lo IOM, fondato nel 1966, conta oggi più di 200 membri ai quattro angoli del globo e diverse sedi negli Stati Uniti e in Australia. Lord *et al.* [1968] pubblicano il libro *Statistical theories of mental test scores*, nel quale per primi trattano in maniera sistematica ed esaustiva i principi della IRT, e che costituirà un manuale di riferimento per importanti studiosi come Samejima, Bock e Birnbaum, che avranno un ruolo attivo negli sviluppi successivi di tale teoria. La trattazione degli sviluppi della IRT saranno oggetto di ulteriore approfondimento nel capitolo 5.

3 Il quadro europeo della ricerca valutativa

"L'Europa non nascerà di getto, come città ideale. Essa si farà; anzi si sta già facendo, pezzo per pezzo, settore per settore".

(Kempfer H. H., 1955)

"Parità, autonomia, valutazione sono gli strumenti semplici per permettere un miglioramento dal basso della scuola operato dai non pochi che hanno ancora passione e voglia di insegnare."

(Vittadini G., su "Il Giornale" del 14.5.2008)

Nelle Conclusioni del vertice di Lisbona, i capi di Stato e di governo hanno riconosciuto il ruolo fondamentale di istruzione e formazione per la crescita e lo sviluppo economico: *"Occorre che ogni cittadino possieda le competenze necessarie per vivere e lavorare in questa nuova società dell'informazione imparando ad apprendere per tutto l'arco della vita"*. Il programma di lavoro *"Istruzione e formazione 2010"* dell'Unione europea definisce il quadro di riferimento della cooperazione politica a livello nazionale ed europeo e un piano d'azione integrato costituisce la base dei programmi finanziati dall'UE nel periodo 2007-2013. In entrambi i casi viene riconosciuto il ruolo centrale che l'istruzione superiore e la formazione professionale svolgono nell'agenda dell'UE per la crescita del capitale umano, dell'occupazione e della trasmissione di valori fondanti per la dimensione sociale. L'istruzione e la formazione sono, infatti, finalizzate primariamente a migliorare la comprensione dei valori della solidarietà, delle pari opportunità, della partecipazione e della coesione sociale.

Data l'importanza centrale dell'apprendimento permanente, l'UE finanzia negli Stati membri, con un apposito programma, una serie di iniziative destinate ad istituti e studenti. La dotazione finanziaria del programma è pari a circa 7 miliardi di euro per il 2007-2013. Il Programma d'azione comunitaria nel campo dell'apprendimento permanente, o *Lifelong Learning Programme (LLP)*, è stato istituito con decisione del Parlamento europeo e del Consiglio il 15 novembre 2006, e riunisce al suo interno tutte le iniziative di cooperazione europea

nell'ambito dell'istruzione e della formazione dal 2007 al 2013. Ha sostituito, integrandoli in un unico programma, i precedenti Socrates e Leonardo, attivi dal 1995 al 2006. Il suo obiettivo generale è quello di contribuire, attraverso l'apprendimento permanente, allo sviluppo della Comunità quale società avanzata basata sulla conoscenza, con uno sviluppo economico sostenibile, nuovi e migliori posti di lavoro e una maggiore coesione sociale, garantendo nel contempo una valida tutela dell'ambiente per le generazioni future. In particolare si propone di promuovere, all'interno della Comunità, gli scambi, la cooperazione e la mobilità tra i sistemi d'istruzione e formazione in modo che essi diventino un punto di riferimento di qualità a livello mondiale. Il Programma di apprendimento permanente rafforza e integra le azioni condotte dagli Stati membri, pur mantenendo inalterata la responsabilità affidata ad ognuno di essi riguardo al contenuto dei sistemi di istruzione e formazione e rispettando la loro diversità culturale e linguistica. I fondamenti giuridici si ritrovano negli art. 149 e 150 del *Trattato dell'Unione* dove si afferma che *"La Comunità contribuisce allo sviluppo di un'istruzione di qualità incentivando la cooperazione tra Stati membri e, se necessario, sostenendo ed integrando la loro azione..."* (art. 149) e che *"La Comunità attua una politica di formazione professionale che rafforza ed integra le azioni degli Stati membri..."* (art. 150).

La struttura si presenta come un insieme composto da:

- un insieme di programmi settoriali (o sotto-programmi);
- un programma trasversale, teso ad assicurare il coordinamento tra i diversi settori;
- il programma Jean Monnet per sostenere l'insegnamento, la ricerca e la riflessione nel campo dell'integrazione europea e le istituzioni europee chiave.

Il sotto-programma *Leonardo da Vinci* promuove la formazione professionale, in particolare i tirocini di giovani lavoratori e formatori in imprese situate al di fuori del loro paese d'origine, così come i progetti di cooperazione che collegano gli istituti di formazione alle imprese. Obiettivo del programma è raggiungere 80 000 tirocini all'anno entro il 2013. L'*Erasmus* finanzia la mobilità degli studenti e la cooperazione universitaria. Circa 150000 studenti beneficiano ogni

Programma Life Long Learning			
Comenius	Erasmus	Leonardo da Vinci	Grundtvig
Istruzione scolastica	Istruzione superiore e alta formazione	Formazione iniziale e continua	Educazione degli adulti
Programma Trasversale			
4 attività chiave – Sviluppo Politico, Apprendimento delle lingue, Nuove tecnologie (ICT), Diffusione dei risultati			
Programma Jean Monnet			
3 attività chiave – Programma Jean Monnet, Istituzioni europee, Associazioni europee			

Figura 3.1: Life Long Learning Program

anno di questo programma. Esiste inoltre il programma Erasmus Mundus, che permette ogni anno a circa 2 000 studenti e docenti universitari provenienti da tutto il mondo di ottenere un master al termine di corsi proposti da consorzi di almeno tre università europee. Il sotto-programma *Grundtvig* sostiene programmi di apprendimento permanente per gli adulti, in particolare partenariati, reti e iniziative di mobilità transnazionali e mira a finanziare la mobilità per 7 000 persone all'anno entro il 2013. Il *Comenius* promuove la cooperazione tra le scuole e i loro insegnanti. L'obiettivo del programma è la partecipazione di tre milioni di allievi tra il 2007 e il 2013. Altri fondi sono inoltre disponibili per promuovere la cooperazione politica, l'apprendimento delle lingue, l'implementazione dell'e-learning nonché la diffusione e lo scambio di pratiche esemplari.¹

L'UE sostiene, inoltre, la formazione professionale attraverso il Cedefop, il Centro Europeo per lo Sviluppo della Formazione Professionale di Salonicco, che apporta conoscenze scientifiche e tecniche in settori specifici e promuove lo scambio di idee tra i diversi partner europei. Istruzione, formazione e apprendimento permanente sono fattori essenziali per un'economia dinamica basata sulla conoscenza, poiché sono la chiave per posti di lavoro di qualità e per una parteci-

¹Per maggiori informazioni si consiglia di visitare il sito:
http://www.programmallp.it/llp_home.php?id_cnt=1

pazione attiva alla società. L'UE, a tale proposito, aveva fissato per il 2010 i seguenti obiettivi relativi all'istruzione:

- Raggiungimento dell'85% dei ventiduenni che abbiano portato a termine la propria istruzione secondaria superiore;
- Percentuale inferiore al 10% dei giovani tra i 18 e i 24 anni che abbandonano la scuola, prima di aver completato la propria istruzione secondaria superiore, la propria formazione professionale o di altro tipo;
- Aumento del 15% del numero dei laureati in matematica, scienze e tecnologia, con una diminuzione del livello di disparità tra i sessi;
- Diminuzione della percentuale di quindicenni con bassi livelli di capacità di lettura, almeno del 20% rispetto al 2000;
- Raggiungimento di almeno il 12% della popolazione nella fascia di età fra i 25 e i 64 anni che siano impegnati nell'apprendimento permanente;
- Riduzione a non più di quattro mesi del tempo che intercorre dal termine della scuola all'ingresso nel mondo del lavoro o in esperienze di tirocinio, o di seguire una formazione supplementare o attività alternative volte a accrescerne le possibilità di occupazione.

Uno degli obiettivi di primaria importanza per il programma di lavoro *"Istruzione e formazione 2010"* dell'Unione europea, finalizzato alla garanzia dei principi di libera circolazione delle merci, delle persone, dei servizi e dei capitali, è quello di individuare un metodo unico, chiaro e condiviso per registrare le competenze possedute dai cittadini dell'Unione Europea.

Nei documenti *Europass* messi a punto dall'UE, infatti, figurano in un unico formato le competenze e le qualifiche acquisite in diversi paesi dell'UE. Tali documenti sono: il Curriculum Vitae Europass, il Portfolio Linguistico Europass, la Mobilità Europass, il Supplemento Europass al Diploma per l'istruzione superiore e il Supplemento Europass al Certificato per la formazione professionale.

L'azione dell'UE nel campo dell'istruzione è essenzialmente volta a promuovere la cooperazione tra paesi che devono far fronte a sfide comuni, nel rispetto delle diversità nazionali. Promuovere la convergenza e l'equiparabilità dei titoli di istruzione superiore, ad esempio attraverso il coordinamento delle riforme, l'introduzione di sistemi compatibili e azioni comuni, permette sia di favorire lo sviluppo personale sia di sfruttare nel modo migliore il potenziale economico dell'Unione europea.

Il 30 novembre 2002, i Ministri dell'Istruzione di 31 paesi europei e la Commissione europea hanno adottato la Dichiarazione di Copenaghen sulla promozione di una maggiore cooperazione europea in materia di istruzione e formazione professionale. Tuttavia è ancora oggi in via di perfezionamento un sistema europeo di trasferimento dei crediti educativi e per facilitare il riconoscimento reciproco e la trasferibilità delle qualifiche professionali. In questo ambito rientra il progetto di un "quadro europeo delle qualifiche per l'apprendimento permanente". I paesi dell'UE stanno, inoltre, lavorando al progetto di un Istituto Europeo della Tecnologia, destinato a diventare un nuovo polo di eccellenza nei settori dell'istruzione superiore, della ricerca e dell'innovazione. La Commissione europea sostiene ancora la modernizzazione delle università in una forma tale da creare sinergie con l'obiettivo di una maggiore crescita ed occupazione, tanto nelle discipline insegnate quanto nel modo di incentivare la mobilità degli studenti. L'UE, a tal fine, promuove la cooperazione nel settore dell'istruzione, gli scambi e i corsi sull'integrazione europea con circa 80 paesi di tutto il mondo, dalla Mongolia al Messico, dall'Algeria all'Australia. È proprio a tale scopo che la Fondazione Europea per la Formazione, un'agenzia dell'UE con sede a Torino, promuove l'adozione di pratiche innovative in tema di sviluppo delle risorse umane nelle società e nelle economie in transizione dell'area del Mediterraneo, dell'Europa orientale e sud orientale e dell'Asia centrale.

Le politiche rivolte ai giovani, tuttavia, non si limitano solo all'istruzione. Nel marzo 2005, infatti, i capi di Stato e di governo hanno adottato il *Patto europeo per la gioventù*, che definisce una serie di principi comuni riguardanti la creazione di opportunità di lavoro per i giovani, le competenze di base che i sistemi di istruzione devono assicurare e la necessità di un equilibrio tra attività professionale e

vita familiare. Il Patto europeo per la gioventù si articola, infatti, su tre assi:

1. Occupazione, integrazione e promozione sociale;
2. Istruzione, formazione e mobilità;
3. Conciliazione tra vita familiare e professionale.

Il programma “Gioventù” dell’UE, ancora, promuove la partecipazione attiva alla cittadinanza europea, attraverso progetti volti a rafforzare nei giovani il sentimento stesso di cittadinanza europea e a svilupparne lo spirito d’iniziativa, la creatività e lo spirito imprenditoriale. L’UE investirà complessivamente circa 0,9 miliardi di euro in queste attività tra il 2007 e il 2013.

3.1 Europa2020... il post Lisbona 2010: gli obiettivi dell’istruzione in Europa

La Commissione Europea al Consiglio di Lisbona del 23 e 24 marzo 2000 lanciava la “Strategia di Lisbona”, un piano di sviluppo per i paesi dell’Unione per il periodo 2000-2010. L’obiettivo complessivo dell’Agenda mirava a rendere l’UE, entro il 2010, *l’economia basata sulla conoscenza più competitiva e dinamica del mondo, in grado di realizzare una crescita economica sostenibile, con nuovi e migliori posti di lavoro ed una maggiore coesione sociale*.

La garanzia di istruzione, obbligatoria e di qualità per tutti e per ciascuno, è la base fondamentale richiesta per costruire una vera Europa della conoscenza. Questo era uno dei contributi richiesti ai sistemi educativi dalla strategia di Lisbona per far sì che la nuova economia europea diventasse entro il 2010 la più competitiva e dinamica al mondo.

Per rispondere a questa sfida, che coinvolge tutti i Paesi europei, e incoraggiare la cooperazione in quest’area, il 12 febbraio 2001, il Parlamento e il Consiglio hanno adottato una raccomandazione riguardante, specificatamente, la valutazione della qualità dell’istruzione scolastica. Lo studio comparato eseguito da Eurydice ha contribuito senza dubbi a fornire una maggiore comprensione sul tema della valutazione in

Europa e della situazione europea attuale. Lo studio indica che la valutazione della qualità degli istituti scolastici assume varie forme in quanto ogni paese ha elaborato una modalità che corrisponde sia al proprio metodo di gestire e organizzare il sistema educativo, sia ai propri obiettivi. Al di là di queste specifiche diversità, è forte una generale presa di coscienza del fatto che esista un bisogno cruciale in tutti gli ambiti di un controllo di qualità e di implementazione dei miglioramenti. La cooperazione europea in campo educativo è essenziale ed è necessario che si sviluppi una vera e propria cultura della valutazione che sia condivisa. La valutazione della qualità dell'istruzione scolastica, sta quindi al centro degli obiettivi assegnati ai sistemi di istruzione e formazione. L'Unione Europea si era data, in merito, nel Consiglio di Lisbona del marzo 2000, l'obiettivo della creazione di un'economia della conoscenza finalizzata a garantire più elevati livelli di crescita economica sostenibile, un maggior tasso di occupazione e una maggior coesione sociale. La strategia di Lisbona ha visto impegnati gli Stati Europei a perseguirne gli obiettivi entro il 2010.

L'edizione 2006 del rapporto annuale della Commissione europea sui progressi compiuti dagli Stati membri per il conseguimento degli obiettivi di Lisbona nell'istruzione e nella formazione, metteva in evidenza come, nonostante i risultati positivi registrati in alcuni settori, i progressi realizzati dai sistemi europei di istruzione e formazione, nell'ottica del pieno raggiungimento degli obiettivi di Lisbona, fossero ancora scarsi. Il commissario all'istruzione, formazione, cultura e multilinguismo Ján Figel', ha allora puntualizzato che occorre compiere passi avanti più decisi, perché senza un impegno più immediato, intenso e concreto da parte di tutti gli Stati membri dell'UE "la competitività dell'Europa non potrà progredire". Lo stesso commissario ha poi aggiunto: "L'investimento nel capitale umano è nettamente un investimento vitale per l'avvenire dell'Europa". Il rapporto evidenzia come, in base ai dati del 2005, sono circa 6 milioni i giovani tra i 18 e i 24 anni che hanno abbandonato gli studi prematuramente. Per raggiungere gli obiettivi fissati a Lisbona - 10% al massimo di giovani che abbandonano gli studi - sarebbe necessario che almeno 2 milioni di essi riprendessero il loro percorso educativo. Per raggiungere, entro il 2010, l'obiettivo dell'85% di studenti che avessero concluso positi-

vamente gli studi superiori, sarebbe stato necessario un incremento di 2 milioni di giovani. In alcuni paesi, però, questo obiettivo è già stato raggiunto e superato. In Slovacchia il dato raggiunge il 91,5%, in Slovenia il 90,6% e nella Repubblica Ceca il 90,3%.

L'attuazione dell'Agenda *Lisbona 2010* è stata prevista utilizzando il metodo del coordinamento aperto e i risultati conseguiti dai diversi stati membri sono stati diversi. Si sono verificate difficoltà e ritardi che hanno reso necessarie la rielaborazione e il rilancio della strategia di Lisbona, in particolare per quanto riguarda la crescita e l'occupazione. Le Istituzioni europee e i Governi dei Paesi dell'UE dopo le opportune verifiche dovranno, quindi, definire nel dettaglio la Strategia post 2010, per una crescita sostenibile e per un'occupazione di qualità che rafforzi la dimensione sociale. Il Rapporto sullo sviluppo umano dell'UNDP (Programma di Sviluppo delle Nazioni Unite) del 1999 mette in evidenza che non tutti gli abitanti del nostro pianeta appartengono al "villaggio globale" dell'economia della conoscenza, in quanto ancora una larga percentuale della popolazione rimane in condizioni di sottosviluppo e, anche laddove le infrastrutture tecnologiche lo consentono, si rileva tuttavia un'elevata percentuale di cittadini che non possiedono neanche le competenze di base. In relazione a ciò Foray [2006], mette in evidenza che anche in Europa, nonostante la qualità dei programmi di istruzione e formazione, alcuni strati della popolazione rimangono sostanzialmente esclusi dalle competenze legate all'economia della conoscenza. Maggiori preoccupazioni nascono, per quel che riguarda il nostro Paese, se ci proiettiamo al futuro poiché dai dati delle ultime indagini Ocse-Pisa emerge un basso livello di istruzione degli studenti italiani rispetto alla media dei paesi OCDE. In Italia un adolescente su 5 non va a scuola e non lavora. Appena il 22,2% dei giovani di età compresa tra i 15 e i 19 anni hanno una scolarizzazione secondaria. Un quadro allarmante se si considera l'alta correlazione tra il possesso di istruzione e i titoli di studio e l'opportunità di trovare un lavoro più qualificato. Ai primi posti per livello di istruzione, Corea (97% dei giovani ha un diploma di scuola secondaria), Norvegia (95%), Giappone e Slovacchia (94%). Agli ultimi posti, Italia, Turchia, Messico, Portogallo, Usa e Gran Bretagna. Questi dati fanno cogliere una stretta correlazione tra istruzione e crescita economica, in quanto proprio la Corea e il Giappone hanno avuto

un'esponenziale crescita economica dovuta al grande investimento in Capitale Umano e alla scelta di basare l'organizzazione economica su un'economia fondata sulla conoscenza.

Ma la formazione scolastica da sola non basta per rispondere efficacemente alle sfide lanciate dal mondo del lavoro: nel vecchio continente si spende ancora poco per i programmi di aggiornamento professionale. In Usa, Finlandia, Svezia e Svizzera più del 40% dei lavoratori ogni anno partecipa a programmi di aggiornamento, mentre la percentuale scende al 10% nel caso di Italia, Spagna, Grecia, Portogallo e Ungheria. L'Italia, però, si mantiene in testa alla classifica per quanto concerne le ore di insegnamento agli studenti di età compresa tra i 7 e i 14 anni, circa 8.000 ore, contro una media globale di 6.852 ore. Mentre nel mondo si registra una tendenza al miglioramento del tasso di scolarizzazione (il 75% delle persone nate dopo gli anni 70 ha finito il liceo, mentre negli anni 40 la percentuale si fermava ad appena il 50%) in Italia assume ancora dati rilevanti il problema della dispersione scolastica. Dai dati trasmessi nel 2007 dal Ministero della Pubblica Istruzione risulta che gli *"early school leavers"*, coloro che si sono congedati un po' troppo presto dalla scuola, in Italia, nel 2006, risultavano ben 890 mila: ragazzi di età compresa fra i 18 e i 24 anni - pari al 20,6% del totale di quella fascia - in possesso della sola licenza media e che non partecipano a nessuna forma di educazione o formazione, insomma, giovani usciti definitivamente dai circuiti formativi. Siamo preceduti solo da Spagna e Portogallo. La media Ue cala al 14, 9%, mentre Finlandia e Danimarca sono già al di sotto del 10% e Francia, Germania, Olanda, Regno Unito e Grecia sono vicini all'obiettivo di Lisbona. Enorme anche il numero dei ragazzi promossi con debito alle scuole secondarie di secondo grado che risulta essere quasi un milione. Così, l'obiettivo di Lisbona 2010 di scendere al di sotto del 10% appare al momento lontanissimo. Nonostante gli sforzi compiuti non si riesce a contenere il fenomeno della dispersione scolastica. Sono, infatti, quasi mezzo milione gli alunni italiani che, ogni anno, interrompono gli studi o vanno incontro ad una bocciatura. I dati sul preoccupante bilancio della dispersione scolastica in Italia sono stati pubblicati dal Ministero stesso con un dossier dal titolo *"La dispersione scolastica: indicatori di base per l'analisi del fenomeno"*. Un fenomeno che al Paese incide per circa 3 miliardi di euro l'anno di costi diretti e una

cifra non quantificabile di costi indiretti che possono essere molto pesanti. Il Commissario europeo per l'istruzione, la formazione, la cultura e il multilinguismo, Ján Figel, avverte: "Sistemi d'istruzione e di formazione efficienti possono avere un notevole impatto positivo sulla nostra economia e società ma le disuguaglianze nell'istruzione e nella formazione hanno consistenti costi occulti che raramente appaiono nei sistemi di contabilità pubblica. Se dimentichiamo la dimensione sociale dell'istruzione e della formazione, rischiamo di incorrere in seguito in notevoli spese riparative".

Negli ultimi anni si è dato largo spazio al dibattito sull'importanza di avvicinare maggiormente la scuola al mondo del lavoro poiché la scuola è risultata spesso inadeguata nel formare cittadini-lavoratori del domani. Va detto tuttavia che la scuola non deve essere "fagocitata" dal mondo delle imprese perché la funzione primaria a cui deve assolvere è sempre e comunque quella educativa, ovvero quella di assicurare l'apprendimento culturale, la capacità di analisi critica, la flessibilità e la creatività, indipendentemente dall'inserimento in uno specifico settore lavorativo.

Bastano, quindi, pochi numeri per collocare l'Italia ai primi posti in Europa per la dispersione scolastica. Nell'anno scolastico 2004/2005 sono stati 342 mila gli alunni (di cui 289 mila al superiore) andati incontro ad una bocciatura. In più, 100 mila tra ragazzini della scuola media e studenti delle superiori dopo essersi iscritti a scuola si sono ritirati, in molti casi, senza comunicare niente a nessuno. La dispersione scolastica è più alta nei primi anni della scuola media e delle superiori. In particolar modo nelle prime classi degli istituti professionali, dove i tassi di bocciatura raggiungono livelli stratosferici: 42%. Ma è soprattutto in tutte le regioni meridionali che si annida la dispersione con valori superiori alle medie nazionali anche del 30%. Segno che il fenomeno è legato anche ad aspetti economici e sociali. Quindi, se ai bocciati e ritirati si sommano gli evasori totali, che superano il mezzo milione di alunni e se, nella scuola secondaria di secondo grado, si aggiungono i promossi con debito, l'insuccesso colpisce più di metà degli studenti italiani.

Dall'ultima indagine Eurostat risulta peraltro che in Italia la spesa pubblica per l'istruzione primaria, secondaria e universitaria si colloca al 18° posto nell'elenco dei Paesi dell'Unione Europea. Si tratta dei

dati comparabili più recenti (si riferiscono all'anno 2006) basati sugli indicatori messi a punto, con una metodologia comune, da Eurostat assieme a Unesco e Ocse. Dell'Italia ne è emerso un quadro piuttosto infelice, che non fa altro che confermare i dati diffusi dall'Ocse nel rapporto annuale Uno sguardo sull'educazione 2009. Secondo l'indagine Eurostat, in media, i paesi dell'Ue spendono per l'istruzione il 5,05% del Pil. La percentuale maggiore è in Danimarca (7,98%) e a Cipro (7,02%), mentre la più bassa in Slovacchia: 3,79%. L'Italia è al 18° posto con il 4,73%. Per l'istruzione universitaria, addirittura, solo la Bulgaria spende meno dell'Italia. Infatti, i 27 Paesi Ue spendono per l'università l'1,1% del PIL, l'Italia lo 0,8%, mentre la Danimarca spende ben il 2,3% del Pil. La spesa media per l'istruzione di uno studente ammonta a 5.970 euro l'anno. In particolare, 4.921 euro si spendono per un bambino della scuola elementare, 6.038 per un allievo delle scuole superiori e 8.591 per uno studente universitario. Per uno studente universitario in Italia lo Stato spende in media 6.984 euro l'anno, circa la metà di quanto si spende in Svezia, Danimarca e Olanda, tre volte meno di quanto si spende a Cipro. Esigui sono anche i fondi privati a disposizione della scuola italiana. Infatti, mentre in media i fondi privati rappresentano il 12,5% delle risorse complessive a disposizione degli istituti di istruzione, in Italia la percentuale è del 7,7%. In media, il 6% della spesa complessiva finisce direttamente nelle tasche degli studenti e delle famiglie sotto forma di aiuti, quali borse di studio e buoni libro. In Italia, inoltre, gli aiuti coprono solo il 4% della spesa pubblica totale, contro il 16% della Danimarca. Non se la cava bene neanche la scuola privata: la media dei Paesi UE è del 12,5%; la spesa in Italia è soltanto del 7,7%.

Un'altra indagine molto interessante condotta da Eurostat, inoltre, mostra il basso grado di penetrazione di Internet in Italia e come l'uso di Internet sia legato al livello d'istruzione delle persone. L'indagine è interessante perché evidenzia come il cosiddetto *digital divide*, cioè la frattura di competenze e potenzialità che esiste tra chi è in grado di usare le nuove tecnologie dell'informazione e della comunicazione e chi invece no, abbia una stretta correlazione con il livello di istruzione. L'Italia è indietro rispetto agli altri membri dell'UE anche in questa classifica, molto probabilmente a causa del livello medio di istruzione della sua popolazione, che è appunto più basso della media europea.

Se si mettono in correlazione il livello di studio, l'investimento nell'istruzione e nella formazione e la diffusione delle nuove tecnologie si comprende quali siano alcune delle grandi problematiche di fondo che condizionano la crescita economica italiana nel mondo.

Oggi, più che mai, ognuno ha bisogno di comprendere la complessità degli eventi ma, ciononostante, i sistemi di insegnamento tradizionali adottati in grande maggioranza in tutti i Paesi continuano a separare, a disgiungere le conoscenze che dovrebbero invece essere interconnesse, e continuano, troppo spesso, a formare menti unidimensionali ed esperti riduzionisti che privilegiano una sola dimensione dei problemi umani, occultando tutte le altre. Ciò non significa che nell'istituzione scuola non si stia cercando di procedere in questi ultimi anni ad un processo di rinnovamento, ma che sicuramente questo stesso procede troppo lentamente rispetto alla repentinità innovativa che caratterizza l'economia della conoscenza. Affinché tutti siano in grado di essere cittadini attivi in questo momento storico e nel futuro, si rende necessaria una riforma del nostro modo di conoscere, una riforma del nostro modo di pensare, una riforma dell'insegnamento: tre riforme strettamente interdipendenti tra loro. *Formazione, capitale umano* e *learning organizations* sono capitoli importanti della strada che oggi tutti i capitalismi nazionali stanno intraprendendo, affidandosi non solo alle innovazioni di singole imprese ma a qualche forma di intelligenza collettiva [Lévy, 1994]. L'Italia lo deve fare al pari degli altri. La modernizzazione degli anni a venire, infatti, non sarà principalmente un affare di macchine e di tecnologie, che pure hanno la loro funzione da svolgere, ma sarà soprattutto una rivoluzione intellettuale, destinata ad investire senza eccezioni l'intelligenza dell'impresa e quella del lavoro. Tre sono le sfide fondamentali da raccogliere su questo fronte:

1. Rendere reversibili le specializzazioni con una robusta immissione di sapere di tipo generale e di capacità di gestire relazioni;
2. Trasformare le competenze tecniche in competenze intellettuali;
3. Spostare l'accento dalla formazione all'apprendimento.

L'UE chiede di accelerare le riforme, soprattutto nel campo dell'istruzione e della formazione, che sono le strutture portanti per

garantire una crescita futura. Gli Stati membri, dunque, hanno il dovere di accelerare il ritmo delle riforme nei loro sistemi di istruzione e formazione, altrimenti una quota consistente della prossima generazione dovrà confrontarsi con l'esclusione sociale.

Il Presidente della Commissione, Josè Manuel Barroso, dinanzi ai capi di stato e di governo riuniti a Bruxelles l'11 febbraio 2010, ha evocato la necessità di una strategia per lo sviluppo sostenibile e l'occupazione che dovrà essere attuata con il progetto "*Europa 2020*". L'uscita dall'attuale crisi può essere l'occasione per entrare in una fase di nuova e sostenibile economia sociale di mercato in cui la prosperità dell'Europa deriverà dall'innovazione e dal miglior utilizzo delle risorse, dove il sapere costituirà l'input principale. Molteplici studi in campo economico e sociale hanno messo in evidenza che il successo competitivo e il raggiungimento di adeguati standard di benessere e coesione sociale dipendono in larga misura dalla diffusione di livelli elevati di competenze e capacità di apprendimento permanente nella popolazione. La nuova strategia dovrà permettere all'UE di uscire dalla crisi e avviarsi sulla strada della ripresa, concentrandosi sui successi conseguiti dalla Strategia di Lisbona e cercando di superarne i principali difetti.

In questa prospettiva di crescita comune, il 3 marzo 2010 la Commissione europea ha lanciato la strategia *Europa 2020* al fine di uscire dalla crisi e preparare l'economia dell'UE per il prossimo decennio. La Commissione ha individuato tre motori di crescita, da mettere in atto mediante azioni concrete a livello europeo e nazionale:

- Crescita intelligente (promuovendo la conoscenza, l'innovazione, l'istruzione e la società digitale)
- Crescita sostenibile (rendendo la produzione più efficiente sotto il profilo delle risorse, più verde e più competitiva)
- Crescita inclusiva (promuovendo un'economia con un alto tasso di occupazione che favorisca la coesione sociale e territoriale)

Vengono fissati, inoltre, cinque obiettivi da cui si evince quali sono i traguardi che l'UE dovrebbe raggiungere entro il 2020 e in base ai quali saranno valutati i progressi compiuti. I cinque obiettivi proposti dalla Commissione, che non potranno essere considerati "punti di

riferimento”, ma dovranno diventare “obblighi concreti”, tradotti in programmi e realizzati, sono:

- il raggiungimento occupazionale del 75% delle persone di età compresa tra 20 e 64 anni;
- l’investimento del 3% del PIL dell’UE in R&S (ricerca e sviluppo);
- il raggiungimento dei traguardi della riduzione dei gas ad effetto serra e i consumi energetici del 20%, soddisfacendo il 20% del fabbisogno energetico mediante l’utilizzo delle energie rinnovabili;
- la riduzione del tasso di abbandono scolastico, che deve essere inferiore al 10%, e il raggiungimento di almeno il 40% dei giovani che abbiano conseguito un titolo di studi pari alla laurea;
- la riduzione di 20 milioni di persone dal rischio di povertà.

La Comunità europea, nel giudicare gli sforzi di ogni singolo Stato, premierà quelli più virtuosi con incentivi sul fronte dell’accesso ai fondi europei; quelli inadempienti saranno invece oggetto di raccomandazioni finalizzate al raggiungimento degli obiettivi prefissati.

Per realizzare concretamente questi obiettivi occorre prioritariamente lavorare sulla valutazione, che erroneamente può sembrare un punto di arrivo in un processo di insegnamento-apprendimento ma che, al contrario, è il punto di partenza poiché consente avere un quadro ben preciso della situazione in cui si trova oggi il nostro sistema di istruzione. La valutazione deve, infatti, garantire un’accurata analisi della situazione di partenza, tenendo ben presenti le caratteristiche storiche, culturali, sociali ed economiche in cui il processo insegnamento-apprendimento ha luogo; deve prevedere obiettivi a breve e a lungo termine, valutando le ricadute delle competenze acquisite sull’autonoma gestione della vita di ogni individuo; deve accompagnare tutto il processo di formazione e garantire gli opportuni interventi qualora si presentino delle difficoltà da parte di qualcuno a conseguire gli obiettivi prefissati per tutto il gruppo classe; deve essere flessibile e pronta al cambiamento e all’imprevisto per consentire la strutturazione di nuovi percorsi alternativi e di nuove strategie

di insegnamento; deve arrivare alla costruzione di un quadro completo di tutto il processo e dei risultati conseguiti per consentire di programmare sempre nuovi cammini lungo la strada della conoscenza.

A livello di sistema è necessario analizzare accuratamente la situazione attuale per avere la percezione del punto in cui è arrivati, cogliendo in modo esatto i punti forti del nostro sistema di istruzione, prima ancora di sottolinearne i punti deboli; comprendendo che cosa si deve ancora fare per garantire una formazione efficace ed efficiente per tutti e per ciascuno; aiutando i ragazzi a diventare maggiormente consapevoli delle qualità e competenze possedute, nonché delle proprie difficoltà, poiché solo facendo leva sui punti di forza si possono superare incertezze e carenze. Se, in qualità di educatori, riuscissimo, infatti, a rendere gli studenti co-protagonisti della valutazione, daremmo loro l'opportunità di acquisire maggior sicurezza e stima di sé, consentendoli di maturare un più elevato livello di capacità di scelta e di decisione e di sviluppare un pensiero riflessivo, critico e creativo.

3.2 Gli indicatori internazionali della valutazione

L'individuazione di indicatori internazionali dell'educazione è di interesse molto recente. Tale esigenza, infatti, è nata verso la metà degli anni ottanta e si è concretizzata all'inizio degli anni novanta. In realtà, un primo tentativo era già stato effettuato all'inizio degli anni settanta ma esso fu un fallimento in quanto era ormai giunta alla fine "una fede quasi ingenua", atteggiamento molto diffuso negli anni cinquanta e sessanta, nell'applicabilità dell'approccio esplicativo nelle scienze sociali, che presuppone "l'uso del modello quantitativo o positivistico della ricerca" [Bottani, N. e Tuijnman, A., 1994]. Tuttavia, dopo la pubblicazione nel 1983 del rapporto sulla scuola americana, *A nation at risk*, prima negli Stati Uniti e poi negli altri paesi dell'OCSE, si è velocemente sviluppata per l'Europa l'esigenza di avere informazioni comparate sui vari sistemi scolastici al fine di disporre, sostanzialmente, di un insieme di indicatori comuni.

Dopo alcuni anni di lavori e di dibattiti molto vivaci, nel settembre 1991, i delegati, presenti a Lugano all'Assemblea generale del progetto INES (International Indicators of Educational Systems) dell'OCSE, decisero di pubblicare, nel settembre del 1992, una prima serie di

indicatori internazionali, costituita da 36 indicatori. Lo scopo era quello di realizzare un database, una raccolta di dati in grado di fornire informazioni sullo stato e sull'evoluzione di un sistema scolastico. I responsabili dell'OCSE erano, tuttavia, consapevoli del fatto che "ci sarebbero stati degli errori, che molte importanti aree non sarebbero state coperte e che sarebbe stato difficile evitare interpretazioni erranee" [Bottani, N. e Tuijnman, A., 1994, pag.41]. Gli indicatori dell'OCSE sono stati organizzati in tre raggruppamenti che danno informazioni su: Contesti demografico, economico e sociale; Aspetti caratterizzanti i sistemi educativi, quali ad esempio le spese per l'istruzione, la scolarizzazione, le ore di insegnamento, la retribuzione degli insegnanti, il numero medio di alunni per classe, ecc.; Risultati dell'istruzione. Quando si analizzano indicatori relativi ad aspetti diversi dell'educazione, non si ha solo il problema della garanzia di validità della singola misurazione effettuata, ma è di fondamentale importanza anche l'aver cura di incrociare le singole misurazioni degli indicatori tra loro, per comprendere globalmente ed effettivamente il sistema educativo analizzato. Dall'analisi dei primi indicatori strutturati dall'OCSE per la valutazione del sistema scolastico, risulta che la maggior parte di essi dà un gran numero di informazioni circa il contesto economico e sociale del sistema scolastico a cui si fa riferimento ma, per quanto concerne i risultati degli apprendimenti, si possono cogliere poche informazioni rilevanti. La situazione non è sostanzialmente cambiata nel corso degli anni novanta con la pubblicazione da parte dell'OCSE degli aggiornamenti degli indicatori. Questa asimmetria nello sviluppo degli indicatori non è, tuttavia, casuale ma deriva soprattutto dal fatto che mancano ancora gli "strumenti adeguati per misurare i risultati o esiti del sistema educativo e per valutare, ciò che è ancora più difficile, le conseguenze a lungo termine dell'educazione per l'individuo, la famiglia, la comunità locale, il posto di lavoro, l'economia nazionale e globale" [Bottani, N. e Tuijnman, A., 1994, pag.34]. Inoltre, se è vero che si occupa di studi sull'educazione è concorde nel ritenere che una valutazione del successo scolastico non possa e non debba essere ridotta a una valutazione del rendimento dello studente in alcune materie fondamentali, molte difficoltà sono state riscontrate dagli esperti che hanno elaborato gli indicatori OCSE sugli aspetti precisi che dovrebbero essere valutati per essere sicuri che la natura

sfaccettata dell'educazione sia riconosciuta in tutta la sua complessità. Il principio che sta alla base di una valutazione del sistema scolastico è quello di un'analisi approfondita dello stesso, finalizzata ad un'azione regolativa del sistema oggetto di indagine. Il sistema, sostanzialmente, viene analizzato per comprendere quale sia la modalità per migliorare l'efficacia e l'efficienza degli apprendimenti. Tuttavia, però, se l'indagine si basa esclusivamente sulla somministrazione di prove oggettive per la verifica degli apprendimenti che non sono in grado, da sole, di ottenere un quadro chiaro e dettagliato della complessità di un apprendimento. De Landsheere, G. [1973], uno dei maggiori esperti che si è occupato di ricerca valutativa in campo formativo, indica i rischi a cui si può andare incontro se la valutazione degli apprendimenti si limita a valutazioni di tipo oggettivo, quali ad esempio la somministrazione di test a risposta chiusa ed evidenzia che tali rischi si amplificano maggiormente se riferiti ad una valutazione dell'intero sistema scolastico. Se gli insegnanti, le scuole e le famiglie ritenessero significativa, ai fini dell'apprendimento, la sola acquisizione di conoscenze valutabili attraverso l'uso di test, si correrebbe il rischio che tutto il sistema scolastico si possa piegare sostanzialmente all'esigenza di mettere gli studenti esclusivamente in condizione di superare i test. Un atteggiamento simile porterebbe ad un ritorno radicale ad un tipo di insegnamento puramente nozionistico, basato sull'addestramento al superamento di prove di verifica strutturate. La maggior parte dei test si presenta, infatti, come prova di verifica con risposte a scelta multipla e potrebbe determinare nello studente un approccio allo studio di tipo mnemonico e contenutistico, un'applicazione stereotipata alla prova, che non è da sola in grado di dare un quadro chiaro della qualità degli apprendimenti. Questo tipo di prove di verifica, come vedremo, sono vantaggiose se limitate alla misurazione quantitativa di alcuni obiettivi specifici del percorso di apprendimento. Le ricerche più recenti sullo sviluppo del pensiero e sull'acquisizione della conoscenza evidenziano la complessità dell'evoluzione di questi processi e la grande varietà di variabili che ne influenzano l'andamento. Come si evince, infatti, dagli studi di psicologia dell'apprendimento e dell'istruzione della psicologa Mason [1996], i test di apprendimento (le tipiche prove a scelta multipla) sono stati concepiti in riferimento a un'impostazione teorica di tipo comportamentista, secondo cui ciò che si deve imparare

può essere suddiviso in obiettivi distinti e separati, corrispondenti ad abilità altrettanto “discrete” e scomponibili. A loro fondamento viene cioè posto un modello riduzionista e gerarchico dell’apprendimento, che contraddice nettamente le acquisizioni recenti sullo sviluppo del pensiero e la conoscenza. Altrettanto rilevante, in merito all’argomento, è il pensiero espresso da Bottani, R. [2002], ideatore nel 1988 degli Indicatori Internazionali dell’Istruzione predisposti dall’OCSE (OCSE-INES), il quale denuncia la “*dittatura dei test*”, esprimendosi in questi termini: “I modelli di valutazione sono forme di controllo, che inducono i docenti e le scuole a modulare i programmi d’insegnamento in funzione delle prove strutturate. In un regime di competizione, o di quasi mercato, tra scuole è impossibile sfuggire alla dittatura dei test, ossia alla pressione esercitata indirettamente sui docenti per adattare l’insegnamento in funzione del tipo di prove, dato che le scuole non possono permettersi di ottenere medie scadenti ai test. Tra l’altro ci si imbatte sempre nello stesso problema: di quali risultati si parla? Che cosa si vuole misurare? Non è affatto semplice delimitare le prestazioni delle scuole. A scuola, tra l’altro si imparano valori, comportamenti, atteggiamenti che non si possono facilmente misurare. Le prove standardizzate - i cosiddetti test - servono fino a un certo punto, perché i test non permettono di misurare la creatività, il pensiero critico, la perseveranza, la flessibilità di pensiero, la motivazione, l’affidabilità, l’entusiasmo, l’immagine di sé, l’autodisciplina, l’empatia, la coscienza civica, e così via”. Le scuole, quindi, spinte dal desiderio di conseguire ottimi risultati e di ottenere, di conseguenza, una posizione di rilievo nelle classifiche delle scuole di eccellenza, corrono il rischio di cadere nella spirale della “*dittatura dei test*”, andando incontro non ad un’effettiva riqualificazione, ma piuttosto a scelte non innovative, conformistiche, nozionistiche. Questo atteggiamento non può condurre ad una formazione complessa, qual è quella che la società richiede, ma solo a forme di addestramento. Un indicatore è tale se è finalizzato alla restituzione di informazioni corrette sulla qualità dell’azione didattica, evidenziano i punti di forza e di debolezza del sistema. La qualità di un sistema scolastico, nazionale o di istituto, non si può ricavare solo dai risultati ottenuti dagli alunni negli apprendimenti in quanto in tal modo si ottiene solo la descrizione dei risultati nel tempo ma non si perviene ad una misurazione del

fenomeno che possa fornirci una lettura dettagliata delle componenti storiche, sociali, culturali e contestuali specifiche che hanno influenzato il processo di insegnamento-apprendimento. La costruzione di indicatori nazionali in campo educativo riflette particolari presupposti circa le finalità dell'educazione e dettano le direzioni da prendere rispetto alle politiche educative. Analizzeremo in seguito alcuni sistemi di istruzione e valutazione nazionale di diversi stati membri dell'Unione Europea, al fine di comprendere come i modelli organizzativi ed educativi siano molto svariati e come, nella maggior parte degli stati europei, la valutazione di sistema sia affidata ad un corpo ispettivo dipendente dallo stato ed è obbligatoria e sistematica. Un indicatore può essere considerato come un rivelatore della presenza o assenza di un determinato fenomeno che Ferrario E. Ferrario [1995] così riassume: "In termini generali un indicatore dell'insegnamento può essere definito come la misurazione quantitativa degli aspetti essenziali di un sistema scolastico, misura che riveste interesse dal punto di vista della valutazione". La quantificazione della presenza o assenza di un determinato carattere rispetto ad un fenomeno osservato, tuttavia non dà informazioni sulla qualità dell'insegnamento in quanto un indicatore non può essere semplicemente sintetizzato in una quantificazione numerica, ma necessita di un'esplicazione più dettagliata che ne colga tutti gli aspetti più significativi. Un dato statistico, dunque, è un indicatore quando è in grado di fornirci informazioni. Il tasso di scolarità, il tasso di ripetenza, il tasso di abbandoni, sui quali spesso si fondano le valutazioni di sistema, rappresentano d'altra parte solo dati demografici della popolazione scolastica ma non forniscono informazioni sui risultati dell'apprendimento e sui processi ad essi legati. Quindi da sola una statistica reca una informazione limitata e che spesso assume significato solo se posta in relazione con altri indicatori. Per valutare un sistema è necessario, dunque, un insieme coerente di indicatori che diano una visione d'insieme dello stato di salute del sistema scolastico, che siano indicatori di contesto, di risultato, di processo.

A scopo esplicativo si analizzano di seguito i sistemi di valutazione di alcuni stati della Comunità Europea, ponendo particolare attenzione al sistema di valutazione italiano.

3.3 Sistemi di Istruzione e di Valutazione in Europa

3.3.1 Il sistema di valutazione in Francia

In Francia l'obbligo scolastico è di 10 anni, da 6 a 16 anni. Il sistema scolastico è così articolato:

- Scuola Primaria/Ecoles Elementares (6 - 11 anni),
- Scuola Secondaria/Collèges (11 - 15 anni),
- Liceo/Lycées (15-18 anni).

Claude Thélot, già Directeur de l'Évaluation et de la Prospective et de la Performance (DEPP), presso il Ministero dell'Educazione Nazionale Francese, attualmente "conseiller maître" della Corte dei Conti francese, nonché membro del prestigioso Istituto Nazionale di Statistica e Studi economici (l'INSEE), e ancora coordinatore i lavori della Commissione Nazionale incaricata di effettuare la rilevazione capillare delle esigenze di sviluppo e delle proposte di cambiamento del sistema scolastico francese, evidenzia che il discorso scientifico sulla valutazione è stato avviato in Francia solo una quindicina di anni fa, grazie a una forte volontà politica, ma che i Governi devono impegnare maggiori ed adeguate risorse in questo campo poiché un corretto assetto valutativo è alla base di un sistema di istruzione efficace ed efficiente. Secondo Thélot il modello più efficace di valutazione del sistema di istruzione prevede tre fasi: si cominciano a valutare i risultati degli studenti, non degli insegnanti, e l'organizzazione didattica delle scuole, e si rendono pubblici i dati; in un secondo tempo si aggiungono agli indicatori di performance anche gli indicatori di valore aggiunto, quali i progressi compiuti dalla scuola per ampliare l'offerta formativa, l'impegno nella ricerca in campo educativo, l'apertura alla sperimentazione ed all'innovazione; solo in una terza fase, sulla base della serie storica dei dati raccolti, si potranno individuare i metodi migliori di insegnamento e i migliori percorsi formativi.

Alla valutazione, in Francia, viene data una forte valenza formativa e prognostica. Commissioni di esperti, infatti, hanno lavorato, e lavorano tuttora, al perfezionamento di protocolli di valutazione che consentano ai docenti di disporre di strumenti idonei alla valutazione

del livello di preparazione degli studenti, e che li mettano in grado di “riaggiustare” i percorsi di insegnamento allo scopo di migliorare l’apprendimento là dove necessita. La Francia ha un sistema di valutazione fortemente centralizzato poiché l’autorità generale di controllo del sistema e la responsabilità amministrativa sono affidate al Ministero dell’Educazione. Nel 2005, inoltre, è stato istituito in Francia il Consiglio Superiore della Pubblica Istruzione (Haut Conseil de l’éducation), composto da nove membri nominati per sei anni. Tre dei suoi membri sono nominati dal Presidente della Repubblica, due dal Presidente dell’Assemblea Nazionale, due dal Presidente del Senato e due dal Presidente del Comitato Economico e Sociale. Il presidente del Consiglio superiore è nominato dal Presidente fra i suoi membri. L’Alta Commissione di Istruzione esprime un parere e può formulare proposte, a richiesta del Ministro della Pubblica Istruzione, in merito a questioni relative ai programmi nazionali, ai metodi di valutazione delle conoscenze degli studenti, all’organizzazione e alla formazione degli insegnanti. I suoi pareri e le proposte sono rese pubbliche. L’Alta Commissione di Istruzione dà ogni anno al Presidente della Repubblica una relazione, che viene resa pubblica, dei risultati ottenuti dal sistema di istruzione; questa relazione è trasmessa poi al Parlamento. I suoi pareri e le proposte sono rese pubbliche. L’Alta Commissione di Istruzione esprime un parere e può formulare proposte, a richiesta del Ministro della Pubblica Istruzione, in merito a questioni relative ai programmi nazionali, ai metodi di valutazione delle conoscenze degli studenti, all’organizzazione e alla formazione degli insegnanti. La valutazione di sistema è assegnata alla Direction de l’Évaluation et de la Prospective et de la Performance (DEPP), fino al 2006 chiamata DEP (Direction de l’Évaluation et de la Prospective). Nel 1989, in seguito alla legge di riforma del sistema educativo, sono state istituite autorità di controllo ministeriale con specifici compiti: l’IGEN (Inspection Générale de l’Éducation Nationale) e l’IGAENR (Inspection Générale de l’Administration de l’Éducation Nationale et de la Recherche). La DEPP elabora e realizza il programma di valutazione di sistema e verifica la validità delle scelte politiche effettuate dal Ministero. Esso conduce, inoltre, in collaborazione con altri centri di ricerca, lavori di progettazione e gestione del sistema a breve, medio e lungo termine. Tutti i risultati raccolti nelle indagini di sistema vengono organizza-

ti, dalla DEPP stessa, in una grande banca dati e diffusi attraverso pubblicazioni. La DEPP partecipa anche ai progetti europei ed internazionali finalizzati a comparare le modalità di funzionamento e i risultati di apprendimento conseguiti nei diversi sistemi educativi. L'IGEN è composto da insegnanti che abbiano conseguito il Diploma di Dottorato, nominati attraverso decreto governativo direttamente dal Ministero, dietro proposta di uno specifico Comitato formato da membri dei servizi di Ispezione Generale e da insegnanti universitari. I membri dell'IGEN, raggruppati in team permanenti e specializzati, sono coordinati dal decano dei servizi generali. Ogni team si occupa di uno specifico ambito disciplinare (Storia e Geografia, Matematica, Lingue moderne, ...), inoltre è previsto un team per la scuola primaria e uno per le altre Istituzioni educative. Questo corpo valutativo ha il compito di valutare il personale ispettivo, direttivo, docente, educativo, e di predisporre specifiche attività di formazione. Partecipa alla valutazione globale del sistema analizzando, in particolar modo, i programmi, le metodologie didattiche, le strategie educative, le risorse ed i risultati ottenuti dagli allievi negli apprendimenti. L'IGAENR, guidato da un ispettore generale nominato dal Ministero, controlla e coordina il lavoro dell'IGEN e pubblica un rapporto generale annuale dei lavori svolti, includendo indicazioni per il buon proseguo dei lavori. Dal 1990, poi, sono stati istituiti corpi ispettivi territoriali, i Consigli di Valutazione, i cui membri sono dirigenti e docenti nominati dal Ministero per concorso, che hanno il compito di effettuare ispezioni nelle singole scuole. Il sistema di valutazione francese, pur riconoscendo l'importanza della coesistenza di forme di valutazione esterna ed interna, per il carattere fortemente centralizzato del sistema di istruzione, ha compiuto grandi progressi nel primo campo ma necessita ancora di ricerche e sperimentazioni relative ad esperienze di autovalutazione di istituto. La valutazione esterna è affidata quindi all'IGEN, all'IGAENR e alla DEPP, che sono interni al Ministero, ed ai Consigli di Valutazione, costituiti da docenti, capi d'istituto, esperti del settore, politici e sindacalisti, che sono indipendenti. Ogni scuola viene periodicamente sottoposta ad un'ispezione mirata a valutare la realizzazione delle politiche educative, l'organizzazione scolastica, i risultati degli alunni negli apprendimenti, la collaborazione all'interno dei team docenti, il clima pedagogico della scuola e i rapporti scuola-

famiglia. La valutazione esterna, oltre ad ispezionare le diverse scuole per valutare la qualità, l'efficienza e l'efficacia dell'offerta formativa, ha il compito di valutare gli insegnanti. La valutazione degli insegnanti è affidata, infatti, agli IEN (Inspecteurs de l'Éducation Nationale), per la scuola primaria, e agli IA-IPR (Inspecteurs d'Académie), per il personale dirigente, per i direttori dei centri di orientamento e per gli insegnanti di scuola secondaria. Queste istituzioni, che funzionano a livello regionale, redigono periodici rapporti sui singoli docenti che vengono presentati ad un recteur, un Ministro Regionale dell'istruzione. Questi enti, quindi, hanno il compito di valutare le prestazioni degli insegnanti e di svolgere un ruolo di consulenza e programmazione della formazione in servizio degli stessi. L'ispezione si concentra sull'attività svolta dal singolo insegnante, e si basa sull'analisi del materiale prodotto dal docente, sull'osservazione in classe e su colloqui diretti. La relazione finale scritta è consegnata al recteur, alla scuola di appartenenza e al docente interessato e può influire sulla retribuzione economica e sulla progressione di carriera. Vengono svolte ispezioni anche sulla qualità dell'insegnamento di materie specifiche e di temi di interesse particolarmente rilevante, come ad esempio le modalità relazionali del gruppo docente, che vengono selezionati in relazione alla nascita di un particolare bisogno. La selezione iniziale degli insegnanti è, comunque, compito del Ministero dell'Istruzione.

3.3.2 Il sistema di valutazione in Inghilterra

In Inghilterra, come in Galles e in Scozia, l'istruzione obbligatoria ha una durata di 11 anni, da 5 a 16 anni. Il sistema scolastico è così articolato:

- Scuola Primaria/Primary Schools (5-11 anni),
- Scuola Secondaria/Secondary Schools (11-14, 14-16),
- Scuola Superiore/Upper secondary education (16-18 anni) non obbligatoria.

Il sistema di controllo del servizio scolastico riguarda tutte le scuole finanziate con fondi statali: asili nido, scuole dell'infanzia, scuole

primarie, scuole secondarie, scuole per alunni con bisogni educativi speciali. La valutazione del sistema scolastico nel Regno Unito, dall'asilo nido alla scuola superiore, nonché dal 2000 agli enti locali di formazione professionale per gli adulti, è affidata all'OFSTED, Office for Standards in Education, un'agenzia autonoma, guidata dall'Ispettore scolastico capo di Sua Maestà, il cui referente principale è il Ministero dell'Educazione. L'OFSTED, inoltre, dal 2004 controlla anche le modalità con cui gli enti locali (Las) assolvono le loro funzioni di finanziamento e supporto al sistema educativo, fornendo una sintesi annuale della struttura delle ispezioni all'Ispettore scolastico capo di Sua Maestà, che a sua volta redige un rapporto annuale agli organi parlamentari. La funzione dell'istituto è quella di stabilire gli standards didattici e formativi e di monitorare, attraverso ispezioni, i diversi plessi scolastici per valutarne la qualità. L'OFSTED è coadiuvato dalle LEAs (Local Education Authorities), che sono, invece, autorità educative locali direttamente responsabili delle scuole che hanno una gestione locale, cioè senza finanziamenti statali. Per quanto riguarda il ruolo delle LEAs, si può dire che la loro azione è regolata dal Code of Practice on LEA-School Relations (2001), che stabilisce le modalità di monitoraggio e di rilevamento delle prestazioni delle singole scuole. Non esiste un dovere ispettivo ma, piuttosto, un'azione di supporto alle scuole e al lavoro dell'OFSTED, in vista di una crescita qualitativa degli standard didattici e formativi. Le LEAs, a tale scopo, hanno sviluppato gli Education Development Plans (EDP), piani strategici triennali rivolti al miglioramento degli standard educativi. L'ispezione delle singole scuole, effettuata da un'équipe di specialisti, è prevista per legge ed avviene almeno ogni sei anni. L'azione ispettiva si propone di raggiungere i seguenti obiettivi: Informare la scuola, le famiglie, la comunità locale circa la qualità educativa della singola scuola e il livello di raggiungimento degli standard previsti dall'OFSTED; Aiutare la scuola ad elaborare un piano per il miglioramento delle aree considerate deboli o critiche; Fornire, ai decisori politici, informazioni generali circa l'uso delle risorse finanziarie e lo stato e l'evoluzione del sistema scolastico. In ogni singola scuola vengono analizzati aspetti organizzativi e didattici quali, ad esempio, il tempo dedicato all'istruzione, il comportamento mirato o tattico dell'insegnante, le strategie e i modelli di insegnamento e le modalità di relazione insegnante-allievo.

Per quanto concerne gli aspetti relativi all'organizzazione, si dà grande rilevanza all'analisi della leadership, del coinvolgimento degli insegnanti nelle decisioni, del coinvolgimento dei genitori, nelle decisioni del quadro di lavoro, nella definizione dei ruoli e delle responsabilità. Altro aspetto sottoposto ad analisi è quello relativo alla cultura della scuola: la riflessione sui valori e le regole che si danno i membri della scuola, i comportamenti assunti, le attese, la presenza di un clima tranquillo che offra supporto e sostegno, la condivisione del sistema dei valori, l'auto-stima degli insegnanti, il loro impegno e la collaborazione tra essi. Gli ispettori dell'OFSTED, che si fermano nella scuola esaminata almeno una settimana, raccolgono le informazioni principali sulla scuola, quali il numero degli alunni, le modalità organizzative, il curriculum, ecc. e utilizzano diverse fonti di informazione: lo school self-audit, elaborato dalla scuola stessa, sulla base del Frame work for Inspection; le informazioni fornite dal Headteacher che consentono di confrontare le prestazioni degli alunni con i dati nazionali; i risultati della self-evaluation di istituto. Durante l'ispezione l'équipe osserva le lezioni (almeno il 60% del tempo dedicato all'ispezione) e alcuni dei lavori svolti dagli alunni (a campione, su determinate materie), effettua interviste e discussioni con gli alunni sulla comprensione e la conoscenza di materie specifiche e di aree del curriculum, somministra un questionario anonimo per i genitori sul livello di soddisfazione della scuola e della preparazione del proprio figlio. Gli ispettori completano il loro lavoro discutendo delle modalità organizzative con il dirigente scolastico e lo staff della scuola ed esaminando i piani di lavoro degli insegnanti. Al termine dell'ispezione, viene organizzata una riunione con gli insegnanti e il capo d'istituto allo scopo di restituire, oralmente, i primi risultati. Il Registered Inspector (RgI), in seguito, procede alla redazione della relazione finale che contiene: Una descrizione delle principali caratteristiche della scuola; Un'analisi del contesto; Un giudizio sull'efficacia dell'attività educativa della scuola, attraverso l'analisi degli standard e degli obiettivi didattici e formativi raggiunti; Un giudizio sulla qualità della leadership e del management; I risultati della qualità educativa riscontrata nel corso della valutazione ispettiva; Il confronto tra i livelli degli standard raggiunti con quelli previsti a livello nazionale; L'analisi della gestione e dell'uso delle risorse finanziarie; La valutazione delle azioni finalizzate

allo sviluppo spirituale, morale, sociale e culturale degli alunni. Lo scopo del rapporto è quello di delineare i punti di forza e i punti di debolezza della scuola e di formulare eventuali proposte per un piano di miglioramento. Il rapporto ispettivo è diviso in tre parti e presenta inizialmente una breve carta d'identità della scuola (le informazioni sull'istituto, le caratteristiche principali della popolazione studentesca, ecc.). La prima parte della relazione delinea sinteticamente i punti di forza e di debolezza che emergono dalla valutazione complessiva della scuola quali i risultati raggiunti, la qualità educativa riscontrata, le modalità di direzione e di gestione, il punto di vista dei genitori e degli studenti, i miglioramenti necessari da apportare per elevare i livelli e gli standard di insegnamento e di apprendimento, i miglioramenti da operare a livello gestionale, organizzativo e didattico. La seconda parte si sofferma in modo analitico sui risultati della valutazione circa i percorsi compiuti dagli studenti, gli standards raggiunti nell'apprendimento, nelle materie e nei corsi, in rapporto al livello nazionale. In questa parte si descrive la qualità educativa riscontrata nelle modalità di insegnamento e di apprendimento (Teaching and learning), l'attuazione del curriculum, la costruzione del senso civico e il rispetto delle regole (Care), l'azione di guida e di supporto svolta nei confronti degli studenti (Guidance and support), le relazioni con i genitori, con le altre scuole e con il territorio (Partnership with parents, other schools and the community), ed infine, l'azione del dirigente e dello staff insegnante e tecnico. La terza parte della relazione si sofferma sui risultati ottenuti nelle diverse materie presenti all'interno del corso di studi (The quality of education in subjects and courses). Per svolgere un lavoro così sistematico e dettagliato, come quello appena descritto, l'OFSTED deve disporre di un'équipe di esperti specializzati in settori diversi. Tra essi troviamo: il dirigente, il team esperto di strumenti dell'informazione e della comunicazione tecnologica, il team esperto di gestione ed organizzazione scolastica, il team di psicologi, il team esperto di scienze, di musica, di tecnologia, di attività motoria, di storia, di speciali bisogni educativi, di lingua, di matematica, di arte e immagini, di geografia. La documentazione prodotta dall'OFSTED è, in genere, pubblica e disponibile sulla rete internet, tuttavia la graduatoria dei punteggi ottenuti, per ogni aspetto o area, dalle singole scuole, non è accessibile a tutti. Dal momento

in cui la scuola riceve la relazione finale, gli organi decisionali della scuola, entro 40 giorni, elaborano e predispongono un piano di azione e di miglioramento. Le scuole che hanno riportato risultati particolarmente negativi vengono tenute sotto stretta osservazione e sottoposte a frequenti controlli da parte dell'OFSTED. Le scuole considerate in difficoltà sono classificate secondo queste tre categorie: Scuola non in grado di fornire standards accettabili e che richiede misure speciali; Scuola che fornisce standards accettabili ma presenta, tuttavia, aree di pesante criticità e debolezza; Scuola che non presenta gravi punti di debolezza, ma viene giudicata *underachieving*, cioè al di sotto del livello di quello che la scuola potrebbe fare. Per le scuole in difficoltà esiste un apposito servizio di sostegno gestito dal SIP, School Improvement Partner, che agisce in qualità di "amico critico" aiutando la dirigenza della scuola a valutare le prestazioni, a identificare le priorità e a progettare cambiamenti efficaci. Altro aspetto della valutazione, che è molto rigoroso nel Regno Unito, è quello legato alla valutazione degli insegnanti. Questi ultimi, per insegnare, devono essere in possesso della certificazione del QCA (Qualifications and Curriculum Authority), l'autorità che, nel Regno Unito, vigila sulla correttezza delle ammissioni nella pubblica amministrazione. Il compito di questo istituto, oltre ad essere quello di fornire una banca dati dei curricula britannici in ogni settore, suggerisce ai sudditi di Sua Maestà quali percorsi di studi intraprendere, a quali imprese rivolgersi per trovare un lavoro, cosa scrivere sui propri curricula e come organizzarli. È al QCA che le scuole si rivolgono per la ricerca del personale che, anche dopo l'assunzione è sottoposto a valutazione sistematica. Gli insegnanti sono, quindi, nominati direttamente dalle scuole con procedure di selezione ed assunzione competitive.

3.3.3 Il sistema di valutazione in Germania

Il sistema scolastico è così articolato:

- Scuola Primaria/Grundschule (5/6 - 10/12 anni),
- Scuola Secondaria: Hauptschule e Realschule (11/13-16 anni),
- Gymnasium (11/13 - 19 anni)

L'istruzione scolastica in Germania è determinata dalla struttura federalista dello Stato. La Costituzione, infatti, stabilisce che la determinazione dei diritti e dei doveri in materia scolastica spetti alle Länder (regioni) che hanno piena autonomia nell'emanazione delle leggi in materia di organizzazione e valutazione scolastica. Lo Stato interviene solo come supervisore. L'autorità suprema è rappresentata dai Ministri dell'Educazione e degli Affari Culturali delle diverse Länder che controllano l'amministrazione dell'istruzione. Gli stessi stabiliscono le direttive in materia di politica sociale, controllano gli organismi, gli Enti e gli istituti sotto la loro autorità. L'amministrazione del sistema scolastico è ripartita su tre livelli:

- livello superiore: Ministeri dell'Educazione e degli Affari culturali delle singole Länder;
- livello intermedio: Dipartimenti Scolastici dei Governi Regionali;
- livello inferiore: gli Uffici Scolastici delle autorità locali.

Lo Stato controlla anche le scuole private che devono conformarsi per qualità di insegnamento ed insegnanti alle normative ufficiali sancite dai Länder di appartenenza territoriale. L'ente preposto al coordinamento delle attività delle singole Länder è la *Ständige Konferenz der Kulturminister der Länder in der Bundesrepublik Deutschlands* (Conferenza permanente dei ministri dell'Educazione e degli affari culturali delle Länder nella repubblica Federale di Germania). Tale Conferenza ha, come funzione primaria, la rappresentazione degli interessi generali per giungere alla definizione di posizioni comuni. L'obiettivo principale è la creazione e la garanzia di una concordanza sulle normative emesse dalle Länder. Tutte le decisioni devono essere prese all'unanimità e hanno valore solo di raccomandazione e i Ministri si impegnano affinché i parlamenti dei diversi Länder le adottino. L'intero sistema scolastico della Germania, quindi, è sottoposto al controllo dello stato, ma allo stesso tempo si possono registrare anche notevolissime differenze fra le diverse Länder, cioè ogni regione ha il potere di modificare, in parte, o di applicare diversamente il sistema scolastico, questa autonomia ha consentito di programmare nelle diverse regioni differenziazioni della durata delle varie fasi dell'istruzione obbligatoria. Allo scopo, comunque, di trovare accordi

comuni ed evitare di adottare provvedimenti che possono essere in conflitto o ostacolarsi a vicenda, i ministri dell'educazione delle Länder collaborano non solo con il governo centrale ma anche fra loro. In Germania, ogni regione sviluppa un proprio programma educativo e stabilisce i propri obiettivi formativi definendo, in tal modo, i propri standard. Sebbene quest'ultimi siano mediamente molto alti, possono variare sensibilmente tra le varie Länder. Al di là di tali differenze, comunque, nelle scuole tedesche i bambini vengono valutati su una serie di materie e ricevono annualmente due pagelle (una alla fine di ogni semestre). I voti vanno dall'1 al 6 secondo il seguente schema:

1. sehr gut (ottimo)
2. gut (buono)
3. befriedigend (discreto)
4. ausreichend (sufficiente)
5. mangelhaft (mediocre)
6. ungenügend (insufficiente)

In riferimento al processo valutativo, il sistema educativo tedesco è piuttosto severo, infatti, se un alunno non raggiunge la sufficienza in due o più materie, riportando il voto di 5 o 6, potrebbe dover ripetere l'intero anno scolastico anche nella scuola primaria, sebbene le regole varino in ogni regione. Poiché lo Stato federale collabora con le Länder nell'esercizio di certe competenze è stata creata la *Bund-Länder-Commission für Bildungsplanung und Forschungsförderung* (Commissione Nazionale per la Pianificazione dell'Istruzione e della Ricerca) che rappresenta un forum permanente. Essa è competente nelle seguenti attività: preparazione, realizzazione, valutazione e finanziamento di esperienze pilota di portata trans regionale. Le esperienze pilota sono esperienze atte a favorire lo sviluppo dell'istruzione e la sperimentazione di nuovi approcci pedagogici e/o organizzativi. Gli organi preposti alla supervisione di queste esperienze sono i *Landesinstitute für Schulpädagogik oder Bildungsforschung* (Istituti di Ricerca Pedagogica). Essi dipendono dal Ministero dell'Educazione e degli Affari culturali e hanno come compiti: la messa a disposizione di programmi,

servizi di sostegno e materiali didattici, la valutazione delle riforme per una loro efficace realizzazione. L'autorità di controllo scolastico esercita una supervisione nei confronti degli insegnanti e del capo d'istituto, per verificare il loro operato. Gli insegnanti vengono valutati periodicamente al fine di progredire nella loro carriera professionale. L'accertamento si basa su visite alle classi da parte del capo d'istituto e dell'ispettore scolastico, su colloqui individuali con gli insegnanti e sulla verifica degli apprendimenti degli studenti.

3.3.4 Il sistema di valutazione in Spagna

L'obbligo scolastico in Spagna è di 10 anni, da 6 a 16 anni. Il sistema scolastico è così articolato:

- Scuola Primaria/ Educacion Primaria (6-12 anni),
- Scuola Secondaria/ Educacion Secundaria (12-16 anni),
- Scuola Superiore/Bachillerato (16-18 anni),
- Formazione Professionale/ Formación profesional (16-20 anni)

In Spagna, che ha avuto una lunga storia di centralizzazione del potere, il cammino di decentralizzazione è stato molto lento. Nel 1990 la riforma della scuola varata dal socialista Gonzales ha lasciato alle autorità statali solo la definizione delle direttive curriculari di base, vale a dire gli insegnamenti minimi che hanno il fine di garantire agli alunni la validità del titolo. Nel 2002 sono stati ampliati ulteriormente i poteri delle autorità scolastiche locali poiché ad esse è stato affidato il compito del reclutamento degli insegnanti e della valutazione di sistema. A partire dal 1990 l'autovalutazione d'istituto è obbligatoria in tutte le scuole al fine di verificare la validità delle decisioni assunte nell'esercizio della propria autonomia ed ogni istituto è libero di definire i propri criteri e le proprie procedure autovalutative. Il compito valutativo è assegnato al Consiglio di Scuola, che deve valutare i progetti curriculari ed extracurriculari, l'organizzazione didattica e i risultati conseguiti negli apprendimenti dagli studenti. La responsabilità di valutare il progetto curricolare e i processi di insegnamento spetta all'Assemblea degli Insegnanti. La valutazione del sistema educativo è gestita

dal *Ministero de Educación y Ciencia* (MEC) attraverso l'IE (Instituto de Evaluación), il quale risponde al Ministero e lavora in collaborazione con le Comunità Autonome, le regioni spagnole, che a loro volta sono responsabili dei compiti valutativi sul territorio. La valutazione di sistema avviene, di norma, ogni quattro anni. I risultati della valutazione esterna sono notificati al Consiglio di Scuola e, tramite il Capo d'Istituto, comunicati all'Assemblea degli Insegnanti. Le competenze relative all'educazione sono distribuite, quindi, secondo un modello decentralizzato che contempla la partecipazione delle Comunità Autonome, delle Amministrazioni Locali e dei Centri Scolastici. Secondo questo modello spetta al governo centrale l'organizzazione generale del sistema, la definizione del curriculum nei suoi elementi essenziali, la cooperazione internazionale in ambito educativo, il fomento della ricerca, la regolamentazione delle qualifiche, l'amministrazione degli istituti all'estero, il controllo e la valutazione del sistema e della sua applicazione. Spettano invece alle Comunità Autonome le responsabilità amministrative relative alle rispettive aree di giurisdizione, l'organizzazione delle istituzioni locali, la gestione del personale, la pianificazione di progetti educativi, l'orientamento ed il finanziamento dei sussidi. Le amministrazioni locali sono invece responsabili delle strutture e del mantenimento delle scuole infantili e primarie, della progettazione di attività supplementari ed extra-curricolari oltre che del monitoraggio degli alunni in età d'obbligo scolastico. L'Istituto nazionale di valutazione l'IE (Instituto de Evaluación), prima del 2006 denominato INECSE (*Instituto Nacional de Evaluación y Calidad del Sistema Educativo*), è responsabile per la progettazione e lo sviluppo del sistema di indicatori, ai quali lavora in collaborazione con le Comunità Autonome. L'IE, inoltre, cura la partecipazione della Spagna alle valutazioni internazionali attraverso lo sviluppo e l'analisi di indicatori internazionali. Il sistema di indicatori è strutturato su indicatori educativi essenziali, ritenuti validi per descrivere e comprendere la realtà del sistema di istruzione, per consentire la definizione di obiettivi formativi e per le politiche più adeguate per il raggiungimento degli obiettivi stessi. Il Ministero della Pubblica Istruzione, Sport e Politica sociale stabilisce le linee guida del sistema di valutazione che si articola in valutazione globale, valutazione periodica e valutazione formativa. Le Comunità Autonome, nel rispetto delle linee guida del Ministero e

dello IE, redigono una loro “legislazione” per la valutazione che tiene conto delle caratteristiche locali. La Spagna è impegnata a partecipare al processo di miglioramento del sistema d’istruzione, in linea con gli obiettivi dell’Unione Europea fissati per il 2010. A tal fine l’IE pubblica annualmente relazioni dettagliate sui progressi conseguiti e sugli ostacoli che restano da superare. Lo IE esamina ed analizza in particolar modo il livello di istruzione della popolazione tra i 25 e i 64 anni, la spesa sostenuta per l’istruzione, il numero di iscritti ai diversi ordini di scuola, il numero di studenti che accedono all’istruzione superiore, il numero degli studenti stranieri integrati nelle classi e la loro distribuzione geografica, la partecipazione nella popolazione adulta ad attività di istruzione o di formazione professionale permanente, l’acquisizione degli apprendimenti degli studenti in scienze, lettura e matematica secondo gli indicatori PISA 2006, la percentuale di dispersione scolastica, la percentuale degli studenti che conseguono una laurea. Il Ministero della Pubblica Istruzione, Sport e Politica sociale rendere accessibili le tutte le informazioni offerte dal sistema di valutazione, in modo da soddisfare l’impegno per la trasparenza.

Occorre precisare che la valutazione degli insegnanti in Spagna è prevista solo su base volontaria e gli esiti possono essere tenuti in conto per lo sviluppo della carriera professionale insieme ad altri elementi, come la formazione, la ricerca e l’innovazione.

3.4 Autonomia scolastica e valutazione di sistema in Italia

In tutti i paesi occidentali, come abbiamo visto, i sistemi di istruzione si sono attivati per predisporre batterie di verifica degli apprendimenti finalizzate a verificare il livello di preparazione conseguito dagli alunni in alcune discipline ritenute fondanti, quali la lingua madre, la matematica e le scienze. Sono stati, inoltre, sottoposti a valutazione i sistemi di istruzione nel loro complesso per fare il quadro della qualità dell’istruzione, valutando l’organizzazione amministrativa e didattica del sistema, la spesa sostenuta per l’istruzione, gli investimenti per l’innovazione e la ricerca e svariati altri parametri. L’approccio usato si fonda sostanzialmente su indagini statistiche, predisposte ed utilizzate per prendere decisioni politiche in merito al sistema di istruzione.

In un paese democratico, tuttavia, è nell'interesse della comunità intera conoscere il livello di qualità del sistema scolastico, poiché la formazione dell'individuo è alla base della crescita sociale, culturale ed economica di tutto il paese. Negli ultimi quindici anni è nata l'esigenza di disporre di informazioni più accurate, più ampie e più fondate scientificamente sul funzionamento dei sistemi scolastici. La sola rilevazione a campione finalizzata alla valutazione degli apprendimenti non risulta, infatti, essere sufficiente per valutare un sistema così complesso qual è quello scolastico. Non si possono, quindi, che condividere le considerazioni degli esperti e dei politici che hanno individuato anche in Italia la necessità di istituire una struttura, il Sistema Nazionale di Valutazione, appunto, assegnandole il compito di fornire informazioni significative sulla qualità sistema scolastico.

Tra la fine degli anni 80 e i primi degli anni 90 del Novecento, grazie alla diffusione dell'autonomia scolastica, si è avvertita anche in Italia la necessità di disporre di un servizio nazionale che, analogamente a quanto avveniva in altri stati, potesse valutare la produttività e l'efficacia del sistema scolastico nel suo insieme e dei singoli istituti scolastici. Fino agli inizi degli anni 90 in Italia c'era un solo istituto che si occupava di valutazioni del profitto scolastico e delle competenze degli studenti su larga scala: il CEDE (Centro Europeo dell'Educazione), insediato a Villa Falconieri a Frascati. Per decenni questo centro è stato diretto dal prof. Aldo Visalberghi che, con la collaborazione di un'équipe di specialisti in psicometria, ha coordinato la partecipazione dell'Italia alle principali indagini internazionali della IEA (International Association for the Evaluation of Educational Achievement). La IEA è un'associazione indipendente, senza scopo di lucro, di centri di ricerca nel campo delle Scienze dell'Educazione (circa 53 paesi), fondata nel 1958, con sede ad Amsterdam, il cui scopo è quello di condurre ricerche comparative internazionali nel campo della valutazione dei risultati scolastici.

Gli specialisti italiani di valutazione su larga scala del profitto scolastico si sono in un certo senso auto-formati in seno al CEDE, partecipando ai progetti internazionali di valutazione comparata[Bottani, N. e Cenerini, A., 2003]. Nonostante l'impegno di Visalberghi e dei suoi collaboratori, questa presenza italiana nelle indagini internazionali di valutazione ha suscitato scarso interesse nella classe politica

italiana, nei sindacati e nei media. I risultati degli studenti italiani nelle prove internazionali che si sono susseguite dal 1964 in poi non sono stati, infatti, mai sufficientemente pubblicizzati, al punto che spesso non ne erano a conoscenza neanche gli insegnanti stessi. Agli inizi degli anni 90, di conseguenza, la cultura della valutazione di sistema, delle scuole e delle conoscenze degli studenti in Italia era molto bassa in tutti gli ambienti che si occupavano, direttamente o indirettamente, di educazione e formazione. Fu proprio nel 1990 che l'allora Ministro della Pubblica Istruzione, Sergio Mattarella, definì due linee strategiche che l'Italia avrebbe dovuto perseguire per equipararsi agli altri paesi europei nel campo dell'istruzione: l'Autonomia delle istituzioni scolastiche e la Costruzione di un sistema nazionale di valutazione.

In quello stesso anno, infatti, da un rapporto elaborato dal CENSIS, emerse che Stati Uniti, Regno Unito, Francia, Svezia ed Olanda, avevano già predisposto un accurato servizio nazionale di valutazione. L'Italia dovette impegnarsi quindi a colmare il gap rispetto agli altri paesi. Solo nel 1993, però, il Ministero della Pubblica Istruzione nominò una commissione, presieduta da Visalberghi, con il compito di studiare un modello di sistema di valutazione. Tra le proposte operate dalla commissione il Ministero decise di istituire un Comitato tecnico-scientifico per definire le linee guida del sistema di valutazione. Soltanto nel maggio del 1997 fu emanata la direttiva n. 307, che assegnava al CEDE, sotto la guida di Benedetto Vertecchi, la realizzazione di un Servizio Nazionale per la Qualità dell'Istruzione (SNQI). Il SNQI diede, inoltre, l'avvio all'Archivio Docimologico per l'Autovalutazione delle Scuole, finalizzato alla creazione di un database al quale le scuole potessero rivolgersi per scegliere gli items da somministrare agli alunni al fine di valutare il livello di preparazione negli apprendimenti. A tutte le ricerche condotte fino a questo momento fu mossa, però, l'accusa di rimanere solo a livello teorico e di non avere una reale ricaduta sul sistema di istruzione.

Nell'autunno del 1997, il Ministero della Pubblica Istruzione, Luigi Berlinguer, sottopose i suoi progetti di riforma della scuola ad un gruppo di esperti internazionali designati dall'OCSE (Organisation for Economic Co-operation and Development), l'organizzazione di cooperazione e sviluppo economici con sede a Parigi che è specializ-

zata nella valutazione comparata delle politiche nazionali nel campo sociale, della scienza, della ricerca, dell'ambiente e dell'educazione. Alla fine dell'esame, il gruppo di esperti presentò una propria relazione che conteneva una serie di raccomandazioni rivolte al Ministro della Pubblica Istruzione italiana. Della relazione intitolata *Esami delle politiche nazionali dell'istruzione- Italia*, riportiamo un breve stralcio che ne coglie i punti più significativi: "... per il momento la struttura disegnata nella direttiva n. 307 non ha una precisa corrispondenza in altri Paesi europei, che hanno seguito diversi approcci nello stabilire i rapporti tra autorità e sistema di valutazione. In alcuni paesi, il compito della valutazione dell'istruzione è affidato ad agenzie collegate al Parlamento o al Governo, ma che sono indipendenti nel determinare i programmi di intervento. È il caso, per quanto riguarda la Gran Bretagna del Qualifications and Curriculum Authority (QCA, precedentemente SCAA), e dell'OFSTED (Office for Standards in Education) o dello Skolverket (Agenzia Nazionale) in Svezia. Altrove, la valutazione viene affidata a servizi governativi interni al Ministero o, in alcuni casi, a un corpo di professionisti incaricati di un'attività di monitoraggio continuo degli standard educativi e del funzionamento del sistema. Esistono anche modelli intermedi, come l'Institut Nacional de la Calidad y Evaluación (INCE) spagnolo, che dipende dal Ministero dell'educazione, ma il cui comitato direttivo è costituito in maggioranza dai rappresentanti delle comunità autonome, o come il Comitato di Valutazione all'interno del SOED (Scottish Office of Educacion) scozzese, dove sono rappresentati i vari attori del sistema educativo". "... il Ministero potrebbe avere maggior forza se il sistema di valutazione fosse indipendente. Vorremmo, per questa ragione, raccomandare che al sistema di valutazione sia accordata una maggiore indipendenza rispetto a quella che ora vediamo nei documenti che abbiamo citato, e ciò specialmente per garantire un flusso continuo di valutazione oggettiva".[OCSE, 1998] A questo giudizio seguirono una serie di raccomandazioni:

- Istituzione di un sistema di valutazione indipendente, la cui attività sia focalizzata sulla definizione di parametri di valutazione tali da mettere le scuole nella condizione di auto-valutarsi;
- Istituzione di un ente indipendente incaricato di svolgere ricerche

indipendenti in materia di istruzione per garantire un parere indipendente sul funzionamento del sistema formativo;

- Coinvolgimento degli ispettori ministeriali della Pubblica Istruzione nel programma di miglioramento delle scuole e nella valutazione dei risultati;
- Creazione di un sistema di testing per valutare gli alunni in determinati momenti del corso di studi o in determinate classi, specialmente al termine della scuola dell'obbligo;
- Pubblicizzazione dei risultati alle singole scuole, ai genitori e alla comunità, al fine di diffondere la conoscenza di quelle che sono risultate le migliori pratiche.

I lavori del Ministero della Pubblica Istruzione sui progetti di riforma andarono avanti e, con il decreto legislativo n.258 del 20 luglio 1999, il CEDE viene trasformato in Istituto Nazionale per la Valutazione del Sistema dell'Istruzione - INValSI, ora Servizio Nazionale di Valutazione del Sistema Educativo di Istruzione e di Formazione. All'INValSI vengono attribuiti i seguenti compiti:

- Effettuare verifiche periodiche e sistematiche sulle conoscenze e abilità degli studenti e sulla qualità complessiva dell'offerta formativa delle istituzioni di istruzione e di istruzione e formazione professionale, anche nel contesto dell'apprendimento permanente;
- Studiare le cause dell'insuccesso e della dispersione scolastica con riferimento al contesto sociale ed alle tipologie dell'offerta formativa;
- Effettuare le rilevazioni necessarie per la valutazione del valore aggiunto realizzato dalle scuole;
- Predisporre annualmente i testi della nuova prova scritta, a carattere nazionale, volta a verificare i livelli generali e specifici di apprendimento conseguiti dagli studenti nell'esame di Stato al terzo anno della scuola secondaria di primo grado;

- Predisporre modelli da mettere a disposizione delle autonomie scolastiche ai fini dell'elaborazione della terza prova a conclusione dei percorsi dell'istruzione secondaria superiore;
- Provvedere alla valutazione dei livelli di apprendimento degli studenti a conclusione dei percorsi dell'istruzione secondaria superiore, utilizzando le prove scritte degli esami di Stato secondo criteri e modalità coerenti con quelli applicati a livello internazionale per garantirne la comparabilità;
- Fornire supporto e assistenza tecnica all'amministrazione scolastica, alle regioni, agli enti territoriali, e alle singole istituzioni scolastiche e formative per la realizzazione di autonome iniziative di monitoraggio, valutazione e autovalutazione;
- Svolgere attività di formazione del personale docente e dirigente della scuola, connessa ai processi di valutazione e di autovalutazione delle istituzioni scolastiche;
- Svolgere attività di ricerca, sia su propria iniziativa che su mandato di enti pubblici e privati;
- Assicurare la partecipazione italiana a progetti di ricerca europea e internazionale in campo valutativo, rappresentando il Paese negli organismi competenti;
- Formulare proposte per la piena attuazione del sistema di valutazione dei dirigenti scolastici, definisce le procedure da seguire per la loro valutazione, formula proposte per la formazione dei componenti del team di valutazione e realizza il monitoraggio sullo sviluppo e sugli esiti del sistema di valutazione.

Non sembra che i suggerimenti dati dagli esperti dell'OCSE siano stati pienamente presi in considerazione se si considera, prima di tutto, la non indipendenza del sistema di valutazione rispetto al Ministero della Pubblica Istruzione, dell'Università e della Ricerca. I compiti assegnati risultano, inoltre, essere eccessivi ed estremamente diversificati in relazione al numero dei componenti l'organigramma dell'istituto di valutazione. A tal proposito è necessario sottolineare che l'INValSI

si avvale di un gran numero di servizi e di consulenze esterne all'istituto stesso, sia per quanto riguarda la ricerca, sia per l'elaborazione e l'analisi dei dati delle prove di valutazione somministrate agli alunni.

Nel 1999 con il DPR n. 257 il Ministero della Pubblica Istruzione stabilì i metodi e le scadenze per rilevazioni periodiche, finalizzate alla verifica del raggiungimento degli standard di qualità fissati dal Ministero stesso. Nel 2001 il Ministro Moratti insediò una commissione di lavoro, presieduta da G. Elias, con il compito di mettere a punto una proposta di Servizio Nazionale di Valutazione. A seguito nacque un Progetto Pilota mirato alla valutazione degli apprendimenti della lingua italiana e della matematica e alla verifica del grado di attuazione del Piano dell'Offerta Formativa elaborato dalle scuole. Il primo anno di attuazione del Progetto Pilota fu l'anno scolastico 2002/2003: in 2500 scuole furono somministrati, alla classe V della Scuola Primaria, alla III della Scuola secondaria di Primo Grado e alla II della Scuola Secondaria di Secondo Grado, items specifici di lingua italiana e matematica. A questo progetto seguirono nei due anni scolastici successivi il Progetto Pilota 2 e il Progetto Pilota 3, utilizzando le stesse modalità di campionamento e di somministrazione. Nell'anno scolastico 2005/2006 la valutazione fu somministrata con la stessa tipologia di prove ma fu estesa obbligatoriamente a tutte le scuole d'Italia. Vista l'entità di lavoro ed organizzazione che comportò questa esperienza di valutazione di tipo censuario, nell'anno scolastico 2006/2007 si tornò alla somministrazione a campione e furono modificate le classi da valutare: la II e V classe della Scuola Primaria, la I e la III della Scuola Secondaria di Primo Grado e la II e la V classe della Scuola Secondaria di Secondo Grado. Nel 2004 il Consiglio dei Ministri ha approvato, con il DL 286, lo schema di decreto legislativo attuativo della legge 53/2003 riguardante la "definizione delle norme generali nell'istruzione e dei livelli essenziali delle prestazioni in materia di istruzione e formazione professionale", che con l'art. 3 prevede la "valutazione degli apprendimenti e della qualità del sistema educativo di istruzione e di formazione". Con il decreto stesso, infatti, viene istituito il Servizio Nazionale di Valutazione del Sistema di Istruzione e Formazione. Con questo decreto l'Italia intende finalmente dotarsi, dopo molte esitazioni, dubbi e tergiversazioni, di un servizio nazionale di valutazione della scuola. I principi e i criteri direttivi del decreto

si possono riassumere come segue: Spetta ai docenti la valutazione periodica e annuale degli apprendimenti e dei comportamenti nonché la certificazione delle competenze acquisite; l'INVALSI ha il compito di effettuare verifiche periodiche e sistematiche sulle competenze e sulle abilità degli allievi, sulla qualità complessiva dell'offerta formativa delle istituzioni scolastiche e formativi "ai fini del progressivo miglioramento della qualità del sistema di istruzione e formazione". Questa legge introduce definitivamente la valutazione sistematica ed obbligatoria degli apprendimenti e della qualità complessiva del servizio scolastico come parte integrante di un sistema.

Sarà comunque solo con il DPR 28 marzo 2013 n. 80 che viene emanato il "Regolamento sul sistema nazionale di valutazione" in materia di istruzione e formazione. Le valutazioni effettuate dall'Istituto Nazionale di Valutazione, sostanzialmente, sono relative alla misurazione del livello di preparazione raggiunto da determinate fasce di alunni in specifici settori disciplinari. Nelle scuole viene somministrato anche un questionario per effettuare una valutazione di sistema che consiste in una rendicontazione delle condizioni strutturali e di contesto in cui le stesse scuole operano, delle scelte organizzative adoperate, della gestione finanziaria e delle strategie attuate per ridurre la dispersione scolastica ed aumentare i livelli di apprendimento degli alunni.

3.4.1 Autonomia scolastica ed autovalutazione d'Istituto

Anche l'iter legislativo del regolamento sull'autonomia organizzativa e didattica delle istituzioni scolastiche, previsto ai sensi dell'art. 21 della legge n.59 del 15 marzo 1997, ha avuto un cammino lento. Nella scuola l'autonomia è regolamentata, infatti, dal DPR n. 275 dell'8 marzo 1999 ed è entrata in vigore nell'anno scolastico 2000/2001. Il decentramento determina uno spostamento del baricentro dell'azione pubblica verso la periferia, mediante un duplice percorso di devoluzione di compiti e funzioni sia verso le autonomie istituzionali (Regioni, Province, Comuni), sia verso quelle sociali (enti, associazioni, comunità). La Scuola partecipa ad entrambi i processi, potendosi essa considerare sia un'istituzione, nel senso di articolazione particolarmente qualificata all'esercizio della funzione pubblica della formazione e dello sviluppo della cultura a tutti i livelli, sia una comunità, ossia

una formazione sociale composta da docenti, allievi, genitori, in cui si promuove la formazione di ogni cittadino, in rapporto costante con le altre comunità sociali, culturali, produttive, del contesto territoriale di riferimento. Il DPR n. 275 assegna, infatti, alle istituzioni locali gran parte delle funzioni di gestione prima svolte dallo Stato a livello centrale o periferico. All'istituzione scolastica autonoma è inoltre riconosciuta personalità giuridica: essa pertanto si configura come autonomo centro di imputazione di conseguenze giuridiche, con titolarità di un proprio patrimonio, con necessità di approvare un proprio bilancio, in definitiva con la possibilità di determinare una propria gestione patrimoniale e finanziaria. Ciascuna unità scolastica, nel rispetto della normativa vigente, ha quindi la possibilità di autodeterminarsi liberamente, dandosi proprie regole e propria organizzazione, reperendo e gestendo liberamente le risorse economiche, definendo autonomamente i curricula ed i programmi di insegnamento, sperimentando nuovi modelli di azione didattica e culturale. Nel perseguimento dei propri obiettivi funzionali, l'istituzione scolastica è chiamata a collaborare stabilmente con le altre realtà, sociali, culturali e produttive del territorio, fornendo adeguata risposta, in termini di qualità e di successo formativo, alla domanda educativa dell'utenza.

Il risultato di questa riforma rappresenta una sfida rivolta alla modernizzazione della scuola, che fa parte di un più ampio progetto di trasformazione della pubblica amministrazione. Con il DPR n.275 si istituzionalizza nelle scuole il Piano dell'Offerta Formativa (POF), uno strumento tramite il quale la scuola significhi se stessa in una carta degli intenti, progettando adeguatamente i percorsi formativi che ritiene qualificante scegliere e mettere in campo. Rappresenta, infatti, la carta d'identità dell'istituzione scolastica nella quale viene compartecipata con le famiglie la programmazione curricolare, extra curricolare, didattica ed organizzativa. L'autonomia delle scuole si esprime nel POF attraverso la descrizione:

- delle discipline e delle attività liberamente scelte della quota di curriculum loro riservata;
- delle possibilità di opzione offerte agli studenti e alle famiglie;

- delle discipline e attività aggiuntive nella quota facoltativa del curriculum;
- delle azioni di continuità, orientamento, sostegno e recupero corrispondenti alle esigenze degli alunni concretamente rilevate;
- dell'articolazione modulare del monte ore annuale di ciascuna disciplina e attività;
- dell'articolazione modulare di gruppi di alunni provenienti dalla stessa o da diverse classi;
- delle modalità e dei criteri per la valutazione degli alunni e per il riconoscimento dei crediti;
- dell'organizzazione adottata per la realizzazione degli obiettivi generali e specifici dell'azione didattica; dei progetti di ricerca e sperimentazione.

Il Piano dell'Offerta Formativa è elaborato dal Collegio dei Docenti sulla base degli indirizzi generali per le attività della scuola e delle scelte generali di gestione e di amministrazione, definiti dal Consiglio di Circolo o di Istituto, tendo conto delle proposte formative offerte sul territorio e delle proposte espresse dai genitori. Il piano deve garantire flessibilità, che riguarda le scelte innovative compiute da ciascuna scuola rispetto alle attività didattiche, organizzative, curricolari ed extracurricolari; integrazione, che riguarda la coerenza progettuale delle diverse iniziative, nonché gli aspetti di relazione costruttiva e funzionale della scuola con le agenzie, le culture, le necessità territoriali, i bisogni formativi e le risorse/opportunità delle comunità locali; responsabilità, che investe tutti i processi decisionali attivati da ciascuna scuola, nell'ambito della propria discrezionalità, e attraverso il coinvolgimento di tutti gli attori scolastici, di tutte le componenti istituzionali ed anche dell'utenza. Il POF rappresenta sostanzialmente una la mappa dell'itinerario delle decisioni relative alle scelte organizzative e didattiche del Circolo o dell'Istituto, quindi è, allo stesso tempo, un piano previsionale e un piano esecutivo nel senso che si programmano tutte le attività che si intendono realizzare nel corso dell'anno scolastico e si verificano i risultati raggiunti. Le scelte didattiche e organizzative sono rese pubbliche implicando un'assunzione di

responsabilità della scuola e un diritto delle famiglie e degli studenti a conoscere in modo assolutamente trasparente l'offerta formativa complessiva della scuola. Il POF elabora un progetto formativo per cui non si fonda solo su un senso di un apprendimento circoscritto all'acquisizione di contenuti disciplinari e alla trasmissione di saperi, ma è l'elaborazione di un percorso complesso denso di esperienze, relazioni, opportunità che danno all'esperienza scolastica una valenza formativa integrale. Il POF è il frutto di un'attività di progettazione di team che ha uno scopo preciso: il successo formativo di ogni studentessa e studente. Per l'insegnante è lo strumento per definire le proprie attività, per lo studente e i genitori lo strumento che li aiuta a conoscere cosa la scuola intende fare e come. La sua costruzione prevede la partecipazione e la responsabilità degli insegnanti, degli studenti e delle studentesse, dei genitori, del personale Amministrativo, del personale Tecnico, del personale Ausiliario, dei soggetti rappresentativi l'ambiente esterno in cui la scuola vive e opera. Il POF è, quindi, uno strumento dinamico che sintetizza diversi bisogni, interessi, attese, responsabilità, è una dichiarazione e presa in carico di azioni concrete.

Il POF è un mosaico di impegni concreti e operativi, composto da tanti tasselli che richiamano le nuove opportunità offerte dall'autonomia (es. riorganizzazione dei tempi e degli spazi, flessibilità del curriculum, ampliamento dell'offerta formativa ecc.), che ruotano intorno alla *mission* che il Circolo o l'Istituto si è data. Per mezzo del POF, l'autonomia trova il suo senso e il suo campo d'azione nell'aula, nella relazione quotidiana tra chi insegna e chi impara, nel miglioramento delle condizioni di apprendimento. È, infatti, nella quotidianità dell'aula che si misura l'efficacia delle azioni e il buon esito delle scelte; è negli interventi didattici di ogni giorno che acquistano senso la flessibilità, l'organizzazione modulare dei percorsi formativi, i curricula personalizzati, ecc.

Al centro dell'autonomia, al centro del POF, c'è l'innovazione didattica: è questa che dà senso e valore alle trasformazioni in atto e sollecita gli insegnanti ad individuare strategie, comportamenti, tecniche, modelli organizzativi, modalità di intervento, strumentazioni utili a rendere più sicuro l'apprendimento, a garantire agli studenti e alle studentesse, ogni giorno e in ogni aula, il massimo sviluppo delle loro potenzialità individuali. Non è sufficiente, tuttavia, dichia-

rare e condividere l'offerta formativa d'istituto ma è indispensabile organizzare forme di monitoraggio che ne verifichino l'andamento e la validità. La valutazione del POF deve veder coinvolti tutti i protagonisti dell'esperienza educativa, docenti, personale ATA, genitori ed alunni al fine di dare a tutti l'opportunità di una riflessione sistematica sul percorso formativo, sull'utilizzo delle risorse disponibili, su possibili miglioramenti. I processi autovalutativi del Piano dell'Offerta Formativa devono garantire il rispetto delle caratteristiche che sono insite nel piano stesse e che sono da ricercare nel carattere formativo, partecipato e pragmatico.

Il monitoraggio del Piano dell'Offerta Formativa potrebbe trovare forme di soluzione innovativa nell'impiego di strumenti statistici per la valutazione della Customer Satisfaction, che consentano di raccogliere informazioni sul livello di soddisfazione di tutti gli utenti coinvolti nel processo di insegnamento-apprendimento: studenti, famiglie, dirigenti, personale educativo, personale ausiliario ed amministrativo e personale docente. Lo scopo di un monitoraggio sistematico è quello di comprendere meglio fenomeni quali la percezione e le attese che tutta la comunità ha nei confronti dell'istituzione scuola o di approfondire le ragioni del fenomeno, purtroppo sempre più diffuso, dell'abbandono scolastico.

La *customer satisfaction* può essere, quindi, considerata come una misura dell'efficacia di un servizio. Attraverso un'indagine di customer satisfaction, infatti, si può misurare la soddisfazione di coloro che hanno utilizzato un determinato servizio e quanto questa discosti dalle aspettative iniziali. Alunni, genitori, insegnanti e personale ATA arrivano al servizio con le proprie attese che possono essere determinate da fattori diversi: essere portatori di particolari bisogni, avere già esperienze pregresse in quel servizio o in servizi simili, aver raccolto il racconto dell'esperienza vissuta da altri. Queste attese hanno un effetto diretto su come ognuno potrà percepire il servizio che gli viene offerto o, nel caso del docente, il servizio prestato. Non deve sembrare strano, quindi, che, di fronte allo stesso servizio, ci siano utenti soddisfatti e utenti insoddisfatti. Attraverso un'indagine di customer satisfaction è possibile, allora, comprendere quali siano gli elementi del servizio offerto che incidono sulla soddisfazione o sull'insoddisfazione degli utenti. Per questo si dice che la customer satisfaction

è un modo per ascoltare i propri utenti: le informazioni che si raccolgono attraverso l'indagine, infatti, possono aiutare le istituzioni a conoscere e comprendere le attese dei destinatari, le loro priorità e, in generale, la loro soddisfazione. Attraverso le informazioni raccolte l'amministrazione può progettare nuovi servizi o ideare interventi migliorativi dei servizi già esistenti. Questa attività di ascolto, inoltre, può essere utile anche per ripensare, più in generale, ad aspetti non solo organizzativi ma anche strategici relativi ai servizi. In questo senso, quindi, l'indagine rientra nel ciclo del miglioramento continuo del servizio e deve essere collegata alle strategie e agli strumenti di programmazione e controllo dei servizi. È importante, però, che la rilevazione sia realizzata seguendo un percorso metodologico corretto ai fini di ottenere informazioni che possano orientare adeguatamente i decisori. Errori metodologici nelle diverse fasi della ricerca incidono profondamente sulla qualità delle informazioni che si raccolgono fino ad inficiarne la validità. Per questo è necessario presidiare ogni fase della ricerca con competenza e rigore.

Per realizzare un'indagine di customer satisfaction è importante seguire rigorosamente tutte le fasi che la caratterizzano:

1. Disegno dell'indagine
2. Indagine preliminare e test del questionario
3. Stesura del questionario
4. Raccolta dei dati
5. Elaborazione ed interpretazione
6. Uso e comunicazione dei risultati

A titolo esemplificativo riportiamo di seguito lo schema dell'impianto di un'indagine di customer satisfaction utilizzabile nella scuola per il monitoraggio de Piano dell'Offerta Formativa (Disegno 3.2).

Nella prima fase è necessario definire le finalità dell'indagine e pianificarne la realizzazione avendo chiare le caratteristiche del contesto di riferimento, le indagini realizzate nel passato, la metodologia che si intende utilizzare, la composizione del gruppo di lavoro, l'organizzazione operativa dell'indagine, le risorse disponibili. Segue la

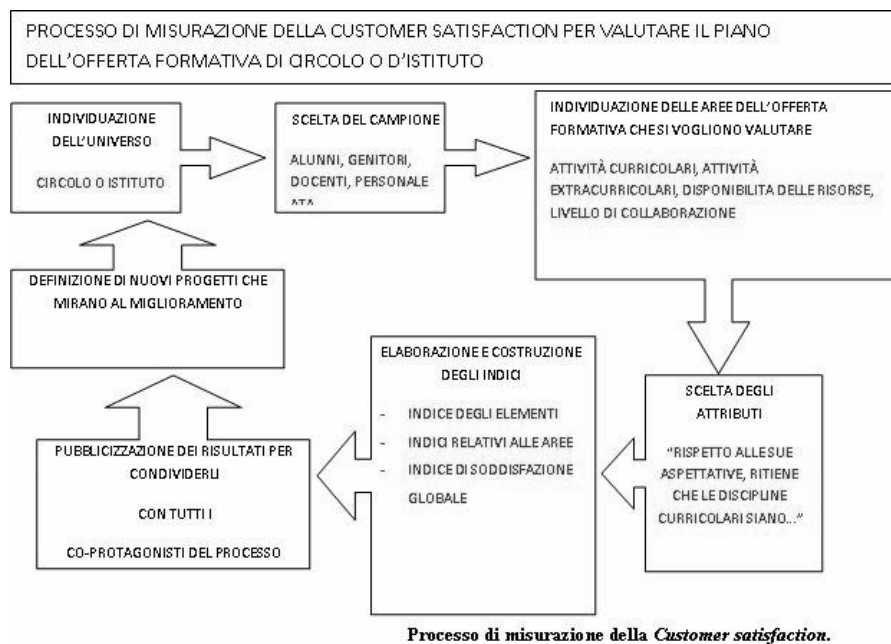


Figura 3.2: Il monitoraggio del POF

fase dell'indagine preliminare che è molto importante per evitare di costruire uno strumento di analisi che sia autoreferenziale. A questo fine è importante conoscere in modo ottimale le condizioni di contesto in cui ci si muove, le caratteristiche generali del servizio oggetto dell'indagine, le strategie utilizzate per attuare il piano dell'offerta formativa e le risorse di cui l'istituzione dispone. Si potrà procedere, quindi, o ad un'indagine di tipo documentale, caratterizzata dalla comparazione con indagini precedentemente realizzate, o dal confronto con la bibliografia esistente in materia; o in alternativa si può procedere ad un'indagine qualitativa attraverso interviste o focus group rivolte a testimoni privilegiati che sono appunto tutti gli stakeholder, ossia i soggetti portatori di interesse, che nel caso della scuola sono docenti, studenti, genitori e personale ATA. Si può quindi procedere alla creazione dello strumento di rilevazione, ovvero il questionario. Questa è una fase delicata nella quale ogni errore commesso incide

sulla qualità dell'informazione raccolta: l'uso di un linguaggio troppo tecnico, la confusione nell'organizzazione delle domande, l'utilizzo di formule tendenziose, possono pregiudicare pesantemente la buona riuscita dell'indagine. Prima di cominciare a stendere il questionario, allora, è importante scegliere:

- come organizzare le domande e con quale modalità porle;
- le scale di giudizio da utilizzare;
- il tipo di modello da adottare per rilevare il grado di soddisfazione dell'utente.

Per quanto riguarda la formulazione delle domande occorre tener presente che è necessario utilizzare un linguaggio adeguato alle persone che dovranno rispondere alle domande, usando una costruzione semplice delle frasi poiché non dobbiamo dimenticare che il questionario, in ambito scolastico deve essere somministrato anche agli alunni. Le domande dovranno, pertanto, focalizzare chiaramente l'argomento ed esplicitarlo in modo facilmente fruibile. Chi somministra i questionari deve fornire istruzioni chiare su come rispondere alle domande. Nella costruzione delle scale è opportuno utilizzare un numero limitato di categorie o gradazioni (di solito non più di otto), scegliendo categorie appropriate, né troppo ampie, né troppo ristrette. È necessario, inoltre, decidere con cura come rilevare le risposte neutrali. I settori oggetto di studio possono essere relativi a:

- Singoli elementi dell'offerta formativa;
- Aree in cui si può scomporre l'offerta formativa;
- L'offerta formativa nel suo complesso.

In letteratura possiamo individuare diversi modelli finalizzati al rilevamento della customer satisfaction, ai quali daremo solo un breve accenno in quanto per l'approfondimento si rimanda a trattazioni specifiche. La varietà dei modelli dipende dalla natura stessa della customer satisfaction che è soggettiva, latente e multivariata. La customer satisfaction è, infatti, influenzata da elementi psicologici, culturali, demografici ed economici. I modelli comunemente più utilizzati sono

il *ServQual* e il *ServPerf*. Il modello *ServQual*, sviluppato da Zeithaml *et al.* [1991], è uno strumento atto a valutare la percezione dei clienti relativa alla qualità del servizio, definita come il giudizio del consumatore sul grado di eccellenza attribuito ad un'entità, e sulla definizione di alcune dimensioni quali: gli aspetti tangibili, cioè gli aspetti delle strutture fisiche, delle attrezzature, del personale e degli strumenti di comunicazione; l'affidabilità, riguardante la capacità di prestare il servizio in maniera affidabile e precisa; la capacità di risposta, in merito alla volontà di aiutare i clienti e alla fornitura del servizio; la capacità di assicurazione, che comprende la competenza e la cortesia dei dipendenti connessa con la capacità di ispirare fiducia e sicurezza; l'empatia, riguardante l'assistenza premurosa e individuale che viene prestata al cliente. Il modello *ServPerf*, di Cronin Jr and Taylor [1992], considera invece le sole percezioni senza la componente delle attese. Dopo aver creato il questionario è necessario testarlo cioè verificare, sul campo, se le domande sono chiare, se le modalità previste per la raccolta delle informazioni sono adeguate, se ci sono ridondanze, ecc. Il test del questionario, o collaudo, è una fase fondamentale che consente di avere indicazioni necessarie al miglioramento dell'organizzazione dell'indagine e del questionario. Questa fase, che dipende dalle scelte fatte nelle fasi precedenti, è estremamente importante per procedere alla pianificazione delle azioni di miglioramento, quali la metodologia di ricerca, il campione selezionato, le domande fatte, ecc., determinano la possibilità di avere certe rappresentazioni piuttosto che altre. Le rappresentazioni grafiche possono essere particolarmente efficaci per individuare le priorità del cambiamento e, di conseguenza, per orientare le scelte dei responsabili del servizio. Il percorso di indagine non si conclude con l'elaborazione e l'interpretazione dei dati bensì con la definizione di nuovi progetti che mirano al miglioramento dei servizi sottoposti all'indagine. A questo, infatti, servono le indagini di customer satisfaction. L'importante è riuscire ad inserire il percorso del miglioramento continuo all'interno di una logica strategica coerente con la mission dell'organizzazione. Infine, è importante decidere come e a chi gli esiti dell'indagine debbano essere comunicati, poiché la comunicazione è un atto dovuto per garantire la trasparenza stessa dell'indagine.

La customer satisfaction valuta, quindi, statisticamente la qualità

del servizio offerto all'utente, quindi è necessario tener ben presenti i diversi aspetti della qualità stessa:

- *Qualità attesa*: qualità che l'utente si aspetta di ricevere dal servizio di cui usufruisce, per soddisfare le sue esigenze espresse ed implicite;
- *Qualità progettata*: insieme delle caratteristiche di qualità specifiche di un servizio, nonché della strutturazione dei processi e delle risorse coinvolte, definite come interpretazione della qualità attesa;
- *Qualità erogata*: insieme delle caratteristiche di qualità di un servizio, alla fine del processo di erogazione del servizio;
- *Qualità percepita*: ciò che viene percepito dal cliente in termini di soddisfazione rispetto alle proprie aspettative.

La soddisfazione del cliente è data dalla differenza fra qualità percepita e qualità attesa; gli stati base della soddisfazione sono tre: *delight*, *soddisfazione*, *insoddisfazione*. Quando la qualità percepita supera la qualità attesa si registra una condizione di "delight"; quando la qualità percepita è in linea con la qualità attesa si registra soddisfazione; quando la qualità percepita disattende le attese si registra insoddisfazione.

Il monitoraggio della customer satisfaction deve essere condotto: - *ex ante*, come una leva importante da valorizzare per consentire una progettazione efficace degli obiettivi che l'istituzione scolastica intende conseguire. Lo strumento di valutazione *ex ante* utilizzato dalle scuole è da individuare nel POF attraverso il quale le istituzioni, in riferimento alla propria mission e alle risorse disponibili, programmano il piano d'intervento formativo.

- *in itinere*: per monitorare costantemente la qualità del servizio.

- *ex post*, come uno strumento adatto a misurare la qualità dei servizi erogati.

Per misurare la customer satisfaction le tecniche più frequentemente usate, quindi, sono due. La più semplice consiste nel chiedere il grado di soddisfazione e di importanza, per ogni area di prestazione sottoposta ad indagine, tramite scale Likert di valutazione. L'altra

tecnica invece misura lo scarto tra le aspettative e il livello di prestazione percepito. Agli intervistati si chiede cioè di indicare, per ciascuna area esaminata, se ritengono di avere riscontrato una prestazione più elevata, inferiore o in linea rispetto alle aspettative. In tal modo si concentra l'attenzione su ciò che gli utenti dell'istituzione vorrebbero ma di cui non trovano corrispondenza nell'offerta formativa. Per la misurazione della customer satisfaction si utilizzano, sostanzialmente, medie statistiche poiché il calcolo della media ha proprio lo scopo di rappresentare con un solo numero un insieme di dati, indicando quindi l'ordine di grandezza del fenomeno che si studia.

3.4.2 La valutazione degli insegnanti

La valutazione di un sistema scolastico deve interessare tutti gli elementi che lo caratterizzano, attualmente, però, in Italia la valutazione dei docenti è limitata alle procedure di selezione del personale e alla valutazione del primo anno di servizio. Si tratta di una valutazione iniziale mirata a determinare se l'insegnante possieda le qualità necessarie a svolgere la funzione docente, ma tuttavia non esistono forme di valutazione in servizio. Valutare la prestazione professionale, in modo diretto, comporta un notevole coinvolgimento da parte del valutato e può diventare un elemento di motivazione, purché ci sia forte omogeneità fra chi valuta e chi è valutato per quanto attiene ai valori di riferimento del gruppo professionale (i principi, la missione, la responsabilità, la dimensione di servizio).

Al momento dell'ingresso nella professione, oltre a misurare la competenza sui contenuti qualificanti la disciplina, in conformità a parametri predefiniti a livello nazionale, il Comitato di Valutazione nominato dal Collegio dei Docenti tra i suoi componenti, attribuisce valore al metodo didattico, al modo di comunicare, alla capacità di lavorare per obiettivi e, anche, ad esperienze di lavoro in ambiti diversi dall'insegnamento.

Il problema del reclutamento dei docenti è sempre stato un nodo cruciale della politica scolastica italiana. Una cattiva programmazione di assunzione e le modalità con cui sono stati reclutati i docenti hanno determinato, nel corso degli anni, la nascita del precariato. Una prima risposta concreta alla domanda potrebbe essere quella di monitorare

costantemente gli indici demografici e di programmare il numero delle assunzioni in relazione al numero degli obbligati previsti per un certo lasso di tempo. In verità, dagli anni '70 in avanti, considerata la crescita esponenziale della domanda di istruzione, è aumentato costantemente anche il numero dei docenti e si è creato un alto tasso di occupazione senza effettuare indagini adeguate di previsione della domanda. Infatti, se si analizzano i dati statistici relativi al numero dei docenti e degli studenti, si può constatare una crescita inversamente proporzionale dei due valori, poiché ad un aumento del personale docente è corrisposta una diminuzione di quello degli alunni, causata anche da un notevole calo demografico delle nascite.

Prima del 1974 l'insegnante era un esecutore di prescrizioni programmatiche definite a livello nazionale dal Ministero e l'insegnamento era di tipo trasmissivo. Nei venticinque anni successivi ai decreti delegati del '74 l'insegnante è diventato un interprete delle indicazioni ministeriali che, lavorando in team con i colleghi, elaborava programmazioni educative - didattiche maggiormente congruenti con le realtà locali. Di conseguenza al docente non necessitava solo non era richiesta solo la conoscenza della disciplina ma anche competenze psico-pedagogiche e capacità relazionali. Con l'avvento dell'autonomia il docente è divenuto sempre più protagonista del processo di insegnamento-apprendimento e ha maturato competenze progettuali complesse. Fino al 2000 l'assunzione avveniva per superamento di concorso bandito a livello nazionale che si basava su una prova scritta e un colloquio orale. Poi sono arrivate le scuole di specializzazione o SSIS, che prevedevano, oltre all'approfondimento dello studio di specifiche discipline, un periodo di tirocinio nelle scuole e un esame di Stato finale che rilasciava l'abilitazione all'insegnamento.

Dopo una formazione iniziale e un esame di abilitazione, concorsuale o di Stato, il docente assunto deve superare un anno di prova, al termine del quale presentare una dettagliata descrizione delle attività svolte da relazionare ad una commissione composta dal comitato di valutazione di Circolo o d'Istituto, composto da alcuni docenti e dal dirigente scolastico. Superato l'anno di prova al docente viene confermata l'immissione in ruolo e fino al termine della sua carriera non sarà più sottoposto ad alcuna valutazione. La progressione di carriera avviene in relazione agli anni di servizio prestati e non è legata

neanche all'obbligo di formazione continua. Dal CCNL del comparto scuola del 2003, infatti, la formazione in servizio per i docenti non è più un diritto-dovere, ma solo un diritto poiché gli insegnanti sono liberi di scegliere se aggiornarsi o se non farlo.

L'Italia è uno dei pochi paesi in cui i docenti non vengono regolarmente sottoposti a valutazione e in cui si teme che l'efficacia dell'insegnamento possa essere verificata esclusivamente valutando gli apprendimenti degli studenti. Per quest'ultimo motivo i docenti intraprendono lotte al Sistema Nazionale di Valutazione.

Investire sulla formazione in servizio e valutare costantemente il personale significa investire sulla qualità del sistema scolastico. Così il corpo docente italiano si spacca tra chi si impegna molto, si aggiorna regolarmente e lavora con passione e chi investe ben poco nella sua professione. Ma tutti, però, vengono retribuiti e progrediscono nella carriera allo stesso modo. Negli ultimi anni anche quei pochi incentivi dovuti all'autonomia sono stati distribuiti quasi sempre a pioggia, generando forme di protesta da parte dei docenti e scontentando tutti.

In Italia fino ad oggi è esistita una forte opposizione delle rappresentanze sindacali di categoria verso forme effettive di valutazione delle competenze e delle capacità pedagogiche, didattiche e disciplinari dei docenti. Soprattutto sussiste una netta contrarietà a differenziare i trattamenti economici e le prospettive di carriera degli insegnanti sulla base di esiti di procedimenti di valutazione. L'ostacolo maggiore da superare è proprio quello della reticenza degli insegnanti ad essere valutati; occorre quindi lavorare proprio sulla classe docente e educarla all'idea che la valutazione degli insegnanti è una necessità ed una responsabilità che dà garanzia di qualità al sistema scolastico. La valutazione dei docenti non deve, quindi, essere vissuta come un tabù. L'ideologia egualitaristica che ha fin qui ispirato i comportamenti negoziali dei sindacati degli insegnanti italiani è stata passivamente accolta dal mondo politico, producendo, così, situazioni di scarsa equità e di notevole inefficienza. Innanzitutto perché i docenti più bravi e quelli meno bravi sono trattati nello stesso modo, con l'ovvia conseguenza che quelli, essendo privi di qualsiasi forma di incentivazione, tendono ad appiattirsi sulle non eccelse prestazioni di questi" [Schizzerotto, A. e Barone, C., 2006]. Mariani [1991] afferma, rispetto alla valutazione dei docenti che "Le prospettive delineate devono

spesso considerare la presenza di alternative non sempre antinomiche ma comunque di difficile coesistenza: valutazione ed autovalutazione, valutazione collettiva e valutazione individuale, autovalutazione come rifugio o come competenza, riciclaggio delle professioni presenti ed apertura a professioni in divenire, valutazione come dramma o come strategia. Un solo problema ci sembra possa essere sciolto a priori, prescindendo da ogni studio e discussione: la valutazione dell'insegnante deve consistere unicamente in una verifica pedagogica, per cui la gestione delle interazioni tra le potenziali capacità, le inclinazioni ed i bisogni del docente e la qualità, la quantità e soddisfazione della sua attività è finalizzata, e tale deve restare, unicamente alla razionalizzazione delle decisioni educative".

Il concetto di "qualità dell'insegnante" si estende lungo cinque dimensioni:

- a) conoscenze degli specifici settori disciplinari e del contenuto dei programmi;
- b) competenze didattiche, cioè padronanza di un repertorio di strategie didattiche e la capacità di applicarle;
- c) capacità di riflessione e di auto valutazione;
- d) empatia o capacità di identificarsi negli altri e riconoscere la loro dignità;
- e) competenza gestionale, dato che gli insegnanti devono assumersi responsabilità di gestione, dentro e fuori dell'aula.

"Queste dimensioni della qualità degli insegnanti non dovrebbero essere considerate come competenze strettamente comportamentali, ma piuttosto come atteggiamenti di base. Le diverse dimensioni lungo le quali si misura la qualità del docente sono interdipendenti nel senso che una gran competenza in una di esse facilita il miglioramento o l'acquisizione di altre competenze. Inoltre, il contesto istituzionale, locale e nazionale, può favorire o frenare l'insieme delle competenze o alcune di esse" [Margiotta, 1999].

Conseguenza che si dovrebbe definire quasi "naturale" di una valutazione sistematica del corpo docente deve essere quella della

progressione di carriera. Attualmente, però, all'interno dell'istituzione scolastica, il docente non ha possibilità di crescita professionale e nel corso della propria carriera gli scatti stipendiali funzionano secondo una scala di livello in relazione agli anni di servizio maturati. La popolazione docente è così divisa tra coloro che vorrebbero una riforma del sistema che garantisse una più gratificante progressione di carriera, legata non solo alla maturazione dell'anzianità di servizio ma anche al riconoscimento dei meriti, e coloro che ritengono che l'insegnante abbia il compito di valutare e che non debba essere valutato. Alcuni tentativi sono stati fatti con il CCNL del '95, nel quale si parlava della realizzazione di progetti di valutazione sperimentale della qualità delle prestazioni, e nel CCNL del '99, nel quale si faceva riferimento ad un trattamento economico accessorio connesso allo sviluppo della professione docente. Tuttavia sono rimasti scritti solo sulla carta poiché non è stato organizzato un sistema di valutazione in grado di determinare i criteri di attribuzione dei compensi accessori citati.

La valutazione degli insegnanti richiede l'analisi attenta di una diversità di fattori, pertanto deve garantire determinate caratteristiche. Prima di tutto la valutazione deve essere effettuata durante tutta l'esperienza lavorativa del docente per osservare il processo di insegnamento nel succedersi degli anni e deve essere istituzionale, ossia effettuata direttamente dal Ministero tramite personale specializzato, al fine di garantire una più equa e gratificante progressione di carriera. Deve essere effettuata in relazione al contesto specifico in cui il docente opera, poiché la realtà socio-economica-culturale del nostro paese è molto eterogenea ed alcuni docenti si trovano a lavorare in condizioni altamente disagiate rispetto ad altri, trovandosi a supportare un carico di lavoro che va oltre il solo insegnamento soprattutto là dove i giovani non hanno supporti familiari validi. La valutazione si deve concretizzare in un'osservazione diretta, poiché il processo di insegnamento-apprendimento si esplica in una dinamica relazionale che vede coinvolti alunni, genitori, colleghi e personale ATA e non può limitarsi, quindi, ad un'analisi delle conoscenze contenutistiche della disciplina che si insegna. Occorre verificare anche le competenze programmatiche del docente, attraverso la valutazione delle modalità con cui agisce per garantire il rispetto delle specificità di tutti e di ciascun alunno, soprattutto nel caso in cui il docente si trovi a relazionarsi

con alunni con speciali bisogni formativi. Deve, inoltre, essere pragmatica, poiché deve osservare la validità delle scelte metodologiche e didattiche in relazione alle competenze acquisite dallo studente.

Un insegnante competente deve “saper essere”, caratteristica che si esplica prima di tutto attraverso la disponibilità all’ascolto. Siamo geneticamente predisposti ad udire ma non tutti possiedono la capacità di saper ascoltare, di prestare un’accurata attenzione alle richieste e alle specifiche condizioni del contesto nel quale si opera. Un ascolto attivo non passa solo attraverso il canale uditivo ma presuppone l’attivazione di una complessità di canali che ci consentono di metterci in relazione con l’altro. Occorre, quindi, la capacità di entrare in empatia con l’altro, con lo studente, le famiglie, i colleghi, il personale scolastico, attivando tutti i canali ricettivi. Occorre avere la capacità di leggere i messaggi che ci vengono inviati nella loro complessità, avendo cura di coglierne l’aspetto verbale, para-verbale e cinestesico. La relazione di insegnamento-apprendimento si esplica in una pluralità di figure che vede coinvolta una rete di individui, andando ben oltre il rapporto diretto di docente-discente.

L’insegnamento, passando attraverso i canali relazionali ed affettivi, si fonda prima di tutto sull’affettività e sulla capacità di stimolare la motivazione e valorizzare l’autostima.

Un altro aspetto che deve essere sottoposto a valutazione è il “sapere” in quanto all’insegnante è richiesto di dominare ad un buon livello i saperi disciplinari che deve trasmettere ai suoi allievi. Non si tratta di possedere una cultura disciplinare stantia, ma della predisposizione ad un rinnovamento continuo del proprio bagaglio di conoscenze attraverso forme di aggiornamento individuale ed istituzionale, che tengano in conto anche delle nuove tecnologie dell’informazione e della comunicazione. La qualità del “sapere” necessita di un rinnovamento continuo del proprio bagaglio di conoscenze attraverso percorsi di aggiornamento continuo.

Altri indicatori appartengono, invece, al “saper fare” dell’insegnante e costituiscono la sua competenza professionale. Un ricco bagaglio di competenze in campo metodologico e didattico sono necessarie per applicare le strategie di insegnamento più idonee ad ogni specifico contesto. La creatività, la capacità di saper re-inventare, l’abilità di modularizzare ed individualizzare, le connessioni inter e trans-

disciplinari sono elementi che rappresentano un valore aggiunto al “saper fare”.

Grande importanza assume, inoltre, il “saper valutare”. La valutazione è il momento centrale sul quale progettare tutto il processo di insegnamento-apprendimento. Non è da intendersi, come nella prassi spesso avviene, come momento nel quale si tirano le somme rispetto ad un determinato percorso o periodo ma al contrario come il punto di partenza, in quanto è opportuna un’accurata analisi ex-ante per rilevare le risorse, le specificità del contesto e le peculiarità dei singoli alunni con i quali si instaura la relazione di insegnamento-apprendimento; un monitoraggio continuo per valutare l’andamento dell’evolversi del processo di apprendimento e, solo infine, di una valutazione ex-post per tirare le somme e verificare i risultati.

Ribadendo l’assoluta necessità della formazione in servizio, le prime azioni che lo Stato dovrebbe compiere sono di investimento nella formazione per garantire al sistema una riqualificazione del personale docente. Attualmente le attività di formazione del personale docente sono molto sporadiche ed indirizzate ad un numero limitato di docenti, pertanto la maggior parte della formazione in servizio si basa sul senso di responsabilità, sull’interesse ad una crescita professionale, nonché sulla convinzione che impegnare direttamente le proprie risorse economiche sulla formazione sia un investimento per la propria crescita personale e professionale.

4 Approccio olistico alla persona: valutare le competenze

“La scuola non sempre dispone di strumenti per riflettersi e riflettere su se stessa: i ballerini, che praticano la loro arte alla perfezione, dispongono di specchi per osservare i loro movimenti. Dove sono i nostri specchi?”

(E. W. Eisner)

4.1 Le Indicazioni per il curricolo: un lungo cammino lungo per il primo ciclo della scuola italiana

Sulla Gazzetta Ufficiale n.30 del 5 Febbraio 2013 sono state pubblicate le nuove *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione* (il Dm n. 254 del 16 novembre 2012). Il nuovo testo delle *Indicazioni Nazionali per il curricolo* è arrivato alla versione definitiva, dopo una prima bozza del 30 maggio 2012, sottoposta alla consultazione di scuole, associazioni ed esperti, una seconda del 23 luglio 2012 che acquisiva alcune di dette osservazioni e una terza del 4 settembre 2012, anch'essa assoggettata a pochi piccoli ritocchi prima di diventare definitiva con questa ultima versione di novembre, emanata con il Regolamento recante indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione (DM del 16/11/2012).

Si giunge al testo definitivo dopo un decennio dominato da una pesantissima contrapposizione politico-pedagogica tra governi di centrodestra e di centrosinistra che ha pesato notevolmente sul settore di istruzione e formazione, determinando spesso confusione e incertezze. All'insediamento del governo di centrodestra, la riforma Berlinguer (L. n. 30/2000) viene sospesa e poi sostituita con la riforma Moratti (Legge n.53/2003 e D. Lgvo n.59/2004), cui sono allegate le Indicazioni Nazionali. Arrivato di nuovo al governo il centrosinistra, il ministro Fioroni, pur mantenendo la riforma già avviata, azzerava rapidamente le Indicazioni Nazionali sostituendole con le *Indicazioni per il curricolo*.

Ritorna il centrodestra, e il ministro Gelmini, pur non abolendo le nuove Indicazioni, prevede la revisione in termini di armonizzazione tra le Indicazioni Nazionali (Moratti) e le Indicazioni per il curricolo (Fioroni), con l'evidente intento di formulare un testo nuovo con cui far valere la primitiva ispirazioni della propria parte politica relativizzando o vanificando quella degli avversari. Quando prende in mano la situazione il ministro Profumo, per evitare una forzatura di armonizzazione ed integrazione tra i due documenti, sceglie di mantenere le Indicazioni in vigore, quelle del 2007. La scuola si è sentita travolta da un'intollerabile contrapposizione politica che ha pesato gravemente su studenti, famiglie e docenti. Con le *Indicazioni per il curricolo*, pubblicate con D.M. del 31 luglio 2007 e regolamentate con direttiva n. 68 del 3 agosto 2007 [Istruzione, 2007], alla scuola dell'autonomia è stata data la possibilità di progettare i percorsi formativi che ritiene più opportuni per il raggiungimento dei traguardi per lo sviluppo delle competenze. L'aspetto innovativo delle Indicazioni del 2007 sta proprio nel fatto che, contrariamente a quanto avvenuto in passato, il Ministro della Pubblica Istruzione non detta programmi prescrittivi da applicare in tutte le scuole, ma dà appunto delle indicazioni per la costruzione di curricoli formativi validi e soprattutto strutturati in relazioni alle reali esigenze e risorse di ogni istituzione scolastica. Il *curricolo* pubblicato nel 2007, si propone come uno strumento che le istituzioni scolastiche sono state invitate a sperimentare nel corso degli anni scolastici 2007-2008 e 2008-2009, per suggerire al Ministero stesso le possibili modifiche e integrazioni. Le Indicazioni, quindi, hanno rappresentato il quadro di riferimento per l'elaborazione del curricolo e hanno indicato traguardi per lo sviluppo delle competenze di tutti gli alunni. Il testo è stato distribuito nelle scuole in tempo utile per essere conosciuto, letto, studiato, discusso e sperimentato nel corso del biennio 2007-2009. Un modo nuovo di pensare la scuola dal basso, di progettare i percorsi educativi e formativi da parte di chi vive quotidianamente la realtà scolastica e ne conosce tutte le piccole sfumature. L'idea di una scuola che si fa a scuola e che, nell'apertura al dialogo e al confronto, invita al termine della sperimentazione a dare un contributo concreto alla formulazione delle Indicazioni. Una scuola pensata e progettata da tutti coloro che hanno la diretta responsabilità educativa e formativa nella scuola, al fine di garantire allo

stesso tempo l'autonomia e la specificità progettuale delle scuole e un quadro di riferimento complessivo del sistema a livello nazionale. Le *Indicazioni* del 2007 sono introdotte da una prima parte a carattere generale che presenta la cornice culturale nella quale la scuola italiana si trova ad operare, segue una prima sezione riferita alla scuola dell'infanzia e ai relativi cinque campi di esperienza ed una seconda sezione riferita al primo ciclo, primaria e secondaria di primo grado, con l'indicazione delle tre aree disciplinari. Ne emerge un'idea di scuola innovativa che per il suo stesso progresso punta sul capitale umano e rispetta le diversità che caratterizzano ogni istituzione scolastica. La prima fase sperimentale di attuazione durerà fino al 2009 e consentirà alle scuole di conoscere e sperimentare le nuove Indicazioni e al Ministero di raccogliere suggerimenti, valorizzare le buone pratiche e favorire processi di condivisione e di sostegno. Dall'anno scolastico 2009-2010 le Indicazioni per il curriculum entrano definitivamente a regime, accompagnate da apposito Regolamento. Il Ministro, nel pieno rispetto dell'autonomia e del conseguente decentramento del sistema di istruzione, non detta programmi normativi che devono essere applicati in modo prescrittivo ed in ugual misura su tutto il territorio nazionale, ma consente alle singole istituzioni di costruirsi i propri percorsi formativi in relazione agli specifici bisogni delle realtà locali. La scuola dell'autonomia valorizza lo studente, ponendolo al centro del percorso formativo con le sue specificità, in relazione alle quali struttura il curriculum. Le Indicazioni valorizzano la figura del docente che, non sentendosi più un burocrate che esegue programmi, è messo nelle condizioni di vivere l'autonomia scolastica da protagonista. Le *Indicazioni per il curriculum* del 2007 vengono introdotte da un documento molto significativo "Cultura Scuola Persona", nel quale è disegnato il nuovo scenario culturale e sociale nel quale la scuola stessa si trova ad operare; uno scenario caratterizzato da più ricchi stimoli culturali, da molteplici cambiamenti e innovazioni, ma al tempo stesso più instabile e talvolta contraddittorio. La scuola in questo scenario ha l'obbligo di garantire a tutti e a ciascuno pari opportunità di un pieno sviluppo della persona umana e l'effettiva partecipazione di tutti gli alunni al percorso educativo, valorizzando appunto le esperienze e le conoscenze di ciascuno. Fondamentale è il principio della centralità della persona in quanto per progettare

un percorso formativo idoneo occorre prima di tutto leggere i bisogni formativi, le attitudini personali, le capacità e le fragilità, le aspirazioni e le motivazioni del singolo per mettere in campo uno stile inclusivo, un progetto di accoglienza, un processo di empatia con gli alunni e con la classe. Le discipline costituiscono strumenti educativi e culturali per fornire valori e criteri che orientino i bambini e i ragazzi verso l'impegno civile di ciascuno per la costruzione di una società migliore. La scuola, attraverso il superamento della frammentazione delle discipline, crea le condizioni per l'elaborazione dei saperi necessari per comprendere l'attuale condizione dell'uomo planetario. La scuola deve costruire, quindi, un contesto idoneo a promuovere apprendimenti significativi e a garantire il successo formativo per tutti gli alunni. Il curriculum è lo strumento attraverso il quale la scuola, in piena coerenza con il POF, costruisce il percorso formativo che deve essere progressivo, collegiale, trasversale e verticale, per la fascia d'età dai 3 ai 14 anni. Ciò implica una continuità più concreta e responsabile tra i diversi ordini di scuola, mirata alla costruzione di un impianto formativo unitario. Agli insegnanti compete la responsabilità della valutazione e la cura della documentazione, nonché la scelta dei relativi strumenti nel quadro dei criteri deliberati dai competenti organi collegiali. La valutazione assume una preminente funzione formativa, di accompagnamento dei processi di apprendimento e di stimolo al miglioramento continuo, poiché precede, accompagna e segue i percorsi curricolari. Alla scuola viene dato il compito di lavorare in stretta collaborazione con le famiglie e con le altre agenzie educative per fornire supporti adeguati affinché ogni persona sviluppi un'identità consapevole ed aperta. Le famiglie rappresentano il contesto più influente per lo sviluppo dei bambini e, pur nella loro diversità, sono sempre portatrici di risorse che devono essere valorizzate, sostenute e condivise nella scuola. Le famiglie devono partecipare alle fasi di informazione e formazione, seguire la predisposizione del curriculum e del POF, confrontarsi con la scuola e sostenerla nelle scelte curricolari, attraverso una condivisione responsabile che consente di far propri i principi educativi e i valori della scuola. Le famiglie sono chiamate a collaborare concretamente con i docenti per garantire la conoscenza delle caratteristiche relazionali, cognitive, fisiche, spirituali dei propri figli contribuendo alla costruzione comune di un clima che favorisce

il benessere grazie ad un progetto educativo condiviso. Non vi sono piste già tracciate, ma occorre costruire insieme, docenti e genitori, la strada, nell'ascolto reciproco e nel rispetto dei diversi punti di vista dei ruoli e delle competenze di ciascuno. Scuola e genitori, affrontando insieme i grandi temi dell'educazione e proponendo ai bambini un modello positivo di dialogo coerente e rispettoso delle diverse scelte individuali, cooperano affinché la scuola diventi luogo dell'approfondimento culturale e di reciproca formazione tra genitori e insegnanti. Il sistema educativo si fonda, quindi, sul patto educativo fra scuola e famiglia, che diviene l'elemento portante della nuova modalità di fare scuola poiché la scuola non può realizzare il proprio compito di educare senza la condivisione della famiglia. Condividere il progetto educativo significa tracciare un percorso comune che favorisca una crescita globale ed armonica dei bambini e dei ragazzi. La continuità tra i rapporti scuola-famiglia è fondamentale per la costruzione di un progetto educativo condiviso e continuo. I rapporti, infatti, non possono come sovente succede limitarsi ad affrontare situazioni problematiche, ma devono essere tesi a costruire un percorso formativo basato sulla corresponsabilità.

La mission della scuola non si può limitare alla trasmissione di saperi, all'acquisizione di abilità e alla maturazione di competenze ma deve essere quella di formare la persona nel suo complesso, sul piano cognitivo, culturale, affettivo e relazionale, affinché diventi in grado di affrontare positivamente la mutevolezza e l'incertezza degli scenari presenti e futuri. Le Indicazioni illustrano, al termine dei più significativi snodi del percorso curricolare, i traguardi per lo sviluppo delle competenze, che rappresentano riferimenti significativi per gli insegnanti, indicando sentieri da percorrere per consentire all'alunno uno sviluppo integrale della persona.

Nelle Indicazioni del 2012 è pienamente mantenuta l'idea di continuità del percorso di apprendimento dai 3 ai 14 anni. Scuola primaria e scuola secondaria sono costantemente associate e considerate insieme, come un'unica scuola, il primo ciclo, senza capitoli o paragrafi distinti. Così la parte introduttiva generale dedicata alle tematiche della costruzione del curricolo, la presentazione della scuola del primo ciclo, infine le impegnative introduzioni alle singole discipline. Solo all'interno del paragrafo dedicato alla "Alfabetizzazione culturale di

base” figura una specificazione del carattere proprio dei due gradi di scuola, che occupa distintamente alcuni brevi capoversi. Un criterio di continuità che esclude, tuttavia la scuola secondaria di secondo grado e che non investe quella di primo grado del ruolo di tramite in un continuum che favorisca il pieno e armonico sviluppo della persona. Occorre, inoltre, considerare che la strutturazione, la storia e l’approccio metodologico-didattico della scuola primaria e della scuola secondaria di primo grado sono profondamente diverse e che l’introduzione delle nuove indicazioni comporta un profondo cambiamento strutturale ed organizzativo, il quale ha forzatamente avuto avvio con la progressiva generalizzazione degli Istituti Comprensivi, nell’ottica della costituzione di una *scuola unitaria di base*. Gli Istituti Comprensivi hanno il compito di creare una nuova realtà che riesca ad integrare le profonde differenze tra la scuola primaria, nata come scuola popolare per contrastare l’analfabetismo, e la scuola secondaria di primo grado, che porta ancora in sé l’eredità ginnasiale.

Rispetto al testo del 2007, le Indicazioni presentano alcune significative novità. Prima fra tutte l’integrazione delle otto competenze europee per l’apprendimento permanente, l’introduzione di un *Profilo delle competenze dello studente al termine del primo ciclo*, l’integrazione di tre importantissimi riferimenti ministeriali per l’azione educativa e didattica delle scuole:

- la via italiana per la scuola interculturale e l’integrazione degli alunni stranieri;
- le Linee guida per l’integrazione scolastica degli alunni con disabilità;
- le Linee guida per il diritto allo studio degli alunni e degli studenti con disturbi specifici di apprendimento.

Le Indicazioni Nazionali per il Curricolo del 2012 sono centrate sulla figura dello studente, che è posto al centro dell’azione educativa in tutti i suoi aspetti: cognitivi, affettivi, relazionali, corporei, estetici, etici, spirituali, religiosi. Grande importanza viene data alla formazione della classe come gruppo che coopera nell’integrazione e nella gestione dei conflitti. L’idea è quella di una scuola che promuova e favorisca

il ben-essere dell'individuo, accompagnandolo e sostenendolo soprattutto nei momenti di passaggio della crescita e dell'apprendimento. Il decreto attuativo delle Indicazioni prevede anche la costituzione di un *Comitato Scientifico Nazionale* per l'attuazione delle indicazioni nazionali e il miglioramento continuo dell'insegnamento. Il Comitato sarà incaricato di indirizzare, sostenere e valorizzare le iniziative di formazione e di ricerca in modo da aumentare l'efficacia dell'insegnamento secondo gli obiettivi previsti dalle indicazioni e nel costante rapporto con le scuole e le loro esperienze. Grande enfasi viene posta, inoltre, sull'importanza della valutazione quale elemento portante dell'azione educativa, in quanto deve precedere, accompagnare e seguire l'intero processo formativo. A tale proposito è stato approvato in via definitiva il Regolamento che istituisce e disciplina il Sistema Nazionale di Valutazione delle scuole pubbliche e delle istituzioni formative accreditate dalle Regioni. L'Italia si allinea così, almeno a cominciare dalla documentazione cartacea e dagli intenti, agli altri Paesi Europei sul versante della valutazione dei sistemi formativi pubblici, e risponde agli impegni assunti nel 2011 con l'Unione europea, in vista della programmazione dei fondi strutturali 2014/2020. Il Sistema Nazionale di Valutazione (S.N.V.) ha lo scopo di dare al Paese un servizio fondamentale per poter aiutare ogni scuola a tenere sotto controllo gli indicatori di efficacia e di efficienza della sua offerta formativa ed impegnarsi nel miglioramento, di fornire all'Amministrazione scolastica le informazioni utili a progettare azioni di sostegno per le scuole in difficoltà e di valutare i dirigenti scolastici e offrire ai decisori politici la dovuta rendicontazione sulla effettiva identità del sistema di istruzione e formazione (Consiglio dei Ministri n. 72 del 8 Marzo 2013). Il S.N.V. si basa sull'attività dell'Invalsi (Istituto nazionale per la valutazione del sistema di istruzione e formazione), che ne assume il coordinamento funzionale; sulla collaborazione dell'Indire (Istituto nazionale di documentazione, innovazione e ricerca educativa), che può aiutare le scuole nei piani di miglioramento; sulla presenza di un contingente di Ispettori con il compito di guidare i nuclei di valutazione esterna.

Ogni singola scuola costruirà il proprio rapporto di autovalutazione secondo un quadro di riferimento comune e con i dati messi a disposizione dal sistema informativo del MIUR, dall'INVALSI e dalle

stesse istituzioni scolastiche. Il percorso si concluderà con la predisposizione di un piano di miglioramento e la rendicontazione pubblica dei risultati. Previste anche le visite dei nuclei esterni di valutazione. Sono oltre 1300 le istituzioni scolastiche che stanno già seguendo in via sperimentale questo percorso. Le istituzioni formative accreditate dalle Regioni verranno valutate secondo priorità e modalità stabilite in sede di Conferenza Unificata.

Gli assi portanti del S.N.V. sono quattro:

- a) autovalutazione delle istituzioni scolastiche, sulla base di un fascicolo elettronico di dati messi a disposizione dalle banche dati del sistema informativo del Ministero dell'istruzione, dell' INVALSI e delle stesse istituzioni scolastiche, che si conclude con la stesura di un rapporto di autovalutazione da parte di ciascuna scuola, secondo un format elettronico predisposto dall'Invalsi e con la predisposizione di un piano di miglioramento;
- b) valutazione esterna da parte di nuclei coordinati da un dirigente tecnico sulla base di protocolli, indicatori e programmi definiti dall'Invalsi, con la conseguente ridefinizione dei piani di miglioramento da parte delle istituzioni scolastiche;
- c) azioni di miglioramento con l' eventuale sostegno dell'Indire, o di Università, enti, associazioni scelti dalle scuole stesse;
- d) rendicontazione pubblica dei risultati del processo, secondo una logica di trasparenza, di condivisione e di miglioramento del servizio scolastico con la comunità di appartenenza.

4.2 Valutare per competenze

La Raccomandazione del Parlamento Europeo e del Consiglio dell'Unione Europea il 18 dicembre 2006 hanno fissato le *"competenze chiave per l'apprendimento permanente"*, che sono ritenute essenziali per garantire la realizzazione e lo sviluppo personali, la cittadinanza attiva, l'inclusione sociale e l'occupazione. Le competenze chiave sono state così delineate:

1. Comunicazione nella madrelingua;

2. Comunicazione nelle lingue straniere;
3. Competenza matematica e competenze di base in scienza e tecnologia;
4. Competenza digitale;
5. Imparare a imparare;
6. Competenze sociali e civiche;
7. Spirito di iniziativa e imprenditorialità;
8. Consapevolezza ed espressione culturale.

Tutti i documenti ufficiali nazionali ed europei ribadiscono continuamente l'importanza della promozione dello sviluppo delle competenze, ma paradossalmente non è ancora stata data una definizione univoca che sappia spiegare che cosa effettivamente sia la competenza. Eppure uno degli obiettivi che si propongono i sistemi di istruzione a livello mondiale è proprio quello di strutturare curricula fondati sull'acquisizione di competenze e finalizzati ad una certificazione delle competenze acquisite che sia condivisa e spendibile nella società tutta e nel mondo del lavoro in particolare.

Per avere un'idea più chiara del concetto di competenza possiamo, nuovamente, chiedere aiuto all'etimologia: l'aggettivo competente deriva dal tardo latino "competente", participio presente di competere, che significava in origine "concordare" e più tardi "addirsi", cioè essere adatto, quindi competente. Dunque alle origini di questo concetto c'è soprattutto un'idea di attività adeguata, idonea, adatta, proporzionata, opportuna. Si tratta, quindi, non solo di fare riferimento ad un contenuto, costituito dalle conoscenze e dalle abilità possedute da una persona, ma anche di sottolinearne l'adeguatezza rispetto ad una determinata situazione. L'individuo competente possiede conoscenze e saperi procedurali, ma al tempo stesso deve "saper essere" nell'esercizio della sua competenza. La competenza presuppone, quindi, una consapevole capacità di far interagire i saperi e le abilità possedute all'interno di un contesto in cui assumono uno specifico significato il sapere, il saper fare, il sentire, il saper agire e il saper essere.

Il D.M. del 22 agosto 2007, relativo all'estensione dell'obbligo formativo fino ai 16 anni, detta le *otto competenze chiave di cittadinanza* che tutti devono acquisire entro i 16 anni e che sono ritenute necessarie per la costruzione del sé, lo sviluppo di corrette e significative relazioni con gli altri, una positiva interazione con la realtà naturale e sociale. Le competenze chiave individuate sono:

1. Imparare ad imparare;
2. Progettare;
3. Comunicare;
4. Collaborare e Partecipare;
5. Agire in modo autonomo e responsabile;
6. Risolvere problemi;
7. Individuare collegamenti e relazioni;
8. Acquisire ed interpretare informazioni.

Una definizione di competenza che accontenti tutti gli addetti ai lavori, e che corrisponda in pieno ai vari aspetti che il termine possiede, non è ancora stata trovata [Sgalambro, 2002]. Manca ancora una definizione univoca in quanto ogni settore, per specifiche esigenze, definisce il termine competenza in modo diverso poiché differente è il modo di approcciarsi alla competenza stessa, sia dal punto di vista concettuale che applicativo. La Babele delle definizioni e delle applicazioni non deve necessariamente, però, essere vista in modo negativo, anzi potremmo osare affermare che consiste in una ricchezza grazie alla quale diversi settori, scuola, formazione professionale, mondo del lavoro, società, possono confrontarsi e completarsi a vicenda. Volere necessariamente trovare una definizione univoca risulterebbe, di fatto, una forzatura che nega l'aderenza alle specificità delle esigenze che si riscontrano nei diversi settori interessati [Petracca, 2000]. Del resto l'apprendimento stesso non rappresenta un semplice accumulo di conoscenze, ma soprattutto un interiorizzazione delle stesse e il transfert del loro uso a fini comportamentali. Socrate nel Teeteto [Platone, 1999]

sostiene appunto che la vera conoscenza è quella che si trasforma in coscienza di sé e degli altri.

Il mondo del lavoro investe sempre di più nel Knowledge Management, ovvero nella gestione della conoscenza, in un processo di organizzazione e sistematizzazione di quel vasto insieme di competenze, esperienze ed informazioni che possono garantire un buon livello di competitività a livello mondiale. Le organizzazioni per far fronte alla globalizzazione, ai repentini cambiamenti, all'incertezza che caratterizza il presente ed il futuro, devono maturare un atteggiamento proattivo, poiché non è più sufficiente avere solo la capacità di essere reattivi ai mutamenti, ma occorre prevederli, anticiparli e guidarli nella direzione più vantaggiosa per l'organizzazione stessa. La competenza che deve maturare il capitale umano, in questo contesto, è da ricercare nella capacità di essere al tempo stesso specializzati, polivalenti, dinamici e aperti alla condivisione. La chiave della crescita del capitale intellettuale sta proprio nella capacità di condividere poiché dalla condivisione delle competenze dei singoli nasce un patrimonio collettivo la cui crescita è praticamente illimitata. Nel mondo del lavoro è maturata la concezione dell'importanza del lavoro di gruppo poiché si è compreso che le competenze del gruppo stesso vanno ben oltre la somma delle competenze dei singoli individui che lo compongono poiché sono il risultato di un processo di integrazione, collaborazione e negoziazione continua.

Il mondo della formazione professionale ha compreso che il suo punto di forza deve essere quello di promuovere la formazione per tutto l'arco della vita, pertanto si preoccupa di formare non solo i giovani che desiderano immettersi nel mondo del lavoro, ma anche coloro che avvertono la necessità di stare sempre al passo con i tempi o ancora che si trovano ad affrontare dei periodi transazionali che richiedono un cambiamento del proprio ruolo in seno ad un'organizzazione o in nuovi contesti lavorativi. L'approccio al problema della competenza, in questo settore, è mirato soprattutto allo studio delle modalità più idonee alla certificazione delle competenze stesse attraverso la definizione di standard di competenze che possano essere documentati e spendibili. La formazione professionale ha, quindi, il compito di formare professionalità capitalizzabili che possiedano competenze di base necessarie a garantire una cittadinanza attiva, competenze specia-

lizzate in specifici settori professionali e ancora competenze trasversali, quali ad esempio il saper lavorare in gruppo, il saper analizzare ed affrontare l'imprevisto, il saper decentrare il proprio punto di vista.

Nel mondo della scuola il dibattito sulla competenza ha visto un altalenarsi di prese di posizione diverse e spesso contrastanti tra loro. Il termine competenza è relativamente nuovo al mondo della scuola, poiché già negli anni settanta, all'interno del dibattito sulla pedagogia per obiettivi, si discuteva se questi dovessero indicare le *performance*, intese come risultati scolastici, o le *competence*, intese come abilità, capacità ed atteggiamenti degli studenti. Oggi prevale la scelta di lavorare alla costruzione di curricula più complessi ed organizzati, mirati ad assicurare l'acquisizione di competenze di gestione dei processi di sviluppo posti in essere dalla persona, pertanto la finalità dell'azione educativa è quella di formare persone che siano in grado di comprendere l'aspetto metacognitivo dei propri apprendimenti e che siano consapevoli della propria identità personale e culturale. Le competenze indicano, quindi, la capacità di usare con responsabilità ed autonomia conoscenze, abilità e attitudini personali e sociali, in situazioni di lavoro, nella formazione professionale, nello sviluppo personale, nello studio.

La scuola dell'autonomia possiede uno strumento essenziale alla costruzione di un curriculum che aderisca pienamente alla realtà, che è il Piano dell'Offerta Formativa, il quale consente di strutturare percorsi formativi mirati all'acquisizione di competenze di base, specifiche e trasversali fortemente contestualizzate, in relazione ai vincoli e alle risorse di cui il territorio dispone, e al tempo stesso dinamiche e trasferibili a contesti diversi. "Nella nuova cornice dell'autonomia quindi, gli insegnanti, nelle vesti di maggiori protagonisti del processo di formazione, acquisiscono un ruolo determinante, in qualità di operatori socio-culturali che svolgono attività finalizzate alla mediazione culturale e alla promozione della formazione, non solo in direzione degli allievi, ma anche verso la società civile, sulla base di un piano di offerta formativa, tanto ampio e articolato da soddisfare una pluralità di esigenze formative" [Crivellari, 2004]. Nel campo dell'istruzione scolastica, e non meno nel settore della formazione e del mondo del lavoro, non si può prescindere dal considerare la persona nella sua olistica integrità poiché, come ci ricorda [Rogers, 2012] riferendosi

all'apprendimento, "nessuno apprende dal collo in su". Non si può, quindi, settorializzare la persona e curarsi di valutarne solo l'aspetto cognitivo o comportamentale o relazionale, e così di seguito. Possiamo sicuramente considerare una competenza come uno sviluppo graduale di competenze più semplici che progressivamente si integrano tra di loro per giungere alla maturazione di una competenza più complessa. Tuttavia non dobbiamo mai perdere di vista il fatto che è la persona, nella sua completezza e complessità, che matura una specifica competenza, che apprende con tutto il suo essere che è sì cognitivo, ma è anche affettivo, relazionale, motivazionale, corporeo, e che porta con sé uno specifico bagaglio culturale, esperienziale e sociale.

La competenza può allora essere vista, in questa prospettiva, come la punta di un iceberg in cui la parte del curriculum emersa è sicuramente più facilmente osservabile, documentabile, valutabile, certificabile, ma profondamente influenzata e determinata dalla sua parte latente, che è la parte sommersa dell'iceberg e che rappresenta tutto il mondo interiore della persona.



Figura 4.1: Iceberg di Spencer & Spencer "Il curriculum esplicito e il curriculum latente"

Spencer and Spencer [2008], appunto, mettono in evidenza che la competenza costituisce una caratteristica intrinseca di un individuo,

che si compone di motivazioni, tratti, immagine di sé, ruoli sociali, conoscenze ed abilità. La competenza è, quindi, un prodotto complesso nato dall'integrazione di obiettivi, contenuti, metodi e valutazione, che potrebbe così fare riferimento al cosiddetto "curricolo esplicito", mentre sotto la superficie, dentro la parte sommersa dell'iceberg, troverebbe posto il cosiddetto "curricolo latente", o implicito, fatto cioè di tutto ciò che viene trasmesso dalla famiglia, dalla scuola, dalle agenzie educative e dal contesto socio-culturale più in generale, senza piena consapevolezza e in modo non programmato ed informale. Allo scopo di mettere in luce tutti i fattori di processo formativo che condiziona la realizzazione della competenza, occorre individuare gli strumenti più idonei per la valutazione sia del curricolo esplicito che di quello latente.

Quando l'insegnante assegna una prova di verifica allo studente per valutare il conseguimento di un apprendimento, la prestazione fornita nell'esecuzione dello specifico compito assegnato, si configura come indicatore del possesso della competenza segnalata da un descrittore. Il risultato conseguito però non da informazioni immediate su quello che succede all'interno della persona affinché si produca questa prestazione. L'esecuzione di un compito mette in moto processi di apprendimento che possono essere facilitati, ottimizzati e velocizzati tramite la messa in opera di opportune strategie, che consentano un'interazione efficace tra i processi impliciti e l'esplicitazione degli apprendimenti. Ogni studente in relazione ai propri stili personali di apprendimento attiva le proprie strategie cognitive che gli consentono di acquisire nuove conoscenze. Queste stesse strategie, tuttavia, sono condizionate dal curricolo latente, costituito da aspetti molto profondi della persona, spesso inconsci, relativi alla capacità di linguaggio, di comunicazione, di cultura, alla percezione del sé, alla motivazione. Il curricolo latente, quindi, condiziona in modo decisivo la realizzazione del curricolo esplicito.

L'insegnante, somministrando una prova di verifica, valuta il risultato di un apprendimento, un prodotto, ma non ha informazioni immediate sul processo che a quel prodotto sottende. Per valutare un processo formativo non sono sufficienti pertanto le sole prove di verifica strutturate ma ad esse va associata una valutazione qualitativa che, con l'uso di metodologie specifiche, possa renderlo esplicito,

comprensibile e valutabile.

4.3 Le competenze e il curricolo scolastico

Pellerey [2004] definisce la competenza come la capacità di mettere in moto e di coordinare le risorse interne possedute, conoscenze, abilità, attitudini, con quelle esterne per affrontare positivamente situazioni sfidanti. La competenza emerge allora come “un’orchestrazione di un insieme di schemi cognitivi, sociali, culturali, affettivi, prasseologici” [Varisco, 2004] che, integrando le esperienze passate, si compenetrano e accrescono il loro livello di complessità e le loro inter-correlazioni.

Capacità, conoscenze, abilità, competenze, sono termini molto utilizzati nella letteratura pedagogica, tuttavia nel mondo della scuola permangono alcune difficoltà nell’univoca interpretazione dei termini stessi che emergono se si analizzano le programmazioni annuali di molti docenti o i giudizi finali di valutazione. Occorre, pertanto soffermarsi un momento a riflettere sul loro significato onde evitare ambiguità nella trattazione dei concetti stessi. La capacità è rappresentata dalle potenzialità a fare, pensare, agire che sono proprie dell’essere umano. È una propensione attiva che ogni essere umano ha in quanto tale; tutti gli esseri umani possiedono, infatti, potenzialità che sono comuni a tutta la specie e, al tempo stesso, ciascuno ne possiede di specifiche, particolari e determinate che lo rendono un essere unico. Il compito di chi si occupa e si preoccupa di fare di un bambino un uomo, dai genitori, agli insegnanti, agli educatori, è quello di fare in modo che ciascuno espliciti al meglio e in modo armonico tutte le sue potenzialità, valorizzandole nella loro specificità poiché *ciascuno di noi è solo se stesso: un’irripetibilità* [Bertagna, 2006]. Le capacità di un essere umano non sono entità statiche che ciascuno possiede e che aspettano solo di esplicitarsi naturalmente, ma sono dinamiche ed in continua evoluzione e devono essere stimulate in modo opportuno affinché si traducano in comportamenti osservabili. Un bambino nasce con la capacità della parola ma non emetterà alcun suono se lo si lascia crescere nella solitudine e nel silenzio. Affinché possa divenire un adulto capace di esprimere le proprie opinioni con senso critico e rispettoso di quelle altrui, deve essere guidato nel lungo processo

che va dalla lallazione alla dialettica. La conoscenza riguarda il sapere, teorico e pratico:

- i saperi dichiarativi, cioè le informazioni con carattere di generalità, stabilità, capacità generativa di altre conoscenze; il sapere che cosa (*know that*);
- le conoscenze condizionali che sono espresse da principi, regole, dall'etica individuale e collettiva, dai valori; il sapere che cosa si deve fare e perché lo si deve fare (*know-where/when/why*);
- la padronanza delle conoscenze tecnico-procedurali, all'interno di un processo che sia contemporaneamente sviluppo di conoscenza e strategia di esecuzione; il sapere che cosa e sapere come si fa (*know that e know how*).

A scuola i docenti verificano se un allievo possiede conoscenze e a che livello, attraverso la misurazione degli apprendimenti. L'abilità si riferisce al saper fare: non solo al fare, quindi, ma appunto anche al sapere le ragioni e le procedure di questo fare. In altre parole, anche al sapere perché operando in un certo modo e rispettando determinate procedure si ottengono certi risultati piuttosto di altri. È al tempo stesso un sapere descrittivo, esecutivo e critico. Quando un insegnante misura l'acquisizione di un apprendimento, al tempo stesso, valuta le abilità, poiché per superare positivamente una prova di verifica non basta possedere le sole conoscenze ma occorre saperle applicare ed essere in grado di effettuare delle scelte. Viceversa, quando si vuole verificare un'abilità, come ad esempio la stesura di un progetto, si valutano anche i contenuti e le procedure. L'insieme delle conoscenze e delle abilità si può dire che costituisce la cultura.

La competenza è l'insieme delle capacità potenziali di ciascuno portate effettivamente al miglior compimento e perfezionamento; indicano cioè quello che, in termini di atteggiamenti e comportamenti osservabili, nell'unità della nostra persona, siamo effettivamente in grado di fare, pensare e agire, dinanzi alla necessità di affrontare e risolvere situazioni problematiche diverse delle quali facciamo esperienza. Le competenze sono al tempo stesso specifiche, poiché indirizzate alla risoluzione di una situazione concreta, e integrate, poiché ogni competenza, anche se altamente specializzata, mobilita sempre anche

tutte le altre. Le competenze, infatti, sono organizzate secondo una logica sistematica e creativa. Si può quindi affermare che è competente chi dà sempre tutto il meglio di se stesso nell'affrontare ed eseguire un compito specifico e, a questo scopo, mobilita in maniera integrata e unitaria tutte le proprie risorse interne (le capacità, conoscenze, abilità, abitudini, attitudini) ed esterne (i colleghi di lavoro o i compagni, le persone di riferimento, i luoghi e strumenti di documentazione che ha a disposizione), conferendo loro unitarietà. Chi è competente possiede un pensiero astratto che gli consente di effettuare transfert e di cogliere l'aspetto metacognitivo dell'apprendimento. Una competenza si dimostra soltanto esercitandola in un contesto dato, attraverso un'effettiva prestazione, pertanto è osservabile, verificabile e certificabile.

Le *Nuove Indicazioni per il Curricolo* in più punti ribadiscono l'importanza del ruolo della scuola nello sviluppo delle competenze legate al sapere (conoscenza dichiarativa), al saper fare (abilità che dipendono dalla conoscenza procedurale), al saper essere (sapere derivante da cultura, personalità, caratteristiche individuali) e al saper apprendere (consapevolezza del proprio stile di apprendimento). Le finalità della formazione autentica devono essere individuate proprio a partire dal soggetto che apprende, tenendo sempre conto della singolarità e della complessità di ogni persona. Lo studente è, quindi, posto al centro dell'azione educativa e il docente, al fine di garantirgli un adeguato processo formativo, deve progettare gli obiettivi, le metodologie e le strategie di intervento educativo a cominciare da un'accurata valutazione delle competenze possedute e delle aree prossimali di sviluppo che possono venir attivate. La valutazione, infatti, come abbiamo più volte ribadito, precede, accompagna e segue i percorsi formativi.

Quando si è chiamati a valutare una competenza, si deve verificare una complessità che non riguarda solo il prodotto, cioè risultato finale atteso al termine di un percorso di formazione, ma che coinvolge tutto il percorso stesso, cioè il processo attraverso cui si realizza gradualmente l'apprendimento. Nella valutazione delle competenze fondamentale è il ruolo che assume lo studente, poiché è egli stesso che deve divenire protagonista consapevole del proprio processo formativo e prenderne coscienza.

L'autovalutazione è, come vedremo, il cardine su cui si costruisce un portfolio delle competenze. L'autovalutazione non deve, tuttavia

assumere i connotati di una chiusura nei confronti di giudizi esterni: il giudizio personale ha anzi bisogno di confrontarsi con quello dei propri pari, dell'insegnante e, volendo, anche di esaminatori esterni. Si tratta di verificare le proprie percezioni di competenza con quelle risultanti dal giudizio degli altri. La scuola deve porsi come obiettivo del curriculum scolastico la maturazione di competenze cognitive, che dimostrino che lo studente possiede un apprezzabile quadro di conoscenze ben organizzato, operative, che dimostrino abilità nello svolgimento di un compito concreto, affettive e motivazionale, poiché l'apprendimento è un processo attivo che necessita di volontà, di persistenza, di capacità di affrontare anche l'insuccesso. Una programmazione curricolare strutturata per il conseguimento di competenze deve, quindi, mirare all'acquisizione da parte dell'alunno di maturare la capacità di usare consapevolmente ed efficacemente le conoscenze in rapporto a contesti significativi, che non riguardano solo prestazioni riproduttive, ma anche la soluzione di problemi concreti.

Le Indicazioni parlano di traguardi di sviluppo della competenza, quindi la competenza, è considerata come qualcosa che può essere incrementata. I traguardi sono proposti come terminali, si riferiscono cioè alla fine della scuola dell'infanzia, della scuola primaria e della scuola secondaria di primo grado; quindi viene riconosciuta una stretta correlazione tra le tappe di sviluppo cognitivo e il raggiungimento di un dato livello di competenza. La strutturazione del curriculum è delineata unitariamente e prevede una definizione graduale che parte dai campi di esperienza per la scuola dell'infanzia per arrivare alle aree disciplinari per la scuola primaria e la scuola secondaria di primo grado. I campi di esperienza e le discipline costituiscono la cornice entro la quale vengono definiti i traguardi di sviluppo delle competenze, gli obiettivi di apprendimento e le indicazioni per la valutazione. Gli obiettivi si strutturano come intermediari per assicurare lo sviluppo delle competenze. Nelle Indicazioni sono suggerite metodologie didattiche che possono essere interpretate come misure per garantire il passaggio dagli obiettivi alle competenze, evitando la frammentazione dei saperi. Alle singole scuole spetta l'impegnativo compito di tradurre il quadro proposto dalle Indicazioni in realtà educativa. Viene sottolineata l'importanza di un insegnamento disciplinare non frammentato, ma capace di far cogliere le interconnessioni tra i diversi

saperi e di avviare gli alunni ad una visione unitaria della conoscenza. Si favorisce, quindi, l'inter-disciplinarietà e il lavoro collegiale tra insegnanti di discipline diverse. La lista degli obiettivi, rispetto ai programmi precedenti, è stata sfoltita e presenta linee più generali senza andare nei dettagli. Essa fa riferimento agli orientamenti europei, ma è anche attenta ai contenuti più significativi della nostra tradizione culturale. La didattica delle Indicazioni, senza trascurare una solida competenza strumentale, intende accrescere l'autonomia di pensiero, di studio e di apprendimento dell'alunno.

Le nuove Indicazioni intendono essere rispettose dell'autonomia degli insegnanti: definiscono i criteri che una buona proposta didattica deve rispettare, ma non prescrivono in modo dettagliato e minuzioso come devono lavorare gli insegnanti. I criteri sostituiscono le "ricette" poiché sono rivolti ad una scuola più matura e consapevole che, proprio per la libertà di cui può disporre con l'autonomia, viene ulteriormente responsabilizzata.

4.4 Le scale di misura nella valutazione scolastica

Per misurare in modo efficace gli apprendimenti e garantire trasparenza, lo strumento di misura deve essere adatto allo scopo e il concetto stesso di misura deve essere condiviso da docenti, genitori ed alunni. Una misura è tale se è il risultato del confronto di un dato osservato con una posizione identificabile su una scala. L'operazione preliminare alla misurazione vera e propria, pertanto, è l'esplicitazione della scala di cui si serve chi valuta. La teoria della misurazione consente di distinguere tra quattro tipi di scale, che sono, dalla più semplice alla più complessa, la scala nominale, la scala ordinale, quella a intervalli, e quella di rapporti. Quest'ultima scala che, partendo da un punto di valore zero, ci consente di determinare se una quantità posta su essa è multipla di un'altra, tuttavia, non è utilizzabile in campo educativo per misurare il raggiungimento di obiettivi cognitivi, sia perché risulta inverosimile determinare la totale assenza di conoscenza in un soggetto, sia perché al comportamento umano, caratterizzato da una varietà di differenze individuali, è pressoché impossibile attribuire una misura assoluta alla quale la letteratura in materia riconosca

concordemente un significato quantitativamente univoco[Di Battista, T., 2012].

La scala nominale

La *scala nominale* è chiamata così proprio perché si costruisce attribuendo “nomi” a determinate qualità che vengono assunte come criterio di riferimento. Largamente usata nelle indagini in campo formativo, essa consente di operare classificazioni sulla base della presenza o dell’assenza di una qualità considerata, senza dar luogo a graduazioni che stabiliscano la misura maggiore o minore della qualità stessa. Questo tipo di scala è utile per classificare gli studenti in relazione al possesso o meno di una certa caratteristica; un esempio può essere la classificazione in funzione del genere, della composizione del nucleo familiare, del titolo di studio pregresso, ecc. Questo tipo di scala, quindi, consente la collocazione in una categoria ma non la registrazione di possibili gradazioni di una specifica qualità. La scala nominale permette, quindi, di registrare *caratteri qualitativi sconnessi*, cioè non ordinabili, quali ad esempio la cittadinanza, il credo religioso, il titolo di studio dei genitori e così via. Valutazioni di questa natura trovano largo impiego nello studio di relazioni tra variabili esaminate, si pensi ad esempio all’analisi dei livelli di apprendimento in un’ottica di genere, ma non danno indicazioni della gradualità, cioè delle differenze nel possesso di una conoscenza tra studenti diversi o rispetto ad un livello considerato standard di riferimento. In una scala nominale si registreranno soltanto le frequenze assolute, ossia la numerosità dei casi che rientrano in ciascuna classe.

Quando questo tipo di scale viene utilizzato per una valutazione scolastica può, inoltre, essere utile calcolare anche in che proporzione, frequenza relativa, o in che percentuale gli studenti valutati rientrano in una categoria piuttosto che in un’altra. Come misura di tendenza centrale la scala nominale utilizza esclusivamente la moda che individua la categoria con frequenza più alta o percentuale più alta. Questi dati possono essere utili se utilizzati per effettuare confronti con gruppi anche eterogenei, in questo caso è necessario calcolare indici di relazione tra le variabili di un fenomeno come la distribuzione binomiale, quella del chi quadro che permette di verificare se siano

o no significative le differenze di frequenza registrate in due o più campioni, ad esempio in classi parallele di uno stesso istituto.

La scala ordinale

La *scala ordinale* consente, rispetto a quella nominale, di stabilire delle graduatorie d'ordine, cioè delle relazioni di maggioranza e di minoranza che si riferiscono ad una determinata qualità osservata. Le diverse categorie prese in esame dall'indagine, se distribuite su una scala ordinale, non staranno tutte su uno stesso piano, ma saranno ordinate gerarchicamente in relazione ad un valore che viene arbitrariamente attribuito rispetto alla proprietà considerata. Un esempio in campo valutativo può essere quello di ordinare gli studenti di una classe in relazione al voto ottenuto in una disciplina, espresso con aggettivi o con quantità numeriche. Il fatto che le categorie siano ordinate non ci assicura, però, che le distanze tra esse siano uguali. Questo è un dato di fatto di non poca rilevanza in campo educativo poiché i sistemi di valutazione utilizzati, indipendentemente se siano voti numerici, come prevede la normativa vigente espressi da 0 a 10, o giudizi sintetici (non sufficiente, sufficiente, buono, distinto, ottimo), ci danno indicazioni su un ordinamento del livello di apprendimento fortemente contestualizzato ma non garantiscono scientificamente quale sia la distanza dell'uno dall'altro o, ancora, quale valore attribuisca ad ognuno di essi il singolo docente. In parole povere, allo stato attuale delle cose, nella scuola italiana l'attribuzione di giudizio non è supportata da verifiche oggettive sottoposte ad un processo di standardizzazione, pertanto non è possibile determinare né se tutti i docenti attribuiscono ad un voto lo stesso valore, né se abbiano la medesima percezione circa la distanza tra i singoli. A livello di scala ordinale, infatti, possiamo applicare la proprietà dell'ordinamento tra le categorie, ovvero possiamo dire che, rispetto alla caratteristica misurata, una persona che in graduatoria ha una posizione x , ha un valore più elevato rispetto ad una persona in posizione $x - 1$, e che quest'ultima ha un valore più elevato rispetto ad una persona in posizione $x - 2$. Inoltre se $x > x - 1$ e $x - 1 > x - 2$, se ne deduce che $x > x - 2$, nota come proprietà transitiva. In questa scala, però, non siamo in grado di quantificare la distanza tra il valore x e il valore

$x - 1$, e non siamo in grado di dire se tra x e $x - 1$ da un lato, ed $x - 1$ e $x - 2$ dall'altro vi sia la stessa distanza.

La scala rappresentata dai voti scolastici ha assunto connotazioni molto particolari poiché l'ordine che gli studenti assumono nella valutazione degli apprendimenti è altamente influenzato da componenti soggettive e non è costante, ma varia in relazione a diversi fattori: al peso che chi valuta dà ad ogni voto, al tipo di prova di verifica, a fenomeni che possono influenzare il giudizio, al contesto socio-culturale, ecc. Usualmente l'insegnante attribuisce un voto più alto alla prestazione migliore e un voto basso ad una prestazione più scarsa, tuttavia un primo limite sta nel fatto che non tutti gli insegnanti utilizzano lo stesso range di voti, che solo teoricamente va da 0 a 10, inoltre occorre chiedersi se, ad esempio, tra il voto 8 e il voto 9 esiste lo stesso scarto che c'è tra il voto 5 e il voto 6. Quando siamo in grado di dare questa informazione, infatti, siamo in presenza di un livello di misurazione che può essere garantito dalla scala ad intervalli. L'utilizzo di una misurazione su scala ad intervalli presuppone, però, procedure di standardizzazione dei punteggi poco note al corpo docente, se non addirittura completamente sconosciute. Di conseguenza nella prassi quotidiana della valutazione il docente attribuisce un voto scolastico esclusivamente basandosi sulla propria percezione di giudizio, con la convinzione di utilizzare una scala metrica che consente attraverso un semplice calcolo di media aritmetica di esprimere un giudizio globale su una specifica disciplina. Questa prassi consolidata nella tradizione scolastica italiana comporta distorsioni importanti dell'atto valutativo che non trovano alcun sostegno né scientifico né epistemologico, allorché si utilizzano pratiche statistiche inadeguate ed inopportune. Il solo fatto di utilizzare variabili categoriali ordinate, quali sono i voti scolastici, per definire indici di sintesi come la media aritmetica comporta un errore grave e grossolano nell'utilizzo della disciplina statistica, eppure i Sistemi di Istruzione hanno regolamentato e fondato la valutazione scolastica su queste pratiche. È nell'impossibilità della scala ordinale di definire intervalli fissi, che rende assolutamente erroneo il suo attuale utilizzo nella valutazione scolastica, eppure alla fine di ogni quadrimestre e di ogni anno scolastico, in tutte le scuole di ogni ordine e grado, viene attribuito un giudizio finale espresso da un unico voto, quale risultato della media dei voti conseguiti

negli apprendimenti durante il corso dell'anno e dopo farraginosi arrotondamenti degli ancor più orrendamente diffusi "+", "++", "-" e "--". Quest'uso distorto dell'applicazione di una misura statistica alla valutazione è, purtroppo, largamente diffuso in ambito scolastico, e ancor più grave è il fatto che venga utilizzato come sintesi di una misurazione di caratteri eterogenei, come, ad esempio, i risultati conseguiti in una prova scritta e in una prova orale, o ancora tra discipline epistemologicamente assolutamente diverse. Sostanzialmente fare una media aritmetica tra voti espressi con numeri non è concettualmente diverso dal fare la stessa cosa con aggettivi qualificativi: che media c'è tra un sufficiente e un buono? Per contrastare questa pericolosa equivocità, la valutazione scolastica dovrebbe, come anticipato poc'anzi, utilizzare prove di verifica oggettive basate esclusivamente su procedimenti metodologici rigorosi, ma questo comporterebbe un importante investimento nella formazione valutativa dell'intero corpo docente che, viste le attuali politiche volte al risparmio in campo di istruzione, risulterebbe quantomeno utopico.

La scala ad intervalli

La *scala ad intervalli*, nella valutazione degli apprendimenti, risulta essere la più valida poiché possiede la caratteristica di ordinamento e il vantaggio che l'intervallo fra i valori distribuiti su essa rimane costante per tutta la sua estensione. Un'altra caratteristica delle scale ad intervalli, a vantaggio dell'utilizzo in campo formativo, è che lo zero non rappresenta la mancanza della qualità che si va ad osservare, ma un livello di riferimento convenzionalmente definito. Per stabilire a quale punto della scala corrisponde un punteggio grezzo ottenuto in una prova di verifica, il modo più semplice è quello di procedere alla standardizzazione dello stesso attraverso la definizione dei *punti z*. La standardizzazione è, quindi, un procedimento che consente di normalizzare una variabile, riconducendo una variabile aleatoria distribuita secondo una media μ e varianza σ^2 , ad una variabile aleatoria con distribuzione "standard", ossia di media zero e varianza pari a 1. Si tratta di un procedimento abbastanza semplice da attuare da qualsiasi docente ed è particolarmente utile nel caso della variabile casuale normale per il calcolo della funzione di ripartizione e dei

quantili con le tavole della normale standard. Il procedimento prevede di sottrarre alla variabile aleatoria la sua media e dividere il risultato ottenuto per la deviazione standard, utilizzando la formula utile a trovare i punti zeta (z-score o standard score):

$$z = \frac{X - \mu}{\sigma}$$

Questa tipologia di scala consente di determinare con maggior precisione in che misura uno studente possieda una specifica abilità o conoscenza. Il vantaggio principale, inoltre, sta nel fatto che, trasformando i punteggi grezzi e collocandoli su una scala ad intervalli, si possono effettuare confronti dettagliati fra i risultati conseguiti dagli alunni di classi diverse o in differenti prove.

Possiamo dunque affermare che la scala dei voti scolastici, utilizzata quotidianamente dai docenti nell'atto valutativo, rappresenti una sorta di scala ibrida che, pur presentando alcune delle caratteristiche delle succitate scale, tuttavia non è paragonabile in senso stretto a nessuna di esse. Non è una scala ad intervalli poiché è impossibile stabilire se la distanza tra un voto e il suo precedente sia identica a quella tra lo stesso voto e il successivo, motivo per cui spesso gli insegnanti utilizzano una serie di mezzi punti o simboli per indicare le variazioni di intervallo tra un voto e l'altro. Si potrebbe, quindi, dire che si tratta di una scala ordinale poiché le valutazioni si presentano in un ordine progressivo crescente, ma occorre precisare che la scala di valutazione utilizzata in ambito scolastico presenta particolarità ben precise che statisticamente parlando non le consentono di rientrare a pieno titolo neanche nella scala ordinale. L'intervento di fattori soggettivi che influenzano la valutazione scolastica, infatti, determina una non unicità del valore attribuito ai voti da insegnanti diversi, e ancora bisogna tener presente che ogni atto valutativo è fortemente influenzato dal contesto nel quale si esplica. La scala dei voti può quindi essere definita come una scala ordinale speciale, ossia una scala ordinale non assoluta. In un siffatto sistema l'attendibilità e la validità della valutazione sono strettamente correlate al livello di professionalità impiegato dal docente nell'atto di valutare. I docenti, nel modo più obiettivo possibile, dovrebbero valutare l'andamento delle diverse verifiche al fine di analizzare il decorso del processo formativo

e, soprattutto, dovrebbero esimersi dall'errore di basare il giudizio finale su medie aritmetiche applicate a dati categoriali ordinati, che falsano i risultati e rendono inefficaci gli interventi regolativi attuati in relazione ad essi.

4.5 Metodi qualitativi della valutazione

La CM 84 del 10 novembre 2005 mette in evidenza che accertare e certificare la competenza di una persona richiede strumenti caratterizzati da accuratezza e attendibilità che, a differenza di quelli utilizzati per valutare soltanto la padronanza delle conoscenze e delle abilità, eccedono, senza escluderle, le consuete modalità valutative scolastiche disciplinari (test, prove oggettive, interrogazioni, saggi brevi, ecc.), ma richiedono anche osservazioni sistematiche prolungate nel tempo, valutazioni collegiali dei docenti che coinvolgano anche attori esterni alla scuola, a partire dalla famiglia, autovalutazioni dell'allievo, diari, storie fotografiche e filmati, coinvolgimento di esperti e simili. Il livello di accettabilità della competenza manifestata in situazione scaturisce dalla somma di queste condivisioni e coinvolge nella maniera professionalmente più alta i docenti che si assumono la responsabilità di certificarla.

Oggi le attività di valutazione nella scuola, pur fondate prevalentemente sulla base dei dati quantitativi assunti con la verifica e la misurazione, coinvolgono una gamma più ampia di operazioni di analisi, di considerazioni e di interpretazioni qualitative dei fattori che hanno prodotto gli esiti registrati con gli strumenti di rilevazione usati. La ricerca metacognitiva evidenzia che la maggior parte degli apprendimenti si attua per mezzo della trasmissione culturale ad opera della famiglia, della scuola e delle altre agenzie educative, che fungono da mediatori culturali tra il soggetto che apprende e la realtà in cui avvengono gli apprendimenti. Gli insegnanti devono attuare strategie di intervento mirate a presentare uno stesso concetto sotto diverse forme, costruendo percorsi formativi inter e pluri-disciplinari, in tal modo lo studente è messo in condizione di interiorizzare e generalizzare i concetti acquisiti. Gli apprendimenti per essere significativi devono essere ancorati alle conoscenze possedute ed agire sulla motivazione ad apprendere. Altro aspetto, non meno importante, a garanzia di

un apprendimento di qualità è quello metacognitivo che presuppone una riflessione sulle conoscenze che gli alunni stessi hanno del proprio funzionamento cognitivo, in tal modo si favorisce la capacità di autovalutazione.

La scuola quando valuta il percorso formativo di un alunno non può, quindi, limitarsi a misurare il livello di padronanza degli apprendimenti, che si può ottenere da una valutazione di tipo quantitativo, ma deve attivare procedure specifiche mirate alla valutazione di atteggiamenti quali la motivazione, la curiosità intellettuale, il desiderio di riuscita, il pensiero creativo, il senso di responsabilità, l'impegno, l'accettazione della diversità, e di competenze trasversali quali la capacità di sfruttare le informazioni per risolvere problemi, di esercitare il proprio pensiero critico, di mettere in opera il proprio pensiero creativo, di sviluppare la propria identità personale, di saper lavorare in collaborazione con altri. Queste competenze necessitano di una valutazione di tipo qualitativo.

Accanto alla ricerca nomotetica che, come abbiamo visto, cerca di pervenire a leggi generali e che, attraverso l'uso di procedure statistiche, indaga e misura le caratteristiche degli apprendimenti, si è affermata la ricerca ideografica mirata alla valorizzazione dell'unicità dei singoli individui, in relazione agli specifici contesti. Ogni volta che si esprime un giudizio sull'efficacia, sugli stili di apprendimento, sulla sfera affettiva e motivazionale, sulle competenze sociali e relazionali, non è sufficiente utilizzare i soli metodi quantitativi della valutazione. Sono necessari, allora strumenti valutativi diversi, quali l'osservazione, l'intervista, la verbalizzazione, il racconto ed altri che analizzeremo nel dettaglio.

I metodi della valutazione qualitativa si distinguono in non strutturati e strutturati. I metodi non strutturati costituiscono il primo livello della valutazione qualitativa poiché sono completamente aperti e permettono un'analisi di tipo descrittivo della situazione, che ha lo scopo di individuare gli aspetti e i parametri importanti da considerare. Sono tipicamente la descrizione analitica e non valutativa, la registrazione, la videoregistrazione, l'intervista, il racconto, la verbalizzazione. La valutazione qualitativa si ferma a questo livello se lavora con bambini molto piccoli poiché con loro è importante essere aperti a tutte le indicazioni che emergeranno dall'evoluzione dinamica

della situazione. Per cogliere completamente gli aspetti vari di una situazione complessa, i pedagogisti consigliano che la persona che registra o descrive non sia la stessa che conduce l'attività: questo perché chi conduce l'attività ha necessariamente un programma in mente o almeno delle aspettative e quindi tende ad applicare una chiave di lettura non completamente priva di preconcetti. Il primo livello di valutazione qualitativa può essere applicato utilmente nella fase di avvio di un percorso formativo per compiere una prima classificazione all'interno del gruppo degli studenti. Inoltre sarebbe bene tornare sempre a questo livello quando emergono aspetti nuovi ed imprevisti, che richiedono un'analisi non ancora codificata. anche se questo capita nel mezzo di un'attività ben collaudata o con una classe che si pensa di conoscere bene. Può essere utile nel primo periodo di scuola mantenere un diario di bordo per registrare in modo informale le attività svolte, le risposte degli alunni, le prime impressioni. Questo strumento consente di individuare quelli aspetti che necessitano di osservazioni e descrizioni più dettagliate e strutturate con metodo scientifico.

I metodi strutturati, invece, richiedono di limitare il campo di osservazione in modo da approfondire solo alcuni aspetti di interesse specifico, in base a quanto emerso dall'analisi non strutturata o da conoscenze già acquisite. Il vantaggio di scendere più in profondità consente di far cogliere più facilmente eventuali relazioni, operare confronti fra situazioni diverse o apprezzare l'evoluzione nel tempo. Poiché c'è sempre il rischio di perdere di vista aspetti importanti, oppure di non cogliere aspetti nuovi, non previsti nell'impostazione iniziale, è importante reinserire l'analisi non strutturata ogni volta che si capisce di essere di fronte a qualche aspetto rilevante non previsto. Analizzeremo nel paragrafo 4.5.1 i metodi strutturati più diffusi nella valutazione qualitativa.

4.5.1 L'osservazione

L'osservazione è un'attività che comunemente svolge ogni individuo, è l'attività che quotidianamente ognuno di noi compie, in forma non intenzionale il più delle volte, per l'analisi delle informazioni che i nostri organi di senso raccolgono. In campo educativo, invece,

L'osservazione deve assumere carattere di scientificità per consentire il raggiungimento di specifici obiettivi. Pertanto deve essere programmata e registrata sistematicamente, consentendo verifiche e controlli per attestare la validità e la verificabilità dei dati raccolti. L'osservazione in campo educativo non può essere occasionale né tantomeno soggettiva, poiché le informazioni ricavate risulterebbero lacunose, imprecise e fortemente influenzate da interferenze e generalizzazioni, dovute a stereotipi e pregiudizi di chi osserva. L'osservazione sistematica ha lo scopo di analizzare un fenomeno per descriverne il carattere, senza intervenire direttamente per modificarlo. Questo tipo di osservazione viene utilizzata intenzionalmente all'interno del processo di insegnamento-apprendimento, utilizzando schemi di riferimento che consentano classificazioni e categorizzazioni dei fenomeni osservati.

Nella ricerca psicopedagogica e sociale si è fatto ricorso a forme svariate di osservazioni sistematiche che possono essere differenziate sia in relazione alle strategie utilizzate per la raccolta delle informazioni, sia al livello di coinvolgimento dell'osservatore. In ogni caso è essenziale che l'osservazione sia organizzata e metodica, che individui bene il fine a cui mira e che utilizzi strumenti adeguati. Il docente che osserva nella scuola sa che la realtà da lui osservata è stata programmata secondo un piano educativo complesso, sa che questa realtà ha come meta il conseguimento di determinati obiettivi educativi e sa inoltre che l'osservazione stessa fa parte di questa realtà programmata. L'osservazione rappresenta una fase fondamentale del processo di insegnamento-apprendimento, che non coincide con un unico specifico momento ma che precede e accompagna tutto il processo stesso. L'osservazione pre-didattica è utile per la costruzione del profilo iniziale di ciascun allievo; intesa come fase conoscitiva preliminare della programmazione, può indirizzare, infatti, i docenti nella scelta delle metodologie e degli strumenti più adeguati per realizzarla e nell'individuazione dei criteri di verifica e di valutazione del processo educativo stesso. Quella intermedia, invece, supporta l'intervento educativo dell'insegnante che viene messo nelle condizioni di arricchire e potenziare le capacità dei soggetti. L'osservazione in ambito scolastico assume carattere di scientificità quando serve, appunto, alla realizzazione di obiettivi, quando è programmata, registrata sistematicamente, messa in relazione con l'ipotesi iniziale, e soggetta a prove e

a controlli.

L'osservazione può essere diretta o indiretta: nella prima, l'osservatore è nel campo di azione del soggetto osservato e registra direttamente i comportamenti oggetto di analisi; nella seconda tra l'osservatore e il fenomeno da osservare si frappongono strumenti di osservazione (test, interviste, indagini tramite altre persone ecc.). Nella scuola il docente utilizza entrambe poiché, soprattutto quando è ad un primo approccio con il gruppo classe, ha la necessità di osservare direttamente gli studenti in situazioni organizzate e in momenti di attività libera, di impostare un dialogo con la famiglia e di usare strumenti opportunamente strutturati. L'osservazione si può effettuare in svariati campi, ad esempio può interessare uno specifico comportamento dell'alunno, il momento dell'inserimento, la relazione con i genitori, la socializzazione tra pari, o ancora accompagnare specifiche attività di apprendimento o ludiche. L'osservazione può avvenire negli ambienti scolastici in cui abitualmente si verificano i fenomeni che si intendono osservare (aula, palestra) o in un ambiente artificiale, ossia in una situazione precostituita per analizzare aspetti specifici, per esempio in laboratorio.

Ogni insegnante che si accinga ad attuare processi valutativi deve conoscere quei fattori della propria personalità che potrebbero influenzare l'atto valutativo e quindi determinare la non obiettività dello stesso. L'osservazione sistematica è diretta da un progetto, ha strumenti per tenere sotto controllo possibili errori o distorsioni, e il momento della registrazione non deve essere mai troppo lontano da quello dell'osservazione. Per prima cosa è necessario dire che è impossibile osservare tutto, consapevolmente o inconsapevolmente l'osservatore effettua una selezione dovuta a fenomeni percettivi, culturali, socio-affettivi. Per ovviare o comunque limitare questa preventiva selezione dei fatti osservati è necessario delimitare il campo di indagine, ovverosia individuare il problema o l'ipotesi che ci spinge all'osservazione. Inoltre il lavoro in équipe, il confronto di punti di vista diversi, il principio di triangolazione possono dare un sostanziale contributo all'oggettività delle osservazioni.

L'osservazione in campo educativo per essere uno strumento di lavoro deve possedere, come abbiamo anticipato, caratteristiche di sistematicità, ripetibilità, comunicabilità, abbandonando la casualità

e l'improvvisazione a vantaggio di una proposta educativa più consapevole e coerente. L'insegnante per effettuare l'osservazione può utilizzare una serie di strumenti scritti caratterizzati da livelli diversi di strutturazione che sono tipicamente utilizzati nella ricerca osservativa e sperimentale. Tra gli strumenti scritti analizzeremo *l'osservazione descrittiva carta e matita*, le *check-list*, le *scale di valutazione*, il *differenziale semantico*, la *griglia di osservazione*, i *test* e *prove di profitto*. Esistono, inoltre, strumenti di osservazione orale largamente utilizzati in campo educativo nella ricerca sperimentale di tipo qualitativo, come i questionari, l'intervista e l'intervista biografica, che però in campo scolastico non trovano un grande utilizzo poiché necessitano nell'organizzazione, nella somministrazione e nell'analisi di una notevole preparazione psicologica dell'intervistatore. Questi ultimi, pertanto, non rientreranno nel nostro campo di indagine.

L'osservazione descrittiva carta e matita è uno strumento di rilevazione di tipo narrativo e non strutturato che consiste nella rilevazione di comportamenti che vengono descritti senza alcun sistema di decodifica. L'osservazione può interessare un singolo alunno, un piccolo gruppo o l'intera classe. La registrazione può avvenire nell'esatto momento in cui si osserva un comportamento o un'abilità, oppure a posteriori. In ogni caso deve essere strutturata rispettando alcune indicazioni ben precise: occorre prima di tutto registrare la data, l'orario di inizio e di fine, il luogo di registrazione, l'attività svolta, l'esatta cornice di riferimento. I dati in una prima fase devono essere registrati ed in seguito analizzati e valutati, poiché la fase di raccolta dati e quella della loro interpretazione non devono coincidere. Nella registrazione dei dati occorre rilevare le relazioni sociali con adulti e coetanei, il rapporto con gli oggetti, le interazioni linguistiche e il comportamento non verbale.

Le *check-list*, come gli strumenti seguenti, rientrano invece tra i "metodi chiusi", ossia strutturati in relazione ad una pre-selezione rigida dei dati da registrare. Si tratta sostanzialmente di elenchi composti dalla descrizione di una serie di comportamenti dei quali si rileva, o meno, la presenza. È uno strumento di facile utilizzo, poco intrusivo e quindi molto adatto all'osservazione diretta in classe. La *check-list* consente di ottenere una prima descrizione del fenomeno osservato e di individuare gli aspetti per i quali è opportuno approfondire l'in-

dagine, tuttavia non consente di stabilire in che misura e con quale costanza si verifica un comportamento.

Le *scale di valutazione* superano il limite che ha la check-list di non prendere in considerazione alcun aspetto di ordinamento, poiché permette, accanto alla rilevazione della presenza di un fenomeno, di operare su di esso una classificazione ed una valutazione. Sostanzialmente si tratta di un'operazione di stima che, utilizzando procedure standardizzate, permette di valutare il grado con cui un fenomeno presenta un determinato attributo. Le più comuni utilizzano scale di tipo numerico o grafo-numerico. Si tratta di utilizzare modelli di scala Likert, ideata dallo psicometrico americano Likert [1932] con lo scopo di elaborare un nuovo strumento, più semplice rispetto ad altri, per la misurazione di opinioni e atteggiamenti. Questo tipo di scala è molto diffusa e utilizzati in campo formativo, sia per l'osservazione di abilità e la rilevazione di livelli di competenza, sia per la raccolta di opinioni all'utenza rispetto ad un servizio offerto. La scala è costituita da una serie di affermazioni (item) semanticamente collegate agli atteggiamenti su cui si vuole indagare: ciascun item rileva lo stesso concetto sottostante, per questo motivo è una scala unidimensionale. Gli item sono presentati agli intervistati sotto forma di batterie. L'intervistato è chiamato ad esprimere il suo grado di accordo/disaccordo, gradimento/non gradimento, ecc. con ciascuna affermazione scegliendo, solitamente, tra quattro-sette modalità di risposta che esprimono, in modo crescente o decrescente, un certo ventaglio di opinioni. A ciascuna modalità di risposta viene attribuito un punteggio e la somma (media) dei punteggi alle risposte di ciascun individuo sull'intera batteria rappresenta la posizione dell'individuo sul concetto indagato. Per questo motivo la scala Likert è una scala additiva. Per fare un piccolo esempio concreto in campo formativo osserviamo quanto segue:

- Nel gioco di squadra rispetta i ruoli assegnati:

0 per niente
1 poco
2 abbastanza
3 molto

4 moltissimo

Il *differenziale semantico* è uno strumento di semplice utilizzo molto pratico da utilizzare per la somministrazione ad un intero gruppo-classe. Si basa sulla misurazione della percezione emotiva che gli alunni hanno rispetto ad alcune parole chiave. Il differenziale semantico si struttura nel modo seguente: si individua una parola chiave e si stabiliscono una serie di aggettivi, che non descrivono le caratteristiche del termine scelto, contrapposti tra loro e disposti alle estremità di una scala di sette intervalli. L'alunno deve indicare quanto la percezione emotiva di uno specifico aggettivo, in relazione al termine dato, si avvicina all'uno o all'altro polo della scala. L'intervallo centrale della scala indica la non risposta. La *griglia di osservazione* è uno strumento

"La scuola per me è..."		
	1 2 3 4 5 6 7	
dolce	-----	aspro
buono	-----	cattivo
brutto	-----	bello
debole	-----	forte
resistente	-----	fragile
stabile	-----	instabile
rapido	-----	lento
vago	-----	preciso
prevedibile	-----	imprevedibile

Figura 4.2: Esempio di differenziale semantico

di semplice costruzione e di compilazione immediata. Consiste sostanzialmente in una tabella a doppia entrata nella quale sono indicati i comportamenti da osservare in relazione all'elenco degli alunni della classe. Il docente compila la griglia secondo un criterio prestabilito che può andare dalla semplice verifica di presenza o assenza di un comportamento, registrabile semplicemente con l'uso di crocette, alla valutazione di un comportamento osservato in relazione ad una scala di valori. Si può chiedere, ad esempio, agli alunni di preparare un breve argomento da esporre ai compagni, anche con l'uso di materiale di supporto.

	CONTENUTO			ESPOSIZIONE					ASCOLTO		
	Esaustivo	Corretto	Approfondito	Frammentaria	Scorrevole	Termini specifici	Uso di supporti	Supporti esaurienti	Prende appunti	Chiede chiarimenti	Chiede approfondimenti
NOME											
—											
—											

Tabella 4.1: Esempio di una scheda di osservazione.

Per fare un esempio concreto, consideriamo la scheda in tabella 4.5.1. Questa scheda permette al docente di prendere velocemente nota del lavoro svolto dall'alunno rispetto alle conoscenze acquisite, delle capacità espositive, anche per quanto riguarda l'uso di supporti quali lucidi, schemi, immagini esemplificative, della partecipazione e dell'attenzione prestata al lavoro dei compagni e quindi dell'interesse per il percorso svolto e della capacità di comprendere, approfondire e confrontare con il proprio lavoro quanto detto da altri. Può essere compilata dall'insegnante con una scala di valori da 1 a 5 (1 = non sufficiente; 2=sufficiente; 3=buono; 4=distinto; 5 = ottimo). La stessa scheda può essere consegnata anche agli alunni per effettuare un'esperienza di co-valutazione.

I *test attitudinali* consistono in osservazioni rigidamente strutturate che consentono di valutare aspetti specifici che caratterizzano l'individuo. I test possono essere di tipo psicometrico, di abilità intellettuale, di tipo sociometrico. Le prove di profitto consentono di verificare il livello di apprendimento in specifici settori. Queste ultime, che consistono nelle prove di verifica strutturate degli apprendimenti, sono state già affrontate nella trattazione della valutazione di tipo quantitativo, pertanto adesso ci soffermeremo solo sui test. Il test prevede una serie di items che riguardano uno specifico aspetto della personalità dell'individuo e che possono essere somministrati a tutta la classe. In campo scolastico risulta molto utile l'utilizzo del test sociometrico per strutturare il sociogramma che individua la situazione relazionale del gruppo classe. Si tratta sostanzialmente di fotografare, attraverso un'analisi delle preferenze nella scelta di alcuni compagni (non più di tre) per svolgere specifiche attività (lo studio, lo svago, una gita, ecc.), il quadro relazionale e sociale del gruppo classe. Questo test consente di individuare se esistono alunni che hanno un ruolo da leader, se la classe è frammentata in sottogruppi, se esistono casi di isolamento dal gruppo classe. È molto interessante proporre il test in fasi diverse dell'esperienza scolastica di un gruppo classe, ad esempio ogni anno scolastico, per verificare come si modificano all'interno di esso le dinamiche relazionali con il passare del tempo.

4.5.2 La valutazione autentica

Il movimento della valutazione autentica o alternativa è sorto negli Stati Uniti agli inizi degli anni novanta come contrapposizione critica alla forma di valutazione diffusa di orientamento comportamentista fondata soprattutto su test standardizzati, strutturati per lo più a scelta multipla. La prospettiva di una valutazione alternativa in sostituzione di quella tradizionale è stata proposta, infatti, per la prima volta da Wiggins [1993]. Wiggins la definì valutazione autentica, poiché la finalità di questo tipo di valutazione è quella di verificare non solo ciò che uno studente sa, ma ciò che “sa fare con ciò che sa”, in quanto la valutazione si fonda sulla richiesta di una prestazione reale e adeguata dell'apprendimento.

A volte capita che alcuni studenti riescano bene nei test a scelta multipla o rispondano correttamente alle domande dell'insegnante in un'interrogazione orale, mostrando anche scioltezza nel ripetere ciò che si è appreso ma, quando viene richiesto loro di dimostrare ciò che sanno in una prestazione concreta, sembrano confusi e dimostrano una competenza da principianti. Capita di assistere addirittura a colloqui d'esame all'università nei quali il candidato espone gli argomenti come una robot ma a ad una richiesta di un esempio concreto si sentono smarriti. Proprio di recente ad un esame di statistica una studentessa dimostrava una preparazione impeccabile citando formule e concetti con spigliatezza e velocità, ma alla richiesta pratica del docente di ricavare semplici dati descrittivi in relazione ai voti ottenuti da soli tre alunni in una prova di verifica si è generato un imbarazzante silenzio. Molti studenti, infatti, in ogni ordine di scuola dalla primaria all'università, fondano il loro studio individuale su un apprendimento mnemonico, che gli consente di ottenere ottime performance in prove di verifica standardizzate ma che non ha efficacia nel permanere delle conoscenze a lungo termine e nella loro applicabilità pratica. Questo indica come la valutazione di concetti e di fatti isolati non dimostri le reali capacità di ragionamento, di creatività e di soluzione di problemi in situazioni concrete di vita.

Verificare che uno studente sia in grado di operare in contesti reali con prestazioni concrete che gli consentono di conseguire determinati obiettivi, fa comprendere molto di più sul processo del suo stesso

apprendimento che non ciò che egli dimostra in prove di verifica standardizzate, che ci danno un'indicazione di quanto sa e di come si colloca rispetto ai compagni nella quantità di conoscenze possedute. La valutazione autentica o alternativa si fonda, quindi, sulla convinzione che l'apprendimento scolastico non si dimostra con l'accumulo di nozioni, ma con la capacità di generalizzare, di trasferire e di utilizzare la conoscenza acquisita a contesti reali. Per questo nella valutazione autentica le prove sono preparate in modo da richiedere agli studenti di utilizzare processi di pensiero più complesso, più impegnativo e più elevato, in quanto si ritiene che, verificando con maggior autenticità l'apprendimento, si possano far raggiungere livelli più elevati di prestazione e preparare meglio gli studenti a un inserimento di successo nella vita reale. La valutazione autentica, infatti, cerca di promuovere e di valorizzare la specificità di ciascuno attraverso la cooperazione con il gruppo dei pari e con i docenti, affinché tutti e ciascuno siano messi in condizione di diventare cittadini attivi, autonomi, responsabili e consapevoli del fatto che il vero patrimonio che consente la crescita futura di un paese è il capitale umano, che si fonda sulla gestione e condivisione della conoscenza.

Il Professor Mario Comoglio, docente di didattica all'Università Salesiana di Roma, autore di diversi testi fondamentali per chi si accosta all'apprendimento cooperativo è un fervente sostenitore dell'introduzione nella scuola di una valutazione autentica. Nel suo intervento al Convegno Nazionale sull'apprendimento cooperativo, che si è tenuto a Torino nel gennaio del 2006 ha evidenziato che "ciò che distingue soprattutto la valutazione tradizionale da quella autentica è la sua tendenza a cercare la misura solo della comprensione scolastica di un contenuto o dell'acquisizione di un'abilità da parte dello studente e non della capacità con la quale quest'ultimo dà senso ai problemi di vita quotidiana o risolve problemi reali utilizzando le conoscenze che possiede". La valutazione autentica mira, infatti, a verificare non solo gli apprendimenti dell'alunno, le sue conoscenze, ma ciò che l'alunno sa fare con ciò che sa, attraverso la soluzione di compiti reali e concreti.

La valutazione autentica è strettamente correlata al concetto di competenza poiché questa rappresenta appunto la capacità di far fronte ad un compito attivando tutte le proprie risorse, cognitive, motivazionali, creative, mettendole in correlazione con quelle che si rendono concre-

tamente disponibili all'esterno. La valutazione autentica mira, quindi, alla verifica di ciò che l'alunno conosce, di ciò che sa fare e al tempo stesso di come costruisce autonomamente il proprio agire. Và, pertanto, oltre la pura valutazione scolastica per entrare nel campo di una valutazione reale, caratterizzata da concretezza, praticità e socialità, poiché spesso la risoluzione di un compito concreto presuppone un lavoro di cooperazione di un gruppo di studenti.

Sin dalla seconda metà degli anni '90, i sostenitori della valutazione autentica si pongono il compito di verificare le abilità degli studenti in progetti operativi reali, mettendo autenticamente gli studenti in condizione di dare prova delle competenze cognitive e meta-cognitive acquisite. Con questa forma di valutazione alternativa si intende verificare, come anticipato, non solo ciò che lo studente sa, ma ciò che sa fare con le risorse intellettuali di cui dispone. La scuola è chiamata alla costruzione di un curriculum finalizzato alla maturazione negli studenti di competenze necessarie allo svolgimento di compiti reali, affinché diventino in grado di essere cittadini attivi e consapevoli. I docenti devono, quindi, fissare le prestazioni che gli studenti dovranno effettuare per dimostrare le loro capacità e su questa base si costruisce il curriculum, che diventa mezzo per lo sviluppo della competenza richiesta per assolvere ad un compito.

La valutazione autentica intende dare agli studenti accurati e dettagliati feedback sui processi usati per apprendere e sulla qualità e quantità dei loro apprendimenti, nonché indicazioni dettagliate sulle strategie che lo studente stesso può applicare per migliorarsi. Affinché la valutazione sia efficace e significativa, è necessario che gli studenti comprendano le procedure stesse della valutazione, le condividano e investano in prima persona energie per valutare i loro stessi processi di lavoro, in tal modo gli studenti si appropriano della loro stessa valutazione e, sentendosene protagonisti, sono più motivati a mostrare i loro lavori e parlare di se stessi con altri. La valutazione autentica, inoltre, consente di agire sulla modificazione dell'apprendimento, aggiustando il tiro quando è necessario, poiché permette di comprendere e correggere i propri errori e di colmare le distanze che vengono rilevate negli apprendimenti, consentendo di avanzare ad un livello successivo di conoscenza e abilità. Per McTighe and Ferrara [1996] l'esecuzione di un compito permette agli studenti di dimostrare la

loro conoscenza e abilità in un modo che si avvicina a quello che dovrebbero fare quando sono fuori dalla classe. Quando per esempio agli studenti viene chiesto di convincere un destinatario specifico, scrivendo in modo deciso le proprie argomentazioni, devono saper usare la loro conoscenza sull'argomento, le loro abilità di scrittura nel modo di più adeguato per raggiungere l'obiettivo dato. Questo è uno dei tanti compiti che si possono assegnare ad uno studente per promuovere la cittadinanza attiva in quanto il compito è congruente con le richieste della vita reale. Questo modo di valutare le conoscenze e le abilità degli studenti, è la prospettiva della valutazione autentica, utile per potenziare la motivazione intrinseca. Valutando le abilità e le conoscenze degli studenti in un contesto reale, gli studenti apprendono ulteriormente il "come" applicare le loro conoscenze e abilità in compiti e contesti diversi.

La valutazione autentica non incoraggia, quindi, l'apprendimento meccanico, l'apprendere passivamente in vista del compito in classe, ma si focalizza sulle competenze di analisi degli studenti; sull'abilità di integrare ciò che apprendono; sulla creatività; sull'abilità di lavorare collaborando; sullo scrivere e parlare in modo competente. Per misurare conoscenze profonde e abilità, in un contesto autentico, occorrono strumenti specifici quali, ad esempio, le rubriche ed il portfolio.

Wiggins and McTighe [1998] definiscono una rubrica come uno strumento per valutare un prodotto oppure una prestazione. Una rubrica è composta, infatti, da criteri valutativi, da una scala definita, e da indicatori descrittivi per differenziare i livelli di comprensione, di abilità, di qualità. Si tratta, quindi di uno strumento che rappresenta una guida all'attribuzione di un singolo punteggio per valutare le prestazioni dello studente, basandosi su un insieme di criteri che vanno da un livello minimo ad uno massimo. Solitamente l'insegnante assegna ad un gruppo un compito da portare a termine che può essere un prodotto, una prestazione, oppure può richiedere delle ampie risposte scritte a una domanda, che comportano da parte degli studenti l'applicazione di abilità come il pensare critico, creativo o l'abilità di analisi ed organizzazione delle informazioni. Le rubriche dovrebbero organizzare ed esplicitare in modo chiaro i criteri di assegnazione del punteggio, così che due insegnanti diversi, applicando la rubrica ad uno stesso lavoro dello studente, generalmente siano in grado

di arrivare al medesimo punteggio per ridurre la soggettività nella valutazione. Una rubrica ha diversi componenti, ognuno dei quali contribuisce alla sua utilità, che sarebbe opportuno comunicare prima agli studenti per dare loro un'idea precisa del compito che devono andare a svolgere.

Gli elementi che compongono una rubrica sono:

- uno o più tratti, o dimensioni, che sono il riferimento per giudicare il lavoro degli studenti;
- delle definizioni per chiarire il significato di ogni tratto o dimensione;
- una scala di valore con la quale stimare ogni dimensione;
- standard di eccellenze per specificare i livelli di prestazione accompagnati da modelli o esempi per ogni livello.

Le rubriche mirano ad educare e migliorare le prestazioni degli studenti, divenendo altresì una guida nell'essenzialità dei saperi disciplinari. Tendono quindi a superare l'idea di una verifica meramente di controllo sulle attività svolte e sulle conoscenze possedute. Heidi Goodrich considera la rubrica un processo che, oltre a consentire all'insegnante di riflettere sulle sue metodologie e strategie di insegnamento, fornisce allo studente la possibilità di partecipare attivamente alla costruzione degli apprendimenti. Utilizzando la rubrica si attiva, quindi, un meta-processo attraverso il quale vengono definiti i livelli di qualità dell'apprendimento in relazione ai quali lo studente sarà valutato e al tempo stesso allo studente viene dato uno strumento di semplice utilizzo per comprendere più precisamente il suo livello di apprendimento, in modo che possa essere in grado di auto-valutarsi in relazione alla scala dei livelli e agli standard di eccellenza. Solitamente si definisce principiante o novizio lo studente che esegue procedure meccanicamente senza tener conto delle variabili di contesto poiché è privo di esperienza e manca di flessibilità; il principiante avanzato, poiché ha maturato una maggiore esperienza, comincia a tener conto delle variabili di contesto e quindi ad essere più flessibile; con il continuo maturare dell'esperienza, grazie al fatto di dover superare difficoltà, gestire imprevisti e correggere errori procedurali, lo studente diventa competente e dimostra una maggior oggettività e un maggior distacco

	<i>Completamente raggiunto</i>		<i>Raggiunto</i>	<i>Parzialmente raggiunto</i>	<i>Da raggiungere</i>
Comprensione del problema	Comprensione totale del problema utilizzando conoscenze anche non matematiche	Comprensione totale del problema utilizzando parzialmente conoscenze anche non matematiche	Buona comprensione del problema ma non interpreta in modo corretto alcune parti utilizzando parzialmente conoscenze anche non matematiche	Non interpreta in modo corretto la maggior parte del problema	Interpreta in modo non completamente corretto il testo del problema
Individuazione e selezione dati	Chiara identificazione di tutti i dati necessari alla soluzione del compito	Identificazione della maggior parte dei dati necessari alla soluzione del compito	Identificazione di dati essenziali per la soluzione del compito, ma alcuni dati non vengono trovati	Identificazione di alcuni dei dati necessari alla soluzione del compito, ma alcuni dati sono letti in modo non del tutto corretto	Identificazione e di solo alcuni dati necessari alla soluzione del compito
Elaborazione di strategie	Elabora strategie diversificate (più rapide, originali) che portano alla soluzione corretta	Elabora una strategia che porta alla soluzione corretta	Elabora una strategia che porta alla soluzione corretta con lievi omissioni o errori procedurali	Elabora una strategia parzialmente corretta con errori procedurali	Elabora una strategia non completamente e corretta

Figura 4.3: Esempio di rubrica

emotivo nel prendere in tempi più immediati le decisioni opportune; a questo punto lo studente, facendo sempre più ricorso alla memoria esperienziale relativa a procedure e strategie usate in passato, matura un approccio olistico ed intuitivo ai problemi che lo aiuta a comprendere con più immediatezza quale sia la soluzione migliore. In questa fase lo studente si definisce abile competente; lo studente diventa, quindi, sempre più in grado di intuire la soluzione più idonea e immediata poiché ha maturato un bagaglio esperienziale che lo rende completamente autonomo, oggettivo, razionale e flessibile. Lo studente è ora competente esperto poiché è in grado di decontestualizzare le esperienze, attraverso un processo di astrazione, e di compiere transfert grazie alle competenze acquisite.

L'insegnante per effettuare una valutazione autentica utilizzando la rubrica deve, dopo aver scelto con accuratezza il compito da assegnare, mostrare ai propri allievi esempi concreti di lavori simili a quello richiesto. Gli esempi di lavori devono comprendere sia quelli ben eseguiti, sia quelli più scadenti in modo da consentire di identificare, attraverso la loro analisi, le caratteristiche che distinguono un buon lavoro da uno scadente. L'insegnante proprio attraverso quest'analisi,

descrivendo nei particolari tutte le caratteristiche che deve avere un lavoro di qualità, ha la possibilità di esplicitare le motivazioni che, rispetto allo standard di eccellenza, pongono altri lavori ad un livello basso o medio. In questo modo si articola la struttura stessa della rubrica e se ne realizza una prima stesura. Gli studenti hanno ora il compito di confrontare la prima stesura della rubrica con gli standard di eccellenza prefissati per verificare se la rubrica impostata rispetta adeguatamente i livelli di qualità o se necessita di modifiche, che saranno quindi apportate per pervenire ad una stesura finale. Gli studenti lavorano alla realizzazione del compito assegnato in gruppo e, dopo una prima fase del lavoro, l'insegnante chiede loro di fermarsi, invitandoli a valutare quello che stanno facendo attraverso la rubrica, per rendersi conto di come stanno procedendo. Subito dopo, si invitano i compagni a svolgere la stessa verifica sui lavori dei loro pari, fornendo delle indicazioni eventuali su come correggere o migliorare il lavoro. Ad ogni studente, individualmente, viene concesso del tempo per rivedere il proprio lavoro, anche basandosi sulle osservazioni di feedback dei compagni. Completato il compito o il prodotto, l'insegnante usa la medesima rubrica che hanno utilizzato gli studenti, per valutare definitivamente i loro lavori. La rubrica è, quindi, un evento dinamico e perciò può esser rivista e aggiornata sia attraverso l'apporto della classe che di altri insegnanti, affinandosi e modificandosi con il variare dei componenti la classe e delle necessità dei compiti e prodotti, nonché con il raggiungimento del livello superiore da parte degli studenti. Le rubriche sono uno strumento che consente agli studenti di costruire le proprie competenze e di verificare costantemente i propri progressi in quanto l'insegnante esplicita in modo chiaro le attese in relazione alla qualità richiesta dal compito. Lo studente in questo modo si sente protagonista della propria formazione e, il fatto di conoscere nel dettaglio a priori le caratteristiche qualitative che deve avere il prodotto finale, lo fa sentire più responsabile della sua stessa formazione. Inoltre, la possibilità che dà la rubrica di ricevere feedback continui sui propri punti di forza e di debolezza consente una presa di coscienza delle proprie capacità ed abilità, nonché dei propri limiti, che come abbiamo visto non è un fatto irrilevante nella costruzione delle proprie conoscenze.

Lo strumento privilegiato per la valutazione autentica è il *portfolio*

delle competenze che esplicita la funzione certificativa, formativa e documentativa della valutazione stessa. Il portfolio valorizza sia il processo che il prodotto della formazione poiché consente di valutare al tempo stesso le prestazioni finali, i processi, le strategie messe in atto e i progressi compiuti in un preciso contesto, favorendo il docente in una valutazione sistematica e continua e guidando lo studente all'autovalutazione. Il portfolio dell'alunno è uno spazio in cui raccontare biografie ed autobiografie, tracciare idee ed eventi per comprendere il senso dell'esperienza scolastica e delle relazioni con la vita familiare e sociale. Il portfolio è un documento inter-soggettivo e processuale che accompagna lo studente nel suo percorso formativo. Esso, infatti, nasce da un forte progetto di cooperazione scuola-famiglia mirato a far vivere l'esperienza formativa come un'inter-relazione di studenti, genitori e docenti che hanno la possibilità di vivere e condividere l'intero percorso formativo da co-protagonisti.

Il portfolio nasce in campo artistico e indica una cartella contenente tutti i lavori più significativi di un'artista; si diffonde poi in campo professionale per accompagnare il curriculum vitae esplicitando in modo dettagliato le proprie performance, le esperienze più significative del percorso formativo e lavorativo dell'individuo, organizzate in una documentazione ragionata. Si tratta, sostanzialmente, di una raccolta cartacea utilizzata regolarmente di diversi Stati per sostenere l'apprendimento e lo sviluppo professionale, attraverso l'accreditamento di esperienze e competenze. Il portfolio è stato individuato dal Consiglio europeo di Strasburgo quale strumento più idoneo per gli studenti per la certificazione delle proprie competenze linguistiche, una sorta di "diario autocosciente" della loro carriera scolastica da poter esibire soprattutto nel corso degli scambi culturali tra una nazione e l'altra. Il portfolio è un approccio all'apprendimento e alla valutazione basato su una visione globale della persona e dei processi formativi che la caratterizzano, che si fonda su una visione integrata basata sulla raccolta di certificazioni, elaborati, esperienze significative, interessi ed attitudini, documentate liberamente dallo studente, dai docenti e dai genitori perché ritenuti particolarmente significativi per esplicitare le specificità che caratterizzano quel individuo in particolare. Il portfolio è, quindi, una collezione significativa di artefatti, prodotti durante il periodo formativo dello studente, accompagnati da riflessioni che mostrano la

significatività degli stessi in relazione al soggetto e i motivi della scelta. Gli artefatti possono essere di tipologie diverse e memorizzati su supporti differenti (cartacei, magnetici, ottici), possono essere compiti, o consegne realizzate, descrizioni delle stesse, osservazioni, relazioni, narrazioni di un percorso effettuato. Il focus è posto sul processo di apprendimento e sul potenziale dello studente, per garantire la personalizzazione e la specificità dell'esperienza formativa. I campioni selezionati del percorso dello studente riflettono la crescita lungo il tempo attraverso una logica diacronica e narrativa, che permette di delineare un profilo integrato dello studente e fa emergere la capacità di elaborare procedure e strategie in situazioni complesse, ovvero sia le competenze. Il portfolio, oltre a consentire una valutazione del processo formativo degli studenti, consente proprio a questi ultimi di riflettere e di autovalutarsi, stimolando al contempo il confronto e il dialogo tra studenti, insegnanti e genitori.

Le funzioni del portfolio, quindi, si possono individuare nella valutazione e nell'orientamento. Per quanto riguarda la valutazione abbiamo già messo in evidenza che il portfolio consente di esplicitare le competenze possedute dallo studente in quanto è uno strumento che traccia tutto il processo formativo, evidenziando le attività e le modalità con cui si è costruita la conoscenza in situazioni autentiche. Le competenze esplicitate possono, di conseguenza, essere valutate dal docente, eterovalutazione, e dallo studente stesso, autovalutazione. La finalità intrinseca del portfolio è, infatti, quella di attivare la riflessione sulle proprie performance e sul processo formativo che ad esse sottende. Per quel che riguarda l'orientamento possiamo dire che il portfolio offre allo studente, all'insegnante e alla famiglia, indicazioni sul modo personale di operare, sulle risorse possedute e sulle finalità da porsi per il personale progetto di vita. Per costruire un portfolio non è sufficiente assemblare artefatti ma occorre seguire una precisa metodologia: prima di tutto è necessario avere ben chiara la finalità, lo scopo per cui si intende procedere alla costruzione del portfolio, attraverso la determinazione di obiettivi generali e specifici. Occorre, quindi, stabilire quali materiali sia opportuno introdurre e preparare una presentazione motivata degli stessi, riflettendo criticamente sui prodotti realizzati. Nella fase progettuale occorre, quindi, individuare finalità, modalità di implementazione, struttura organizzativa, i dos-

sier da cui selezionare gli artefatti, gli indicatori per la selezione, le modalità di selezione, le modalità di riflessione e di valutazione. Il portfolio organizza in sé le rubriche, di cui abbiamo parlato sopra, che in relazione ad una scala di punteggi e ad un elenco di linee guida, specificano gli elementi che contraddistinguono la qualità di una specifica prestazione. Le diverse rubriche devono essere messe in correlazione attraverso l'uso di mappe e narrazioni ed integrate con interviste, diari, osservazioni. Il portfolio, infatti, può contenere una serie di strumenti di documentazione che rendono il quadro di riferimento della formazione dello studente più complesso e completo. Tra essi è opportuno menzionare il *dossier*, il *diario* e la *biografia*.

Il *dossier* è spesso confuso con il portfolio ma non si identifica con esso, pur essendone una parte fondamentale poiché rappresenta proprio la raccolta delle prestazioni più significative selezionate dallo studente e dall'insegnante, si può trattare di testi prodotti sia in forma scritta che audio o video, a testimonianza del livello raggiunto in specifici apprendimenti.

Il *diario* è uno strumento molto diffuso in campo valutativo che serve per narrare se stessi, i propri interessi, le proprie attitudini, la percezione del *locus of control*, le frustrazioni di fronte all'insuccesso, la stima di sé e la maturazione delle competenze. Può essere scritto dallo studente stesso al fine di consentirgli, attraverso la rilettura nel tempo, una riflessione critica sui cambiamenti negli stili di apprendimento, nella motivazione, nell'impegno e nell'interesse. Può ancora essere scritto dall'insegnante come diario di bordo per documentare il processo di insegnamento-apprendimento o una specifica attività progettuale.

La *biografia* rappresenta, invece, la storia del proprio apprendimento in riferimento ad una particolare disciplina con riferimento alle abilità orali e scritte, alla metodologia e all'approccio adottato. Contiene anche griglie di valutazione e schede di autovalutazione. In questa sezione si inseriscono i certificati, titoli ottenuti e la descrizione del futuro programma di apprendimento.

Una forma innovativa di portfolio è l'*e-portfolio* che rappresenta la versione digitale del portfolio stesso. Attualmente esistono diversi software programmati per la gestione dell'*e-portfolio* o addirittura alcuni portali on-line che consentono di pubblicare direttamente sul

web la documentazione del proprio percorso formativo. Grazie alle caratteristiche del mezzo utilizzato, è possibile inserire nella documentazione anche grafica, video e suoni, rendendo più completo l'intero quadro dello studente. L'e-portfolio, inoltre, ha il vantaggio di poter essere condiviso on-line senza vincoli di spazio e tempo. Si capisce che grande vantaggio può essere questo soprattutto per la valutazione di crediti formativi o per proporsi al mondo del lavoro.

4.5.3 Motivare alla valutazione: l'autovalutazione

Mialaret [1977] afferma che l'apprendimento è facilitato quando il soggetto conosce i risultati a cui mira, a questo si potrebbe aggiungere l'importanza della riflessione critica sui propri progressi e sulle difficoltà incontrate, al fine di rendere lo studente protagonista della propria esperienza formativa e quindi di motivarlo maggiormente ad apprendere. Tessler [2002] parla di "valutazione del soggetto-allievo sul suo essere soggetto-protagonista nel suo ambiente-contesto di formazione". L'allievo contemporaneamente, quindi, valuta se stesso e l'ambiente che contribuisce a costruire, sulla base di un principio fondamentale: indipendentemente dall'età, un soggetto apprende solo se è soggettivamente consapevole del senso e del valore personale del sapere che scopre e che costruisce, integrando il nuovo con il conosciuto.

La consapevolezza del senso e del valore di ciò che si apprende presuppone capacità riflessiva sulla propria azione, insieme metacognitiva e autovalutativa. Il percorso dalla dipendenza valutativa all'autonomia può essere facilitato con la tecnica della co-valutazione. L'autovalutazione aiuta ad esercitare un pensiero critico sul proprio modo di apprendere e ad ascoltare le idee degli altri in relazione alle proprie. Questo cammino comporta un delicato lavoro di distacco dalla sola valutazione esterna, o eterovalutazione, dove l'allievo rischia comunque di rimanere oggetto e non soggetto di valutazione, per integrarsi con processi autovalutativi, attraverso tecniche come vedremo di co-valutazione. La co-valutazione è una tecnica insieme formativa e valutativa, in cui sia l'insegnante che l'allievo valutano individualmente la medesima prestazione, e successivamente confrontano le loro risposte affinandone il giudizio. Con le procedure di co-valutazione,

quindi, l'allievo impara ad affinare l'osservazione e a differenziarne gli oggetti, a selezionare gli elementi di rilevazione, a relativizzare i giudizi propri e altrui, ad attribuire significato, senso e valore alla sua azione.

La capacità di autovalutazione necessita di una profonda conoscenza dei propri stili cognitivi, delle proprie strategie mentali, delle proprie capacità e dei propri limiti. L'alunno deve essere guidato a maturare questa capacità, muovendo i primi passi nella condivisione della valutazione che deve essere vissuta come un momento importante del processo formativo e non con ansia. Se si vogliono sviluppare competenze trasversali negli allievi, occorre tener presente l'interezza della persona, la sua dimensione cognitiva, affettiva, sociale e meta cognitiva. Significa prestare attenzione allo sviluppo globale della persona, lavorando in stretta collaborazione con i colleghi docenti, per andare oltre la settorialità della propria disciplina. Significa dar modo allo studente di vivere da protagonista il proprio percorso formativo, renderlo cosciente dei progressi compiuti, delle difficoltà incontrate, delle strategie necessarie a superarle. L'autovalutazione è una stima, una riflessione critica del valore di certe idee, lavori, situazioni, percorsi, progressi educativi partendo da criteri determinati dallo studente stesso. Se un allievo riesce a maturare un locus of control interno, manifestando la capacità di auto valutare obiettivamente il proprio grado di preparazione e il proprio rendimento in una prestazione, diventa meno dipendente dalla valutazione di un'altra persona e si sente più responsabile dei propri successi o insuccessi. L'autovalutazione, infatti, mira sia alla costruzione delle conoscenze, sia alla formazione globale della persona, poiché aiuta ad esercitare un pensiero critico sul proprio modo di apprendere, ad ascoltare le idee degli altri in relazione alle proprie, per comprenderle meglio.

4.6 I metodi della valutazione quantitativa

"Di tutto conosciamo il prezzo, di niente il valore."
(Friedrich Nietzsche)

La valutazione quantitativa è storicamente la metodologia di verifica degli apprendimenti più largamente utilizzata nella scuola italiana.

La ricerca docimologica, infatti, per anni si è interessata alla predisposizione di strumenti di misurazione degli apprendimenti che garantissero oggettività nella valutazione.

Prima di entrare nel dettaglio della trattazione occorre fare una puntualizzazione sul significato dei termini *verifica* e *valutazione*, che spesso vengono erroneamente utilizzati come sinonimi.

Quando utilizziamo l'espressione *verifica* degli apprendimenti, ci riferiamo all'insieme delle procedure con cui determiniamo se e in che misura, o con quale ampiezza, siano stati conseguiti determinati obiettivi didattici prefissati.

Con il termine *valutazione*, invece, si indica l'operazione che conclude il processo di controllo di *verifica* degli apprendimenti e che consiste, sostanzialmente, nell'attribuzione di un giudizio sulle prestazioni precedentemente misurate, determinandone in tal modo l'appartenenza ad una specifica classe, collocata su una scala, nominale o ordinale. L'atto del valutare, quindi, è seguente rispetto a quello del verificare, poiché la valutazione è il processo mediante il quale si attribuisce, come abbiamo più volte ribadito, valore ad un oggetto, ad un'azione o ad un evento, che sono stati precedentemente sottoposti a *verifica*.

Verifica e *valutazione*, sono quindi da considerare come due momenti concatenati di un unico processo, poiché, se è vero che la valutazione è strettamente collegata e determinata dalle procedure, dai metodi e dagli strumenti utilizzati per effettuare la *verifica*, è pur vero che l'attendibilità del giudizio è, a sua volta, interconnessa e condizionata dalla scelta degli strumenti di *verifica*. Lo scopo primario della ricerca docimologica è, appunto, quello di studiare i metodi più idonei a pervenire ad un'espressione di giudizio che sia attendibile dal punto di vista concettuale (ricerca di strategie che assicurino oggettività) e strumentale (individuazione degli strumenti che siano garanzia di rigore ed oggettività).

La valutazione quantitativa degli apprendimenti trova il suo fondamento nell'approccio teorico del *modello funzionalista*. Precisiamo che quando parliamo del modello teorico di riferimento, consideriamo una particolare modalità di approccio alla valutazione costituita da una propria filosofia di pensiero, da uno substrato teorico specifico, da principi e concetti a cui richiamarsi, in modo esplicito o implicito, per affrontare empiricamente la valutazione. Il sistema di riferimento ci

orienta nella scelta dei criteri, dei soggetti, degli oggetti, delle misure, degli strumenti.

Il modello funzionalista, nato dall'esigenza di abolire azioni didattiche basate sulla soggettività e sull'ambiguità, si caratterizza per un approccio alla valutazione di tipo quantitativo. Si fonda, quindi, sulla verifica della congruità tra obiettivi programmati e risultati conseguiti. Nella valutazione quantitativa, infatti, l'interesse è centrato sulla verifica della conformità tra obiettivi e risultati, e non sull'analisi del processo che sottende a questa relazione. Il focus della valutazione è centrato sull'organizzazione dell'azione formativa.

A questo modello si contrappone il *modello fenomenologico*, nato dall'esigenza di approfondire le conoscenze sul processo di apprendimento al fine di migliorare l'offerta formativa. Il focus è, quindi, spostato sugli effetti prodotti dall'azione formativa, sulla relazione didattica e su come sono state percepite dagli studenti le attività proposte dai formatori. Si basa sostanzialmente sull'accertamento del valore sociale ed educativo dell'azione formativa.

Sinteticamente possiamo dire che nell'approccio funzionalista si privilegiano gli aspetti oggettivi dell'azione formativa, mentre in quello fenomenologico gli aspetti che caratterizzano le specificità del soggetto e del contesto.

In un approccio alla valutazione di tipo funzionalista l'insegnante deve tener conto di alcuni passaggi fondamentali:

- individuazione delle finalità del proprio insegnamento;
- scelta delle esperienze educative e didattiche più idonee a raggiungere le finalità prefissate;
- scelta delle metodologie e delle strategie necessarie ad organizzare e realizzare le esperienze programmate;
- progettazione degli strumenti e delle modalità di verifica per valutare la conformità tra obiettivi e risultati.

L'insegnante che progetta il suo intervento con queste modalità, presta grande attenzione alla progettazione educativo-didattica e pone al centro di essa la valutazione, che precede ed accompagna tutto il lavoro di progettazione. L'analisi della fattibilità, in relazione alle risorse

e al contesto operativo, rappresenta la fase di avvio della valutazione, che si esplicherà per mezzo di una strutturazione temporale delle verifiche. La valutazione quantitativa garantisce, in tutte le tutte le sue fasi, trasparenza nella pratica della valutazione ed è conclamata dalla letteratura quale modello efficace per la verifica degli apprendimenti disciplinari.

Utilizzare esclusivamente questo tipo di approccio in campo formativo può presentare, tuttavia, dei limiti poiché questo tipo di valutazione non restituisce alcuna informazione circa il processo formativo che ha portato a conseguire un determinato risultato. Per questo motivo sarebbe opportuno favorire un'integrazione metodologica, associando ad un approccio nomotetico, indirizzato cioè all'*oggettività* e alla *generalizzabilità*, un approccio ideografico, che miri alla comprensione dei risultati conseguiti progressivamente dal singolo studente, attraverso una documentazione dei processi che sottendono ai risultati ottenuti, con un'attenzione particolare alla specificità e all'originalità delle singole situazioni e dei singoli soggetti.

Occorre sempre tener presente che con la somministrazione di una prova di verifica vengono sottoposti a valutazione alcuni aspetti specifici dell'apprendimento di una particolare disciplina ma non l'apprendimento nel suo complesso. Per valutare l'apprendimento nella sua complessità, infatti, occorrerebbe effettuare una valutazione integrata di tutti i processi cognitivi attivati dall'allievo per apprendere. Il risultato di una prova di verifica, tuttavia, è importante in ambito scolastico per valutare l'acquisizione di conoscenze che, se correttamente integrate tra loro, consentono l'evoluzione dell'apprendimento.

La valutazione quantitativa, come abbiamo anticipato, si esplicita pertanto attraverso la somministrazione di prove che devono essere opportunamente strutturate e calibrate. Vedremo di seguito come per ottenere risultati affidabili sia necessario seguire rigorosamente un iter specifico, che si realizza attraverso la *misurazione*, la *verifica* e la *valutazione*.

Uno degli strumenti più utilizzati per la valutazione degli apprendimenti è sicuramente il questionario, strutturato in *item*, ossia composto da un certo numero di quesiti che svolgono la funzione di stimolo ad una risposta cognitiva. La misurazione rappresenta la registrazione quantitativa della prova, che si realizza attribuendo un

punteggio ad ogni item secondo una regola logicamente accettabile e prefissata. La verifica è costituita, quindi, da due momenti: il primo coincide con la somministrazione della prova, durante la quale lo studente risponde allo stimolo proposto in relazione alle proprie conoscenze; il secondo nella verifica vera e propria dell'esattezza delle risposte. Segue, infine, il momento della valutazione che consiste nell'attribuzione di un valore al risultato della prova.

Occorre ancora evidenziare quali siano le caratteristiche che rendono affidabile una valutazione di tipo quantitativo.

La valutazione quantitativa deve:

- garantire *validità*, cioè deve rispettare la congruità tra ciò che si vuole valutare e la misura che si utilizza per valutare. "Per ogni funzione o obiettivo specifico della valutazione va impiegato uno strumento di verifica omologo e congruente con quella funzione o con quel obiettivo"[Domenici, G., 1993]. La validità indica, quindi, la precisione con cui la prova misura l'obiettivo prefissato ed è indicazione dell'efficacia della prova. La valutazione deve essere fortemente ancorata al curriculum effettivamente svolto e gli item proposti devono risultare coerenti con gli obiettivi, i contenuti, il tipo di scuola, la metodologia utilizzata. Gli stimoli proposti negli item devono essere strutturati per attivare la risposta cognitiva, in modo da consentire agli studenti di esprimere la padronanza degli obiettivi sottoposti a verifica. Per garantire la validità di misurazione, nella strutturazione di una prova di verifica degli apprendimenti, è prima di tutto necessario che sia formulata in modo chiaro, che il linguaggio non sia ambiguo e che il livello di difficoltà sia adeguato e coerente con gli obiettivi che si intendono verificare.
- garantire *attendibilità*, ossia deve garantire esattezza nella misurazione indipendentemente da chi effettua la misurazione stessa, dal contesto nel quale si opera e dalla scansione temporale. L'attendibilità si fonda, quindi, sulla costanza della rilevazione. Questo significa che il risultato della misurazione deve rimanere invariato al variare del somministratore, del periodo temporale nel quale si somministra e del contesto nel quale si opera.

- garantire *oggettività*, cioè deve esprimere la caratteristica di controllare e limitare tutti gli aspetti soggettivi che possano inficiare la qualità della verifica. Le prove strutturate consentono di stabilire a priori il peso da attribuire a ciascun item, a seconda che la risposta sia esatta, errata o omessa, di conseguenza una stessa prova corretta da valutatori differenti otterrà sempre la stessa valutazione. A garanzia di oggettività, tutti coloro che devono sostenere una stessa prova devono essere informati in modo chiaro ed univoco circa la consegna e devono avere il medesimo tempo a disposizione per lo svolgimento.

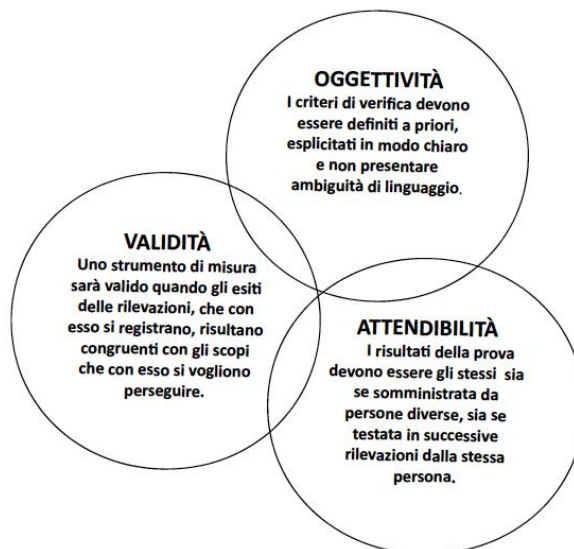


Figura 4.4: Attendibilità-Oggettività-Validità

Le misure che si ottengono dalla somministrazione di una prova di verifica si definiscono, quindi, valide se il grado di corrispondenza tra la misura e l'oggetto misurato può essere considerato valido, ossia capace di fornire informazioni dettagliate per assumere decisioni efficaci, efficienti e pertinenti in merito alle qualità indagate. Al contrario

se la prova di verifica risulta strutturalmente poco accurata si ottiene un quadro informativo di scarsa validità. Le misure che si ottengono dalla somministrazione di una prova di verifica si definiscono, invece, attendibili quando risultano costanti sia al variare del soggetto che le rileva, sia se somministrate in condizioni ambientali diverse e ancora al ripetersi della prova stessa in una successiva somministrazione. Le misure che si ottengono dalla somministrazione di una prova di verifica si definiscono, in fine, oggettive se l'esattezza delle risposte e i punteggi ad esse relativi sono predeterminati, rispetto ai momenti della somministrazione e della correzione della prova, in maniera non ambigua.

Attendibilità, validità ed oggettività di una prova possono essere, inoltre, maggiormente garantite se viene favorita una metodologia di lavoro che preveda il coinvolgimento di un gruppo di insegnanti che abbiano, oltre ad una buona competenza disciplinare, anche una consolidata esperienza proprio nell'insegnamento del settore disciplinare considerato; solo così è possibile garantire un'analisi corretta e completa delle tante e complesse variabili connesse alle reali difficoltà che una prova di verifica comporta per gli allievi. La cura della progettazione di prove di verifica in una scuola, dovrebbe prevedere l'organizzazione dei docenti in gruppi di lavoro costituiti dagli insegnanti del team, in dipartimenti per aree disciplinari, e per progetti trasversali. Tali azioni comprendono attività di progettazione, di pianificazione, di produzione dei materiali, di gestione dei gruppi di studenti, ed anche della valutazione. La scuola dell'autonomia comporta, infatti, una valutazione plurale e condivisa, non individualistica. La valutazione progettata e strutturata individualmente da ogni docente per la sua specifica disciplina potrebbe, infatti, essere troppo autoreferenziale, comportando distorsioni nella valutazione globale dello studente. Queste distorsioni, tuttavia, possono essere superate grazie al principio della *triangolazione dei punti di vista*. Il *principio di triangolazione* sostiene che per ridurre la soggettività valutativa è opportuno che l'osservazione sia effettuata da almeno tre persone. Tale principio, tuttavia, non intende sfiduciare le osservazioni individuali, bensì valorizzarle nel confronto e nel raccordo tra colleghi, per ottenere un risultato caratterizzato da un importante valore aggiunto. "Osservare qualcosa da più punti di vista significa abbracciarla, prenderla dentro, comprenderla

meglio. Valutare qualcosa da più punti di vista significa condividere significati e valori, ad un livello di profondità irraggiungibile da un singolo valutatore"[Tessaro, 1997].

Occorre tuttavia precisare che, quando si parla di triangolazione, non ci si riferisce esclusivamente ad un gruppo formato da tre docenti, ma che il tre è considerato il numero minimo per garantire l'attendibilità di una valutazione. Aumentando, quindi, il numero di docenti, si pensi ad esempio ad un consiglio di classe, aumentano i punti di vista e di conseguenza si incrementano gli incroci e le comparazioni, garantendo una valutazione globale di più alta qualità. Dopo tutto il senso degli scrutini di un consiglio di classe è quello della condivisione di un processo formativo, quindi non può basarsi solo sulla comunicazione tra colleghi dei singoli voti attribuiti individualmente ad un alunno.

La misurazione dei risultati ottenuti nelle prove di verifica di tipo quantitativo, deve far ricorso a metodi statistici che consentano di stabilire quale peso abbia il risultato conseguito da un allievo in relazione ai risultati conseguiti nella stessa prova da soggetti. Vedremo in seguito che per far ciò occorre stabilire prima di tutto la tendenza centrale dei risultati in una distribuzione di punteggi, proprio attraverso una misurazione statistica, e il tenore medio della dispersione o deviazione intorno ad essa; in questo modo sarà possibile indicare la posizione di ogni risultato rispetto a tali valori. La misurazione degli apprendimenti conseguiti, che si ottiene somministrando prove di verifica, dà al docente la possibilità di ottenere informazioni diverse sulle conoscenze acquisite dallo studente, poiché i risultati ottenuti possono essere messi a confronto con altri valori. Se la valutazione è riferita al sé, ossia al singolo studente, l'insegnante può valutare il progresso operato dall'alunno, rispetto alla situazione di partenza; se invece è riferita a norma, l'insegnante confronta i risultati raggiunti da ciascun alunno con i risultati raggiunti dal gruppo, inteso quale campione rappresentativo di riferimento. L'insegnante può, ancora, effettuare una valutazione riferita a criterio, confrontando i risultati raggiunti dagli alunni con gli obiettivi prefissati nella programmazione educativo-didattica, in tal caso gli obiettivi programmati, ossia i risultati attesi, assumono la funzione di criteri per valutare se i traguardi siano stati raggiunti ed in quale misura.

4.6.1 Le prove oggettive di valutazione

L'elaborazione di una prova di verifica di tipo quantitativo impone un'attenta analisi degli ambiti disciplinari per i quali si decide di verificare l'acquisizione di conoscenze. Ogni prova di verifica deve, infatti, essere costruita in relazione alla struttura concettuale di ogni singola disciplina. Le tipologie di prove somministrabili ad uno studente possono essere classificate in relazione al loro grado di strutturazione, cioè al modo in cui vengono presentati sia gli stimoli che le risposte. Avremo, quindi:

- *prove non strutturate* (stimolo aperto – risposta aperta): sono le prove più diffuse e tradizionali, caratterizzate da una forte componente soggettiva nell'esecuzione e nella valutazione, come l'interrogazione, il tema, il riassunto, la relazione, l'articolo;
- *prove semi-strutturate* (stimolo chiuso – risposta aperta o stimolo aperto – risposta chiusa): sono quelle in cui le risposte non sono univoche ma pre-determinabili grazie ai vincoli posti negli stimoli, se questi vincoli risultano correttamente impostati la prova garantisce oggettività nella valutazione;
- *prove strutturate* (stimolo chiuso – risposta chiusa): sono i quesiti vero/falso, scelta multipla, corrispondenze, completamente con risposte possibili, caratterizzate da un forte oggettività nell'esecuzione e nella valutazione.

La scelta di una prova di verifica deve essere ben ponderata, poiché per ogni funzione e/o obiettivo specifico della valutazione va impiegato uno strumento di verifica omologo e congruente con quella funzione e/o con quell'obiettivo[Domenici, G., 1993].

Le caratteristiche strutturali delle prove di verifica tradizionali, stimolo aperto-risposta aperta, determinano spesso la stesura di giudizi poco affidabili che non danno nessuna garanzia di rappresentare in maniera esauriente e fedele le reali competenze degli allievi. Quando il docente durante un'interrogazione propone una domanda aperta deve essere assolutamente certo di essersi espresso in modo chiaro e congruente alla conoscenza che desidera valutare. Ciò nonostante non

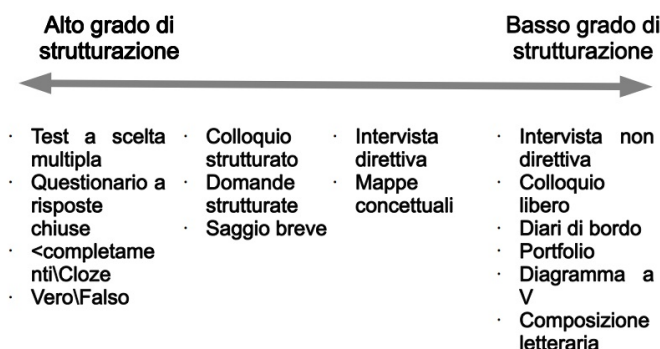


Figura 4.5: Il livello di strutturazione prove

si può avere la certezza che l'interpretazione della richiesta risulti univoca in allievi diversi. Altro limite delle prove di verifica tradizionali è che il giudizio espresso dall'insegnante non può essere conseguente a criteri oggettivi stabiliti a priori, ma soltanto alla scelta soggettiva di una soglia minima di accettabilità dell'argomentazione, pertanto lo stesso allievo che risponde in modo identico alla medesima domanda posta da docenti diversi, può ottenere giudizi molto differenziati tra loro. Ciò dimostra che il paradigma interpretativo dell'insegnante risulta, quindi, altamente soggettivo e che le informazioni che derivano da una prova di verifica non strutturata non consentono di delineare un quadro conoscitivo esauriente e completo delle abilità e/o delle lacune di ciascuno e di tutti gli allievi, in riferimento ad ognuno e a tutti gli obiettivi della procedura didattica considerata. Questo tipo di prova di verifica è il più utilizzato nella valutazione scolastica poiché consente agli insegnanti di valutare la prontezza nelle risposte, la capacità argomentativa, la capacità di compiere correlazioni ed individuare nessi logici, e di comprendere quali siano gli stili di apprendimento e la metodologia di studio utilizzata dallo studente. È proprio in relazione al quadro che si fa un'insegnante del singolo alunno e del contesto classe che si assumono decisioni in merito alla diversificazione della proposta d'istruzione, all'avvio di eventuali strategie di recupero, sostegno o potenziamento, e, più in

generale, su come portare avanti la proposta educativa. Altro aspetto a favore della predilezione per questo tipo di prova da parte degli insegnanti è da ricercare nel fatto che non richieda tempi di preparazione troppo impegnativi. Tuttavia possono essere fissati criteri oggettivi di giudizio che consentano, attraverso la compilazione di una griglia di osservazione, di ottenere una valutazione oggettiva anche di questa tipologia di prove.

Spesso i docenti muovono forti accuse alle prove di verifica standardizzate poiché ritengono, a causa di una formazione professionale fortemente influenzata da componenti psico-pedagogiche, che le conoscenze di una persona non possano essere ridotte a numeri, tuttavia, come puntualizza Vertecchi [1993], nelle prove di verifiche tradizionali la valutazione è effettuata sulla base di un'informazione incerta e destrutturata, mentre nelle prove oggettive sulla base di un'informazione che si cerca di uniformare a scelte e termini di confronto esplicitamente individuati e perciò, in quanto tali, obiettivi.

Le prove oggettive di profitto, caratterizzate dalla chiusura degli stimoli e da un diverso livello di chiusura delle risposte, consentono, quindi, di superare i limiti di indagine che derivano dalla soggettività interpretativa. L'ambiguità interpretativa diminuisce, e di conseguenza risulta più univoca, in relazione alla precisione, chiarezza e circoscrizione degli stimoli proposti. La strutturazione della prova costituisce, pertanto, una condizione necessaria, se pur non sufficiente, perché risultino garantiti i requisiti della validità e di attendibilità delle rilevazioni, delle misurazioni e delle valutazioni degli apprendimenti. Le prove strutturate e semistrutturate sono, infatti, definite oggettive proprio perché il punteggio totalizzato informa in modo univoco sulle abilità possedute dallo studente e perché la somministrazione risulta equa, in quanto a tutti gli allievi sono richieste le medesime prestazioni, nelle stesse condizioni di contesto e di tempo. Le prove oggettive possono essere classificate ancora come:

- *Informali*: quando sono realizzate dagli insegnanti di una scuola per verificare l'acquisizione di conoscenze in relazione ad una specifica programmazione educativo-didattica e vengono utilizzate, solo all'interno della scuola, per accertare lacune o abilità dei propri studenti.

- *Standardizzate*: quando vengono utilizzate, per confrontare i risultati ottenuti da un campione di studenti di una determinata fascia d'età con valori ritenuti standard di riferimento.

Nell'analisi delle procedure necessarie alla costruzione di una prova di verifica ci riferiremo, in questa trattazione, sia a prove oggettive di tipo informale, sia a prove standardizzate, con lo scopo di poter fornire alcuni suggerimenti ai docenti che desiderassero strutturare autonomamente prove di verifica, anziché utilizzare quelle "preconfezionate" che si trovano su molte guide o riviste didattiche.

Volendo costruire una prova è necessario stabilirne preliminarmente gli obiettivi, che non devono essere di carattere generale (del tipo: la formazione del cittadino, lo sviluppo della personalità, la promozione del senso critico, ecc.), ma devono riguardare specifici settori educativi. Si possono valutare, quindi, aspetti affettivo-sociali che riguardano sia gli elementi psicologici, quali il sé, l'autostima e la motivazione, sia quelli relazionali, come le modalità di rapportarsi a coetanei ed adulti, dentro e fuori del contesto scolastico; obiettivi cognitivi che riguardano l'apprendimento di nozioni, di metodi, di modelli, di tecniche, ecc.; e ancora, obiettivi psico-motori, relativi alla conoscenza e rappresentazione del proprio corpo, all'individuazione di punti di riferimento nello spazio rispetto a se stessi e ad altri.

La prova deve essere, quindi, centrata sui punti nodali del percorso formativo e riservare spazi ad aspetti o approfondimenti particolari, per questo motivo è fondamentale definire con precisione a priori l'area delle competenze che si vogliono accertare. In questa prospettiva sono sorte varie teorie tendenti a definire le tipologie di abilità che si vorrebbero far acquisire agli allievi.

Un esempio, largamente utilizzato, è costituito dal modello tassonomico di Bloom [1990], che fa riferimento a livelli di obiettivi ordinati gerarchicamente: *conoscenza, comprensione, applicazione, analisi, sintesi, valutazione*. Bloom distribuisce tutti gli obiettivi in ognuna delle tre aree che costituiscono la complessità del processo formativo: *Area Cognitiva, Area affettiva, Area Psicomotoria*.

- La *conoscenza* è intesa come capacità di trattenere le informazioni relative a fatti, nomi, definizioni.

- La *comprensione* consiste nella capacità di tradurre in diversa forma i dati conosciuti.
- L'*applicazione* si ha con l'estensione dei concetti posseduti a nuove situazioni.
- Il livello dell'*analisi* è costituito dalla capacità di discriminare elementi di informazione, individuare ipotesi, separare qualitativamente i dati.
- La *sintesi* consiste nell'organizzare funzionalmente e in modo produttivo gli elementi di conoscenza posseduti.
- La *valutazione*, è la capacità di esaminare criticamente una situazione, formulando giudizi appropriati e rappresenta il livello più elevato degli obiettivi.

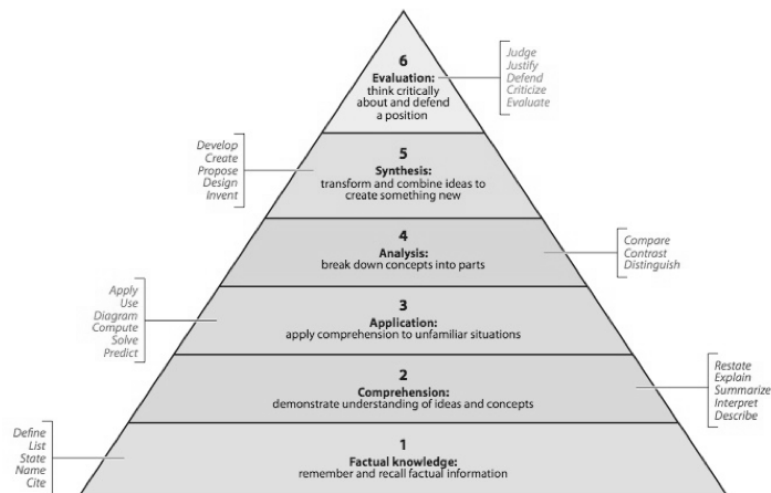


Figura 4.6: La tassonomia di Bloom in Wood [2009]

La *conoscenza* è il livello tassonomico più semplice e rappresenta quasi un prerequisito, per padroneggiare i livelli tassonomici successivi; essa

è misurabile e verificabile con prove strutturate e consente valutazioni oggettive di aspetti specifici dell'apprendimento. Anche la *comprensione* e l'*applicazione* possono essere verificate con test e concorrono ad avere una quadro più completo della valutazione degli apprendimenti. *Analisi, sintesi e valutazione*, invece, sono processi mentali superiori e più difficili da valutare con strumenti oggettivi senza che si faccia ricorso a metodi statistici di analisi che consentano la misurazione di tratti latenti e indaghino la relazione tra le variabili prese in esame. Per procedere alla valutazione dei processi superiori, quindi, vanno individuati criteri epistemologici di riferimento e modelli statistici di indagine.

La tassonomia di Bloom è solo uno dei tanti esempi che si possono scegliere nel campo della ricerca pedagogica per definire in modo sistematico le abilità e le conoscenze che potrebbero, in relazione ad ogni specifico percorso formativo, essere oggetto di valutazione.

Per la costruzione e la gestione di una prova oggettiva, una volta determinati gli ambiti disciplinari oggetto di verifica, occorre ancora stabilirne le funzioni valutative alle quali deve assolvere la prova di verifica stessa. Occorre quindi tener presente se si tratta di una *verifica iniziale*, mirata all'individuazione di pre-requisiti disciplinari e/o ad analisi di contesto; di una *verifica in itinere* per monitorare l'andamento del processo formativo; di una *verifica sommativa* per valutare complessivamente le conoscenze acquisite al termine di un percorso formativo.

Quando ci riferiamo alla valutazione degli apprendimenti, occorre tener presente che ogni classe di studenti è caratterizzata da un forte eterogeneità nelle capacità intellettive, nell'uso di strategie di pensiero, nell'interesse e nell'impegno allo studio, pertanto è necessario definire obiettivi di apprendimento minimi che imprescindibilmente tutti devono conseguire. Si tratta, dunque, di stabilire quale sia il livello di accettabilità, necessario a ritenere sufficiente la prova esaminata. Nel caso in cui, per attribuire un giudizio o voto di sufficienza, l'insegnante faccia riferimento al raggiungimento degli obiettivi minimi programmati per la classe a cui appartiene alunno, l'accettabilità della prova si definisce funzionale; se invece si tiene conto del livello medio conseguito dagli alunni in seguito alla somministrazione di una prova di verifica oggettiva ed a tale livello si pone la soglia di accettabilità,

allora si parla di accettabilità statistica.

4.6.2 Tipologie di prove strutturate

La tipologia della prova da somministrare, come abbiamo anticipato, deve essere scelta in relazione agli obiettivi che si intendono sottoporre a verifica. Vengono di seguito presentati alcuni esempi esplicativi delle diverse tipologie di prove oggettive di verifica di cui un insegnante può disporre. Le prove di verifica sono classificate in relazione al loro livello di strutturazione.

	STIMOLO CHIUSO	STIMOLO APERTO
RISPOSTA CHIUSA	Scelta multipla Vero/Falso Corrispondenze Tabelle Cloze <i>Appartengono alla categoria delle prove strutturate</i>	Serie di domande che richiedono risposte brevi, a volte monosillabiche (sì, no, errato, non sono d'accordo, ecc.) <i>Appartengono alla categoria delle prove semi-strutturate (la struttura è nella risposta)</i>
RISPOSTA APERTA	Serie di quesiti ben precisi cui si richiede di dare risposte elaborate autonomamente dall'allunno, sulla base di criteri direttivi ben definiti. <i>Appartengono alla categoria delle prove semi-strutturate (la struttura è nella domanda)</i>	Tema Interrogazione orale Relazioni Lavori di gruppo Prove pratiche <i>Appartengono alla categoria delle prove non strutturate</i>

Figura 4.7: Classificazione delle prove di verifica in riferimento al livello di strutturazione dello Stimolo/Risposta

Come si evince dalla figura 4.7, le prove di verifica più rigidamente strutturate sono quelle che prevedono una risposta chiusa ad uno stimolo chiuso. Vedremo, quindi, più nel dettaglio le caratteristiche delle diverse tipologie di prove che appartengono a questa categoria.

- *Scelta multipla*: è la prova di verifica strutturata più largamente diffusa in ambito formativo, si utilizza, infatti, in tutti gli ordini e gradi di scuola. È costituita da una serie di item, ognuno dei quali presenta una domanda stimolo ed alcune risposte fra le quali scegliere quella esatta. Le possibilità di scelta devono essere non meno di 3 e non più di 5. Tra esse occorre inserirne una risposta assolutamente vera, almeno una assolutamente falsa e alcune attendibili, ossia che non siano né sicuramente vere, né sicuramente false. In quest'ultimo caso le risposte fungono da *distrattori* e costringono lo studente ad un esame attento di ciascun item. A titolo esplicativo consideriamo il seguente esempio:

1 La media e la varianza della variabile statistica X = "punteggio nel test di conoscenza della lingua inglese scritta" valgono rispettivamente 11,36 e 2,5.

La variabile statistica Y = "punteggio conoscenza della lingua inglese parlata" è legata ad X dalla relazione $Y = 2,6 + 3X$.

La media di Y vale

- a) 36,88;
- b) 34,08;
- c) 36,68*;
- d) 13,96;
- e) Non so.

La domanda, come si può osservare, deve focalizzare un solo problema o concetto, esprimendolo con precisione per evitare incertezze e confusioni nella scelta della risposta.

I distrattori devono essere plausibili rispetto al problema considerato e, per contenuto e natura, devono risultare in qualche modo collegati alla domanda. I distrattori e la risposta corretta, inoltre, si devono presentare nella stessa forma, ossia devono essere relativamente simili, sia per struttura sintattica, sia per lunghezza. È essenziale evitare che le risposte esatte e i distrattori riprendano termini o formulazioni usate nella domanda per non condizionare la scelta, poiché se lo studente individua parte

della domanda nella risposta è portato d’impatto a fare quella scelta senza riflettere. Se, ad esempio, la risposta b fosse stata “La media di Y vale 34,08”, lo studente sarebbe stato indotto d’impeto a questa scelta senza effettuare calcoli poiché sarebbe stata l’unica risposta a richiamare la parola “media”.

L’item deve essere costruito con una forma semplice ed esplicita, e formulato per mezzo di una frase in forma interrogativa o affermativa. I diversi item, inoltre, non devono concatenarsi o essere l’uno conseguente dell’altro, pertanto si deve fare in modo che i quesiti siano tra loro indipendenti. Non bisogna cadere ancora nell’errore di usare risposte esatte alle domande precedenti, come distrattore o come risposta corretta in altri quesiti. Nei quesiti relativi alla comprensione della lettura, il linguaggio deve essere simile a quello del testo al quale ci si riferisce e gli elementi grammaticali, o la struttura stessa della frase non devono favorire l’individuazione della risposta esatta. Se si dovesse ritenere necessario che le alternative di risposta inizino con lo stesso gruppo di parole, è preferibile includere il gruppo di parole nel corpo della domanda e porre nella risposta solo la parte conclusiva della frase. L’uso di alcune espressioni può influenzare la risposta: bisogna evitare, quindi, di utilizzare espressioni come “sempre”, “tutti”, “dappertutto”, che inducono a credere nell’inesattezza della risposta che le contiene ed evitare ugualmente termini come “quasi sempre”, “spesso”, “qualche volta” che inducono a credere nell’esattezza delle alternative. Unica eccezione a quanto detto è il caso in cui tali espressioni siano usate in scala graduata: “sempre-quasi sempre-talvolta-quasi mai-mai”. È necessario, inoltre, evitare di inserire negazioni semplici o doppie nel corpo domanda, se fosse proprio indispensabile, bisognerebbe avere l’accortezza di evidenziare la negazione in neretto, con sottolineatura o usando la lettera maiuscola. In fine, quando si propongono risposte che indicano cifre o quantità è opportuno disporle in ordine crescente o decrescente.

Oltre alle regole citate che occorre seguire per costruire una prova di verifica a scelta multipla bisogna ancora prestare atten-

zione al rispetto di alcune convenzioni grafiche: le alternative di risposta dei quesiti che terminano con il punto interrogativo devono cominciare con la maiuscola; i distrattori dei quesiti a completamento devono iniziare con la minuscola; alla fine di ogni alternativa di risposta si deve usare il punto; il corpo della domanda nella forma a completamento non deve terminare mai con i due punti o con i puntini sospensivi; le alternative di risposta vanno elencate e non disposte di seguito una dietro l'altra. Il punteggio da assegnare ad ogni item, nella pratica docimologica, viene individuato seguendo lo schema in tabella 4.2. Tuttavia molti esaminatori attribuiscono percentuali di punteggio da attribuire ai diversi distrattori in base a quanto essi si avvicinano alla risposta esatta.

Punti	Risposta
$k - 1$	<i>Risposta esatta</i>
$-\frac{1}{(k-1)}$	<i>Risposta errata</i>
0	<i>Risposta non data</i>

Tabella 4.2: Esempio punteggi per quesiti a risposta multipla

dove k indica il numero di alternative.

Il fattore caso, per questa tipologia di prove, è molto ridotto. Se si struttura una prova costituita da 40 item e si presentano quattro risposte tra cui scegliere, considerando come livello di sufficienza la soluzione corretta del 70% delle domande, la probabilità di ottenere tale risultato fornendo risposte casuali si mantiene entro valori del tutto trascurabili, essendo pari allo 0,1%. Il numero ottimale delle scelte, per ridurre la possibilità di una scelta casuale, è di 4 o 5 poiché in tal caso la probabilità di rispondere correttamente ad un'intera prova di verifica composta ad esempio da 40 item sarebbe dell'ordine di $1/4^{40}$ o $1/5^{40}$.

Un accorgimento per ridurre l'incidenza della scelta casuale delle risposte è quello di prevedere la collocazione della risposta corretta in modo che nell'insieme dei quesiti, le risposte esatte

non corrispondano tutte alla stessa lettera; questo accorgimento è previsto espressamente anche dalla normativa sulla terza prova dell'esame di stato. I quesiti a scelta multipla, in conclusione, consentono di verificare, rispetto alle altre tipologie di prove strutturate, un ampio spettro di conoscenze e di capacità e di rendere trascurabile la probabilità di risposta corretta casuale. Questo tipo di prove di verifica si sta diffondendo notevolmente, non solo nelle valutazioni in itinere da parte degli insegnanti, ma anche nell'esame di maturità con l'introduzione della terza prova, nelle valutazioni nazionali dell'INVALSI, negli esami universitari e in ambito professionale.

- *Vero/Falso*: questo tipo di prova, utilizzato preferibilmente per verificare conoscenze semplici, comprende tutti quegli item che propongono una scelta fra due alternative, di cui una sicuramente falsa e una sicuramente vera. Oltre alle opzioni di scelta vero/falso, si possono utilizzare sì/no, giusto/errato, ed altre che comunque garantiscano la scelta tra due alternative contrapposte tra loro. Anche per questa tipologia di item occorre utilizzare alcuni accorgimenti.

Per quanto concerne la formulazione linguistica è opportuno utilizzare affermazioni che facciano riferimento ad un solo concetto, che deve essere espresso in forma chiara e concisa, evitando dettagli secondari e negazioni nel corpo della domanda. Il linguaggio utilizzato deve essere adeguato al livello dei destinatari e delle tematiche da valutare e deve evitare l'uso di quantificatori quali "sempre", "spesso", "qualche volta", "mai", ecc., in quanto il loro inserimento può indurre a ritenere più giusta o errata l'affermazione che li contiene a prescindere dal contenuto del quesito.

Per quanto concerne l'organizzazione dei quesiti è opportuno controllare l'indipendenza delle diverse affermazioni, poiché in caso contrario, si potrebbe verificare che un'affermazione possa facilitare o pregiudicare la risposta ad un'altra. Il numero delle affermazioni vere e di quelle false dovrebbe bilanciarsi nel totale dei quesiti. Per l'attribuzione dei punteggi possiamo considerare la tabella 4.3. Una limitazione dei quesiti di tipo

Punti	Risposta
$k - 1 = 2 - 1 = 1$	<i>Risposta esatta</i>
$-\frac{1}{(k-1)} = -1$	<i>Risposta errata</i>
0	<i>Risposta non data</i>

Tabella 4.3: Attribuzione punteggi Quesiti Vero/Falso

vero o falso è la possibilità di tirare ad indovinare la risposta giusta. Anche senza riflettere sull'item, infatti, l'alunno ha il 50% di possibilità di dare la risposta esatta. Questa limitazione può essere superata includendo un ampio numero di risposte del tipo vero/ falso e presentando più quesiti per individuare una stessa abilità, in questo modo la possibilità di risposte esatte sono progressivamente ridotte. Altro accorgimento può essere quello di prevedere la penalizzazione dell'errore, che approfondiremo in seguito.

- *Corrispondenze o relazioni*: con questi termini si raggruppano quei quesiti che prevedono di stabilire delle relazioni tra i termini di due insiemi distinti, solitamente distribuiti su due diverse colonne. Per fare un esempio, osserviamo la tabella 4.8. Questo tipo di prova si può utilizzare per verificare conoscenze molto specifiche e di contenuto immediato. Il numero delle corrispondenze da inserire nella prova dovrebbe essere compreso tra 5 e 15. È opportuno evitare di costruire una corrispondenza perfetta tra le due colonne, cioè le due liste dovrebbero essere disomogenee per numero, al fine di portare lo studente a non rispondere per esclusione e a dover riflettere su tutte le relazioni tra gli elementi dei due gruppi. Se il numero degli elementi fosse uguale, inoltre, l'ultima associazione non avrebbe alcuna possibilità di scelta. Il numero degli elementi della colonna delle "rispost"e al massimo deve essere il doppio di quelle a cui chiede di associare. In questo tipo di prova il fattore caso è trascurabile e la correzione risulta semplice ed oggettiva. Per l'attribuzione del punteggio si da un punto per ogni relazione esatta.

Metti in relazione ogni capitale con il rispettivo Stato, collegando in modo opportuno gli elementi presenti nei due gruppi:

Roma	Svizzera
Parigi	Germania
Vienna	Italia
Madrid	Polonia
Londra	Scozia
Atene	Inghilterra
Lisbona	Portogallo
Ginevra	Grecia
Bonn	Spagna
Varsavia	Romania
	Francia
	Austria

Figura 4.8: Esempio: Corrispondenza o relazione

- *Tabelle*: è una prova di verifica strutturata, costituita da tabelle a doppia entrata nelle quali, secondo le indicazioni date nella consegna, lo studente deve compilare le caselle lasciate libere, inserendo i dati mancanti. La correzione è immediata e si attribuisce un punto ad ogni completamento esatto.
- *Cloze o Completamento*: lo studente deve completare un esercizio inserendo le parti mancanti, che possono essere liberamente inserite o scelte da un elenco predefinito che può contenere anche alternative errate per fungere da distrattori.

Il termine è stato coniato, nella letteratura psico-linguistica, dal giornalista Wilson Taylor [Taylor, 1953] come metodo di misurazione della leggibilità dei testi. Taylor trasse ispirazione dai “closure tests” (test di chiusura) già usati dagli psicologi della Gestalt. I Gestaltisti proponevano figure semplici, in genere di tipo geometrico, con qualche elemento mancante da inserire per completare il quadro. I Gestaltisti attribuivano a questa esperienza

Completa, in tutte le sue parti, la seguente tabella, inserendo in modo opportuno i dati mancanti

	Perimetro	Area	Diagonale
Quadrato di lato..... cm			$2\sqrt{2}$
Quadrato di lato 3 cm			
Quadrato di lato..... cm		cm ² 249	

Figura 4.9: Esempio: Compilazione di tabella

una dimensione totalitaria in grado di indurre la strutturazione e la relazionalità dei dati osservati nel tutto da realizzarsi. Taylor per primo pensò di utilizzare questo strumento in campo linguistico.

I cloze necessitano di alcune regole precise per la loro strutturazione. Prima di tutto i periodi iniziali e finali dovrebbero rimanere integri, per consentire di comprendere correttamente il significato del testo. Nel corpo del testo si omettono, quindi, dei termini che lo studente dovrà inserire. Non è opportuno eliminare più di 2 o 3 parole per periodo per evitare eccessive ambiguità. Bisogna, inoltre, evitare di costruire quesiti con più di 15 completamenti. Per non facilitare troppo la prova, la lunghezza degli spazi deve essere sempre la stessa e non proporzionale al termine da inserire e i termini da inserire devono essere elencati seguendo un ordine alfabetico.

Per fare in modo che la verifica consenta di valutare con accuratezza il livello di conoscenza posseduto dallo studente su un dato argomento, si può rendere la prova più articolata inserendo per ogni parola da completare, oltre alla risposta esatta, almeno un'alternativa. Tuttavia il numero di parole da inserire deve

- a) Completa inserendo al posto dei puntini i seguenti termini: *piogge acide, ossigeno, inquinamento, anidride carbonica, fotosintesi clorofilliana, linfa elaborata*.

Le piante sono gli unici esseri viventi capaci di combattere l'_____ dell'aria, infatti, attraverso la funzione della _____ utilizzano l'anidride carbonica e restituiscono _____. Purtroppo i veleni presenti nell'aria, cioè le sostanze inquinanti, si mescolano con la pioggia formando le _____ che possono danneggiare gravemente le piante e addirittura rovinare i monumenti.

Figura 4.10: Esempio: Cloze

essere al massimo doppio rispetto agli spazi.

Il docente prima di somministrare la prova deve verificare la coerenza sintattica e grammaticale tra testo da completare e termini presentati in calce, evitando di utilizzare sinonimi tra i termini da inserire. Esistono diverse tipologie di cloze, in relazione alla loro strutturazione possiamo distinguere:

- Completamento di una frase senza alternative di risposta.
- Completamento di una frase con alternative di risposta.
- Completamento di un testo senza alternative di risposta.
- Completamento di un testo con alternative di risposta.

Per l'attribuzione del punteggio si può proseguire secondo quanto indicato nella tabella 4.4. Per una valutazione oggettiva degli

Tipologia risposta	Punti	Risposta
<i>Scelta da un elenco</i>	+2	<i>Risposta esatta</i>
<i>Scelta da un elenco</i>	−1	<i>Risposta errata</i>
<i>Risposta libera</i>	+2	<i>Risposta esatta</i>
<i>Risposta libera</i>	0	<i>Risposta errata</i>

Tabella 4.4: Attribuzione punteggi con prove di tipo “Cloze”

apprendimenti si possono utilizzare, oltre a quelle elencate sopra, anche una serie di prove di verifica semi-strutturate, che

garantiscono un buon grado di oggettività. La strutturazione può interessare lo stimolo, per cui avremo prove con stimolo chiuso e risposta aperta, o la risposta, in questo caso si tratterà di domande molto articolate che però richiedono una risposta chiusa del tipo “sì”, “no”, “sono d’accordo”, ecc.

- *Questionario aperto*: è quel questionario che dopo ciascuna domanda lascia uno spazio libero per consentire di dare la risposta. Tale strumento non presenta i limiti della casualità poiché richiede una conoscenza precisa dell’argomento, ma bisogna stabilire a priori se si vogliono risposte sintetiche, discorsive o argomentative. Nel primo caso si valuta semplicemente il numero di risposte esatte e il punteggio viene stabilito caso per caso tenendo conto delle difficoltà presentate. La valutazione e il punteggio da assegnare dovranno tener conto della pertinenza, dell’esattezza e della completezza. L’errore non viene penalizzato togliendo punti, ma solo non assegnandone.
- *Breve saggio*: è una delle prove semi-strutturate più difficili da valutare perché più condizionabile dalle interferenze soggettive del docente. Sostanzialmente si chiede allo studente di esporre in un numero di righe predefinito un determinato argomento. Per l’attribuzione del punteggio si può stabilire una scala empirica di punteggi che può essere strutturata come segue:
 - mancanza di risposta: 0 punti;
 - risposta esatta ma troppo essenziale o risposta generica: 1 punto;
 - risposta che mostra comprensione dell’enunciato, ma inadeguata per la qualità dell’esposizione e del contenuto: 2 punti;
 - risposta adeguata allo scopo, per contenuto ed esposizione: 3 punti.
- *Risoluzione di un problema*: è certamente una prova classica di verifica che presenta un buon livello di strutturazione poiché presuppone una risoluzione univoca. La valutazione di un problema

non è certamente una cosa semplice, poiché occorre valutare diversi aspetti, quali l'esatta conoscenza delle regole da applicare, la correttezza del procedimento, l'economia del procedimento, la correttezza dei calcoli, la correttezza delle soluzioni e delle risposte. Per l'attribuzione dei punteggi si può procedere con la costruzione di una scala empirica, come nell'esempio precedente.

4.6.3 La costruzione delle prove di verifica strutturate.

Quando si determinano i diversi item nella strutturazione di una prova di verifica occorre sempre iniziare con i quesiti più semplici e gradualmente andare avanti con i più difficili; in modo tale che lo studente non si blocchi al primo quesito, ma abbia la possibilità di ripercorrere gradualmente il percorso fatto per elaborare le soluzioni richieste. Occorre, inoltre, valutare con cura la lunghezza che deve avere la prova. Solitamente per una valutazione formativa sono necessari 20 – 25 quesiti, mentre per una valutazione sommativa ne occorrono 40 – 50. Un sistema largamente utilizzato per determinare quanto debba essere lunga una prova di verifica è quello di far riferimento al tempo impiegato dall'insegnante per la trattazione dei contenuti oggetto di analisi. Solitamente, infatti, si ritiene che un apprendimento più semplice necessiti di minor tempo rispetto ad un apprendimento complesso, pertanto si dà un peso maggiore a quelle verifiche mirate a valutare conoscenze che abbiano necessitato di tempi più lunghi. Per attribuire correttamente l'indice di importanza ai diversi quesiti, si può utilizzare, quindi, il fattore tempo utilizzato dal docente per portare gli alunni ad acquisire l'obiettivo specifico, che è indagato da un singolo item, e rapportarlo al tempo totale che ha impiegato per far apprendere l'obiettivo generale di conoscenza. Sostanzialmente si determina una proporzione del tipo:

$$\frac{\text{obiettivo specifico}}{\text{obiettivo generale}} = \frac{\text{tempo impiegato per ob. specifico}}{\text{tempo impiegato per ob. generale}}$$

Se allora si considerano i tempi impiegati per la trattazione di un singolo obiettivo specifico, si può determinare il grado percentuale di importanza da assegnarsi agli item che intendono valutare quell'obiettivo,

utilizzando la seguente formula:

$$p_j = \frac{T_{os}}{T_{og}} \cdot 100$$

dove,

p_j è il peso assegnato all'item j – esimo;

T_{os} è il Tempo dedicato a ciascun obiettivo specifico;

T_{og} è il Tempo complessivo dedicato all'obiettivo generale.

Per comprendere meglio facciamo un esempio pratico: consideriamo che per far conseguire un determinato obiettivo generale (X), l'insegnante suddivide questo in una serie di obiettivi specifici ($x_1, x_2, \dots, x_n$) e che per ognuno di essi impieghi tempi diversi. Quando la stessa insegnante struttura la prova di verifica per valutare il conseguimento dell'apprendimento dell'obiettivo X, dovrà valutare tutti i singoli obiettivi specifici, assegnando ad ognuno di essi un peso, in modo da poter stabilire quanti item siano necessari per sondare ogni singolo obiettivo specifico. Per rendere più semplice la determinazione del numero di obiettivi idonei a valutare ogni obiettivo specifico, può essere utile costruire una tabella, come segue (vedi tabella 4.5):

Obiettivo generale (X)	Tempo assegnato (T_{os})	Peso assegnato (p_j)
<i>Ob. specifico x_1</i>	<i>4 ore</i>	<i>8</i>
<i>Ob. specifico x_2</i>	<i>2 ore</i>	<i>4</i>
<i>...</i>	<i>...</i>	<i>...</i>
<i>Ob. specifico x_n</i>	<i>6 ore</i>	<i>12</i>
Tempo assegnato (T_{og})	<i>50 ore</i>	<i>100</i>

Tabella 4.5: Valutazione dell'importanza dell'item

Se l'insegnante per verificare X vuole preparare una prova di verifica composta, ad esempio, da 25 item, dovrà predisporre 2 item per x_1 , che rappresentano l'8% del peso totale della prova; 1 item per x_2 , 3 item per x_n , e così via.

Il passo successivo da compiere è quello della strutturazione dei diversi item per la prova di verifica scelta. L'insegnante, quindi, sceglierà

il tipo di prova strutturata e per ogni obiettivo da valutare costruirà gli appositi item. Si procede, quindi, alla costruzione dei punteggi. Questa fase è essenziale per garantire la validità e la veridicità dei risultati, poiché una delle componenti che garantisce oggettività alla prova sta proprio nella predefinizione e nella condivisione dei punteggi. I punteggi che si attribuiscono agli item si definiscono punteggi grezzi e non sono da confondere con la valutazione finale della prova, poiché sarà necessario eseguire un passo successivo per convertire il punteggio grezzo totale di una prova in voto.

I punteggi ottenuti possono essere riferiti ad un criterio, ovvero ad un punteggio soglia che stabilisce il livello di accettabilità della prestazione, oppure alla norma, ovvero alla prestazione media conseguita dal gruppo in cui il soggetto è inserito.

Abbiamo già visto, durante la trattazione delle tipologie di prove di verifica strutturate, alcuni esempi per l'attribuzione dei punteggi ai singoli item, per tanto non ci dilungheremo oltre nella trattazione di questo aspetto. Un concetto che, invece, è il caso di approfondire è quello della *penalizzazione dell'errore* che ha lo scopo di ridurre l'incidenza sul punteggio finale delle risposte che lo studente ha "tirato ad indovinare". Sostanzialmente se non si penalizzano gli errori, il punteggio grezzo totale sarà dato dalla somma algebrica dei punteggi attribuiti alle risposte esatte ed errate; se, invece, si penalizzano le risposte errate, si dovrà togliere dal punteggio grezzo totale quella parte di esso che potrebbe essere stata determinata dal caso. Per far questo è sufficiente applicare una semplice formula:

$$P_p = R_c - \frac{R_e}{(k - 1)}$$

dove,

P_p è il punteggio penalizzato per ciascun tipo di item;

R_c è il numero delle risposte corrette per gli item di ciascun tipo;

R_e è il numero delle risposte errate per gli item di ciascun tipo;

k è il numero delle alternative delle risposte.

Questa formula è valida per le prove strutturate a risposta multipla, poiché nel caso, ad esempio, di una verifica del tipo "vero/falso" la penalizzazione dell'errore corrisponderebbe al rapporto $-\frac{1}{k-1} = -\frac{1}{2-1} = -1$. Occorre ancora ricordare che se alle diverse prove è

stato attribuito uno specifico peso, quando si penalizza l'errore non ci dobbiamo scordare di prenderlo in considerazione; pertanto la formula precedente diventerà:

$$P_p = R_c - \frac{R_e}{(k-1)} \cdot p_j$$

dove p_j rappresenta, appunto, il peso attribuito a ciascun item.

Procediamo ad un esempio concreto: supponiamo che un allievo sia stato sottoposto ad una batteria di test per valutare l'acquisizione di un apprendimento e che abbia risposto alle diverse tipologie di prove come esposto in tabella 4.6

Tipo di prova	Peso Item	Penalità risposta errata	E risposte esatte	S risposte errate	Punteggio grezzo senza P _p	Punteggio grezzo con P _p
Scelta multipla	3	-1	9	6	27	21
Vero-Falso	1	-1	15	5	15	10
Vero-Falso	1	-1	15	5	15	10
Vero-Falso	1	-1	10	—	10	10
Punt.Tot.					52	41

Tabella 4.6: Esempio di penalizzazione dell'errore

Tra i risultati dei punteggi grezzi totali c'è una notevole differenza se si penalizza, o no l'errore, pertanto se non si è certi della risposta, piuttosto che tentare la sorte, conviene non rispondere!

Da un punto di vista matematico possiamo considerare la penalizzazione dell'errore come il reciproco del numero delle alternative errate proposte dal singolo quesito. Praticamente se il quesito prevede 5 alternative di cui una sola corretta e 4 errate e alla risposta corretta si è deciso di attribuire 4 punti, allora la penalità dovrà essere $-4 \cdot 1/4$, quindi -1 . Applicando una penalità di -1 alle risposte errate, ovviamente, si può incorrere nel caso in cui, se lo studente risponde erroneamente al maggior numero delle risposte o addirittura le sbaglia tutte, si possono ottenere punteggi negativi. Ad esempio su 40 item a 5 alternative con 35 risposte errate si ottiene un punteggio pari a -15 , poiché per le 5 risposte corrette si attribuisce un totale di 20 punti, per le 35 errate un totale di -35 .

Come già detto in precedenza, la penalizzazione dell'errore viene utilizzata dall'insegnante e preventivamente comunicata agli studenti al fine di scoraggiare la scelta di rispondere in modo casuale alle domande. Tuttavia, se si vuole intendere come livello di sufficienza la soluzione corretta del 70% delle domande, la probabilità di ottenere tale risultato fornendo risposte casuali si mantiene entro valori del tutto trascurabili, essendo pari a $1/1000$. È preferibile, quindi, evitare di penalizzare gli errori, sia per non incorrere in calcoli più complessi, sia per evitare punteggi con segno negativo.

4.7 La valutazione scolastica degli apprendimenti: tradizione e innovazione nello scenario italiano

Abbiamo tanto parlato di valutazione, ma occorre riflettere su quale sia il significato profondo del "valutare". Quali i fini della valutazione oggi, per una scuola che si dichiara "scuola dell'innovazione? Quali i mezzi più idonei e le strategie più sofisticate per valutare in un'epoca così complessa? E ancora ... Chi deve valutare? Chi o che cosa deve essere valutato? Dove? Quando? Come? Perché?

Valutare (dal latino *vàlitus*) significa, etimologicamente, valere, avere prezzo, stimare. La valutazione è uno dei compiti fondamentali dai quali nessun insegnante può esimersi, pertanto è necessario che,

chi è chiamato a valutare, abbia ben chiaro il fine che intende perseguire, l'oggetto al quale deve essere dato un valore e quale valore sia opportuno attribuirgli. Valutare il risultato conseguito da uno studente, misurare la sua preparazione, comprendere l'effettiva validità di quell'apprendimento, non può essere solo un'etichettatura di valore; la scuola non può diventare un supermercato di menti, nel quale il docente si limiti ad etichettare la testa dello scolaro con il "giusto prezzo".

La docimologia (termine coniato dal francese Piéron [1922], da *dokimázo*, esaminare, e *dokimasía*, prova/esame, + *logos*, discorso), è stata affrontata in Italia proprio nel suo significato etimologico di "riflessione scientifica sugli esami", di studio destinato alla critica e al miglioramento delle votazioni scolastiche. De Landsheere, G. [1973] la definisce come la "scienza che ha per oggetto lo studio sistematico degli esami, in particolare dei sistemi di votazione e del comportamento degli esaminatori e degli esaminati". Oggi si preferisce denominare l'area di ricerca della valutazione educativa con il termine di "teorie e tecniche della valutazione", poiché alle ricerche di tipo tradizionale centrate soprattutto sulla valutazione quantitativa degli apprendimenti, si sono affiancate più recentemente quelle di tipo qualitativo. L'approccio alla valutazione nelle scuole, oggi, è quasi esclusivamente di tipo quantitativo e si preoccupa di valutare più il prodotto, attraverso la somministrazione di prove di verifica strutturate, che non il processo, in quanto l'obiettivo prioritario del sistema di istruzione è ancora quello di individuare le modalità più idonee di certificazione di conoscenze, piuttosto che supportare la ricerca nella comprensione dei processi che sottendono all'apprendimento.

In Italia, infatti, per molti anni è stato largamente diffuso un approccio di tipo "nomotetico", volto a sviluppare ricerche mirate a pervenire a leggi generali che, tramite applicazioni e procedure statistiche, riescano a stabilire caratteristiche "medie", ritenute standard di situazioni e comportamenti. Con il progredire della ricerca in campo educativo è emersa, però, la necessità di individuare metodi e strumenti idonei ad analizzare e valutare i processi che consentono apprendimento. Questo nuovo approccio, definito dalla letteratura di tipo "ideografico", mirato al recupero dell'originalità dei singoli nel rispetto della specificità che caratterizza tutti e ciascuno, si pone come finalità

l'analisi dei processi formativi e la valutazione della qualità degli apprendimenti. Valutazione quantitativa e valutazione qualitativa, che spesso nella letteratura pedagogica si sono trovate ideologicamente contrapposte, trovano oggi nella scuola dell'innovazione la possibilità di compenetrarsi per consentire una valutazione più completa dello studente.

I nostri studenti, il futuro della nostra nazione, hanno il bisogno e il diritto di essere adeguatamente educati a vivere come cittadini autonomi, consapevoli, responsabili e competenti nell'affrontare la complessità che caratterizza questa nostra era planetaria. Morin *et al.* [2005], nella sua opera "Educare per l'era planetaria", partendo dall'analisi etimologica del termine "complessità" (dal latino cum *plectere*: intrecciare, collegare), spiega in che misura la complessità, quale ordito di elementi eterogenei, sia la caratteristica più evidente del XXI secolo ed evidenzia che "la missione dell'educazione per l'era planetaria è quella di rafforzare le condizioni che renderanno possibile l'emergenza di una società-mondo composta da cittadini protagonisti, impegnati in modo consapevole e critico nella costruzione di una civiltà planetaria.

Alla luce di quanto detto, risulta evidente che educare è, oggi più che mai, un compito che richiede agli insegnanti elevate competenze e un alto livello di professionalità, poiché essi in prima linea si trovano a dover gestire i repentini cambiamenti che avvengono in campo sociale, economico e culturale, che caratterizzano il nuovo secolo. L'innovazione tecnologica, la messa in discussione di valori fino ad oggi ritenuti inviolabili, la forte interrelazione che si è stabilita tra culture diverse e la globalizzazione hanno determinato repentini e profondi cambiamenti. I docenti delle nostre scuole, formati prevalentemente con metodi legati alla tradizione, si sono dovuti rimettere in gioco per stare al passo con i tempi, per affrontare il cambiamento e gestire l'imprevedibile, salvaguardando al tempo stesso i valori inalienabili dell'umanità.

Il Consiglio Europeo straordinario di Lisbona, tenutosi nei giorni 23 e 24 marzo 2000, è nato dalla volontà di imprimere un nuovo slancio alle politiche comunitarie attraverso la progettazione di un obiettivo strategico per l'Unione finalizzato alla risoluzione del problema occupazionale, alla realizzazione di riforme economiche e alla coesione sociale nel contesto di un'economia basata sulla conoscenza.

Due recenti evoluzioni stanno modificando profondamente l'economia e la società contemporanee:

- la globalizzazione dell'economia che costringe l'Europa ad essere all'avanguardia in tutti i settori nei quali è forte l'intensificarsi della concorrenza;
- l'arrivo repentino e l'importanza crescente delle tecnologie dell'informazione e delle comunicazioni sia nella sfera professionale che in quella privata.

I sistemi di istruzione dell'Unione Europea si sono, di conseguenza, dovuti impegnare ad una revisione completa del sistema stesso e garantire l'accesso, per tutti i cittadini della comunità, alla formazione lungo tutto l'arco della vita. Il compito che si è dato l'Unione Europea è quello di modellare questi cambiamenti in modo coerente con i propri valori. Il Consiglio Europeo di Lisbona, dunque, ha cercato di formulare orientamenti in grado di cogliere le opportunità offerte dalla nuova economia, prefissandosi, quindi, l'obiettivo strategico di "... diventare l'economia basata sulla conoscenza più competitiva e dinamica del mondo, in grado di realizzare una crescita economica sostenibile con nuovi e migliori posti di lavoro e una maggiore coesione sociale." Quelli di Lisbona sono obiettivi di medio e lungo termine che richiedono un eccezionale impegno politico ed organizzativo. Lo spazio europeo dell'istruzione e della formazione è oggi il punto ideale di raccordo dei progetti culturali, tecnologici e scientifici e punto di partenza di ogni progetto sociale ed economico. L'Italia in questa prospettiva, negli ultimi anni, si è impegnata a rinnovare profondamente scuola, università e strutture di formazione. Qualcosa si è fatto ma molto è ancora da fare.

Il Programma Operativo Nazionale *"La Scuola per lo sviluppo"* è stato fortemente improntato al conseguimento degli obiettivi europei. In particolare il piano di interventi contribuisce in maniera specifica a promuovere l'apprendimento lungo tutto l'arco della vita e ad offrire opportunità concrete di formazione permanente per garantire il passaggio dalla scuola alla vita attiva, attraverso il miglioramento della qualità del sistema scolastico. Il programma mira, inoltre, a favorire il conseguimento di maggiori capacità di adattamento ai cambiamenti

tecnologici ed economici nonché di qualifiche adeguate alle nuove necessità del mercato del lavoro. D'altronde la nuova comunicazione tecnologica mette su un altro piano le categorie spazio temporali poiché oltre a comunicare qui ed ora, si comunica ovunque e sempre. Si sono allargati gli orizzonti del mondo fisico consentendo attraverso quello virtuale di condividere esperienze formative, progetti di ricerca, attività lavorative. Ognuno stando nella propria scuola, sul proprio posto di lavoro e, sempre più, nella propria casa può condividere esperienze comunicative con persone lontane e sperimentare percorsi formativi innovativi. Il quadro storico-economio-sociale nel quale la scuola, di ogni ordine e grado, si trova ad operare oggi vede, quindi, gli insegnanti impegnati in una revisione globale delle finalità, degli obiettivi, delle strategie di intervento, dei sistemi di valutazione.

È tempo di guardare lontano e di trovare i modi e i mezzi più idonei a garantire nelle nuove generazioni lo sviluppo di un pensiero razionale, critico e, soprattutto, creativo che consenta a tutti e a ciascuno di vivere positivamente e serenamente la complessità della nostra era. I docenti di ogni ordine e grado hanno la responsabilità di essere protagonisti attivi nella costruzione di un sistema di istruzione innovativo e competitivo.

Il binomio, valutazione quantitativa – valutazione qualitativa, che in questi ultimi decenni è spesso risultato come dicotomico, alimentando contrapposizioni ideologiche che hanno additato il primo come tradizionalista e il secondo come progressista, deve oggi trovare il modo di mediare le due posizioni poiché in realtà si tratta di due aspetti che, se resi complementari, possono garantire una visione più globale della valutazione, come evidenzia?: *“[tali] tipologie di indagine [e di valutazione]... , l’una quantitativa ... , l’altra qualitativa ... non si differenziano per la presenza o meno di numeri, di descrizioni o di narrazioni ... non dicono cioè le stesse cose in modo diverso, ma conducono a conoscenze diverse [circa fattori indagati e valutati]”*, che devono essere integrate per garantire una visione olistica ed armonica dell’individuo.

Questo lavoro è la ricerca di un possibile percorso di integrazione tra sistemi di valutazione tanto diversi tra loro. L'accusa che la valutazione autentica o alternativa muove a quella tradizionale, basata sulle prove oggettive di verifica, è quella di limitarsi a valutare ciò che un ragazzo sa, ma di non riuscire a cogliere il processo attraver-

so il quale l'alunno è giunto ad una specifica conoscenza e neppure la reale capacità di applicazione, in una determinata situazione, di quanto appreso. L'accusa che la valutazione tradizionale muove alla valutazione autentica è di non essere scientificamente strutturata e basata su prove standardizzate ed obiettive. Lo scopo di questo lavoro è appunto quello di trovare una possibile soluzione per costruire un esempio di valutazione che riesca ad integrare i sistemi di valutazione più tradizionali con quelli più innovativi, poiché la scuola non può guardare solo al risultato finale ma anche al percorso formativo, al processo di insegnamento-apprendimento. Questa dicotomia tra prodotto e processo, tra risultato finale e percorso progressivo, tra esito atteso e sviluppo in itinere è forse più apparente che reale, ma per dimostrarlo occorre riflettere attentamente e trovare un possibile percorso di mediazione. D'altronde non si può pensare il futuro dimenticando il passato, poiché occorre essere consapevoli che la conoscenza è il veicolo per la partecipazione attiva ai mutamenti in corso. Il termine valutazione è oggi sempre più utilizzato, sia nel linguaggio comune che in quello specifico dei diversi settori disciplinari. Si parla di valutazione, infatti, in tutti i settori economici, sociali, politici e più in generale in ogni contesto organizzativo di qualunque natura e genere.

Ma cos'è la valutazione? e soprattutto, quando è opportuno ricorrere a strumenti specifici di valutazione? Se riflettiamo un attimo su quanto ci circonda, ci accorgiamo che è difficile trovare un fenomeno reale o una circostanza che non sia sottoposta ad un criterio di valutazione, fondato sulla nostra esperienza o basato su strumenti specifici.

Quando ci svegliamo la mattina la prima cosa che facciamo è decidere cosa indossare e la nostra decisione è subordinata ad una valutazione, per esempio in relazione alle condizioni atmosferiche o al luogo nel quale ci dobbiamo recare. Così, decidiamo di indossare un abito elegante per andare ad una cerimonia, comodo e pratico per affrontare un lungo viaggio, di lana in caso di freddo, di lino se fa particolarmente caldo, ecc Stessa cosa vale per la scelta della colazione, per il pranzo, per il mezzo di trasporto, ..., insomma ognuno di noi durante la giornata ricorre sovente a valutazioni che condizionano le scelte della vita quotidiana. Si utilizzano ancora criteri di valutazione anche quando si deve fare un acquisto importante, per

esempio la casa o l'auto; quando si valuta una proposta di lavoro, quando si progetta il proprio futuro. E ancora ... chi di noi non è stato sottoposto qualche volta nella sua vita ad una valutazione? Persino le nostre mamme, sin dai primissimi giorni di vita, ci hanno valutato: "Oh! Come è buono!"; " Assomiglia al padre."; "È così capriccioso, non dorme mai!", ecc ... Per non parlare poi di quando si va a scuola ... chi di noi non è mai stato sottoposto ad una prova di verifica, ad un'interrogazione, ad un esame? chi non ha fatto mai il confronto con i voti ottenuti da un compagno?

Oggi ogni organizzazione, pubblica o privata che sia, ha un comitato di valutazione preposto a stabilire l'efficienza e l'efficacia del proprio operato, sia esso rappresentato da servizi o da prodotti. Si parla di valutazione dei dirigenti, del personale, di ascolto del cliente per prendere decisioni (*customer satisfaction*). Insomma sembrerebbe che in ogni dove non si possa fare a meno di ricorrere alla valutazione.

L'atto del valutare è necessario, quindi, ogniqualvolta ci troviamo in una situazione nella quale dobbiamo prendere una decisione. In tutti gli esempi riportati abbiamo parlato di valutazione, ma allora è possibile affrontare il tema della valutazione da un punto di vista scientifico? Esiste un metodo trasversale a diversi ambiti di applicazione, che consenta di osservare in modo corretto un fenomeno reale ed aiuti a prendere la decisione più vantaggiosa? Quando osserviamo un fenomeno reale è necessario avere strumenti che ci consentano un'astrazione formale del fenomeno stesso, mirata alla comprensione profonda di esso. Questa fase della nostra indagine conoscitiva di un fenomeno reale possiamo chiamarla con il termine di *formalizzazione*. Solo dopo aver formalizzato un fenomeno reale, possiamo *individuare il modello* matematico o statistico che ci consente di misurare il fenomeno stesso e di procedere, quindi, ad una sua *valutazione*.

Dopo una disamina di alcuni dei principali problemi connessi alla valutazione, affronteremo l'argomento fornendo un metodo generale che sia valido per una pluralità di casi; riservandoci di trattare metodi specifici solo a casi di studio concreti.

La valutazione è, quindi, un'attività fondamentale che investe trasversalmente le varie fasi dell'azione formativa. È diventata, infatti, una fase imprescindibile dei progetti finalizzati all'educazione e alla formazione. La valutazione è l'atto, e al tempo stesso la conseguenza,

dell'attribuzione di un valore. In campo educativo e formativo, come peraltro sarebbe opportuno in ogni organizzazione, è necessario rendere esplicite le modalità e lo "strumento di misura" impiegati nella valutazione al fine di garantire oggettività, affidabilità e validità alla valutazione, nonché per rendere il processo valutativo replicabile.

Le fasi della valutazione generalmente sono individuabili nel modo seguente:

- valutazione ex -ante (osservazione e analisi delle risorse, sia in termini di capitale umano, sia in merito alle risorse materiali e tecnologiche; elementi di coerenza interna del progetto; grado di innovazione e di trasferibilità);
- valutazione in itinere (monitoraggio continuo del processo formativo; analisi della validità delle strategie e metodologie in atto; verifiche sui contenuti appresi; ecc. . .)
- valutazione ex - post (schede di valutazione; indicatori che ci possano fornire elementi di efficacia e di qualità dell'intervento; partecipazione ad indagini valutative nazionali e/o internazionali).

La valutazione ex-ante risulta essenziale nella fase di progettazione, in quanto serve alla preparazione dei piani, degli interventi e della programmazione, dei quali è parte integrante. Essa è finalizzata a verificare in primo luogo la fattibilità e la pertinenza di un intervento educativo. Per questa ragione la valutazione ex-ante implica l'analisi dei bisogni dei soggetti, del contesto e dei committenti. Durante questo primo momento valutativo si procede alla raccolta dei dati che consentiranno di ottenere informazioni relative ai destinatari dell'azione formativa e alle peculiarità del contesto in cui essa si sviluppa. Questi dati devono, quindi, essere messi in relazione con l'ipotesi di intervento educativo e formativo, valutando la congruità investimento-benefici e la trasferibilità.

La valutazione in itinere è fondamentale per monitorare lo stato di avanzamento dell'intervento formativo. La peculiarità di questa fase della valutazione è da ricercare non solo nella verifica del raggiungimento di uno specifico obiettivo, ma soprattutto nell'analisi dei processi che sottendono al processo di insegnamento-apprendimento.

Il monitoraggio in itinere del processo formativo consente, infatti, di intraprendere percorsi di ri-progettazione e ri-modellazione dell'offerta formativa.

Ai fini di valutare la modifica dei comportamenti e degli atteggiamenti nei confronti dell'apprendimento e della relazione interpersonale, in genere vengono utilizzate dai singoli docenti schede o griglie di osservazione; per quanto concerne la verifica degli apprendimenti solitamente vengono somministrate prove di verifica più o meno strutturate. È fondamentale comunque che i docenti, afferenti alle diverse discipline, si riuniscano periodicamente per confrontarsi al fine di ottenere una visione trans-disciplinare degli apprendimenti. La valutazione consente, infatti, di verificare anche la trasferibilità di una specifica esperienza ad altri contesti. La valutazione in itinere è utile, quindi, al coordinamento per verificare in corso d'opera la corrispondenza tra i risultati e gli obiettivi intermedi dell'azione educativo-formativa, sia sul piano del saper essere (atteggiamenti, comunicazione, relazione interpersonale); sia sul piano del sapere (conoscenze, competenze tecniche), sia sul piano del saper fare (modalità di lavoro, conoscenza pratica).

La valutazione ex-post è, come dice il termine stesso, successiva all'intero ciclo di attività formativa e può riguardare vari aspetti, quali ad esempio l'efficacia della formazione, la ricaduta sociale, l'aumento di produttività, l'incremento di produzione aziendale, ecc. . . . Essa può essere classificata in valutazione interna, che ci consente di verificare la conformità dei risultati ottenuti con gli obiettivi prefissati e valutazione esterna, che tiene conto della congruità tra i risultati ottenuti e i bisogni del territorio o del settore di attività. La valutazione ex-post può risultare estremamente importante per ottenere informazioni utili per la presa di decisione riguardo alle politiche future e per la valorizzazione delle buone prassi.

Il sistema di valutazione dei risultati nell'istruzione è sempre stato inteso come la fase dell'attività formativa dedicata alla quantificazione del grado di competenze acquisite da parte dei soggetti in formazione. Come è noto la disciplina che si occupa di questo aspetto è la docimologia.

Recentemente il mondo scientifico ha aperto un dibattito sul tema, mostrando che i sistemi scolastici che hanno dato maggior enfasi al

controllo degli standard raggiunti dagli studenti, tramite strumenti di valutazione quantitativa, "sistemi di Accountability", si sono contraddistinti per il conseguimento di livelli di qualità degli apprendimenti più elevati. Il fenomeno è più accentuato nelle strutture che hanno una maggiore autonomia nella gestione delle risorse e dei programmi di insegnamento. In altri termini, sembrerebbe che l'aumento della qualità formativa necessita contemporaneamente di politiche mirate all'incremento delle risorse, al controllo dei processi e al miglioramento dei risultati. In sintesi, i sistemi di Accountability possono essere visti come la produzione di informazioni sui risultati.

Le ricerche dimostrano che l'effetto sulla qualità è, tuttavia, strettamente subordinato all'assetto istituzionale a cui ci si riferisce. Mai come nei giorni nostri i governi sentono, infatti, l'esigenza di ridisegnare le linee guida dei settori primari dello stato: la sanità, la sicurezza e l'istruzione. Ciò è dovuto essenzialmente agli alti costi sostenuti in tali settori, che nel quotidiano si manifestano in cattivi servizi, alti costi diretti per l'utente finale, lungaggine burocratica ecc. Ecco che in uno scenario di tale tipo, un governo prima che ponga in essere azioni in qualunque direzione, deve implementare strumenti di comprensione del contesto: raccolta delle informazioni, elaborazione delle stesse e realizzazione di azioni finali, eventualmente correttive. Ma tutte queste fasi, che in linea di massima presentano già di per sé molte problematiche, diventano ancor più spinose se il contesto in analisi è la dimensione sociale che per definizione valuta e misura le capacità dell'individuo. Entra così, anche nel settore dell'istruzione, l'ottica di gestione manageriale, avvalendosi dei ben noti criteri di efficacia ed efficienza e quindi di economicità del servizio. La questione si evolve ulteriormente se indaghiamo sulla qualità, intesa nelle sue due anime: di prodotto-servizio o di processo.

Il contesto internazionale mostra un grosso sforzo per la misurazione del livello raggiunto dagli studenti nelle competenze di base: logico-matematiche, linguistiche e scientifiche, per le diverse età, con l'ausilio di test che garantiscano la comparabilità spazio-temporale.

Ovviamente esiste una relazione molto forte fra i diversi aspetti del processo insegnamento-apprendimento: i processi formativi (qualità del processo), le competenze degli studenti (qualità del servizio), le risorse degli istituti scolastici e la libertà di scelta delle famiglie. L'im-

plementazione di strumenti volti alla valutazione dell'insegnamento e dell'apprendimento non sono scevri da critiche e punti di debolezza, ma l'intento che si sta perseguendo in ambito internazionale, e da alcuni anni anche in Italia, è quello di strutturare ed inserire nei contesti scolastici sistemi di valutazione che siano il più possibile vicini agli studenti e agli insegnanti.

5 Classical Test Theory (CTT) vs Item Response Theory (IRT)

*“Ma la vita è un test tipo vero o falso’ o un test a risposta multipla?”
(Piperita Patty - Charles M. Schulz)*

In letteratura la ricerca in campo educativo ha affrontato il problema della conoscenza di fenomeni reali latenti, quali l’abilità, la capacità o più in generale le attitudini, convenzionalmente indicati con il termine *“latent trait”* [Lord *et al.*, 1968], che sono costrutti non direttamente osservabili, ma desumibili da un insieme di osservazioni (prove attitudinali, questionari, prove fisiche, mentali, test, ecc.) che consentono, appunto, di effettuare una stima dei fenomeni stessi. L’abilità di lettura e di scrittura, l’utilizzo efficace dell’apprendimento, la comprensione delle informazioni, la competenza... sono variabili intangibili che non possono essere osservate direttamente, di conseguenza, per studiarne le caratteristiche, il ricercatore necessita di strumenti di misura e di modelli di analisi che, applicati ad elementi osservabili, consentano di cogliere le peculiarità dei *tratti latenti*. Nel linguaggio di uso comune noi riusciamo facilmente a descrivere questi costrutti, elencandone consapevolmente gli attributi specifici, ma non possediamo uno strumento di misura che sia in grado di stimare direttamente la loro consistenza, in quanto la variabile latente esprime un concetto astratto e non una dimensione fisica concreta. L’obiettivo della ricerca valutativa in campo educativo è proprio quello di validare uno strumento di misura atto a cogliere la quantità di questi tratti latenti, che genericamente vengono indicati nel prosieguo della trattazione come *“abilità”*. Il termine *“abilità”* sarà, quindi, utilizzato in riferimento ad un *costrutto*, ossia ad un insieme di concetti astratti che indicano un aspetto della vita intellettuale del soggetto, non osservabile direttamente ma inferito a partire da una serie di indicatori empirici osservabili. Il costrutto si concretizza per mezzo di una fondata riflessione scientifica, di un’idea che consenta

di categorizzare e descrivere alcuni comportamenti di per sé non direttamente osservabili [Crocker and Algina, 1986], estrapolati da una consolidata teoria ed esplicitati operativamente per essere misurati tramite specifici indicatori. L'*indicatore* è una variabile che descrive il costrutto non osservabile tramite una regola di corrispondenza. Il rapporto tra costrutto e indicatori può delineare il verificarsi di due situazioni:

- gli indicatori riflettono il costrutto, ossia ne rappresentano la manifestazione empirica, come conseguenza della presenza del costrutto stesso (*indicatori riflettivi*), per fare un esempio consideriamo il costrutto "autostima" che si manifesta empiricamente, si riflette, nell'accettazione del sé, nella relazione con l'altro, nell'interazione con il contesto, ecc.
- gli indicatori, al contrario, formano, contribuiscono, determinano o addirittura causano il costrutto (*indicatori formativi*), per fare un esempio consideriamo il costrutto "competenza di lettura" che è determinato da il "comprendere e riconoscere il significato letterale o figurato di parole", "individuare informazioni", "cogliere la relazione logica", ecc.

Quando gli indicatori sono riflettivi, il costrutto è causa degli indicatori, quando sono formativi, essi stessi sono causa del costrutto. Questo comporta una diversa considerazione degli errori, delle correlazioni tra item e delle tecniche statistiche da applicare.

Alla maggior parte delle persone è sicuramente capitato più di una volta nella vita di compilare un test. L'uso dei test, infatti, è largamente diffuso in tutti i settori e se ne possono trovare di tipologie e finalità più svariate in riviste di ogni genere, nelle prove d'esame presso scuole e università, nei colloqui di assunzione e nelle più differenziate prove di selezione concorsuali. Il primato di diffusione e varietà è detenuto dai test psicometrici, che sono facilmente reperibili e che misurano una grande varietà di costrutti, identificabili come variabili psicologiche che non sono direttamente osservabili (es. intelligenza, personalità, motivazione, ecc). Ma come si fa a sapere se un test è in realtà un buon test? Come si fa a sapere se un test misura un costrutto con precisione? E ancora, misura effettivamente ciò che si intende misurare? E con

quale margine d'errore?

Questionario, Valutazione, Verifica, Misurazione... sono alcune delle parole chiave, che sempre più frequentemente vengono utilizzate a baluardo di garanzia di qualità nelle direttive ministeriali, nell'autonomia di scolastica e nella certificazione di abilità e competenze dalla scuola primaria all'università. Migliaia di studenti, delle scuole di ogni ordine e grado, quotidianamente sono sottoposti alla somministrazione di questionari e sovente i risultati ottenuti sono finalizzati all'ammissione a specifici percorsi formativi o attività di perfezionamento. La diffusione smoderata dell'uso dei test, tuttavia, non è supportata, in campo di istruzione e formazione, da una *cultura della valutazione* ma è frutto di una corsa all'adeguamento dei sistemi di istruzione al contesto europeo e globale, che spesso ha degenerato nel fenomeno che Bottani, R. [2002] definisce "Dittatura dei Test". Lo stesso Bottani [2011] ha mosso un'ampia critica all'impianto dei test costruiti dal Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) per selezionare i dirigenti scolastici, soprattutto in riferimento alle modalità di costruzione dei questionari, evidenziando la necessità di una collaborazione tra esperti in campo educativo e statistico. Possiamo pensare ai più diversi scenari in cui un test finisce per misurare un costrutto diverso da quello che si era prefissato inizialmente, soprattutto quando si indagano variabili che riguardano fenomeni complessi quali aspetti relativi all'apprendimento, all'intelligenza, all'autostima, ecc. Altro scenario possibile è quello in cui un test non sia in grado di cogliere nella sua interezza il costrutto che si intende analizzare, ma consenta la misurazione solo di alcune sotto-variabili. In campo formativo, poi, potrebbe verificarsi che scarsi risultati conseguiti in un questionario per la valutazione degli apprendimenti non corrispondano esclusivamente ad un basso livello di abilità possedute, ma potrebbero essere influenzati da una costruzione del test non conforme alle regole tipiche di strutturazione dello stesso o non adeguate al contesto di riferimento.

La *Teoria Classica dei Test*, in seguito indicata con il più diffuso acronimo inglese CTT (Classical Test Theory), nasce e si sviluppa proprio con il duplice interesse di misurazione di un costrutto e di validazione dello strumento utilizzato nella misurazione stessa. Negli ultimi anni, come vedremo, questa teoria è stata integrata e largamente

sostituita nel mondo scientifico da una metodologia di indagine più complessa ed organica, che è appunto quella della *Teoria della Risposta all'Item*, alla quale di seguito ci riferiremo indicandola con l'acronimo IRT (Item Response Theory).

Lo scopo di questo capitolo è di sottolineare la necessità di un approccio rigoroso nella costruzione di prove strutturate e di analizzare i limiti dei modelli più diffusi e più utilizzati in campo docimologico (Classical Test Theory) in favore di approcci basati sui modelli probabilistici (Item Response Theory). La IRT è stata sviluppata ampiamente in campo psicometrico e sociologico, ma nel settore della ricerca educativa le applicazioni in materia di istruzione non sono ancora state pienamente esplorate. Nella ricerca educativa, infatti, la IRT rimane ancora confinata alla sperimentazione didattica e la misurazione dei risultati scolastici solo su indagini condotte a livello nazionale. Lo scopo di capitolo è quello evidenziare l'importanza dell'applicazione di modelli IRT nella ricerca educativa e suggerisce anche alcune indicazioni per la ricerca futura.

La valutazione degli apprendimenti è una delle funzioni centrali che il docente, in ogni ordine e grado di istruzione, è chiamato a svolgere quotidianamente, ma che troppo spesso è condotta senza l'uso di un metodo scientifico e rigoroso. I Collegi dei Docenti, i gruppi di lavoro, i comitati di valutazione e i singoli docenti impiegano ore ed ore di lavoro per costruire test per la valutazione degli apprendimenti nella convinzione che il semplice utilizzo di uno strumento oggettivo sia garanzia di una corretta valutazione. Nei paragrafi che seguono utilizzeremo, come esempio dei metodi di misura tipici della CTT e della IRT, i dati ricavati dalla somministrazione di questionari somministrati da scuole diversi per la valutazione degli apprendimenti di Istituto. La scelta dell'uso di questi questionari è mirata ad evidenziare la necessità di una formazione specifica del corpo docente in campo valutativo e della ancor più cogente esigenza della maturazione di una effettiva e consapevole cultura della valutazione.

Quello che si vuole evidenziare è che i questionari per la valutazione degli apprendimenti costruiti quotidianamente dai docenti vengono progettati esclusivamente in relazione al proprio bagaglio esperienziale e/o sulla scia delle prove di valutazione degli apprendimenti somministrate a livello nazionale dall'INVALSI, senza però che sia utilizzato

alcun modello che consenta effettivamente di stimare le misure delle abilità degli studenti e i livelli di difficoltà e discriminatività degli item di un questionario. Una scuola, che intenda mettersi in linea con le esigenze richieste dalla società globale e che si ponga l'obiettivo di rendere i propri studenti cittadini competenti e competitivi, deve investire nella formazione del docente-ricercatore per evitare generalizzazioni non scientifiche e rigorose.

Possiamo identificare la teoria dei test, sia essa affrontata con un approccio secondo la CTT sia con le metodologie proprie della IRT, come una branca della matematica applicata e della statistica, che necessitano per un corretto utilizzo di consistenti competenze epistemologiche ed empiriche. I test per la valutazione degli apprendimenti sono strumenti di misurazione di cui deve essere stabilito il livello di attendibilità e di validità. La quasi totalità dei questionari che vengono giornalmente somministrati dai docenti, invece, sono costruiti senza alcun supporto esperto da parte di statistici, psicometrici e docimologi, di conseguenza non vengono valutati nei termini delle loro proprietà metriche. Non si ritiene possibile acquisire una conoscenza dei concetti fondamentali della CTT e della IRT e applicarli in modo rigoroso a situazioni concrete senza comprendere, almeno in certa misura, le basi matematiche su cui si fondano le teorie dei test nelle diverse articolazioni. I test per la valutazione degli apprendimenti devono essere usati in maniera appropriata, soprattutto in considerazione del fatto che i risultati dell'applicazione dei test influiscono sulla certificazione delle competenze di uno studente e, di conseguenza, possono influenzare gli sviluppi futuri nel proseguo degli studi o la selezione rispetto all'assegnazione di uno specifico incarico. Consideriamo di seguito alcuni esempi per chiarire il tipo di domande a cui la teoria dei test cerca di dare una risposta, secondo quanto evidenziato da McDonald [1999]; quando un insegnante struttura un questionario per verificare i livelli di apprendimento dei propri studenti, dovrebbe porsi le seguenti domande:

- (a) Un punteggio complessivo è sufficiente per valutare la prestazione degli studenti oppure sarebbe più opportuno scomporre l'abilità in sotto-categorie e calcolare punteggi differenziati per ognuna di esse?

- (b) Gli item che compongono il questionario sono tutti efficaci per la valutazione della preparazione degli studenti, oppure alcune domande presentano criticità?
- (c) Quanto è attendibile il risultato ottenuto da una somministrazione quale indice della preparazione dello studente?
- (d) Il numero di item utilizzati nel questionario è sufficiente per valutare la preparazione degli studenti, oppure sarebbe opportuno utilizzare un numero di item differente?
- (e) Se fosse stato utilizzato un questionario differente, si sarebbe ottenuta la stessa valutazione della prestazione degli studenti?
- (f) La prestazione degli studenti con abilità molto basse e con abilità molto elevate viene misurata con la stessa accuratezza della prestazione degli studenti che si collocano più in prossimità della media del gruppo?
- (g) Il questionario utilizzato misura effettivamente la conoscenza che si intende stimare, oppure sono presenti distorsioni sistematiche dovute all'influenza di conoscenze relative ad ambiti estranei al campo di interesse?

Tutte queste domande non possono trovare una risposta appropriata nelle "a-metodologie" usate oggi nella scuola italiana per la valutazione degli apprendimenti.

Interessanti ricerche condotte da Marzano, R.J [2006] mettono in evidenza la centralità del docente nella qualità scolastica. Lo stesso autore evidenzia che la valutazione rappresenta una "delle armi più potenti nell'arsenale di un insegnante" per promuovere e favorire il successo scolastico negli studenti. I risultati emersi in un'indagine condotta in differenti contesti scolastici evidenziano che l'apprendimento è certamente influenzato da diversi fattori, ma che la differenza è data soprattutto dalla qualità dell'intervento del docente, soprattutto del docente-ricercatore che possiede competenze specifiche in campo valutativo. Quanto detto è esplicitato nell'immagine 5.1, che evidenzia l'incremento delle abilità degli studenti a seguito di esperienze educative sperimentali condotte incentivando forma di *valutazione partecipata*.

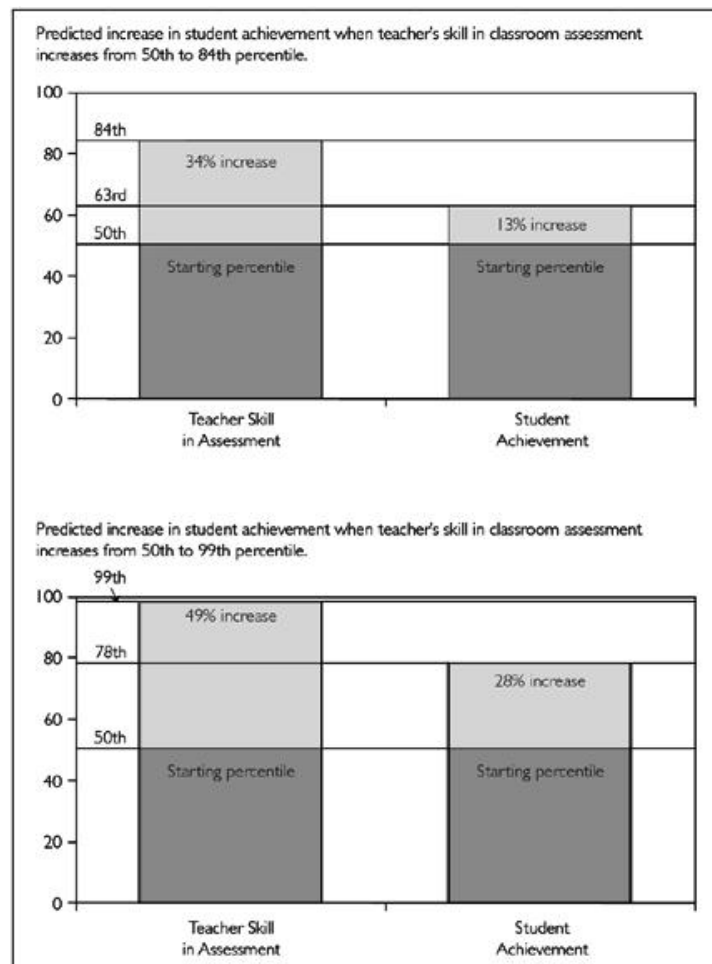


Figura 5.1: Classroom Assessment and Grading That Work [Marzano, R.J, 2006, pag.4]

5.1 Classical Test Theory (CTT): principi e assunti di base

La Teoria Classica dei Test, comunemente abbreviata nel mondo scientifico come CTT quale acronimo dell'espressione inglese Classical Test Theory, nasce all'inizio del XX secolo con la formalizzazione dell'equazione fondamentale del modello da parte di Spearman [1904], che ipotizza una relazione lineare e additiva tra il punteggio osservato di un test, la misura della variabile latente "abilità" e la componente casuale dell'errore. Uno dei contributi più importanti per l'implementazione della teoria della CTT si trova nell'opera di Guilford [1950], "Fundamental Statistics in Psychology and Education". Nel corso del Novecento la CTT viene sperimentata ed approfondita per essere definitivamente formalizzata alla fine 1960 ad opera di Lord & Novick in "The axioms and principal results of classical test theory" [Lord, 2005]. La CTT parte dall'assunto che le differenze esistenti tra punteggi diversi ottenuti da individui esaminati per mezzo di uno stesso questionario, dipendono solo dalle diverse abilità possedute dai soggetti e tutte le possibili interferenze, siano esse dovute a fattori esogeni e/o endogeni, sono assunte come costanti [van der Linden and Veldkamp, 2004].

Alla base della teoria c'è l'idea che il punteggio totale osservato tramite la somministrazione di un test è costituito in parte dal *punteggio vero*, in parte da una *componente erratica*. Quanto affermato si può sintetizzare nell'equazione fondamentale della teoria, che consente di definire formalmente il concetto di errore di misurazione come differenza tra il punteggio osservato e il suo corrispondente punteggio vero; in pratica essa formalizza l'impossibilità da parte di un particolare punteggio osservato di eguagliare il punteggio vero a causa di disturbi casuali. L'equazione è definita come segue:

$$X = T + E \quad (5.1.0.1)$$

dove

X è il punteggio osservato;

T è il punteggio vero della variabile osservata;

E rappresenta l'errore non sistematico nella misurazione.

La CTT presuppone, quindi, che ciascun punteggio osservato (X) incorpori al suo interno una componente vera (T) e una componente di errore non sistematico (E). Il punteggio osservato (X) sarà, dunque, un punteggio teorico, ossia atteso, che ciascun fenomeno reale possiede rispetto alle sue specificità e che, a causa degli errori di misurazione, non può essere osservato direttamente, ma può essere stimato attraverso la misurazione reale. Per questo esso è concepito come una quantità ipotetica, non osservabile, che non può essere direttamente misurata. Naturalmente se la procedura di misurazione fosse perfettamente affidabile e valida, il punteggio X risulterebbe esente da errore e, di conseguenza, coinciderebbe con il punteggio T .

Quando si misura un costrutto in campo formativo, educazione e psicologico si possono verificare errori non sistematici, ossia errori casuali non dovuti direttamente all'errore sistematico di misurazione. L'errore di misurazione casuale e non sistematico, non è una proprietà della caratteristica misurata ma è conseguente alla misurazione effettuata sull'oggetto. Esso, quindi, è correlato in modo inversamente proporzionale all'affidabilità, poiché maggiore è la componente di errore, minore sarà l'affidabilità della misurazione. Gli errori non sistematici potrebbero essere dovuti alle più svariate cause, per esempio distrazioni dell'individuo al di fuori della situazione della prova, la condizione di benessere fisico del candidato, elementi di disturbo ambientale, ecc. Molte sono, infatti, le componenti che possono influenzare un candidato al momento di svolgere una prova. Talvolta queste influenze possono avere un effetto positivo sul risultato del test, altre volte un influsso negativo, ma in ogni caso provocano un certo errore intorno al punteggio vero. Gli errori casuali sono supposti distribuiti normalmente e il valore atteso dell'errore, dato dalla media della distribuzione degli errori su un numero infinito di prove, assumerà quindi valore 0. La CTT, inoltre, assume che gli errori casuali non siano correlati tra loro, in quanto non sottendono a nessun modello sistematico, né al punteggio vero (T), in quanto anche in questo caso non vi è relazione sistematica tra il punteggio vero e il fatto che l'individuo ottenga risultati positivi o negativi.

La CTT presuppone che le misurazioni siano uniformemente disperse attorno alla media, di conseguenza, il punteggio vero è in realtà un punteggio medio. Se effettuassimo su uno stesso fenomeno

ripetute misurazioni di una stessa caratteristica, utilizzando sempre la stessa procedura si può ipotizzare, quindi, di ottenere, a causa della presenza dell'errore, risultati diversi che più o meno si avvicinano a quello vero. La teoria classica della misurazione assume che la distribuzione di tutti i valori così registrati sia normale e che il valore vero sia quello che presenta la frequenza più alta o, meglio, che tale valore sia quello con la probabilità più alta di avvicinarsi a quello vero. In altre parole le misure rilevate si distribuiscono con maggiore frequenza intorno al valore del punteggio vero e simmetricamente al di sopra e al di sotto dello stesso, assumendo la tipica forma della normale gaussiana. In pratica gli errori positivi compenseranno gli errori negativi e l'ampiezza dell'errore dipenderà dall'oscillazione dei punteggi osservati intorno al punteggio medio, considerato appunto stima del punteggio vero. Questo significa che si può assumere che all'aumentare del numero delle misurazioni, tali errori si annullino fino ad avere media zero. Se ne deduce, allora, che la distribuzione di frequenza, e quindi di probabilità, dell'errore di misurazione ha la stessa forma della distribuzione di frequenza di probabilità del punteggio osservato; conseguentemente la funzione di distribuzione dell'errore e la funzione di distribuzione del punteggio osservato saranno uguali. Questo avviene però solo dal punto di vista teorico, in quanto nella realtà non necessariamente gli errori in una direzione compensano gli errori nell'altra direzione, ovvero la distribuzione degli errori non è necessariamente simmetrica, soprattutto se l'oggetto di indagine è un fenomeno latente relativo ad aspetti della personalità di un individuo. In conseguenza di quanto detto, possiamo affermare che i test sono, quindi, strumenti di misura imprecisi. Se, infatti, il punteggio osservato non corrisponde alla reale abilità posseduta dallo studente esaminato, significa che potrebbe variare nel caso in cui uno stesso individuo fosse sottoposto allo stesso test in modo ripetuto. Teoricamente la deviazione standard della distribuzione degli errori casuali per ogni individuo, dà informazioni sulla grandezza dell'errore di misura. Solitamente si assume che la distribuzione degli errori casuali sia la stessa per ogni individuo. La CTT, infatti, usa la deviazione standard degli errori come base per la misura dell'errore e la definisce *standard error of measurement*. In pratica la deviazione standard dei punteggi osservati e la misura dell'attendibilità (*reliability*) sono utilizzate

per stimare l'errore standard della misurazione [Kaplan and Saccuzzo, 1997]. Più alto risulterà l'errore standard di misurazione più bassa sarà l'accuratezza con la quale il test misura una specifica abilità. Viceversa un valore molto piccolo dello standard error di misurazione ci dice che il punteggio osservato e il punteggio reale, relativi alla misura di una specifica abilità, sostanzialmente coincidono. Per la definizione stessa dell'errore casuale, se potessimo fare infinite misurazioni della variabile, la sommatoria degli errori casuali tenderebbe a zero e il punteggio vero giungerebbe a coincidere con la media. Solitamente nella valutazione della validità di un test la teoria classica utilizza l'analisi del punteggio totale ottenuto, la frequenza delle risposte corrette per indicare il livello di difficoltà, la frequenza delle risposte per esaminare i dimostratori e la correlazione totale tra i vari item per valutare la capacità di discriminazione [Plake and Impara, 1997]. Tuttavia, in questo modo, il sistema di valutazione è condizionato dallo specifico campione sul quale si somministra il questionario, ma queste critiche non sono rilevanti laddove il campione sia rappresentativo e non vari nel tempo [Hambleton, 2001], come ad esempio nel caso di una valutazione sistematica e periodica che accompagni un gruppo classe durante la frequenza ad un determinato ordine e grado di istruzione.

L'errore sistematico, invece, relativo ad esempio alle caratteristiche del test utilizzato o al contesto specifico in cui si svolge la prova, interesserà tutte le misure allo stesso modo. Questo significa che se è presente un errore di strutturazione, o di formulazione delle domande o dei distrattori, in uno o più degli elementi del questionario somministrato ad un gruppo di candidati, il peso dell'influenza dell'errore sul risultato sarà lo stesso su tutti i candidati. In realtà, quindi, l'equazione alla base della CTT necessita di un'ulteriore specificazione:

$$X = T + E + B \quad (5.1.0.2)$$

dove B indica l'errore sistematico che, influenzando i punteggi osservati, li rende meno affidabili.

Poiché i test di valutazione degli apprendimenti o di misurazione di componenti psico-emotive sono utilizzati principalmente per determinare le differenze individuali e le specificità dei singoli, l'influenza dell'errore sistematico non è considerata di grande importanza. Questo è il motivo per cui nell'equazione fondamentale della CTT l'errore

sistematico viene omesso. Tuttavia, è importante notare che se si effettua un confronto tra le prestazioni di candidati la cui prova è affetta da errori sistematici con le prestazioni di candidati che hanno completato una prova libera da tali errori, il confronto non sarà propriamente corretto. Questa, come vedremo, è una delle critiche mosse alla CTT dalla IRT, che affronta il problema dell'equità nei test. Ipoteticamente, se somministriamo una prova ripetutamente ad un candidato per un numero infinito di volte, il range dell'errore di misura sarà uguale al range del punteggio osservato. Il punteggio vero di una persona può essere trovato, secondo la CTT, considerando il punteggio medio che l'individuo otterrebbe sulla stessa prova se avesse compilato un questionario un numero infinito di volte. La media dei punteggi grezzi è considerata, infatti, la migliore stima del punteggio vero (T). Ovviamente nella realtà è impossibile effettuare infinite ripetizioni. Inoltre, nella maggior parte dei costrutti psicologici e dei processi di apprendimento, somministrare più volte uno stesso questionario sarebbe poco significativo in quanto l'effetto della memorizzazione degli item da parte degli individui assume un'influenza sistematica ed indesiderabile sulle prestazioni. Potrebbe accadere, infatti, che l'individuo possa ricordare la precedente sessione di un test e sia indotto a rispondere in modo simile, o potrebbero nelle somministrazioni successive capire come risolvere alcuni problemi incontrati in quelle precedenti e quindi eseguire meglio la prova, questo è particolarmente vero per le prove di abilità. Se si somministra un medesimo questionario ad un numero elevato di individui, abbiamo invece il problema che le specifiche abilità di ognuno possano avere pesi diversi sul risultato medio dell'intera popolazione.

Il modello classico di misurazione si fonda su alcuni postulati:

- a.** il valore atteso dell'errore di misurazione è uguale a zero, ovvero

$$\mathbb{E}(E) = 0$$

- b.** il valore atteso del punteggio osservato è uguale al valore atteso del punteggio vero; infatti

$$\mathbb{E}(X) = \mathbb{E}(T) + \mathbb{E}(E)$$

ma essendo $\mathbb{E}(E) = 0$ allora $\mathbb{E}(T) = \mathbb{E}(X) = T$;

- c. la correlazione e la covarianza tra punteggio vero ed errore sono nulle

$$r_{te} = 0 \quad \text{cov}_{te} = 0$$

- d. la varianza dei punteggi osservati è maggiore della varianza dei punteggi veri.

La CTT si preoccupa, come vedremo meglio nel dettaglio, dello studio dei questionari con il duplice scopo di testare l'affidabilità e la validità del questionario e di misurare tratti latenti di un individuo, che nel nostro caso specifico si riferiranno ai risultati scolastici del capitale umano in istruzione. Nei prossimi paragrafi vedremo appunto in che modo la CTT affronta questi tre aspetti.

5.1.1 Analisi delle risposte ad un questionario

Nell'analisi di un questionario, seguendo l'approccio della CTT, lo scopo primario è quello di indagare, in relazione alle risposte date dagli studenti, la strutturazione del test.

Il primo semplice passo consiste nell'analisi dell'*indice di difficoltà* di tutte le domande, ossia nello studio del comportamento degli item rispetto alla resistenza che lo stesso oppone alla sua risoluzione [Vertecchi, 2003]. Per calcolare l'indice di difficoltà è sufficiente calcolare il rapporto tra il numero di risposte errate ottenute dalla somministrazione del questionario e il numero complessivo delle risposte. Sarà quindi

$$Df_i = \frac{nE_i}{N} \quad (5.1.1.1)$$

dove, nE è il numero delle risposte errate date dall'individuo i – esimo e N è il numero totale delle risposte del questionario. L'indice può oscillare da un valore 0, se tutti gli studenti rispondono correttamente alla domanda, a un valore 1, se tutti rispondono in maniera errata. La CTT identifica come troppo difficili le domande che abbiano una percentuale di risposte sbagliate superiore al 70%.

Lo stesso indice può essere espresso in termini di facilità, ossia considerando il numero delle risposte corrette sul totale delle risposte

date

$$Fc_i = \frac{nC_i}{N} \quad (5.1.1.2)$$

dove, nC è il numero delle risposte corrette date dall'individuo i – esimo e N è il numero totale delle risposte del questionario. In questo caso l'indice varierà sempre da 0 a 1, ma assumerà il massimo del valore su tutti gli studenti hanno risposto in maniera corretta; secondo lo stesso principio diciamo, quindi che un item è troppo facile se ad esso ha risposto oltre il 70% del collettivo preso in esame. Talvolta l'indice di facilità è definito anche "*indice di attrazione*"

Da quanto detto si evince che esiste la relazione:

$$Fc_i = 1 - Df_i$$

poiché i due indici sono complementari.

Se ci interessa indagare la difficoltà/facilità dell'item, per rendere più agevoli e veloci i calcoli, si può procedere alla trasformazione delle risposte in variabile dicotomica, attribuendo valore 1 alla risposta corretta e valore 0 alla risposta errata. Tuttavia questo procedimento in caso di item a risposta multipla sottostima l'indice e non è attuabile se intendiamo verificare anche la facilità/difficoltà dei singoli distrattori, in quanto la dicotomizzazione comporta, nell'operazione di sintesi, la perdita delle informazioni rispetto alle singole risposte all'item.

In figura (5.2) possiamo osservare i risultati dell'analisi dell'indici di difficoltà dei dati del collettivo preso in esame. Se leggiamo la colonna delle risposte errate (0) ci esprimiamo in termini di livello di difficoltà, se leggiamo la colonna delle risposte corrette ci esprimiamo in termini di livello di facilità. Gli indici evidenziati con un valore > 0.70 sono item troppo facili, quelli che assumono valori intorno allo 0.60 sono gli item che possiedono un livello buono di facilità e quelli con valori < 0.50 sono item troppo difficili. Se il test, come nel nostro caso, è composto, per la maggior parte, da item facili, allora non sarà in grado di discriminare tra rispondenti con livelli di abilità diversi, in quanto quasi tutti i rispondenti saranno in grado di fornire una risposta corretta alla maggioranza degli item. Se il test è composto da una maggioranza di item difficili risulterà, invece, altamente selettivo poiché occorrerà avere un alto livello di abilità per saper rispondere. In generale è buona pratica costruire test composti da item che coprano

Proportions for each level of response:		
	0	1
X1t	0.1548	0.8452
X2t	0.0595	0.9405
X3t	0.0119	0.9881
X4t	0.2738	0.7262
X5t	0.0476	0.9524
X6t	0.3571	0.6429
X7t	0.0595	0.9405
X8t	0.1786	0.8214
X9t	0.0357	0.9643
X10t	0.1548	0.8452
X11t	0.2738	0.7262
X12t	0.4762	0.5238
X13t	0.8690	0.1310
X14t	0.4167	0.5833
X15t	0.3095	0.6905
X16t	0.3929	0.6071
X1i	0.4881	0.5119
X2i	0.2024	0.7976
X3i	0.2619	0.7381
X4i	0.5952	0.4048
X5i	0.2976	0.7024
X6i	0.2262	0.7738
X7i	0.2857	0.7143

Figura 5.2: Analisi livello difficoltà degli item secondo la CTT

tutti i livelli di difficoltà. Secondo la CTT, la scelta che viene usualmente fatta è quella di una dispersione moderata e simmetrica del livello di difficoltà attorno ad un valore leggermente superiore al valore che sta a metà tra il livello $1/n$, dove n è il numero delle alternative di risposta all'item, e il punteggio pieno. Nel nostro caso, poiché gli item hanno quattro alternative di risposta, il livello del caso per ogni item è pari a $1,00/4 = 0,25$. Il livello ottimale di difficoltà media sarà, quindi, $\frac{0,25+1,00}{2} = 0,62$. Nel caso di item con sole due alternative di risposta (es. Sì/No, Vero/Falso, ecc.) il livello ottimale di difficoltà media è $\frac{0,50+1,00}{2} = 0,75$. Più in generale, item con livelli di difficoltà superiore a 0,90 o inferiore a 0,20 dovrebbero essere utilizzati con cautela.

Gli indici di difficoltà relativi al questionario per la comprensione del testo in lingua italiana evidenziano che il test presenta un numero troppo elevato di item facili, quindi non è costruito in modo equilibrato. Dall'analisi di differenti questionari costruiti per la misurazione degli apprendimenti in diverse scuole è emerso che nella grande maggioranza dei casi i questionari sono risultati troppo semplici.

A questa metodologia utilizzata dalla CTT per lo studio del livello di difficoltà degli item rileviamo alcuni aspetti critici:

- un primo limite si riscontra nel fatto che, procedendo ad una trasformazione dicotomica delle risposte, si perdono informazioni circa tutte le alternative di risposta, di conseguenza non siamo in grado di stimare se l'imputazione del livello di difficoltà sia dovuta ad una non corretta strutturazione nella scelta delle possibili risposte all'item;
- un secondo limite sta nel fatto che le risposte mancanti vengono considerate come errori in quanto ad esse viene attribuito un punteggio pari a zero;
- le prime due osservazioni portano a riflettere su un limite più generale che è quello dovuto al fatto che, per definire il livello di difficoltà, si proceda sostanzialmente ad una attribuzione di punteggio alla prova pervenendo ad un indice non generalizzabile a contesti differenti da quello nel quale il questionario è stato somministrato. D'altra parte, riflettendo sui punteggi, potrebbe accadere che ottengano lo stesso punteggio soggetti che abbiano

risposto solo a risposte facili o che abbiano risposto anche a risposte difficili e, tuttavia, ai due individui la CTT attribuisce lo stesso livello di abilità.

L'attribuzione di punteggi si rende sicuramente necessaria nel momento in cui lo scopo dell'indagine sia quello di misurare il livello di un apprendimento e, soprattutto, quando si voglia indagare la capacità di una prova di discriminare i differenti livelli rispetto ad un'abilità specifica. La CTT propone quale *indice di discriminatività* il calcolo della *correlazione punto-biserial*.

Il coefficiente di correlazione punto-biserial (r_{pb}) è un coefficiente di correlazione utilizzato quando una variabile è dicotomica. La variabile può essere effettivamente dicotomica, come il genere, o può essere dicotomizzata artificialmente. Una delle critiche mosse alla CTT sta proprio nel processo di dicotomizzazione in quanto un processo artificiale di trasformazione non elimina la continuità di fondo quando ci troviamo in presenza di questionari a scelta multipla. In questo caso, infatti, sarebbe più opportuna una correlazione biserial.

La correlazione punto-biserial è matematicamente equivalente alla correlazione di Pearson che, in presenza di una variabile continua X e una variabile dicotomica Y , sarà $r_{xy} = r_{pb}$. Per dimostrare quanto detto si possono assegnare due distinti valori numerici alla variabile dicotomica. Per calcolare r_{pb} , si suppone che la variabile dicotomica Y abbia i due valori 0 e 1. Se dividiamo i dati in due gruppi, il gruppo 1 che contiene le y alle quali è stato attribuito il valore 1 e il gruppo 2 su che contiene le variabili alle quali è stato assegnato il valore 0, allora il coefficiente di correlazione punto-biserial è calcolato come segue

$$r_{pb} = \frac{M_1 - M_0}{\sigma_n} \sqrt{\frac{n_1 n_0}{n^2}} \quad (5.1.1.3)$$

dove σ_n rappresenta la deviazione standard dell'intera popolazione:

$$\sigma_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5.1.1.4)$$

M_1 è il valore medio della variabile continua X per tutti i dati del gruppo 1, rappresenta quindi la media dei punteggi dei soggetti che

hanno risposto in maniera corretta all'item; M_0 è il valore medio delle X per tutti i dati del gruppo 2; n_1 è la numerosità del gruppo 1; n_0 è la numerosità del gruppo 2 ed n è la dimensione totale del campione.

Questa formula stabilisce una correlazione tra due variabili, una dicotomica che consiste nell'aver risposto in maniera esatta o errata ad uno specifico item, l'altra continua che consiste nel punteggio complessivo ottenuto da coloro che hanno risposto esattamente all'item. Il valore può variare da -1 a $+1$. Un'alta correlazione, superiore a $0,25$, sta ad indicare che tendenzialmente coloro che rispondono esattamente quella domanda hanno dei punteggi elevati all'intero test. Una correlazione vicina allo zero, sta invece ad indicare che il l'item non discrimina efficacemente i soggetti più abili da quelli meno abili. Infine, una correlazione negativa, indica che chi risponde correttamente a quella domanda tendenzialmente risponderà peggio alle altre. L'interpretazione complessiva di questi indici può dare numerose informazioni. Ad esempio, un basso indice di discriminatività sta ad indicare che la domanda non è utile all'interno del test e che potremmo eliminarla o sostituirla con un'altra.

5.1.2 Analisi dell'affidabilità di un test

Il problema dello studio dell'affidabilità, o attendibilità, di un questionario riguarda l'analisi della coerenza fra i punteggi ottenuti dalla somministrazione di una prova se la prova viene somministrata in momenti successivi. Mira a verificare, quindi, se i punteggi ottenuti siano indipendenti dalle variazioni accidentali, sia nelle caratteristiche dei soggetti testati che nelle modalità di somministrazione, e da fattori che possono costituire fonti di errore nella misurazione. Con questo termine si intende in primo luogo la coerenza tra i punteggi di un candidato, quando esegue uno stesso test più volte. Il punteggio che ottiene un individuo, tuttavia, può variare molto se lo stesso questionario per la misurazione degli apprendimenti viene compilato più volte in un breve lasso di tempo, in quanto il soggetto apprende ad ogni somministrazione (questo concetto peraltro è alla base dell'idea della somministrazione dei test in modo adattivo finalizzato appunto all'apprendimento). I diversi punteggi dei test potrebbero, quindi, essere influenzati da eventi casuali, piuttosto che dal reale livello di abilità

posseduta dal soggetto. Un test è tanto più attendibile quanto più fornisce misure accurate, cioè immuni da errori casuali. L'affidabilità può, allora, essere pensata come il grado di precisione con cui si misura un'abilità, dove per precisione si intende la stabilità del punteggio osservato di un individuo in ripetute e identiche somministrazioni dello stesso test. Di conseguenza diremo che quante più variazioni nei punteggi si ottengono a causa di errori casuali fra diverse somministrazioni di uno stesso test allo stesso gruppo di soggetti, tanto meno il test è da ritenere attendibile. Se la somministrazione del questionario viene fatta ad un gruppo di soggetti otteniamo una distribuzione di valori e quindi una variazione dei punteggi. Questa varianza sarà dovuta a componenti vere e a componenti d'errore, cioè dipenderà dal diverso grado di possesso di una certa caratteristica da parte dei soggetti e da fattori occasionali estranei al test e non controllabili. Possiamo affermare che la varianza totale della distribuzione dei punteggi equivale alla somma della varianza delle componenti vere e di quella delle componenti d'errore, ossia

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (5.1.2.1)$$

dove:

- σ_X^2 corrisponde alla varianza dei punteggi osservati;
- σ_T^2 corrisponde alla varianza del punteggio vero;
- σ_E^2 corrisponde alla varianza delle componenti d'errore.

L'attendibilità di un test può essere definita come la proporzione di varianza vera rispetto alla varianza totale di una distribuzione di punteggi

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2} \quad (5.1.2.2)$$

dove:

- ρ_{XT}^2 è il coefficiente di attendibilità del test;
- σ_T^2 è la varianza delle componenti vere;
- σ_X^2 è la varianza totale dei punteggi.

Possiamo dare, quindi, un'interpretazione dell'affidabilità come misura dell'ammontare della variazione di punteggi osservati attribuibile

alla variazione dei punteggi veri. Scrivendo la 5.1.2.2 come

$$\rho_{XT}^2 = 1 - \frac{\sigma_E^2}{\sigma_X^2} \quad (5.1.2.3)$$

si ricava che

$$\sigma_E = \sqrt{\sigma_X^2(1 - \rho_{XT}^2)} \quad (5.1.2.4)$$

Da quest'ultima espressione si comprende la relazione inversa che lega l'affidabilità all'errore: in casi estremi se $\rho_{XT}^2 = 1$ tutta la variazione dei valori osservati è attribuibile ai punteggi veri, mentre se $\rho_{XT}^2 = 0$ tutta la variazione degli punteggi osservati è attribuibile all'errore.

Nella CTT, quindi, l'affidabilità è definita come ρ_{XT}^2 , che è il quadrato del coefficiente di correlazione tra i valori osservati e i valori veri; ρ_{XT}^2 è chiamato coefficiente di affidabilità della misurazione.

Poiché la varianza dei punteggi osservati può essere considerata in termini di somma della varianza del punteggio vero e della varianza dell'errore, possiamo esprimere la 5.1.2.2 come segue:

$$\rho_{XX'}^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} \quad (5.1.2.5)$$

da cui si deduce che il quadrato della correlazione tra i punteggi osservati e i punteggi veri è uguale alla correlazione tra gli i punteggi osservati di due misurazioni parallele. Possiamo ancora dedurre che l'affidabilità dei punteggi diventa maggiore se la percentuale di varianza dell'errore è piccola e viceversa. La radice quadrata dell'indice di affidabilità rappresenterà la correlazione tra i punteggi veri e i punteggi osservati.

Il coefficiente ρ_{XT}^2 serve a definire il concetto di affidabilità (reliability) mentre il coefficiente $\rho_{XX'}^2$ serve a valutarla empiricamente. L'indice di affidabilità tuttavia fornisce indicazioni circa il livello di attendibilità dell'intero questionario, tuttavia, non dà alcuna informazione in merito all'affidabilità dei singoli item; questo limite viene superato, come vedremo, dall'IRT.

La CTT studia l'attendibilità di un questionario in modo non univoco. Come abbiamo visto, infatti, la considera come la proporzione della varianza vera sulla varianza totale dei punteggi, come la correlazione del test con se stesso attraverso successive applicazioni, o come

correlazione tra forme parallele di uno stesso test. Il coefficiente di attendibilità è, di conseguenza, stimato in vari modi:

- come correlazione tra i punteggi conseguiti da un gruppo di soggetti in due metà di uno stesso test (metodo dello split-half);
- come correlazione tra le due distribuzioni di punteggi ottenute applicando due volte uno stesso test ad uno stesso gruppo di soggetti (metodo test-retest);
- come correlazione tra le due serie di punteggi ottenute somministrando allo stesso gruppo di soggetti due forme parallele dello stesso test (parallel form);
- come studio della coerenza o omogeneità fra gli item (alpha di Cronbach).

Lo split-half

Il questionario oggetto di studio viene suddiviso in due parti che vengono trattate come se fossero due test distinti e paralleli. Si definiscono così due distinte distribuzioni di valori: la distribuzione A sarà costituita dai punteggi ottenuti negli item della prima parte; la distribuzione B sarà costituita dai punteggi ottenuti nella seconda parte.

La suddivisione delle variabili in due gruppi deve essere effettuata in modo che esse risultino omogenee tra loro in modo che siano correlabili tra loro. Spesso i questionari per la valutazione degli apprendimenti però sono costruiti in modo tale che i diversi item abbiano un livello di difficoltà crescente di conseguenza una suddivisione a metà in relazione all'ordinamento potrebbe determinare che nella parte più semplice si ottengano risultati migliori rispetto alla parte più complessa. Per ovviare a questa problematica solitamente il ricercatore rileva i punteggi conseguiti negli item dispari, identificando così la distribuzione A, e quelli pari, che costituiranno la distribuzione B. Questo metodo è conosciuto nella letteratura docimologia con "odd-even". Con una simile divisione, anche nel caso in cui le prove siano disposte secondo una difficoltà crescente, potrebbero risultare due metà abbastanza equivalenti. Lo scopo è quello di ottenere due sub-test

aventi medie e varianza abbastanza simili, come se si trattasse di due test paralleli. Ottenute le due distribuzioni di punteggi si procede allo studio della loro correlazione. Occorre considerare, però, che il coefficiente di correlazione risultante esprime il grado di attendibilità solo di metà test. Per ovviare al problema e stimare il livello di attendibilità dell'intero test, è necessario correggere il coefficiente di correlazione per mezzo del metodo di Spearman-Brown:

$$\rho_{XX'}^* = \frac{n\rho_{xx'}}{1 + (n-1)\rho_{xx'}} \quad (5.1.2.6)$$

dove,

$\rho_{XX'}^*$ indica la stima del coefficiente di attendibilità dell'intero test;

$\rho_{xx'}$ è l'indice di affidabilità del test preso in esame;

n indica il numero delle volte in cui il test viene "allungato".

Il metodo di Spearman-Brown trova ampio uso in campo psicometrico sia per stimare il livello di affidabilità di un test sia per effettuare previsioni di affidabilità al modificare della lunghezza del questionario. Nel caso specifico dello split-half n è pari a 2, in quanto è come se il test fosse raddoppiato nel numero di item, poiché se poniamo $n = 2$, raddoppiamo la lunghezza del questionario, aggiungendo elementi con le stesse proprietà di quelle del test sottoposto ad analisi.

La formula, quindi, stima l'affidabilità di un nuovo test replicando la prova in esame n volte. Per valori di n inferiori ad 1 si stima l'effetto sull'affidabilità nel caso in cui si riducano le domande del test. La formula può anche essere utilizzata per predire il numero di repliche necessarie per ottenere questionario affidabile:

$$n = \frac{\rho_{xx'}^*(1 - \rho_{xx'})}{\rho_{xx'}(1 - \rho_{xx'}^*)} \quad (5.1.2.7)$$

Il modello IRT, come vedremo, fornisce un mezzo molto più preciso di prevedere cambiamenti nella qualità della misurazione con l'aggiunta o la rimozione di singoli item ad un questionario.

Il metodo Test-Retest

Il metodo del Test-Retest, largamente utilizzato per ricerche in campo psicometrico e sociologico, consiste nello studio della correlazione

fra due distribuzioni di misura ottenute somministrando due volte lo stesso test allo stesso gruppo di soggetti dopo un certo intervallo di tempo. Il coefficiente di correlazione calcolato viene identificato come coefficiente di attendibilità del test.

Il coefficiente di attendibilità ottenuto esprime, quindi, il grado di stabilità nel tempo del test e di generalizzabilità dei risultati in caso di somministrazioni diverse. Quanto più alto risulterà il coefficiente di attendibilità, tanto minore sarà l'influenza delle variazioni accidentali sui punteggi, a parità di caratteristiche dei soggetti e dell'ambiente in cui il test viene somministrato.

Il Parallel form

Quando ci riferiamo a *forme parallele* intendiamo versioni equivalenti, ossia il più possibile omogenee e sovrapponibili, di uno stesso test. Lo stesso gruppo di soggetti viene sottoposto, quindi, in tempi diversi alle due differenti forme dello stesso test.

Il coefficiente di correlazione fra le due serie di punteggi rappresenterà il coefficiente di attendibilità del test.

Poiché, come abbiamo visto l'affidabilità non può essere stimata direttamente, in quanto ciò richiederebbe di conoscere i punteggi veri, presupposto che, secondo la teoria classica dei test, è impossibile, le stime di affidabilità possono essere ottenute in vari modi costruendo un test parallelo. La proprietà fondamentale di un test parallelo è quella di riprodurre lo stesso punteggio vero e la stessa varianza del punteggio osservato per ogni individuo, esattamente come la prova iniziale. Se consideriamo due prove parallele X e X' , allora questo significa che

$$E(X_i) = E(X'_i) \quad (5.1.2.8)$$

e che

$$\sigma_{E_i}^2 = \sigma_{E'_i}^2 \quad (5.1.2.9)$$

Sotto queste assunzioni, possiamo affermare che la correlazione tra i punteggi dei test paralleli corrisponde all'indice di affidabilità [Lord, F. M. & Novick, M. R., 1968, cap.2]

$$\rho_{XX'} = \frac{\sigma_{XX'}}{\sigma_X \sigma_{X'}} = \frac{\sigma_T^2}{\sigma_X^2} = \rho_{XX}^2 \quad (5.1.2.10)$$

che corrisponde appunto alla 5.1.2.5.

Specifichiamo che, se le due versioni del test sono state somministrate a distanza di tempo, tale coefficiente di correlazione diventa un indice sia della stabilità nel tempo, sia della coerenza delle risposte date a diversi campioni di prove che presentano contenuti simili. Se, invece, le due forme vengono somministrate consecutivamente, il coefficiente di correlazione trovato esprimerà il grado di attendibilità fra le due versioni, ma non tra i due tempi dell'applicazione. La letteratura, sia in campo pedagogico che in quello psicologico, sconsiglia tuttavia la somministrazione delle due forme in breve successione o in tempi molto ravvicinati, per evitare problemi di affaticamento e di demotivazione degli individui sottoposti a test. Tuttavia l'utilizzo di test paralleli per la stima dell'affidabilità è raramente presa in considerazione dai ricercatori perché test realmente paralleli sono molto difficili da costruire. Molto più frequentemente viene utilizzata, per stimare l'affidabilità, una misura di coerenza interna conosciuta come α di Cronbach, che sarà trattata nel paragrafo che segue.

L'alpha di Cronbach

L'attendibilità di un test può essere studiata in funzione della coerenza od omogeneità degli item all'interno di un test (inter-item consistency o internal consistency reliability), ossia per mezzo dell'analisi delle inter-correlazioni fra gli item per la valutazione della coerenza interna di un questionario. A tal proposito Cronbach [1951] ha predisposto una formula generale per valutare il grado di affidabilità di un test a partire dai singoli item, espresso da un indice che egli ha chiamato "coefficiente alpha". L'indice originariamente è stato ricavato da Kuder and Richardson [1937] per item dicotomici e poi generalizzato da Cronbach [1951] per item a risposte ordinali di qualunque tipo.

Il coefficiente α è definito come:

$$\alpha = \frac{K}{K-1} \left(1 - \frac{\sum_{i=1}^K \sigma_{Y_i}^2}{\sigma_X^2} \right) \quad (5.1.2.11)$$

dove,

K è il numero degli item ("K – items");

σ_X^2 è la varianza del punteggio totale osservato;
 $\sigma_{Y_i}^2$ indica la somma delle varianze di ogni item del test;
 σ_X^2 è la varianza totale del test.

Il coefficiente α è legato concettualmente alla stima dell'affidabilità usata da Spearman-Brown, poiché entrambi gli indici si fondano sull'idea che l'attendibilità di un test possa essere espressa come il rapporto tra le varianze del punteggio vero e del punteggio totale (5.1.0.2):

$$\rho_{XT} = \frac{\sigma_T^2}{\sigma_X^2}$$

si confronta con l'equazione (5.1.2.2). Il coefficiente α varia in un range da 0 a 1, in quanto risultato del rapporto tra due varianze. Tuttavia, in relazione alle procedure di stima utilizzate, il valore del coefficiente α potrebbe assumere un qualsiasi valore inferiore o uguale a 1, comprendendo quindi anche valori negativi, anche se hanno senso solo i valori positivi [Ritter, 2010]. Un questionario è tanto più affidabile quanto più elevato è il valore assunto dal coefficiente α ; solitamente si ritiene buono un valore uguale o superiore a 0,70 su un campione significativo [Cortina, 1993]. Nel caso di test psicometrici, la letteratura considera accettabili questionari con un coefficiente α compreso tra 0,75 e 0,83 [Nunnally, J. C., 1978, pag.245].

Il valore dell' α di Cronbach generalmente aumenta all'aumentare delle inter-correlazioni tra gli item di prova, in quanto ciò denota una crescente presenza di coerenza interna nel questionario. Occorre tener presente che, poiché le inter-correlazioni tra gli item di una prova sono massimizzati quando ognuno di essi misura esattamente lo stesso costrutto latente, allora l' α di Cronbach è ampiamente utilizzata per studiare l'*unidimensionalità* del questionario rispetto al tratto latente indagato. Tuttavia, la inter-correlazione media tra gli item di un questionario è influenzata, come qualsiasi altra media, dai *valori outlier*, ossia quei dati che all'interno di una distribuzione presentano valori anomali. Così, mentre la inter-correlazione modale tra gli item di un test sarà uguale a zero quando l'insieme degli item misura costrutti latenti diversi e indipendenti, la inter-correlazione media potrebbe risulterà invece maggiore di zero [Schmitt, 1996, pag. 350]. Alcuni ricercatori, infatti, hanno dimostrato che α può assumere valori molto

elevati anche quando l'insieme degli item che compongono un questionario misura diversi costrutti latenti indipendenti [Mcdonald, 1999; Zinbarg, 2006].

Una regola comunemente accettata per misurare la coerenza interna con l' α di Cronbach è la seguente [George, D. e Mallery, P., 2003]: Il

α di Cronbach	Coerenza interna del test
$\alpha \geq 0,9$	<i>Eccellente</i>
$0,8 \leq \alpha < 0,9$	<i>Buono</i>
$0,7 \leq \alpha < 0,8$	<i>Accettabile</i>
$0,6 \leq \alpha < 0,7$	<i>Discutibile</i>
$0,5 \leq \alpha < 0,6$	<i>Povero</i>
$\alpha < 0,5$	<i>Inaccettabile</i>

Tabella 5.1: Misura della coerenza interna di un test (Cronbach's α)

coefficiente alpha può essere visto come un'estensione della formula di [Kuder and Richardson, 1937], che è una misura equivalente per risposte dicotomiche. Occorre tener presente che l'alpha di Cronbach non è robusta in presenza di dati mancanti.

5.1.3 Analisi della validità di un test

La validità rappresenta il grado o la precisione con cui un test misura esattamente la variabile per la quale è stato costruito. Un test potrebbe, tuttavia, risultare molto attendibile ma al tempo stesso non risultare necessariamente valido. La validità può anche essere intesa come l'assenza di errore, che si esplica nell'appropriatezza dello strumento di misurare il tratto che si propone di misurare. La validità può essere misurata in riferimento ad una norma o ad un criterio. La teoria classica dei test basa l'interpretazione del punteggio di un test sul confronto della prestazione di un rispondente rispetto alle prestazioni degli altri rispondenti appartenenti alla stessa popolazione di riferimento. Le misure riferite ad una norma vengono utilizzate per differenziare tra loro i corrispondenti. Se il test è costituito da item con una risposta giusta e una risposta sbagliata, si selezionano gli item che producono un valore di probabilità di risposta corretta

prossimo a 0,50 allo scopo di massimizzare la variabilità. Nel caso di test riferiti a criterio si pone il problema che se il valore di probabilità di risposta corretta è molto basso allora dobbiamo porci il problema di comprendere perché questo si verifica. Un valore basso potrebbe essere dovuto al fatto che l'insegnante non spiega bene, o al fatto che i testi non erano adeguati, o al fatto che gli studenti non avevano studiato abbastanza, che non avevano le conoscenze pregresse adeguate, ecc. In base alla CTT il punteggio ottenuto da un rispondente si può interpretare solo come un punteggio riferito ad una norma. Le misure riferite ad una norma indicano la posizione della rispondente in riferimento alla distribuzione di un campione di grandi dimensioni e rappresentativo della popolazione di riferimento. La CTT per stimare la validità, quindi, effettua una procedura di normalizzazione dei punteggi e procede all'analisi della correlazione tra i punteggi osservati e i punteggi normati. Il procedimento per la verifica della validità di un test consiste essenzialmente nel mettere in relazione le misure in un test con altre misure raccolte indipendentemente da questo, ma relative alla stessa variabile valutata dal test. In pratica, si procede calcolando il coefficiente di correlazione tra due serie di misure, la prima costituita dai punteggi ottenuti nel test (variabile X) e la seconda formata dai punteggi normalizzati (variabile Y). Il coefficiente così ottenuto si chiama coefficiente di validità e viene indicato dal simbolo r_{XY} . Solitamente la CTT utilizza il coefficiente r di Pearson, in relazione al quale il test sarà tanto più valido quanto più alto sarà il coefficiente di correlazione ottenuto. Occorre considerare che non esiste un livello generale considerato come norma per cui il coefficiente di validità possa essere considerato accettabile. Il valore di r deve, in ogni caso, essere sufficientemente alto da risultare significativo ad un livello di probabilità fissato a 0,01 oppure a 0,05.

5.1.4 La misurazione dei risultati dell'apprendimento e la correzione per guessing

La CTT, nella valutazione degli apprendimenti, attua una procedura di somma lineare dei punteggi grezzi ottenuti in seguito alla somministrazione di un questionario, quindi effettua un ordinamento degli studenti in base al punteggio conseguito.

Abbiamo già accennato nel paragrafo 4.4 alla questione della normalizzazione dei punteggi grezzi in riferimento al fatto che la popolazione scolastica si assume sia distribuita normalmente in relazione al fenomeno dell'apprendimento. La costruzione di uno strumento di misura per la verifica di livelli di apprendimento assume che ogni item del test discrimini fra due soggetti lungo un continuum di difficoltà, ossia che gli item misurino lo stesso tratto, ma abbiano differenti gradi di difficoltà e che ogni item differenzi senza errori nella distribuzione lungo il continuum di difficoltà. La misurazione, tuttavia, è affetta da errori, dei quali è utile stimare l'entità. Occorre considerare, inoltre, che quando si valutano le prestazioni di un gruppo di soggetti estratto casualmente da una popolazione assunta come normalmente distribuita, la loro distribuzione raramente ha forma normale, cioè tale campione raramente è rappresentativo della distribuzione normale della popolazione. In questi casi, per poter operare matematicamente e statisticamente su questi dati, bisogna "normalizzarne" la distribuzione, ovvero trasformare i punteggi grezzi in punti standard normalizzati.

Uno dei vantaggi della trasformazione dei punteggi grezzi in *punti z* è di rendere possibile il confronto fra le prestazioni di un soggetto in due test diversi. Ovviamente l'interpretazione dei punti *z* dipende dall'andamento della curva della loro distribuzione. Quanto più una distribuzione di punteggi si avvicina alla normale tanto più facilmente possono essere interpretati i punti *z* corrispondenti. La standardizzazione ha lo scopo di rendere i dati direttamente confrontabili, tra soggetti e tra prove diverse, caratteristica che i dati grezzi non possiedono se vengono mantenuti nella forma originale. I punti *z* indicando la posizione dei dati in termini di distanza dalla media μ in relazione all'unità di misura della deviazione standard σ

$$z_j = \frac{x_j - \mu_x}{\sigma_x} \quad (5.1.4.1)$$

I punti *z* godono delle seguenti proprietà:

1. La somma algebrica dei punti *z* è nulla:

$$\sum_{j=1}^n z_j = 0$$

2. La media aritmetica dei punti z è nulla:

$$\mu_{z_j} = \frac{\sum_{j=1}^n z_j}{n} = 0$$

3. La sommatoria dei punti z al quadrato corrisponde al numero del collettivo preso in esame:

$$\sum_{j=1}^n z_j^2 = n$$

infatti

$$\sum_{j=1}^n z_j^2 = \frac{\sum_{j=1}^n (x_j - \mu_X)^2}{\sigma_X^2} = \frac{n \sum_{j=1}^n (x_j - \mu_X)^2}{\sum_{j=1}^n (x_j - \mu_X)^2} = n$$

4. La deviazione standard dei punti z è uguale ad uno:

$$\sigma_z = \sigma_z^2 = 1$$

infatti

$$\frac{\sum_{j=1}^n (z_j - \mu_z)^2}{n} = \frac{\sum_{j=1}^n z_j^2}{n} = \frac{n}{n} = 1$$

Se i dati originali hanno una distribuzione normale, si ha una corrispondenza biunivoca costante fra tutti i punti z e i ranghi percentili $R_{(p)}$, che sono anch'essi indipendenti da unità di misura. Per l'attribuzione di un voto scolastico, o in ogni caso di un qualsiasi punteggio, può essere utile ovviare al problema che i punti z possono essere anche negativi. Possiamo, in tal caso, effettuare una trasformazione lineare dei punti z , rendendoli interi positivi:

$$f(z) = a + bz_j \quad (5.1.4.2)$$

dove $f(z)$ è il nuovo punteggio standardizzato, a e b sono costanti assegnate e corrispondono rispettivamente alla media e alla deviazione standard della nuova scala e z_j è il punteggio standardizzato già calcolato. Una delle critiche che da sempre viene mossa ai test a scelta multipla è che, anche chi non ha la minima idea di quale possa

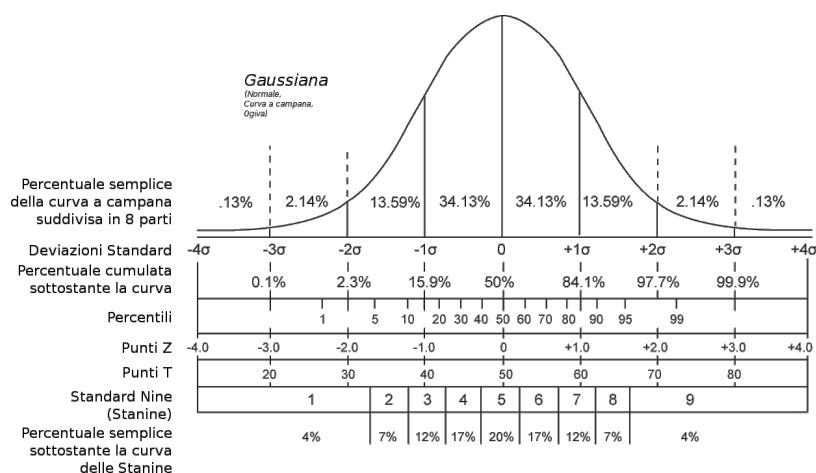


Figura 5.3: Distribuzione normale standardizzata

essere la risposta corretta ad un item, può comunque avere una certa probabilità di indovinarla scegliendo a caso tra le risposte, dato che è certamente una tra le alternative elencate. Frary [1988], in “Formula Scoring of Multiple-Choice Tests (Correction for Guessing)”, evidenzia come, le stesse persone, a parità di tutte le altre condizioni, tendono ad ottenere punteggi più alti nei test a scelta multipla rispetto a test con domande a risposta aperta. Per evitare questa distorsione, la CTT utilizza una procedura di scoring delle risposte agli item a scelta multipla che penalizzi uno stile di risposta casuale. Si tratta della *correzione per guessing*, che probabilmente è fin troppo pretenzioso, dato che sembrerebbe assumere che tutti i soggetti adottino la strategia del “tirare ad indovinare”, quando invece è perfettamente possibile che qualcuno non lo faccia, ricevendo però la correzione in ogni caso.

La procedura di correzione per guessing utilizzata dalla CTT è data da:

$$X_g = C - \frac{E}{(k - 1)} \quad (5.1.4.3)$$

dove X_g è il punteggio corretto per guessing, C è il numero di risposte corrette, E il numero di risposte errate e k il numero di alternative di risposta per ogni item. Se uno studente deve rispondere ad un que-

questionario composto da 10 item ognuno dei quali ha 5 possibili risposte, avrà la probabilità di indovinarne due risposte su dieci rispondendo a caso, la correzione per guessing fa sì che il punteggio corretto atteso più probabile in caso di stile di risposta completamente casuale sia uguale a zero.

$$X_g = 2 - \frac{8}{(5-1)} = 0$$

La formula, però, non tiene conto delle risposte omesse o non date. Questo significa che due soggetti con lo stesso numero di risposte corrette, ma un diverso numero di errate ed omesse, non otterranno lo stesso punteggio corretto.

Procediamo con un semplice esempio di quanto potrebbe accadere in una stazione reale. Consideriamo il caso di Tonio e Riccardo, due esaminandi ad una prova d'esame. Il questionario è composto da 40 item a 5 alternative di risposta, quindi ogni studente rispondendo a caso potrebbe indovinare 8 risposte su 40. Tonio ottiene 32 risposte corrette, 2 errate e 6 omesse, Riccardo ottiene 32 corrette, 6 errate e 2 omessa, quindi sommando semplicemente le risposte corrette entrambi gli studenti otterrebbero un punteggio grezzo di 32. I loro punteggi corretti saranno, invece:

$$X_{\text{Tonio}} = 32 - \frac{2}{(5-1)} = 31,50 \quad X_{\text{Riccardo}} = 32 - \frac{6}{(5-1)} = 30,50$$

In pratica, per effetto della correzione per guessing a parità di risposte esatte, Tonio ottiene un punteggio superiore rispetto a Riccardo nella graduatoria finale dei punteggi grezzi. L'idea che sta dietro a questo ragionamento è che Riccardo potrebbe aver ottenuto lo stesso numero di risposte corrette di Tonio solo perché ha rischiato di più, nel senso che ha tirato più ad indovinare, come potrebbe testimoniare il numero maggiore di risposte errate, da cui la possibile giustificazione per l'applicazione della correzione.

La correzione, sostanzialmente, corrisponde al risultato che si sarebbe ottenuto se si fossero penalizzate le risposte errate. Spesso, infatti, può capitare che nelle istruzioni del test sia indicato che ogni risposta corretta vale 1, ogni risposta omessa 0 e che sia prevista una certa penalizzazione per le risposte errate. Nella CTT, appunto, la penalizzazione dell'errore viene inserita nei questionari per scoraggiare

la risposta data a caso, e viene calcolata in relazione alla proporzione

$$-\frac{1}{(k-1)}$$

dove k è il numero di alternative di risposta. Questo fa sì che uno studente che applichi totalmente uno stile di risposta casuale, ottenga più probabilmente come punteggio finale 0. Avremo quindi che, se le alternative di risposta sono 4, la penalizzazione dell'errore sarà pari a $-0,33$, se sono 5 sarà $-0,25$, se sono 6 $-0,20$, e così via. Naturalmente, in base a quanto si vuole scoraggiare uno stile di risposta casuale, è possibile variare la penalizzazione anche arbitrariamente. Come sottolinea Frary [1988], poiché la formula per la correzione per guessing può essere applicata al punteggio totale solo se tutti gli item hanno lo stesso numero di alternative di risposta, molti docenti costruiscono questionari per tipologia di risposta, interpretato la questione come una sorta di obbligo a sviluppare test composti da item con lo stesso numero di alternative. In realtà, se si applica la semplice formula $-\frac{1}{(k-1)}$ alla risposta di ogni item e si sommano i punteggi corretti, si otterrà un punteggio totale corretto per guessing anche se gli item hanno un numero di alternative di risposta diverso, ossia in presenza di questionari che inglobano al loro interno risposte a scelta multipla, risposte dicotomiche, risposte di tipo cloze, ecc.

Per fare un semplice esempio concreto, supponiamo che un questionario sia composto da 5 domande Vero/Falso (gruppo A), 4 domande a tre alternative di risposta (gruppo B) e 4 domande di tipo cloze a quattro alternative di completamento (gruppo C). Tonio ottiene 3 risposte corrette al gruppo A di domande, 2 al gruppo B e 3 al gruppo C. Il suo punteggio totale sarebbe quindi pari a 8. Il suo punteggio totale corretto, invece, per ogni blocco di domande sarà:

$$X_{\text{Tonio,A}} = 3 - \frac{2}{(2-1)} \quad X_{\text{Tonio,B}} = 2 - \frac{2}{(3-1)}$$

$$X_{\text{Tonio,C}} = 3 - \frac{1}{(4-1)}$$

da cui un punteggio totale corretto è uguale a

$$X_{\text{Tot}} = X_{\text{Tonio,A}} + X_{\text{Tonio,B}} + X_{\text{Tonio,C}} = 2 + 1 + 2,66 = 5,66$$

La correzione per guessing ha come obiettivo principale quello di rimuovere la componente di risposte casuali dal punteggio totale, in base all'assunzione che uno stile di risposta completamente casuale corrisponda al livello più basso del costrutto misurato dal test. Un recente e interessante studio di Espinosa and Gardezabal [2010] ha confermato come all'aumentare della penalizzazione aumenti anche la proporzione di risposte omesse. I due autori, però, sono andati oltre e hanno simulato il comportamento di soggetti in grado di calcolare esattamente le probabilità che hanno di rispondere correttamente alla domanda. In questo caso, il valore ottimale della penalizzazione dovrebbe essere relativamente alto, ma non inciderebbe sulla correttezza della stima del livello del soggetto nel costrutto. Lo stesso risultato è stato ottenuto anche nel caso della simulazione di soggetti non in grado di valutare il loro grado di conoscenza della risposta corretta o che utilizzano una strategia del tipo scartare alcune risposte come troppo improbabili e scegliere a caso fra le rimanenti. Inoltre, i risultati ottenuti suggeriscono che sebbene la penalizzazione non favorisca i soggetti più prudenti, tale svantaggio è minimo rispetto all'errore di misurazione che si commetterebbe se la penalizzazione non venisse utilizzata. Nel momento in cui il punteggio al test viene interpretato in riferimento a norme, effettuando un processo di standardizzazione del punteggio grezzo, che il punteggio sia stato corretto per guessing oppure no cambia poco, dato che il punteggio assume senso solo nel confronto con quanto ottenuto dal campione normativo. Frary [1988] suggerisce quindi di limitare l'uso della correzione per guessing a quelle situazioni nelle quali il test abbia uno scopo altamente selettivo, come ad esempio un test con un elevato livello di difficoltà, un test in cui sia richiesto un punteggio di minimo di soglia, un test in cui il fattore tempo sia preponderante. Nei questionari costruiti per la valutazione dei livelli di apprendimento, invece, non è opportuno in quanto deve essere data la possibilità di attuare processi mentali di elaborazione delle abilità apprese che portano lo studente ad effettuare una riflessione attenta tra le scelte possibili ad una domanda.

5.1.5 Analisi di un test secondo la CTT: studio di caso

Un questionario per la valutazione dell'abilità di comprensione del testo è stato somministrato a tutti gli studenti delle classi quarte di Scuola Primaria di un Istituto Comprensivo di Chieti Scalo. All'interno dell'istituto sono stati istituiti comitati tecnici con lo scopo di elaborare strumenti e standard di misurazione e valutazione degli apprendimenti degli alunni. Il questionario è stato costruito dai tutti i docenti di lingua italiana delle classi quarte dell'istituto, organizzati in un gruppo di lavoro finalizzato a:

- riflettere sulle indicazioni nazionali sugli apprendimenti in classe quarta in lingua italiana al termine del primo quadrimestre e tenerne conto nella costruzione del questionario;
- prendere visione delle prove somministrate dall'INVALSI (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione) negli anni precedenti al 2012 e su esse effettuare una lettura critica rispetto alla formulazione ed ai contenuti;
- analizzare i quadri di riferimento del Sistema Nazionale di Valutazione (SNV) rispetto alla prova di comprensione del testo;
- costruire il questionario di rilevazione del livello di comprensione posseduto dagli studenti delle classi quarte dell'istituto rispetto alla comprensione del testo in lingua italiana;

riflettere sui risultati ottenuti dagli studenti nella compilazione dei questionari e predisporre i piani d'azione per superare gli aspetti critici e valorizzare i punti di forza.

I comitati tecnici hanno visto gli insegnanti impegnati in incontri sistematici con cadenza bi-settimanale. Come riferimento per la definizione degli obiettivi di misurazione e la costruzione delle prove sono stati analizzati i quadri di riferimento di Italiano per le classi quarte redatti dall'Invalsi. Mossi dal desiderio di trovare un modo di valutazione oggettiva, gli insegnanti hanno strutturato un percorso di auto-aggiornamento volontario mirato al miglioramento della efficienza educativa. Sono stati analizzati i risultati delle rilevazioni Invalsi dell'anno precedente, disaggregati per singola domanda e per

classe, analizzando un ricco set di grafici (fino a 50 grafici per scuola) messo a disposizione dall'Invalsi e corredato da un'analitica guida alla lettura per facilitarne l'interpretazione. Il lavoro dettagliato su tutte queste informazioni ha avuto lo scopo di ricavare con precisione i punti di forza e debolezza nella preparazione degli studenti come punto di partenza per programmare interventi educativi e didattici atti al miglioramento. Contestualmente a queste esperienze di auto-valutazione/auto-formazione, alcuni docenti referenti hanno frequentato un corso di formazione in rete inter-istituzionale "Dalle prove al curricolo, la valutazione esterna per analizzare, interpretare, progettare". Il corso ha fornito, in chiave laboratoriale, strumenti e metodologie utili al lavoro di valutazione, oltre che una lettura più analitica e dettagliata dei risultati delle prove Invalsi nella literacy linguistica degli alunni (analisi ed interpretazione dei risultati delle prove, i dati nazionali, il confronto con standard esterni, il confronto all'interno della scuola, il rapporto di scuola, spunti per un percorso di costruzione di prove, ecc.). Il gruppo ha lavorato esaminando gli aspetti della competenza di lettura da valutare per gli alunni della classe quarta. La riflessione si è concentrata su alcuni aspetti della competenza che le prove Invalsi hanno adottato per la prova di comprensione somministrata nel maggio 2011. Prese in esame le competenze considerate utili e gli item a cui esse facevano riferimento, riflettendo sul curricolo d'istituto e rileggendo attentamente le prove Invalsi degli anni precedenti, i docenti hanno ricavato gli aspetti della competenza di lettura che sembravano più idonei per costruire i quesiti da proporre agli alunni di classe quarta. Rispetto alle prove Invalsi sono stati scartati alcuni aspetti della competenza di lettura ritenuti troppo articolati per gli alunni di classe quarta, come la relazioni di coerenza e coesione testuale, l'integrazione di informazioni per inferenze complesse e la valutazione del contenuto e della forma di un testo riflettendo sulla validità delle informazioni alla luce di esperienze personali. Il gruppo di lavoro ha stabilito i punti di riferimento dell'Invalsi ritenuti più idonei da inserire nelle prove per studenti di classe quarta primaria; dei punti citati in nota sono stati selezionati i seguenti: 1, 2, 3, 5, 6, 7.¹

¹QdR INVALSI ITALIANO 28.02.2011: Aspetti sulla base dei quali viene verificata la competenza di lettura.

Per garantire che la prova copra diverse componenti della competenza di lettura, si

Dopo una attenta analisi di svariati testi, è stato scelto un testo narrativo nel quale gli alunni erano chiamati a distinguere parti fantastiche intrecciate con altre realistiche; il testo, non troppo lungo, permette all'alunno una discreta attività riflessiva e la messa in atto di strategie per la sua interpretazione che vanno al di là della semplice comprensione letterale. Sono stati strutturati 23 item, calibrati secondo le modalità già sperimentate e studiate dalle prove Invalsi.

La somministrazione dei questionari è avvenuta per tutti gli studenti nel medesimo giorno e nelle medesime condizioni ambientali. La cura della somministrazione, della raccolta dei dati, dell'analisi e della valutazione è avvenuta ad opera dei docenti di ambito scientifico delle classi quarte dell'istituto. Le indicazioni date agli studenti, i tempi per lo svolgimento della prova e le modalità di compilazione sono state le medesime utilizzate dal SNV.

La popolazione che ha compilato il questionario è composta da 84 studenti di classe quarta di scuola primaria. Agli studenti è stata somministrata una prova di comprensione del testo strutturata in 23 item a scelta multipla con quattro possibili alternative. Un primo dato emerso dall'analisi della facilità degli item è che la prova presenta

sono distinti sette aspetti, in relazione a cui sono formulati i quesiti. Alcuni di essi chiamano in causa la comprensione di una parte del testo, mentre altri riguardano il testo nel suo insieme, altri ancora l'una o l'altro a seconda dei compiti. L'elenco che segue riassume gli aspetti considerati.

–Aspetti della competenza di lettura per la classificazione dei quesiti:

- 1) Riconoscere e comprendere il significato letterale e figurato di parole ed espressioni; riconoscere le relazioni tra parole.
- 2) Individuare informazioni date esplicitamente nel testo.
- 3) Fare un'inferenza diretta, ricavando un'informazione implicita da una o più informazioni date nel testo e/o tratte dall'enciclopedia personale del lettore.
- 4)Cogliere le relazioni di coesione (organizzazione logica entro e oltre la frase) e coerenza testuale.
- 5a) Ricostruire il significato di una parte più o meno estesa del testo, integrando più informazioni e concetti, anche formulando inferenze complesse.
- 5b) Ricostruire il significato globale del testo, integrando più informazioni e concetti, anche formulando inferenze complesse.
- 6)Sviluppare un'interpretazione del testo, a partire dal suo contenuto e/o dalla sua forma, andando al di là di una comprensione letterale.
- 7) Valutare il contenuto e/o la forma del testo alla luce delle conoscenze ed esperienze personali (riflettendo sulla plausibilità delle informazioni, sulla validità delle argomentazioni, sulla efficacia comunicativa del testo, ecc.)

problemi di calibrazione in quanto gli indici di facilità sono troppo elevati (si veda tabella in figura 5.4). Dall'analisi della correlazione

Proportions for each level of response:			logit
	0	1	
X1t	0.1548	0.8452	1.6977
X2t	0.0595	0.9405	2.7600
X3t	0.0119	0.9881	4.4188
X4t	0.2738	0.7262	0.9754
X5t	0.0476	0.9524	2.9957
X6t	0.3571	0.6429	0.5878
X7t	0.0595	0.9405	2.7600
X8t	0.1786	0.8214	1.5261
X9t	0.0357	0.9643	3.2958
X10t	0.1548	0.8452	1.6977
X11t	0.2738	0.7262	0.9754
X12t	0.4762	0.5238	0.0953
X13t	0.8690	0.1310	-1.8926
X14t	0.4167	0.5833	0.3365
X15t	0.3095	0.6905	0.8023
X16t	0.3929	0.6071	0.4353
X1l	0.4881	0.5119	0.0476
X2l	0.2024	0.7976	1.3715
X3l	0.2619	0.7381	1.0361
X4l	0.5952	0.4048	-0.3857
X5l	0.2976	0.7024	0.8587
X6l	0.2262	0.7738	1.2299
X7l	0.2857	0.7143	0.9163

troppo facile

troppo difficile

Figura 5.4: Analisi facilità degli item di una prova

punto biseriale e dell' α di Cronbach si sono evidenziati item particolarmente problematici (figura 5.5). In particolar modo gli item evidenziati presentano una correlazione punto biseriale inferiore o uguale a 0,2 che segnala che l'item non discrimina correttamente soggetti con livelli di abilità differenti.

Per quanto riguarda l' α di Cronbach si deve confrontare il valore relativo alla prova (all item) con quello dei singoli item: il valore indicato ci dice il valore che assumerebbe l' α se escludessimo l'item j – esimo dalla prova. Gli stessi item evidenziati in giallo segnalano che la loro eliminazione migliorerebbe il valore α della prova. L' α di Cronbach è ritenuta accettabile per valori superiori a 0,70 tuttavia può essere accettata a valori inferiori quando il campione analizzato è di

	Point Biserial correlation with Total Score:	Cronbach's alpha all Items: 0,6733
X1t	0.3883	0.6587
X2t	0.3531	0.6624
X3t	0.1878	0.6716
X4t	0.1241	0.6891
X5t	0.4511	0.6577
X6t	0.4664	0.6513
X7t	0.3531	0.6624
X8t	0.3395	0.6637
X9t	0.4080	0.6613
X10t	0.3277	0.6644
X11t	0.4438	0.6536
X12t	0.2917	0.6748
X13t	0.0294	0.6878
X14t	0.5486	0.6400
X15t	0.4890	0.6481
X16t	0.1561	0.6898
X1l	0.3687	0.6650
X2l	0.5796	0.6379
X3l	0.5285	0.6431
X4l	0.4138	0.6587
X5l	0.3460	0.6657
X6l	0.3919	0.6591
X7l	0.2000	0.6817

Figura 5.5: Analisi facilità della correlazione punto biseriale e dell' α di Cronbach di una prova

basa numerosità (inferiore alle 500 unità statistiche). Il questionario presenta quindi una coerenza interna accettabile, ossia misura una sola abilità rispettando l'assunto di unidimensionalità.

Per valutare il potere discriminante dei singoli item abbiamo utilizzato la correlazione punto-biserial, mettendo a confronto i risultati ottenuti da tutti i soggetti che hanno risposto correttamente a un singolo item con i punteggi di tutti i soggetti nell'intera prova, meno l'item preso in analisi. Il punto biserial, che teoricamente varia da -1 a $+1$, quanto più si avvicina alla zero indica che l'item non possiede potere di discriminare i soggetti che sono andati effettivamente bene alla prova da quelli che sono andati male. Una discriminatività positiva, superiore a $0,2$ indica che l'item discrimina i soggetti più capaci da quelli meno capaci e tanto più si avvicina ad 1 tanto migliore sarà la capacità della domanda di discriminare, ossia di separare chi presenta un buon livello di abilità da chi non lo possiede. Una discriminatività negativa indica invece che fra chi risponde correttamente alla domanda sono più numerosi quelli che sono andati complessivamente peggio alla prova rispetto a chi ha ottenuto punteggi più elevati.

Osservando l'istogramma della distribuzione di punteggi grezzi ottenuti dagli studenti (5.6), si nota che la distribuzione si concentra verso punteggi elevati; questo dato fa ipotizzare o una spiccata prevalenza nel gruppo di soggetti con elevati livelli di abilità o una conferma dell'eccessiva facilità della prova.

È stata, in seguito, condotta un'analisi sui distrattori per verificare la strutturazione dei singoli item e la loro validità. Si riportano i risultati emersi per gli item risultati più problematici (figura 5.7).

L'analisi suddivide i soggetti in tre gruppi per livelli di abilità e valuta la percentuale di risposta corretta riferita ad ogni livello; in questa suddivisione è necessario tenere presente che il numero dei soggetti che rientrano nei due estremi deve essere il medesimo. Nel nostro caso, essendo la popolazione divisibile per tre non si presentano problemi e i tre gruppi avranno la stessa numerosità; nel caso in cui, ad esempio, il collettivo fosse stato di 83 individui il gruppo centrale sarebbe stato composto da 29 individui e i due gruppi estremi da 27, in modo da garantire che avessero la stessa numerosità. Questa suddivisione in tre fasce è definita per terzi, dove ogni fascia racchiude il 33% della popolazione a cui è stata somministrata la prova. Con popolazioni

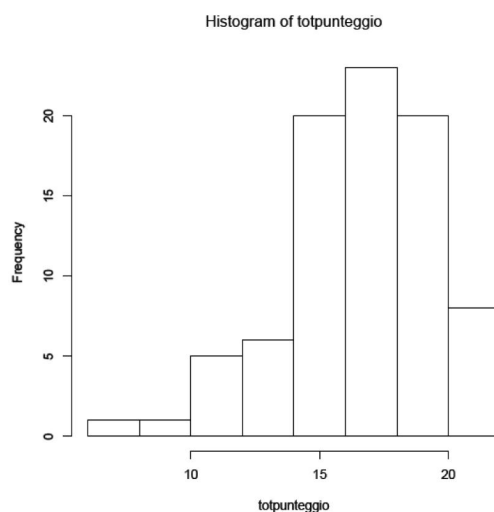


Figura 5.6: Distribuzione dei punteggi grezzi di una prova

molto numerose si può ridurre significativamente l'ampiezza degli estremi, considerando che, secondo l'assunto di distribuzione normale degli apprendimenti, agli estremi sono collocati pochissimi soggetti (le eccellenze e i soggetti con grandi difficoltà).

Si noti come all'item "X4t", ad esempio, ha risposto correttamente una percentuale di studenti con un basso livello di abilità rispetto a coloro collocati ad un livello medio; nella domanda "X13t" addirittura hanno risposto meglio gli studenti con basse abilità rispetto agli studenti con abilità elevate, sempre in questa domanda, inoltre, emerge che la maggior parte delle scelte di risposta è ricaduta sul distrattore "C", ciò significa che è opportuno rivedere la strutturazione dell'item in quanto il distrattore trae troppo in inganno rispetto alla domanda che dovrebbe essere assolutamente vera. Se si osserva il contenuto della domanda "X13t", riportato in figura 5.8, si nota come le risposte siano molto simili tra di loro e difficilmente si può cogliere nell'immediatezza quella assolutamente falsa che secondo le regole di costruzione di un item dovrebbe essere sempre presente.

\$'item_ X4t'	\$'item_ X13t'	\$'item_ X16t'	\$'item_ X71'
score.level	score.level	score.level	score.level
resp. lower middle upper	resp. lower middle upper	resp. lower middle upper	resp. lower middle upper
*A 0.703 0.619 0.846	A 0.378 0.238 0.269	A 0.054 0.000 0.000	A 0.135 0.190 0.077
B 0.027 0.143 0.000	*B 0.135 0.190 0.077	B 0.135 0.048 0.000	*B 0.676 0.667 0.808
C 0.162 0.095 0.154	C 0.351 0.524 0.615	C 0.243 0.381 0.308	C 0.135 0.048 0.077
D 0.108 0.143 0.000	D 0.135 0.048 0.038	*D 0.568 0.571 0.692	D 0.054 0.095 0.038

Figura 5.7: Analisi dei distrattori di una prova

13t. I tempi dei fatti narrati sono:

- A.** Recenti e precisati.
- B.** Passati e non precisati.
- C.** Recenti e non precisati.
- D.** Passati e precisati.

(QDR: Ricostruire il significato globale del testo, integrando più informazioni e concetti, anche formulando inferenze complesse.) Cod. 5b

Figura 5.8: Esempio di un item di una prova che presenta problematicità nei distrattori

Si è passati, quindi all'analisi dei dati applicando il modello logistico a due parametri della IRT (2PL). Il modello sarà spiegato nel paragrafo 5.2.6, pertanto al momento ci limiteremo a dare alcune informazioni che servono al solo scopo di evidenziare le problematiche del questionario.

La curva in figura 5.9 rappresenta graficamente la funzione di informazione dl test (Test Information Function) che contemporaneamente ci permette di dare una lettura del livello globale di difficoltà del questionario e della sua capacità di discriminazione e del livello di abilità degli studenti; si osservi come la prova sia atta a misurare livelli di abilità molto bassi e non sia in grado di discriminare livelli diversi di abilità se non per fasce di livello molto basso. La curva assume quasi una forma a "J" concentrata su valori negativi, che sta ad indicare che circa il 70-80% degli individui è collocato a bassi livelli di abilità, ossia gli studenti hanno ottenuto ottimi voti ma relativi a bassi livelli di abilità.

Di conseguenza, possiamo affermare che il questionario costruito dagli insegnanti non appare idoneo alla valutazione degli apprendimenti scolastici poiché non abbraccia tutti i possibili livelli degli studenti. L'analisi del questionario mette in evidenza come tanto lavoro

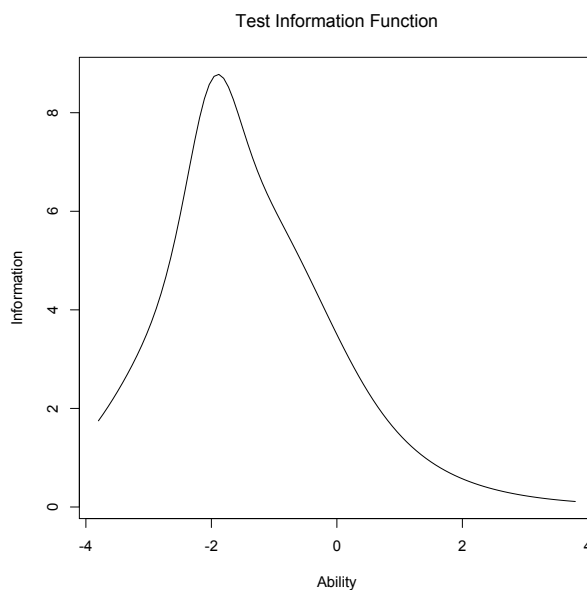


Figura 5.9: Test Information Function di una prova

da parte degli insegnanti e voglia di pervenire a metodi di valutazione più oggettivi è stato vanificato dalla mancanza di rigore scientifico nella costruzione della prova. Gli insegnanti impegnano grandissime risorse personali, in termini di tempo e costi, in percorsi di auto-formazione che purtroppo, come in questo caso, non sempre risultano utili ai fini preposti. L'indagine ha fatto emergere una pressante necessità di percorsi formativi di aggiornamento in servizio che possano consentire la formazione della figura dell'insegnante-ricercatore.

Nel corso degli ultimi anni abbiamo condotto lavori simili in altri due istituti comprensivi pervenendo alle medesime conclusioni. Sono stati sottoposti ad analisi comparate i dati di circa 1000 studenti

della scuola primaria e della scuola secondaria di secondo grado e i questionari sono sempre stati valutati troppo semplici, tanto che la stragrande maggioranza degli studenti consegue altissimi punteggi. In termini di costi/benefici, ci si chiede se sia opportuno impegnare tutte queste risorse umane e materiali per preparare questionari che non abbiano alcuna validità ed attendibilità di misurazione. Riteniamo che sia necessario un profondo cambiamento in termini di cultura della valutazione e che molti docenti debbano maturare la consapevolezza che i risultati ottenuti nei questionari dei propri studenti non possono essere “guidati” al fine di consentir loro di fare la figura del bravo insegnante con alunni preparati, almeno fino a quando i test somministrati non siano effettivamente in grado di misurare i livelli di preparazione degli studenti.

5.2 Item Response Theory (IRT)

I modelli Item Response Theory (IRT), modelli di misurazione a tratti latenti, sono stati definiti come “uno dei più importanti progressi metodologici di misurazione psicologica” [McKinley and Mills, 1989, p. 71]. Il confronto tra la Teoria Classica dei Test (CTT) e la IRT, come vedremo, è inevitabile in quanto la CTT è stato il modello di misurazione più utilizzato fino ad oggi in campo educativo. Seguendo l’esempio di altri Paesi Occidentali, quali Gran Bretagna, Stati Uniti, Australia, ecc. da qualche anno anche in Italia si sta diffondendo, almeno a riguardo al sistema di valutazione nazionale, una cultura della valutazione degli apprendimenti più consapevole e fondata su rigorose metodologie scientifiche. In Italia, il primo approccio ad una valutazione sistematica non è nato direttamente in merito alla misurazione delle abilità e delle competenze degli studenti, quanto piuttosto dei sistemi di istruzione e, più in generale, dei servizi pubblici, a fini non meramente conoscitivi o di controllo ex post, ma anche e soprattutto a fini decisionali [Gori, E. e Vittadini, G., 1999; Chiandotto, B. e Bacci, S., 2005].

Molte applicazioni su modelli a tratti latenti sono stati pubblicate nel corso dell’ultimo decennio, di notevole interesse in campo educativo [Adams *et al.*, 1997b; O’Connell and McCoach, 2008; Fox, 2004, 2005; Fox and Glas, 2001, 2003; Maier, 2001; Swartz and Bakari, 2005].

Sulla base delle ricerche condotte da Adams *et al.* [1997b] e Mislevy [1985], Kamata [2001] ha sviluppato un modello di stima dei parametri del modello IRT per i dati dicotomici, che è stato, inoltre, esteso per i dati politomici in Williams, N. J. e Beretvas, S. N. [2006].

Nella valutazione scolastica, il primo passo da compiere per isolare le variabili latenti di interesse è quello di individuare gli indicatori parziali che meglio rappresentano il tratto latente. Il metodo correntemente utilizzato consiste nel classificare un certo numero di item che siano in grado di coprire l'intero arco del tratto latente considerato (nella pratica della ricerca metodologica e didattica si dovrebbero considerare un centinaio di item per tratto latente). Nell'elaborazione della lista degli item e nella costruzione del questionario, è necessario rispettare alcuni criteri in merito al contenuto, alla lunghezza e alla formulazione che non sono qui oggetto di trattazione e per l'approfondimento dei quali si rimanda alla letteratura specifica.

Abbiamo avuto modo di vedere come la misurazione di un apprendimento possa essere effettuata soltanto in modo indiretto, attraverso le risposte fornite a domande su aspetti parziali che contribuiscono a definire l'abilità complessiva rispetto ad una disciplina. Si comprende, quindi, quanto sia importante l'individuazione di metodi di misura atti a tradurre l'informazione derivante da questo insieme di domande osservabili (item), indicatori parziali della variabile latente, in una misura sintetica e, per quanto possibile, oggettiva della medesima. A tal proposito, i modelli di Rasch, su cui fonda la IRT, costituiscono certamente un adeguato contesto di riferimento, date le proprietà di cui essi godono e che esplicheremo nella trattazione, quali l'unidimensionalità, la sufficienza dei punteggi grezzi degli item e degli individui, l'indipendenza locale degli item e la specifica oggettività. Queste proprietà consentono di trasferire alle scienze educazionali e formative il concetto di misura proprio delle scienze fisiche. Oggetto del presente capitolo è una verifica delle potenzialità dei modelli di IRT, quale metodo di riferimento per la valutazione delle performance degli studenti in riferimento al problema della valutazione scolastica.

Misurare una performance rispetto ad un determinato apprendimento significa collocare il livello di abilità dimostrato lungo un continuum numerico immaginario sul quale sia fissata un'unità di misura convenzionale, in modo tale che sia possibile esprimere un

giudizio quantitativo sulla posizione occupata dall'individuo sia in termini assoluti che relativi.

A tale proposito, i questionari per la valutazione degli apprendimenti rappresentano sicuramente un'utile strumento per la descrizione del profilo qualitativo del fenomeno studiato, ma l'abitudine consolidata e supportata scientificamente dalla CTT, come abbiamo visto, di effettuare la somma dei punteggi conseguiti nelle diverse domande al fine di attribuire un valore di giudizio, se può essere valida per effettuare un ordinamento in relazione al risultato ottenuto, dal punto di vista metrico non rappresenta una misura oggettiva. A titolo di esempio, se uno studente ottiene un punteggio più basso di un compagno nella stessa prova di valutazione, si può soltanto affermare che il primo si trova collocato in un'apposizione precedente su una scala ordinale crescente rispetto al secondo, ma non si è in grado di dare informazioni quantitative in relazione al minor livello di abilità posseduto. In altri termini, i valori numerici assegnati alle possibili risposte alle domande di un questionario (0 o 1, piuttosto che -1, 0 e 1 oppure 1, 2, 3 e 4, ecc.) hanno una natura arbitraria e, quindi, definiscono una scala qualitativa ordinale. Ciò che invece caratterizza il concetto di misura è la possibilità di collocare i conteggi osservati su una scala quantitativa, caratterizzata da intervalli definiti convenzionalmente. La prima difficoltà che si incontra, quindi, nella misurazione di un tratto latente è rappresentata dalla scelta della scala di misura, poiché è comprensibile l'impossibilità di utilizzare una scala metrica convenzionale. L'approccio di misurazione della IRT, come vedremo di seguito nel dettaglio, considera la probabilità di dare una risposta corretta o sbagliata ad un item in relazione al livello di abilità posseduto da uno studente. Ogni item che compone un test avrà, quindi, una sua specifica curva caratteristica (figura 5.10) che descrive la probabilità di rispondere correttamente o erroneamente allo specifico item in relazione all'abilità posseduta da chi viene esaminato.

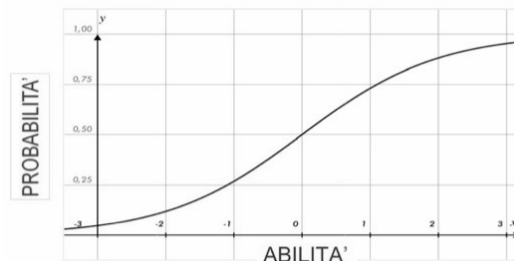


Figura 5.10: curva caratteristica di un item di difficoltà media e probabilità di soluzione dell'item al variare dell'abilità dei soggetti

Abilità e difficoltà vengono considerate congiuntamente e misurate con la medesima scala in considerazione della probabilità che un soggetto, avente una specifica abilità, ha di rispondere correttamente ad un item con un determinato livello di difficoltà:

$$P_j(\theta_i) = \frac{e^{(\theta_i - b_j)}}{1 + e^{(\theta_i - b_j)}} \quad (5.2.0.1)$$

dove

P_j è la probabilità di rispondere correttamente all'item j – esimo;

θ_i è l'abilità del soggetto i – esimo;

b_j è difficoltà dell'item j – esimo;

e è la costante di Nepero (2,71828...). ²

²Il numero "e" può essere definito in uno dei seguenti modi equivalenti:

- come il valore del limite

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

- come la somma della serie

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \dots$$

dove $n!$ indica il fattoriale di n . (Proprio per ottenere, per lo sviluppo in serie della funzione esponenziale, la scrittura compatta $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$, si pone per definizione $0! = 1$).

La caratteristica curva logistica di ogni item, in funzione dell'abilità, rappresenta la delimitazione tra l'area che rappresenta la probabilità di rispondere correttamente e quella di rispondere non correttamente. La curva in figura 5.10 è relativa al modello logistico ad un parametro, il quale assume come uguali le discriminazioni di tutti gli item [Maier, 2001]. La IRT ha l'obiettivo di misurare, quindi, il valore dei tratti latenti di un soggetto esaminato, applicando un modello di tipo probabilistico. Simbolicamente il tratto latente che si intende misurare, l'abilità, è indicato con la lettera θ .

La IRT utilizza approcci differenti alla misurazione dell'abilità in funzione dei parametri che prende in considerazione; di conseguenza vedremo come, all'interno di essa, sia opportuno analizzare

- il modello di Rasch, che ha rappresentato l'origine della IRT;
- il modello ad un parametro (1PL), che prende in considerazione solo il parametro di difficoltà dell'item³;
- il modello a due parametri (2PL), che prende in considerazione il parametro di difficoltà e di discriminatività dell'item;
- il modello a tre parametri (3PL), che rispetto al precedente introduce anche il parametro di *guessing*, ossia della risposta data a caso;
- il modello a quattro parametri (4PL), che rispetto al precedente considera il caso in cui un individuo con abilità molto elevate potrebbe non presentare sempre una probabilità di successo uguale a 1.

Questi modelli sono *nested* fra di loro; infatti si può passare da un modello ad un altro semplicemente vincolando un parametro ad essere uguale a 0 oppure liberandolo. In generale, l'IRT assume che il punteggio che un individuo ottiene ad un test possa essere spiegato in termini di variabili latenti e parametri. La variabile latente è l'abilità

³In molti testi si può trovare indistintamente trattato il modello di Rasch e il modello ad un parametro in quanto diversi autori fanno coincidere i due modelli. Riteniamo doveroso precisare che di fatto i due modelli non coincidono perfettamente ma, poiché anche il modello di Rasch prende in considerazione il solo parametro della difficoltà, possiamo considerare il modello ad un parametro come un modello di Rasch in cui la discriminatività sia posta come una costante per tutti gli item.

che abbiamo interesse di misurare e quantificare mediante l'uso di un questionario ed è quella caratteristica che determina la prestazione al test. I modelli IRT prendono in considerazione anche alcuni parametri, riferiti alle caratteristiche degli item, che contribuiscono a spiegare la prestazione ottenuta in una prova di verifica degli apprendimenti. Nel caso specifico il modello ad un parametro considera che la prestazione sia influenzata, oltre che dall'abilità del soggetto, dalla difficoltà dell'item, che corrisponde al livello di abilità necessario perché il soggetto abbia una probabilità di rispondere correttamente uguale alla probabilità di non riuscirci [Rasch, 1960]. Birnbaum [1969], ha proposto due ulteriori parametri: la *discriminatività* dell'item, ossia la capacità che ha di discriminare individui con diversi livelli di abilità e il *guessing*, ossia un parametro che tenga conto del fatto che in un test a scelta multipla un soggetto, soprattutto a bassi livelli di abilità, potrebbe ottenere una risposta corretta anche tirando ad un indovinare, quindi Birnbaum introduce un *asintoto inferiore* che corrisponde alla probabilità che ha il soggetto di ottenere la risposta corretta solo tirando ad indovinare. Barton e Lord [1981], hanno integrato il modello con l'inserimento di un quarto parametro, la *carelessness*, che rappresenta la probabilità che un soggetto che abbia un elevato livello di abilità possa comunque sbagliare a rispondere, per cui introducono un *asintoto superiore* [De Ayala, 2009].

In ogni caso tutti questi approcci caratterizzano gli item in relazione alla probabilità che chi ottiene buoni o bassi punteggi ad un esame ha differenti livelli di abilità. Il beneficio della IRT sta nel fatto che l'analisi viene condotta su ogni singolo item [Lord, 1980], quindi consente di costruire un questionario inserendo direttamente gli item più idonei; pertanto l'analisi di validità non è condotta al termine del questionario ma su ogni item a-priori. L'analisi mostra quindi l'efficienza dell'item per ogni differente livello di abilità. Un altro vantaggio sta nel fatto che la IRT considera l'invarianza dei parametri degli item rispetto al contesto sul quale si lavora, quindi quest'ultimo non influenza la generalizzabilità dei risultati. Questo significa che una scala uniforme di misura può essere predisposta per il suo utilizzo su differenti gruppi e, al tempo stesso, che differenti gruppi di individui possono essere testati anche con set differenti di item, appropriati ai loro livelli di abilità, garantendo la possibilità di comparare comunque

i risultati [Anastasi *et al.*, 1997].

La IRT, infatti, assume che qualunque abilità possa essere misurabile su una scala che abbia come punteggio medio lo zero e unità di misura 1, su un range che va da $-\infty$ a $+\infty$. Tuttavia, solitamente, la IRT per motivi di più facile gestione dei dati fissa questo range da -4 a $+4$, mantenendo la consapevolezza che esistono comunque valori che vanno oltre questi limiti convenzionalmente fissati.

La Item Response Theory affronta il problema proponendo la traduzione di conteggi discreti osservabili su un continuum latente, ricorrendo ad una soluzione di tipo probabilistico, in merito alla quale, ad esempio, il punteggio grezzo 1, ottenuto in seguito alla somministrazione di un item, viene trasformato nella probabilità attesa di osservare una risposta pari ad 1 all'item stesso, quindi in un valore compreso nell'intervallo continuo $[0, 1]$ e in un termine di errore [Baker and Kim, 2004]. Un'ulteriore trasformazione in logit, quindi in un valore compreso nell'intero asse reale, è poi necessaria per evitare che a soggetti con livelli estremi, ma differenti, del tratto latente sia assegnata la medesima probabilità. Il modello di Rasch, che è a fondamento della IRT, in virtù delle sue specifiche caratteristiche e proprietà, è in grado di garantire il rispetto dei requisiti di misurazione di una variabile latente. Volendo formalizzare quanto detto, possiamo indicare con θ_i l'abilità latente dell'individuo i che compila un questionario per la misurazione degli apprendimenti; se $Y_{ij} = y_{ij}$ è la risposta data dall' i – esimo individuo all'item j – esimo, nel caso di un item dicotomico

$$P(Y_{ij} = 0 \mid \theta_i)$$

sarà la probabilità che non risponda correttamente e

$$P(Y_{ij} = 1 \mid \theta_i)$$

che risponda correttamente.

La probabilità di successo e la probabilità di insuccesso sono eventi complementari.

La probabilità di ottenere una certa sequenza di risposte per l'individuo i è uguale a:

$$P(Y_{i1}, Y_{i2}, \dots, Y_{ik} \mid \theta_i) = \prod_{j=1}^k P(Y_{ij} \mid \theta_i) \quad (5.2.0.2)$$

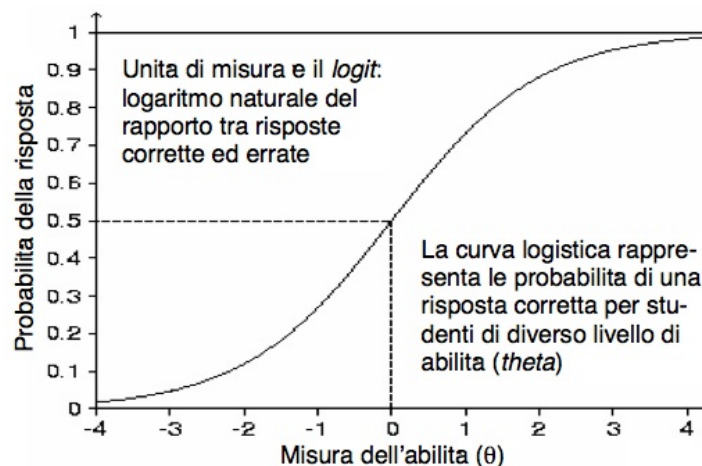


Figura 5.11: La scala di misura nella IRT

È possibile, infine, fare una distinzione all'interno della IRT in funzione del formato degli item, che può essere di *tipo dicotomico* o *tipo politomico*. All'interno degli item dicotomici dobbiamo ancora distinguere gli *item dicotomici naturali*, che possiedono due sole alternative di risposta, di cui solitamente una è giusta ed una è sbagliata; e gli *item dicotomici artificiali* nei quali possiamo classificare quegli item che presentano un ventaglio di alternative nella risposta, tra cui c'è una risposta giusta, e che possono essere ricondotte alla forma dicotomica, attribuendo un valore 1 alla risposta corretta e un valore 0 alle altre risposte. I test di profitto sono un esempio di test che contengono item dicotomici con risposta giusta. Gli item politomici sono invece quegli item che presentano delle alternative di risposta in cui non si può identificare univocamente una risposta giusta in assoluto; è questo il caso di test di personalità o scale di atteggiamento.

5.2.1 Assunti del modello

Una misura, per essere valida ed attendibile, deve possedere due requisiti essenziali e tra loro fortemente correlati: l'*unidimensionalità* e la

specifica oggettività, definita anche *indipendenza locale*, e la *monotonicità*.

Il principio di *unidimensionalità* si riferisce al fatto che il tratto latente, rispetto al quale viene eseguita la misura, deve essere solamente uno, quindi gli item di un questionario devono essere indicatori parziali della medesima variabile latente. La naturale complessità e multidimensionalità dell'essere umano, deve essere semplificata, isolando di volta in volta il tratto latente che si intende studiare. Per formalizzare quanto espresso, diciamo che esiste una entità unidimensionale θ_i , detta abilità latente, associata ad un generico soggetto i , che determina la capacità di superare la prova a cui è sottoposto; la prova è relativa a tale dimensione unica ed è caratterizzata da una difficoltà b_j con $j = 1, 2, \dots, J$.

Da questo principio prende avvio il processo di analisi dei dati per mezzo del modello di IRT. Nel caso in cui dovesse risultare l'assenza di unidimensionalità, non si può procedere all'applicazione del modello, se non individuando sottoinsiemi omogenei di item e conducendo analisi separate per ciascuno di essi. Occorre tuttavia considerare che da un punto di vista concettuale non è ammissibile pensare che esista una sola abilità che indica un soggetto a dare una risposta, ma si può pensare che ce ne sia una dominante che influisce in maniera preponderante rispetto alle altre. Reckase [1979] suggerisce che la unidimensionalità può essere investigata attraverso la metodologia dell'Analisi Fattoriale. Si può affermare, infatti, che un test è unidimensionale se, rappresentando graficamente in ordine decrescente gli autovalori della matrice di correlazione tra gli item, il primo di essi risulta ben discriminato rispetto agli autovalori successivi. Un'altra possibilità è quella di accettare l'ipotesi di unidimensionalità se il primo autovalore della matrice di correlazione degli item è ben distante rispetto agli altri.

La *specifico oggettività*, o *indipendenza locale*, si riferisce, invece, al fatto che il processo di misurazione non deve essere influenzato in alcun modo da caratteristiche dell'individuo che non siano quella che si intende misurare e dalle caratteristiche del test impiegato [Gori *et al.*, 2005]. In altri termini, se uno studente risulta meno abile rispetto ad un altro studente, relativamente ad un certo apprendimento, tale relazione deve rimanere invariata al modificarsi delle caratteristiche del questionario impiegato. L'indipendenza locale garantisce che la

risposta che un soggetto fornisce ad un item è statisticamente indipendente dalle altre.

Rispetto all'assunto di monotonicità diciamo che la probabilità di rispondere correttamente all'item aumenta monotonicamente all'aumentare del livello di abilità; se così non fosse, soggetti con minor livello di tratto avrebbero maggiori chance di rispondere correttamente all'item. Detta funzione caratteristica, quindi, associa la probabilità di rispondere correttamente ad un dato quesito con l'abilità latente del soggetto che risponde al quesito.

Molti sono i modelli teorici possibili che si distinguono fondamentalmente per il numero di parametri associati; il modello di Rasch è quello più semplice dal punto di vista formale poiché è basato su un predittore che include un parametro per ciascun item, corrispondente alla difficoltà dello stesso.

5.2.2 I postulati della Item Response Theory

La IRT si fonda su due postulati fondamentali: la misurazione di tratti latenti e la curva caratteristica dell'item (Item Characteristic Curve, ICC) [Cantrell, C.E., 1999].

I tratti latenti, o semplicemente tratti o abilità, per la loro specifica caratteristica di non diretta osservabilità, possono essere misurati attraverso la misurazione della performance manifestata da un candidato rispetto alla risposta data ad un item di un questionario. Il primo ostacolo nella misura di un tratto latente è rappresentato dalla scelta della scala di misura poiché non è possibile utilizzare una scala metrica. Per ovviare a questo problema la IRT assume che qualunque abilità possa essere misurabile su una scala logistica che abbia come punteggio medio lo zero, come unità di misura e che vada da $-\infty$ a $+\infty$. Tuttavia, solitamente, per motivi di più facile gestione dei dati fissa questo range da -4 a $+4$, mantenendo la consapevolezza che esistono comunque valori che vanno oltre. Poiché si ritiene che ogni candidato abbia una specifica abilità per rispondere correttamente ad ogni specifico item, la scala rappresenta il livello di abilità che un candidato può possedere (θ_i) ed indaga la probabilità di risposta corretta ad un item con specifica difficoltà (b_j), assumendo che tale

probabilità ($P_j(\theta_i)$) sarà bassa in corrispondenza di bassa abilità e sarà crescente a livelli di abilità crescenti.

La curva caratteristica della IRT descrive, contemporaneamente sulla stessa scala, due importanti caratteristiche la difficoltà dell'item e il livello di abilità dello studente.

Occorre precisare che la IRT si fonda sul *teorema di separabilità* in relazione al quale il punteggio complessivo ottenuto da un soggetto ad un questionario composto da un insieme di item è una statistica sufficiente per l'abilità dell'individuo e non dipende dalla difficoltà delle prove; allo stesso modo il punteggio totale su un determinato item è una statistica sufficiente per la difficoltà dell'item e non dipende dall'abilità degli individui [Lovaglio, 2004b].

La ICC consente, per mezzo di una rappresentazione grafica, di ottenere informazioni sia rispetto alle caratteristiche proprie dell'item, sia del livello di abilità dello studente. La ICC, infatti, dà informazioni sulla probabilità di selezionare una certa risposta ad un item, in relazione alla capacità, quindi al tratto latente, della persona [Ostini and Nering, 2005]. Per un fissato livello di difficoltà b_j , al variare di θ_i , si determina la Item Characteristic Curve (ICC), una curva logistica che cresce da 0 (per θ_i che tende a $-\infty$) a 0,5 (quando $\theta_i = b_j$) fino a 1 (per θ_i che tende a $+\infty$).

Nel Modello di Rasch tutte le ICC sono curve logistiche con la stessa pendenza e quindi parallele tra di loro; la sola caratteristica che le distingue è la posizione sulla scala di misura, determinata dalla difficoltà stimata dell'item (figura 5.12).

5.2.3 Il modello di Rasch e il modello logistico ad un parametro (1PL)

Inizieremo ad approfondire i differenti modelli di IRT a cominciare dal modello di Rasch. Nella trattazione affronteremo qui gli aspetti salienti dei modelli IRT e nella trattazione dei modelli a più parametri, per evitare ripetizioni, ci limiteremo a trattare degli aspetti specifici dei modelli stessi.

Il modello di Rasch, come abbiamo anticipato, solo è uno dei possibili modelli di risposta all'item. Esso è chiamato anche Logistic Model, per distinguerlo dai modelli che utilizzano come funzione matematica

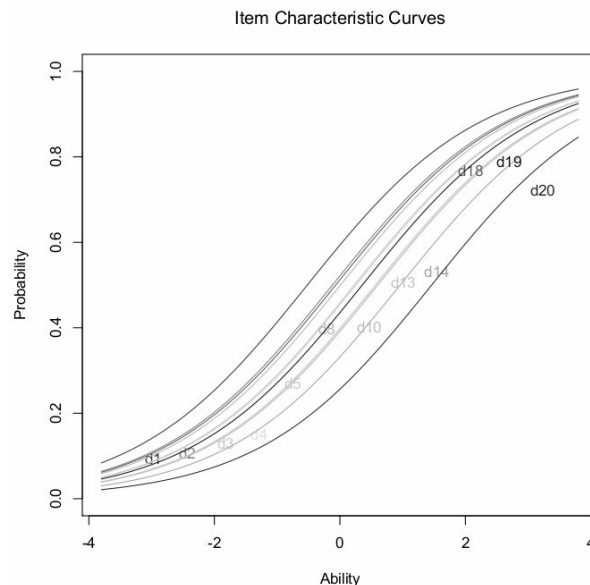


Figura 5.12: ICC con la stessa pendenza (Modello di Rasch)

di riferimento quella ad ogiva invece della logistica. Rasch, infatti, si avvicina con un approccio diverso ai modelli IRT fin dal 1950, usando una funzione logistica anziché quella ad ogiva, già formalizzata da Ferguson [1942]; Lawley [1943]; Mosier [1940b, 1941].

Il modello di Rasch si applica a dati dicotomici, siano essi naturali (come, ad esempio, i risultati ai test di tipo vero/falso) sia che artificiali (ossia ottenuti a seguito di dicotomizzazione delle risposte ad un test a scelta multipla, dove si pone la risposta esatta uguale a 1 e le risposte errate uguali a 0).

Il modello suppone che gli unici due parametri, entrambi rappresentabili sulla stessa scala di misura, che interagiscono per produrre il risultato aleatorio dicotomico X_{ij} , quando un soggetto i risponde ad un item j , siano θ_i , che indica l'abilità latente del soggetto i – esimo, e b_j , che rappresenta la difficoltà incognita dell'item j – esimo.

In via del tutto teorica, è possibile ipotizzare che l'individuo possa

fornire risposte aperte per ogni item e che il soggetto riceva un punteggio uguale ad 1 se risponde correttamente e 0 altrimenti; questo potrebbe aumentare i problemi di interpretazione e, quindi, in generale si preferisce fornire una lista di risposte possibili di cui una sola esatta, mentre le altre, chiamate distrattori, sono errate.

Rasch [1960] afferma che “una persona in possesso di una maggiore abilità rispetto ad un’altra dovrebbe avere una maggiore probabilità di rispondere correttamente a ogni domanda di quel tipo, e analogamente, se una domanda è più difficile di un’altra, ciò significa che per ogni persona la probabilità di rispondere correttamente alla seconda domanda è maggiore”. Per ogni item di un questionario viene rappresentata la curva caratteristica che rappresenta la probabilità di risposta corretta come funzione di un certo tipo di abilità o conoscenze misurate dall’item. Ovvero la probabilità di un soggetto di rispondere correttamente ad un quesito è data dal livello di abilità che possiede in rapporto alla difficoltà del quesito. Questa funzione si può rappresentare per mezzo della curva caratteristica dell’item, *Item Characteristic Curve* (ICC), che è formalizzata come segue:

$$P_j(\theta_i | b) = \frac{\exp(\theta_i - b_j)}{1 + \exp(\theta_i - b_j)} \quad \forall j = 1, 2, \dots, k \quad (5.2.3.1)$$

dalla quale si evince che la probabilità che un soggetto i risponda correttamente all’item j , $P_j(\theta_i)$, dipende dalla differenza tra l’abilità del soggetto (θ_i) e la difficoltà dell’item (b_j). È immediato intuire che la probabilità di una risposta corretta ($X_{ij} = 1$) è uguale a 0,5 solo quando $\theta_i = b_j$, cioè quando l’abilità dell’individuo è uguale alla difficoltà dell’item. Tale proprietà è coerente con l’idea secondo la quale, se un soggetto incontra un item che presenta la stessa intensità relativa alla caratteristica da misurare, la probabilità che il soggetto prevalga sulla prova è uguale alla probabilità che quest’ultima prevalga sul soggetto.

Il *modello logistico ad un parametro* (1PL), introdotto da Birnbaum [1968], equivale sostanzialmente al modello di Rasch con la differenza che introduce il parametro α di discriminatività ponendolo uguale a 1; tale parametro assumerà rilevante importanza nel modello logistico a

due parametri (2PL). La 1PL può essere esplicitata nell'equazione

$$P_j(\theta_i | a, b) = \frac{1}{1 + \exp\{-a_j(\theta_i - b_j)\}} \quad \forall j = 1, 2, \dots, k \quad \text{e} \quad a_j = 1 \quad (5.2.3.2)$$

il parametro discriminazione a è posto a valore costante 1 per tutti gli item, ossia tutti gli item sono vincolati allo stesso potere discriminante.

La differenza $(\theta_i - b_j)$ è potenzialmente infinita, a fronte di $P_j(\theta_i)$ variabile tra 0 e 1, il modello quindi specifica un legame logistico:

$$\text{logit}(P_j(\theta_i)) = \ln \left(\frac{P_j(\theta_i)}{1 - P_j(\theta_i)} \right) = \theta_i - b_j \quad \forall j = 1, 2, \dots, k \quad (5.2.3.3)$$

La trasformazione logit consente di “linearizzare” i punteggi ordinali ottenuti in un test, in modo che intervalli identici corrispondono a quantità identiche della scala misurata, ossia in modo da ottenere una scala di misura continua. In tale ottica, la misura non è proporzionale alla probabilità di risposta ma alla corrispondente trasformazione logit. Si osservi la relazione lineare tra i punteggi logit e i parametri di abilità del soggetto e difficoltà dell'item.

La misura del logit viene definita come il logaritmo naturale del rapporto di verosimiglianza (meglio noto come misura “odds” dall'inglese log-odds unit) definito, a sua volta, dal rapporto tra la probabilità di successo e quella di insuccesso. Partendo dalla probabilità di riuscita e di insuccesso in un item è possibile calcolare, appunto, il punteggio atteso su quell'item in funzione della capacità della persona. La relazione tra il punteggio atteso ad un item e la misura dell'abilità della persona è denominata Item Characteristic Curve (ICC). Gli odds, quindi, definiscono il rapporto che sussiste tra la probabilità di successo e quella di insuccesso. Quando il rapporto di verosimiglianza è minore di uno, allora la probabilità di successo è minore di quella di insuccesso. Il logit è, quindi, un'unità probabilistica definita dal logaritmo naturale, o neperiano, del rapporto di verosimiglianza di un evento.

Da quanto detto si evince che il modello di Rasch stima l'abilità e la difficoltà degli item nella stessa unità di misura, il logit appunto, e assume che un soggetto i abbia una probabilità di rispondere x , superiore alla probabilità di rispondere $x - 1$ sull'item j , solo se la sua

abilità è superiore alla difficoltà della categoria x per l'item j ; qualora tale differenza fosse nulla, allora l'individuo i avrebbe la stessa probabilità di rispondere x ed $x - 1$ sull'item j e l'abilità associata al soggetto coinciderebbe con la stima della soglia di difficoltà che discrimina i punteggi x ed $x - 1$ su tale item. La Curva Caratteristica dell'Item

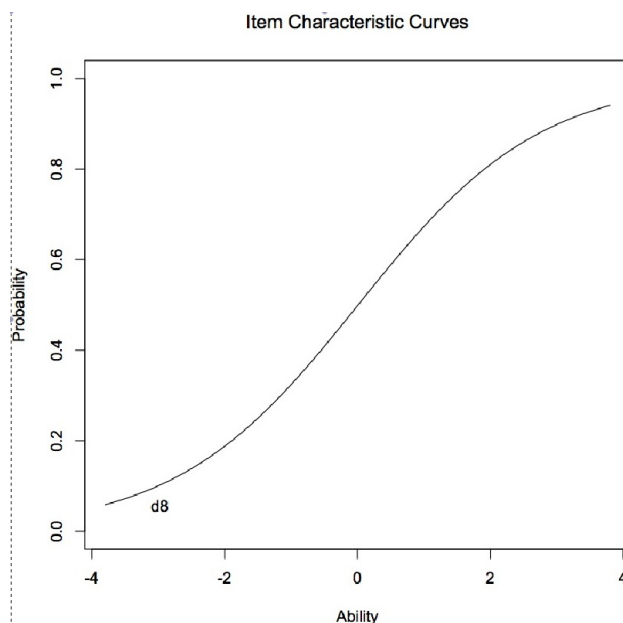


Figura 5.13: Rappresentazione della Item Characteristic Curve (ICC) secondo il Modello di Rasch. La curva presenta un indice di difficoltà di 0,0046, con probabilità di risposta corretta per l'individuo mediano pari a 0,4988

(ICC) appare come un funzione monotona crescente, e rappresenta la probabilità di rispondere correttamente ad un item per i diversi livelli di abilità in funzione della difficoltà dell'item stesso. La curva tende ad approssimarsi asintoticamente a $-\infty$ per valori molto bassi di θ , a cui è associata una probabilità di rispondere correttamente che tende a 0, e a $+\infty$ per valori molto alti di θ , ai quali è associata una probabilità di rispondere correttamente che tende a 1. Anche il parametro difficol-

tà può assumere concettualmente valori compresi tra $-\infty$ e $+\infty$, ma come abbiamo detto convenzionalmente si utilizza anche in questo caso un range più circoscritto, in questo lavoro utilizzeremo un range compreso tra -4 e $+4$. Poiché nel caso di 1PL la difficoltà è l'unico parametro che varia, in quanto α è vincolato ad 1, si avrà che la ICC assume una forma specifica influenzata solo da questo parametro.

La figura 5.13 mostra l'andamento della curva di un item (d8) al crescere del valore di θ_i . L'item presenta un indice di difficoltà pari a 0,0046, con *standard error* pari a 0,1084 e *z value* 0,0426; la probabilità di risposta corretta per l'individuo mediano è pari a 0,4988. L'item al quale abbiamo fatto riferimento, così come gli esempi che seguiranno, è parte di un questionario predisposto per le classi quarte di scuola secondaria di secondo grado per le Olimpiadi Nazionali di Statistica edizione 2013.

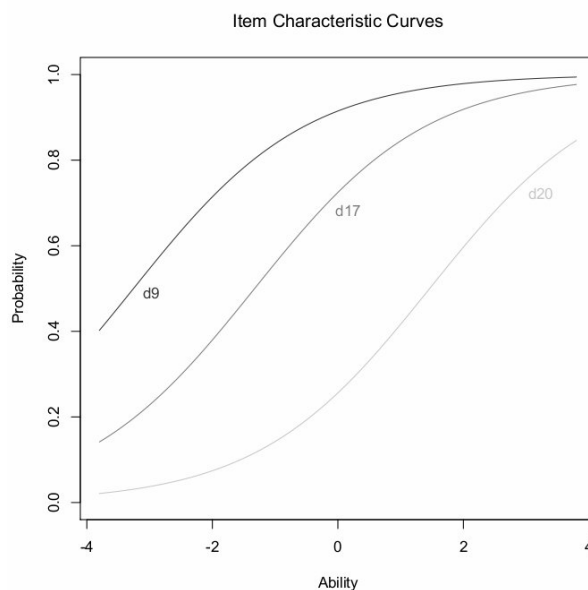


Figura 5.14: Confronto tra la ICC di due item con diverso livello di difficoltà

Item	Dffclt	std.err	z.vals	$P(x = 1 z = 0)$
d9	-2,444	0,1630	-14,9924	0,9201
d17	-1,001	0,1161	-8,6324	0,7314
d20	1,100	0,1183	9,3088	0,2495

Tabella 5.2: Confronto tra item sotto il modello di Rash (figura 5.14)

Una volta stimati i parametri di abilità dei soggetti e la difficoltà degli item e rilevativi standard error, il modello Rasch consente di valutare, attraverso statistiche basate sui residui del modello, l'adattamento globale della struttura teorica ipotizzata sul singolo item e sulla totalità degli item [Wright, 1982].

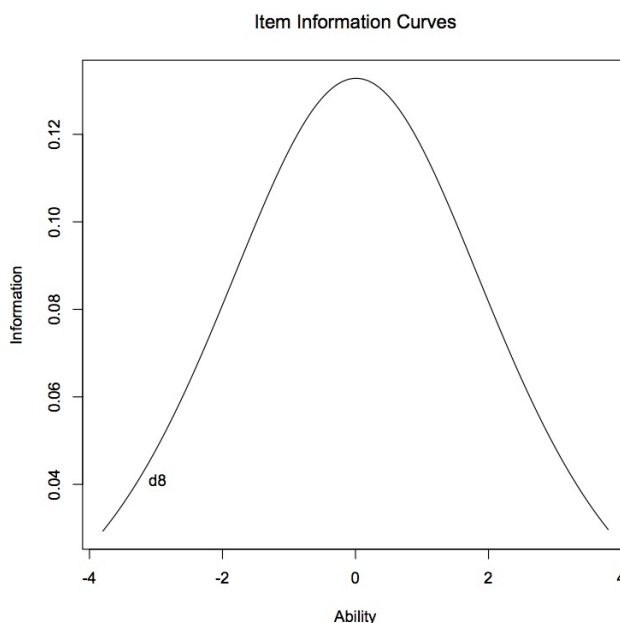


Figura 5.15: Item Information Curve dell'Item n. 8

Se le assunzioni di unidimensionalità e di indipendenza locale vengono rispettate e se il modello si adatta ai dati empirici, allora i parametri stimati godono della proprietà di invarianza. Dal confronto

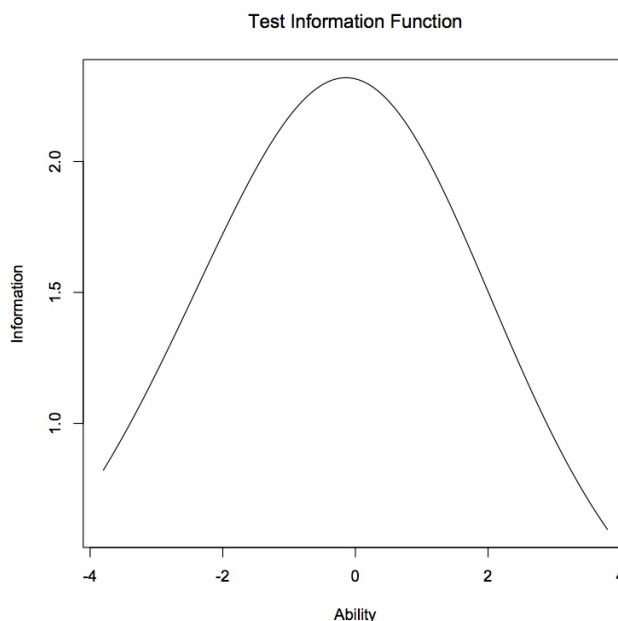


Figura 5.16: Item Information Curve del Test

tra due soggetti risulterà, quindi, che le differenze tra il livello di abilità di due soggetti non cambia qualunque sia la difficoltà dell'item che devono risolvere e, viceversa, la differenza tra le difficoltà di due item non cambia qualunque sia l'abilità del soggetto.

5.2.4 Il metodo di massima verosimiglianza (ML) per le stime del modello

Come abbiamo poc'anzi anticipato, nella IRT si applica per il processo di stima il criterio di *massima verosimiglianza* (maximum likelihood – ML), in quanto il modello non è lineare; l'errore in questo caso diventa il dato stimato che non ha probabilità di essere osservato. Note le risposte degli individui agli item del test, si devono stimare: sia i parametri degli item (item parameters), sia le abilità degli individui (person parameters). In questo lavoro, tra le diverse procedure possibili, si utilizza un metodo di stima congiunta per stimare tutti i

parametri contemporaneamente (massima verosimiglianza). La stima congiunta consiste in un processo sequenziale e iterativo, formato da tre fasi principali. Nella prima si ottiene una stima iniziale delle abilità (θ_i), che si ottiene trasformando i punteggi ottenuti nel test in punteggi z (fase di standardizzazione). La seconda consiste nello stimare con il metodo ML i parametri degli item partendo dalla stima di θ_i , precedentemente calcolata. La terza permette di stimare nuovamente θ_i (con ML) partendo dai parametri degli item precedentemente stimati nella fase due. A questo punto il primo ciclo è terminato. Tuttavia per massimizzare la funzione di verosimiglianza (ovvero massimizzare la probabilità per quel vettore di dati di essere osservato) è necessario ripetere in modo iterativo il processo, fino a che i parametri stimati tra un ciclo ed un altro siano stabili e più precisi (o comunque presentino una variazione trascurabile). Quindi fino a quando ciò non si verifica il processo di stima ricomincia dalla fase due utilizzando come valore di riferimento non il valore stimato nella prima fase, ma il risultato ottenuto nella terza del precedente ciclo. Questo processo iterativo è detto *algoritmo di Newton - Raphson*, e permette di implementare la funzione. Nel processo di stima congiunta di massima verosimiglianza, infine, è possibile calcolare l'errore standard associato alla stima dei parametri. Nel caso della stima dell'abilità, l'errore viene espresso come varianza associata a θ_i^* stimato, e si calcola come il reciproco della funzione informativa del test. Nel caso della stima dei parametri degli item, l'errore viene espresso come matrice di varianza e covarianza dei parametri stimati, ovvero come l'inverso della matrice informativa dei parametri degli item stimati. Si procede, infine, all'esame del fit, considerando tre aspetti: il fit complessivo del modello, il fit di ogni singolo item e il fit del soggetto. L'esame del fit permette di valutare l'effettiva bontà della stima dei parametri, e quindi conseguentemente quanto la funzione di massima verosimiglianza è stata massimizzata. Qualora tale adattamento non si verifichi, ovvero nel caso in cui il modello non si adatti bene ai dati, allora probabilmente il modello utilizzato non va bene e va sostituito. Per la valutazione del fit, in questo lavoro, viene utilizzata la statistica del chi quadrato (χ^2).

Al fine di illustrare il metodo della massima verosimiglianza, si consideri un campione $\{x_i\}_{i=1}^n$ di variabili casuali identicamente e indi-

pendentemente distribuite, con distribuzione normale $x_i \sim N(\mu, \sigma^2) \forall i$.
La funzione di verosimiglianza associata è:

$$\mathcal{L}(\mu, \sigma^2 | \{x_i\}_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\} \quad (5.2.4.1)$$

La massimizzazione della funzione di verosimiglianza è equivalente a massimizzarne il logaritmo:

$$\mathcal{L}(\mu, \sigma^2 | \{x_i\}_i) = \ln \mathcal{L}(\mu, \sigma^2 | \{x_i\}_i) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \quad (5.2.4.2)$$

I parametri μ e σ^2 sono determinati risolvendo il problema di massimo

$$\{\mu, \sigma^2\} = \arg \max_{\mu, \sigma^2} \mathcal{L}(\mu, \sigma^2 | \{x_i\}_i) \quad (5.2.4.3)$$

Le condizioni del primo ordine per un massimo definiscono il seguente sistema di equazioni in μ e σ^2

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2} \sum_i (x_i - \hat{\mu}) = 0 \\ \frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_i (x_i - \hat{\mu})^2 = 0 \end{cases} \quad (5.2.4.4)$$

dove i segni di apice sopra i parametri denotano i loro stimatori. Il massimo della funzione di log-verosimiglianza si ottiene quando le derivate parziali prime si annullano e le derivate parziali seconde sono negative. Dalla prima equazione discende immediatamente lo stimatore di massima verosimiglianza per la media

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5.2.4.5)$$

cioè la media campionaria. La varianza dello stimatore $\hat{\mu}$ è data dalla seguente espressione

$$\text{var}(\hat{\mu}) = \text{var} \left(\frac{1}{n} \sum_{i=1}^n x_i \right) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(x_i) = \frac{\sigma^2}{n} \quad (5.2.4.6)$$

Sostituendo $\hat{\mu}$ nella seconda equazione, si ha lo stimatore di massima verosimiglianza per la varianza

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (5.2.4.7)$$

cioè la varianza campionaria.

L'esempio consente di illustrare alcune proprietà degli stimatore di massima verosimiglianza. È immediato verificare la correttezza $\hat{\mu}$

$$E[\hat{\mu}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \mu \quad (5.2.4.8)$$

D'altra parte, $\hat{\sigma}^2$ non gode di tale proprietà. Ricordando che:

$$\sum_i (x_i - \mu)^2 = \sum_i (x_i - \hat{\mu})^2 + n(\hat{\mu} - \mu)^2 \quad (5.2.4.9)$$

segue che:

$$E[\hat{\sigma}^2] = \frac{1}{n} E\left(\sum_{i=1}^n (x_i - \hat{\mu})^2\right) = \frac{1}{n} E\left[\sum_i (x_i - \mu)^2 - n(\hat{\mu} - \mu)^2\right] = \frac{n-1}{n} \sigma^2 \quad (5.2.4.10)$$

Dunque $\hat{\sigma}^2$ non è uno [stimatore corretto; un tale stimatore sarebbe dato dalla statistica:

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2$$

Vale la pena d'altra parte di osservare che lo stimatore di massima verosimiglianza è comunque uno stimatore asintoticamente corretto; infatti:

$$\lim_{n \rightarrow \infty} E[\hat{\sigma}^2] = \lim_{n \rightarrow \infty} \frac{n-1}{n} \sigma^2 = \sigma^2 \quad (5.2.4.11)$$

In particolare, qualunque stimatore di massima verosimiglianza è asintoticamente corretto e asintoticamente normalmente distribuito.

Una volta ottenute le stime dei parametri, il confronto tra risposte osservate e valori attesi consente di esprimere un giudizio sulla bontà

di adattamento del modello impiegato ai dati osservati. Nell'ambito dei modelli IRT è particolarmente utile valutare la bontà di adattamento, in modo da individuare eventuali violazioni delle ipotesi fondanti il modello.

Le statistiche maggiormente impiegate per la diagnostica del modello di Rasch sono le statistiche Outfit e Infit, basate sul confronto tra risposte osservate per ciascun individuo a ciascun item del questionario e risposte attese sulla base del modello di Rasch stimato [Wright and Masters, 1982].

Indicando con P_{ijx} la probabilità che l'individuo i – esimo scelga la categoria x per il j – esimo item e con X_{ij} la risposta osservata per l'individuo i – esimo e l'item j – esimo, si ha che

$$E_{ij} = \sum_{x=0}^{r_j} xP_{ijx} \quad (5.2.4.12)$$

è il valore atteso della risposta;

$$Y_{ij} = X_{ij} - E_{ij} \quad (5.2.4.13)$$

è il residuo corrispondente;

$$W_{ij} = \sum_{x=0}^{r_j} (X_{ij} - E_{ij})^2 P_{ijh} \quad (5.2.4.14)$$

è la varianza della risposta osservata X_{ij} ed assume valore massimo quando le stime di abilità e difficoltà sono identiche, mentre tende a ridursi all'aumentare della differenza in valore assoluto tra difficoltà dell'item j e abilità dell'individuo i ;

$$Z_{ij} = \frac{Y_{ij}}{\sqrt{W_{ij}}} \quad (5.2.4.15)$$

è il residuo standardizzato avente una distribuzione Normale con media pari a 0 e varianza unitaria. Per valutare l'adattamento complessivo di un item al modello si calcola la media aritmetica semplice o ponderata dei residui standardizzati al quadrato. In particolare, nel

caso di una media aritmetica semplice si ottiene la statistica Outfit (o Unweighted Mean Square statistic)

$$OUT_j = \sum_{i=1}^I \frac{Z_{ij}^2}{I} \quad (5.2.4.16)$$

avente varianza pari a

$$s_j^2 = \sum_{i=1}^I \left(\frac{\frac{\sum_{x=0}^{r_j} (x - E_{ij})^4 P_{ijh}}{W_{ij}^2}}{I^2} \right) - \frac{1}{I} \quad (5.2.4.17)$$

Dal momento che s_j varia in funzione del numero di individui nel campione e W_{ij} varia sia da item ad item che da campione a campione, non è facile determinare un livello di cut-off generale per valutare la bontà di adattamento di un item; di conseguenza, solitamente si procede alla standardizzazione (trasformazione di Wilson-Hilferty), ottenendo una statistica con distribuzione approssimativamente Normale con media 0 e varianza unitaria

$$t_{jOUT} = (OUT_j^{1/3} - 1) \left(\frac{3}{s_j} \right) + \left(\frac{s_j}{3} \right) \quad (5.2.4.18)$$

Poichè la statistica Outfit è il risultato di una media aritmetica semplice, essa risulta particolarmente sensibile a risposte inattese (cioè improbabili) che provengono da individui per i quali l'item j risulta inappropriato, in quanto troppo facile o troppo difficile. Per ovviare a questo problema la statistica Outfit viene solitamente affiancata, o sostituita, con la statistica Infit (Weighted Mean Square statistic) che pondera i residui standardizzati al quadrato con le rispettive varianze individuali

$$IN_j = \frac{\sum_{i=1}^I W_{ij} Z_{ij}^2}{\sum_{i=1}^I W_{ij}} \quad (5.2.4.19)$$

con varianza

$$q_j^2 = \sum_{i=1}^I \left(\frac{\sum_{x=0}^{r_j} (h - E_{ij})^4 P_{ijh}}{W_{ij}^2} \right) / \sum_{i=1}^I W_{ij}^2 \quad (5.2.4.20)$$

Dal momento che la varianza W_{ij} è tanto maggiore quanto più le stime di abilità e di difficoltà sono simili, la statistica Infit dà maggiore peso alle risposte degli individui per i quali l'item j è ben calibrato, cioè ha un livello di difficoltà in linea con l'abilità del soggetto. Anche in tal caso, in pratica, si utilizza la statistica Infit standardizzata

$$t_{jIN} = (IN_j^{1/3} - 1) \left(\frac{3}{q_j} \right) + \left(\frac{q_j}{3} \right) \quad (5.2.4.21)$$

che ad un livello di significatività del 5% assume valori compresi nell'intervallo $[-2; +2]$. La prassi consiste nell'eliminare in una procedura iterativa tutti gli item, o gli individui se le statistiche Infit ed Outfit sono valutate su essi, che presentano un cattivo adattamento al modello, cioè valori che fuoriescono dal suddetto intervallo di significatività. Spesso l'insieme di item esclusi contribuisce a misurare una dimensione separata; nei casi più estremi, invece, può accadere che non sia possibile individuare nessun insieme di item coerenti con le ipotesi del modello di Rasch. Quest'ultima situazione si verifica nel caso in cui il questionario sia mal calibrato oppure se la popolazione degli individui non presenta caratteristiche simili ma è composta da soggetti afferenti a popolazioni diverse. Questa situazione può essere sintomo di un funzionamento diverso degli item in corrispondenza di gruppi di individui distinti: tale fenomeno va sotto il nome di Differential Item Functioning o DIF. L'impatto del DIF sulla validità di un questionario e, conseguentemente, sull'oggettività delle misure dipende sia dal numero di item distorti sia dall'entità delle differenze nei parametri di difficoltà per i vari item tra i diversi sottogruppi della popolazione.

In letteratura esistono varie proposte per la diagnostica del DIF ma quella più diffusa e implementata nei software di uso più comune si basa sull'analisi dei residui tra i sottogruppi individuati rispetto a una o più variabili di aggregazione. Tramite l'inserimento nel modello di un'interazione tra ciascun item (o ciascuna modalità di risposta per ogni item, nel caso di modelli per item politomici) e la variabile di aggregazione, si possono stimare gli effetti differenziali positivi o negativi rispetto alla difficoltà media dell'item e tramite il classico test χ^2 , che valuta la significatività statistica di tali differenze. Nel modello di Rasch con integrazione del parametro di discriminazione fissato a 1

e nel modello a due parametri si può utilizzare il parametro stesso di discriminazione per prevedere in modo esplicito la possibilità che uno stesso item discrimini in maniera diversa per livelli di abilità differenti.

5.2.5 La funzione di informazione

Fondamentale per qualsiasi analisi inferenziale è che le stime conservino una quota rilevante dell'informazione originale. L'informazione in generale rappresenta l'inverso della varianza; quindi, più il valore di informazione è elevato, migliore sarà la precisione nella stima dei parametri del modello.

Se indichiamo con $\sigma^2(\theta_i)$ la varianza dello stimatore dell'abilità, si avrà:

$$I(\theta_i) = \frac{1}{\sigma^2(\theta_i)} \quad (5.2.5.1)$$

che graficamente rappresenta la Funzione di Informazione (Item Information Curve, IIC). Questa funzione dipende dalla posizione delle abilità individuali, nel grafico in figura 5.2.5 possiamo osservare la Funzione di Informazione di diversi item che compongono un test. L'abilità varia tra in un intervallo convenzionalmente fissato tra -4 e +4. Si notino i differenti livelli di difficoltà degli item che sono rilevabili nella posizione di massimo della curva proiettata sull'asse delle ascisse.

Può risultare interessante conoscere l'informazione generale di uno specifico item, in questo caso la funzione si ricava dalla formula:

$$I_j(\theta_i) = P(y_{ij} = 1)P(y_{ij} = 0) \quad (5.2.5.2)$$

che graficamente rappresenta la Funzione di Informazione.

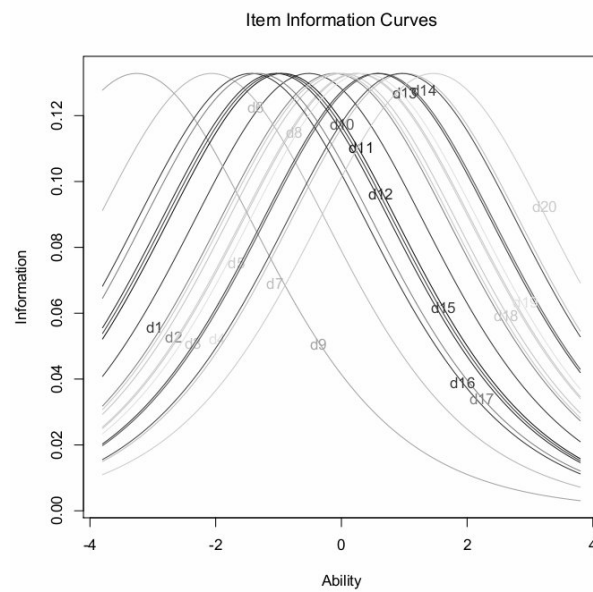


Figura 5.17: Item Information Curves di tutti gli item di una prova

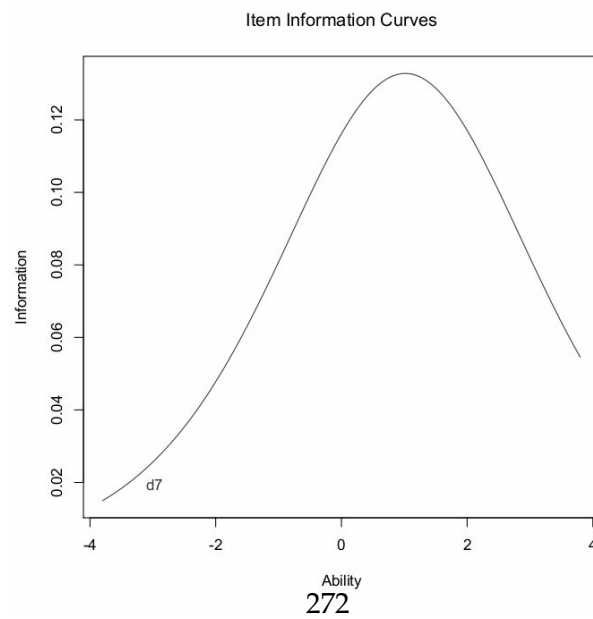


Figura 5.18: Item Information Curve dell'item "d7"

Nel caso in cui la difficoltà dell'item sia pari a 1, si osservi l'item della curva in figura 5.20 che ha difficoltà 1,016, $I_j(\theta_i)$ rappresenta l'informazione fornita dall'item j al livello di abilità θ_i , simmetrica rispetto al valore massimo.

Dai grafici illustrati, si può notare che un item raggiunge il massimo dell'informazione quando abilità e difficoltà si eguagliano, ossia quando $P(y_{ij}) = 0,5$.

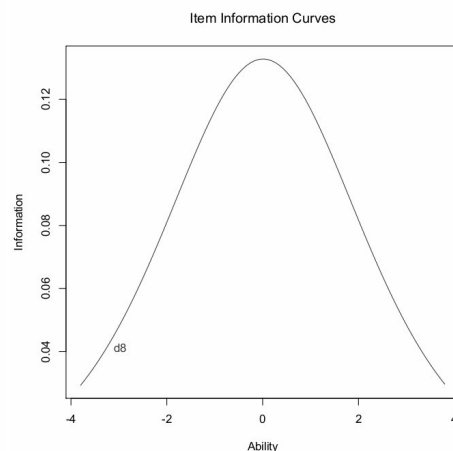
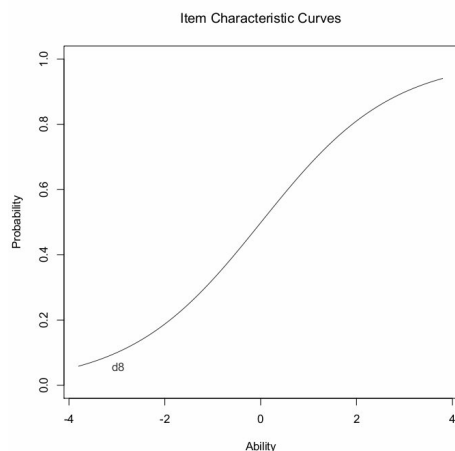


Figura 5.19: ICC dell'item "d8"

Figura 5.20: ICC dell'item "d8"

5.2.6 Il modello logistico a due parametri (2PL)

Abbiamo visto che nel modello di Rasch la funzione che rappresenta la probabilità è una distribuzione logistica cumulata a un parametro, che è rappresentato dalla difficoltà. Una seconda classe di modelli della IRT considera la possibilità che gli item differiscano tra di loro non solo per il livello di difficoltà ma anche per la loro capacità di discriminare tra i soggetti aventi diversi livelli di abilità. La IRT introduce nel modello, quindi, un secondo parametro che definisce discriminatività (*item discrimination*). Nel modello logistico a due parametri (2PL), teorizzato da Birnbaum [1968] da una trasformazione

del modello della Ogiva Normale a due parametri di Lord [1952], la probabilità di una risposta corretta è espressa da

$$P_j(\theta_i|a, b) = \frac{e^{D a_j (\theta - b_j)}}{1 + e^{D a_j (\theta - b_j)}} \quad (5.2.6.1)$$

dove

P_j è la probabilità di rispondere correttamente all'item j – esimo;

θ_i è l'abilità del soggetto i – esimo;

e è la costante Nepero (2,71828...);

D è una costante che moltiplicata per a permette di trasformare la funzione ogiva in quella logistica ($D = 1,7$). La scelta di questo valore per D è dovuta al fatto che tale valore produce equivalenti valori di interpretazione tra i parametri degli item sia sotto il modello normale ad ogiva, sia sotto il modello logistico a due parametri. Spesso molti autori e software statistici pongono per praticità il valore di D a 1;

a_j è il parametro di discriminazione dell'item (convenzionalmente varia tra 0,5 e 2,5 anche se teoricamente può assumere qualsiasi valore);

b_j è difficoltà dell'item j – esimo.

Per interpretare la difficoltà dell'item occorre verificare il punto di ordinata 0,5; la discriminatività, invece, influenza la pendenza della curva.

Osservando le figure 5.21 e 5.22 si osserva come le domande “d3” e “d5” presentano un livello di difficoltà e discriminatività molto vicini ($\text{diff}_{d3} = 0,2419$, $\text{diff}_{d5} = 0,2138$, $\text{dis}_{d3} = 0,7493$, $\text{dis}_{d5} = 0,7603$), infatti le curve quasi si sovrappongono; la situazione è differente se confrontiamo le domande “d7” e “d10” che hanno un livello di difficoltà vicino ma un diverso grado di discriminatività ($\text{diff}_{d7} = 1,2556$, $\text{diff}_{d10} = 1,1166$, $\text{dis}_{d7} = 0,5664$, $\text{dis}_{d10} = 0,3660$), la domanda “d7” è leggermente più difficile e discrimina di più della “d10”. Nella figura 5.23, invece, gli item a confronto presentano discriminatività simile e difficoltà differente, pertanto presentano la stessa pendenza ma la “d2” è più semplice delle altre due. Le ICC che abbiamo visto, sono tutte curve monotone crescenti, ovvero all'aumentare del livello di abilità aumenta anche la probabilità di rispondere correttamente all'item. È possibile però che la curva possa presentare un andamento anomalo, ovvero all'aumentare del livello di abilità si assiste ad una diminuzione della probabilità di rispondere correttamente. In questo caso si parla di discriminazione negativa che si può verificare in due casi: quando

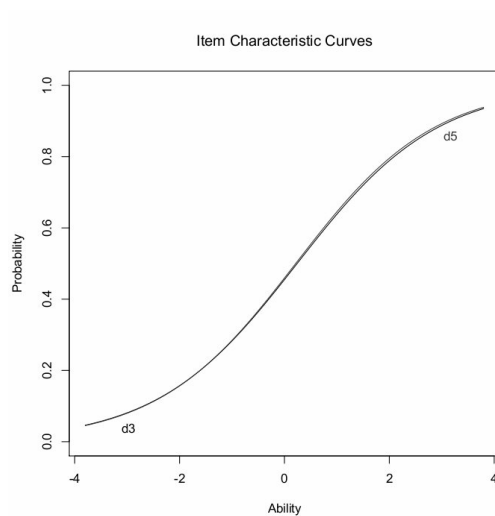


Figura 5.21: 2PL: confronto tra due item con difficoltà e discriminatività simile

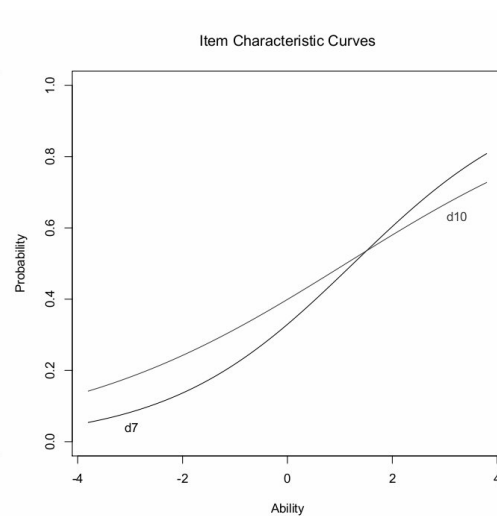


Figura 5.22: 2PL: confronto tra due item con difficoltà simile e discriminatività differente

l'indice di discriminazione è negativo per risposte corrette o quando è negativo per risposte non corrette. In questa ultima situazione, per complementarietà si verifica che le risposte corrette hanno un indice di discriminazione positivo e quindi graficamente si hanno due curve identiche ma con andamento inverso, cioè una monotona crescente (per risposte corrette) e una monotona decrescente (per risposte sbagliate).

Una situazione molto rara, è quella della discriminazione perfetta. Quando un item presenta un potere discriminante perfetto la sua ICC si presenta come in figura 5.24.

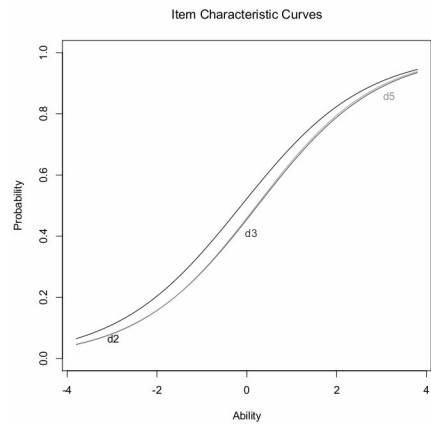


Figura 5.23: 2PL: confronto tra item con discriminatività simile e difficoltà differente

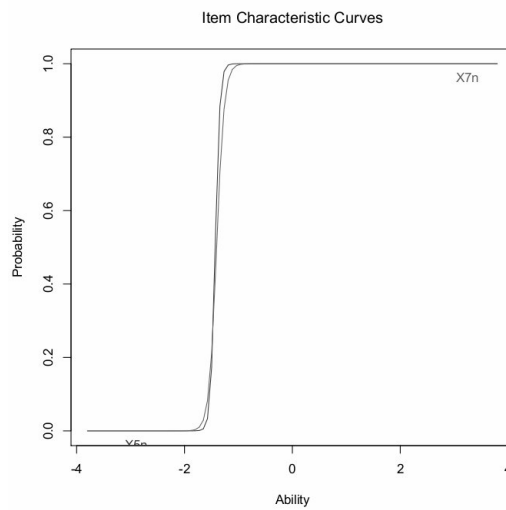


Figura 5.24: 2PL: confronto tra due item con un alto potere discriminante

La formalizzazione della ICC sotto il modello logistico a due parametri è data da

$$P_j(\theta_i | a, b) = \frac{1}{1 + \exp[-Da_j(\theta_i - b_j)]} \quad \forall j = 1, 2, \dots, k \quad (5.2.6.2)$$

dalla quale si evince che la probabilità che un soggetto i risponda correttamente all'item j , $P_j(\theta_i)$, dipende dalla differenza tra l'abilità del soggetto (θ_i) e la difficoltà dell'item (b_j) e dalla capacità di discriminazione dell'item (a_j). La pendenza della curva assumerà valore $a_j = 0,25$ quando sarà $\theta_i = b_j$. Quindi se a_j aumenta, aumenterà anche la pendenza della curva. Nel caso limite in cui a_j cresca all'infinito, la ICC approssima una funzione a salti che vale 0 per $\theta_i < b_j$ e 1 per $\theta_i > b_j$ e che rappresenta la funzione di un item di Guttman con una discriminazione perfetta se $\theta_i = b_j$ e nessuna discriminazione per valori minori o maggiori di θ_i .

5.2.7 Il modello logistico a tre e a quattro parametri

Il Modello Logistico a tre parametri (3PL) è stato ancora formulato da Birnbaum [1968]. Questo modello considera simultaneamente tutti e tre i parametri degli item: difficoltà, discriminatività e guessing. Proprio per questo motivo è particolarmente apprezzato rispetto alla misurazione degli apprendimenti in quanto considera anche la possibilità che uno studente tenti di tirare ad indovinare le risposte (parametro di guessing). La funzione matematica del modello è

$$P_j(\theta_i) = c_j + (1 - c_j) \frac{e^{Da_j(\theta_i - b_j)}}{1 + e^{Da_j(\theta_i - b_j)}} \quad (5.2.7.1)$$

dove

P_j è la probabilità di rispondere correttamente all'item j – esimo;

θ_i è l'abilità del soggetto i – esimo;

c_j è il parametro di guessing;

D è una costante che moltiplicata per a permette di trasformare la funzione ogiva in quella logistica ($D = 1,7$);

a_j è il parametro di discriminazione dell'item (convenzionalmente varia tra 0,5 e 2,5 anche se teoricamente può assumere qualsiasi valore);

b_j è difficoltà dell'item j – esimo;

e è la costante Nepero (2,71828...).

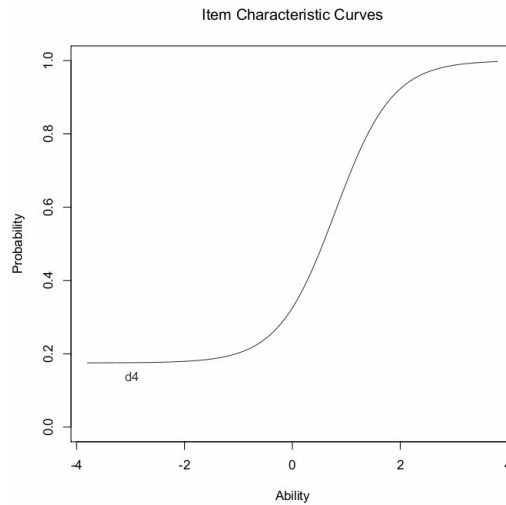


Figura 5.25: 3PL: esempio di item che presenta il parametro di guessing

La formalizzazione della ICC sotto il modello logistico a tre parametri è data da

$$P_j(\theta_i | a, b, c) = c_j + \frac{1 - c_j}{1 + \exp\{-a_j(\theta_i - b_j)\}} \quad \forall j = 1, 2, \dots, k \quad (5.2.7.2)$$

Osservando la ICC della figura 5.25 è possibile notare subito una differenza rispetto alle ICC dei precedenti modelli. L'asintoto orizzontale fissato dal parametro di guessing considera che anche un soggetto con un livello di abilità molto basso, può comunque rispondere correttamente ad un item tirando ad indovinare. Si osservi che il parametro guessing non varia come funzione del livello di abilità, perché in tutti i soggetti si assume la stessa incidenza del caso. Il parametro difficoltà nel 3PL viene interpretato in modo diverso. Infatti non indica più il punto sulla scala di abilità dove la probabilità di risposta corretta è uguale a 0,5, ma è il punto sulla scala dove la probabilità di risposta corretta è uguale a $\frac{(1+c_j)}{2}$.

Quando i modelli logistici vengono applicati in campo valutativo si può prendere in considerazione anche il caso in cui un soggetto con alti livelli di abilità non necessariamente risponda in modo corretto con un'elevata probabilità; in questo caso nel modello può essere inserito un quarto parametro che determina un asintoto superiore. Si tratta del modello logistico a quattro parametri (4PL). Il modello è formalizzato dall'equazione [Barton and Lord, 1981]

$$P_j(\theta_i | a, b, c, d) = c_j + \frac{d_j - c_j}{1 + \exp\{-a_j(\theta_i - b_j)\}} \quad \forall j = 1, 2, \dots, k \quad (5.2.7.3)$$

dove il parametro d_j è definito *carelessness*, ossia inaccuratezza,

5.2.8 Esempio di caso: le Olimpiadi Nazionali di Statistica 2013)

L'applicazione in questione mira a valutare il livello di abilità nella disciplina statistica di un gruppo di studenti di classe quarta frequentanti la scuola secondaria di secondo grado, in particolar modo licei scientifici dislocati sul territorio nazionale. Il questionario è stato somministrato in occasione della terza edizione delle Olimpiadi Nazionali di Statistica, promosse dalla società Italiana di Statistica, che si sono svolte nel marzo del 2013 e che ha visto coinvolti 1465 ragazzi di classe quarta e di classe quinta delle scuole italiane. I ragazzi erano suddivisi in due categorie: la categoria matematica, alla quale hanno aderito studenti che studiano la statistica all'interno del corso curricolare di matematica (528 studenti, di cui 326 di classe quarta e 202 di classe quinta); la categoria statistica, alla quale hanno aderito studenti che frequentano corsi specifici di statistica (937 studenti, di cui 488 di classe quarta e 449 di classe quinta).

L'obiettivo dell'evento è quello di motivare all'apprendimento della statistica per mezzo di un'esperienza fondata su una sana competizione, che consentisse di valorizzare le eccellenze. I questionari sono stati costruiti in modo da consentire un'importante selezione degli studenti con abilità elevate nella disciplina. Per ogni classe e per ogni categoria sono stati strutturati due questionari, uno per la fase delle eliminatorie e uno per la fase delle finali.

Il questionario preso in questione per l'analisi è quello rivolto alla

categoria di statistica delle classi quarte. Il questionario è composto da 20 item a risposta multipla con cinque possibili alternative, una corretta alla quale vengono attribuiti 3 punti, una è sempre il "Non so" che corrisponde alla non attribuzione del punteggio e che risolve il problema della mancata risposta, tre errate che prevedono la penalizzazione di 1 punto. Ai ragazzi sono stati dati quaranta minuti per la compilazione, che è avvenuta on-line tramite la piattaforma e-learning di Moodle. Dal numero iniziale di 488 studenti sono stati esclusi 8 studenti ai quali la piattaforma non ha dato attribuzione di punteggio perché hanno chiuso la prova in maniera errata.

Al fine di valutare nel modo più completo possibile ed in forma comparativa la prova, sono state presi in considerazione alcuni criteri di sintesi, ricavati sia dalla CTT che dall'IRT.

Una prima analisi è stata fatta sulla percentuale di risposte corrette per avere un'idea immediata sul livello di difficoltà di ogni item secondo la CTT. Nella tabella in figura 5.26 si leggono le percentuali di risposte corrette e sbagliate per ogni domanda; si osserva che la prova è abbastanza bilanciata rispetto all'eterogeneità dei livelli di difficoltà per singola domanda in quanto non sono presenti item estremamente facili né item estremamente difficili, inoltre c'è una prevalenza di item che si attestano intorno a 50% di difficoltà. Le domande "d7", "d14" e "d20", risultano essere le più difficili, mentre le domande "d6", "d9" e "d16" le più facili. Il coefficiente di correlazione punto biseriale ci fornisce la misura del grado di coerenza di ciascun item rispetto al test preso nel suo insieme. Nella tabella in figura 5.27 osserviamo che non sono presenti valori critici del coefficiente di correlazione biseriale (valori inferiori a 0,20), sintomo che tutti gli item presentano un accettabile grado di coerenza con il resto del test; tuttavia si osservi che non sono presenti item con valori di coefficiente superiore a 0,50 pertanto i livelli di coerenza sono tendenzialmente bassi. Osservando la colonna relativa all' α di Cronbach, si rileva un valore basso del coefficiente rispetto a tutti gli item ma tuttavia accettabile; sono presenti solo due item ("d14" e "d19") escludendo i quali si otterrebbe un leggero miglioramento dell'indice. La valutazione congiunta dei due indici tuttavia suggerisce di mantenere la prova nello stato originale perché eliminando item non si ottengono vantaggi rilevanti.

Abbiamo verificato la funzionalità dei distrattori rilevando per

Proportions for each level of response:			
	0	1	logit
d1	0.4167	0.5833	0.3365
d2	0.4812	0.5188	0.0750
d3	0.5417	0.4583	-0.1671
d4	0.5958	0.4042	-0.3881
d5	0.5375	0.4625	-0.1503
d6	0.2042	0.7958	1.3605
d7	0.6604	0.3396	-0.6652
d8	0.5021	0.4979	-0.0083
d9	0.1021	0.8979	2.1743
d10	0.5979	0.4021	-0.3968
d11	0.3521	0.6479	0.6099
d12	0.3354	0.6646	0.6838
d13	0.5917	0.4083	-0.3709
d14	0.6521	0.3479	-0.6282
d15	0.3438	0.6562	0.6466
d16	0.2812	0.7188	0.9383
d17	0.2958	0.7042	0.8672
d18	0.4896	0.5104	0.0417
d19	0.5583	0.4417	-0.2344
d20	0.7229	0.2771	-0.9590

Figura 5.26: Esempio di analisi di difficoltà/facilità degli item

	α	r_{pb}
All	0.6786	-----
d1	0.6642	0.4061
d2	0.6641	0.4088
d3	0.6661	0.3915
d4	0.6598	0.4421
d5	0.6632	0.4155
d6	0.6650	0.3841
d7	0.6675	0.3735
d8	0.6658	0.3948
d9	0.6691	0.3304
d10	0.6730	0.3306
d11	0.6757	0.3012
d12	0.6542	0.4866
d13	0.6738	0.3235
d14	0.6841	0.2220
d15	0.6599	0.4391
d16	0.6582	0.4519
d17	0.6629	0.4106
d18	0.6603	0.4391
d19	0.6831	0.2435
d20	0.6725	0.3191

Figura 5.27: Coefficiente di correlazione punto-biserial e α di Cronbach di un test

ogni item le percentuali di scelta ad ogni risposta (figura 5.28, in verde sono evidenziate le risposte corrette). Si osservi come proprio le domande “d7”, “d14” e “d20”, che sono risultate essere le più difficili secondo la CTT, presentano le percentuali più elevate di scelta “E”, che corrisponde alla risposta “Non so”. La percentuale di scelta dei distrattori si distribuisce tra tutte le alternative alla risposta corretta in modo equilibrato e non sono presenti valori anomali, quindi i distrattori sono ben costruiti. Per ogni item sono stati analizzati,

20 items and 480 sample units; 0 missing values

Proportions for each level of response:

	A	B	C	D	E
d1	0.1500	0.5833	0.0583	0.0562	0.1521
d2	0.0833	0.5188	0.1833	0.0812	0.1333
d3	0.4583	0.1042	0.1167	0.0958	0.2250
d4	0.1583	0.1688	0.4042	0.0354	0.2333
d5	0.1562	0.4625	0.0812	0.0792	0.2208
d6	0.0271	0.0104	0.1229	0.7958	0.0438
d7	0.0604	0.3396	0.1438	0.0562	0.4000
d8	0.0438	0.0438	0.0271	0.4979	0.3875
d9	0.8979	0.0229	0.0479	0.0042	0.0271
d10	0.0583	0.4021	0.1146	0.1854	0.2396
d11	0.2354	0.0312	0.6479	0.0208	0.0646
d12	0.0854	0.0458	0.0792	0.6646	0.1250
d13	0.4083	0.2250	0.0833	0.1042	0.1792
d14	0.0688	0.0667	0.3479	0.1083	0.4083
d15	0.0542	0.1042	0.0542	0.6562	0.1312
d16	0.7188	0.0812	0.0646	0.0542	0.0812
d17	0.0771	0.7042	0.0854	0.0667	0.0667
d18	0.1083	0.0562	0.0708	0.5104	0.2542
d19	0.0958	0.0521	0.4417	0.0312	0.3792
d20	0.0896	0.2771	0.1875	0.0417	0.4042

Figura 5.28: Analisi dei distrattori di un test

quindi, congiuntamente i parametri di difficoltà e di discriminatività al fine di consentire una lettura grafica degli indici, nella figura 5.29 si riporta un esempio dell’analisi dei primi quattro item del questionario.

Si osservi l’andamento delle curve corrispondenti alle risposte “E”, che nel questionario equivalgono alle risposte “NON SO”, e si noti l’andamento anomalo delle curve che sta di fatto ad indicare una maggiore probabilità di risposta corretta nei soggetti con alte abilità; questo significa che soggetti con livelli di abilità alta rispondono con

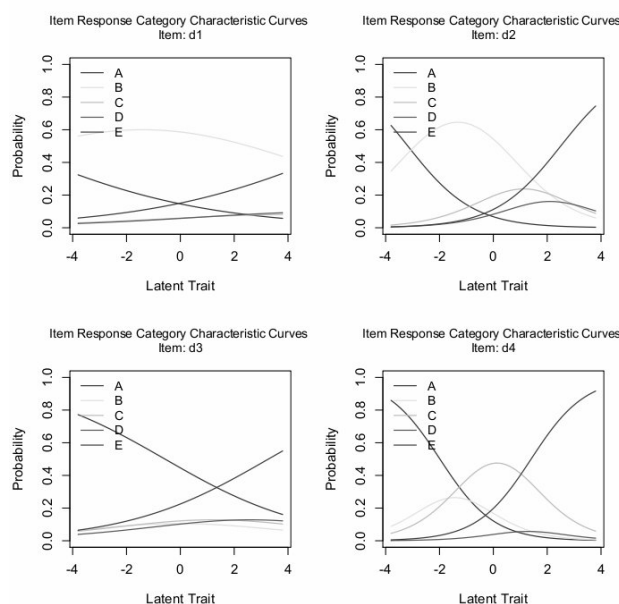


Figura 5.29: Rappresentazione grafica dell'analisi dei distrattori di un test

maggior frequenza di non sapere la risposta per non rischiare la risposta errata ed incorrere nella penalizzazione. L'analisi mette in evidenza che il questionario presenta delle carenze nella costruzione dei distrattori, che necessitano di essere modificati. Nell'item *d1* i distrattori C e D risultano, infatti, problematici in quanto attirano la risposta di soggetti con alti livelli di abilità; il distrattore A, invece, è ben costruito poiché solo i soggetti con bassi livelli di abilità sono indotti a sceglierlo; la risposta corretta, B, risulta essere molto facile. Nell'item *d4*, la risposta corretta è la C, tuttavia i distrattori B e D presentano problematiche nella rappresentazione grafica che invitano ad una correzione degli stessi. Per avere un'idea chiara di come dovrebbe essere rappresentato graficamente un item ben calibrato possiamo osservare il grafico 5.30 relativo ad un item somministrato dall'Invalsi nella prova di italiano per le classi seconde di scuola

primaria nell'anno scolastico 2012-13 Si noti la curva della risposta

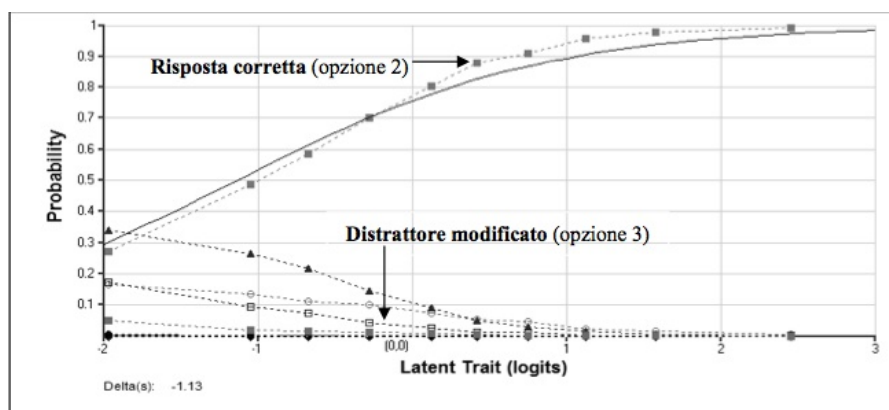


Figura 5.30: Curva caratteristica quesito A13 prova Italiano livello 2 a.s. 2012-13

corretta che risulta essere monotona crescente e che si differenzia da quelle dei distrattori, che risultano essere attrattivi per soggetti con bassi livelli di abilità.

Abbiamo eseguito l'analisi di Rasch per calcolare le difficoltà degli item, i risultati della quale sono riportati nella tabella in figura 5.31.

Il fatto che non venga riportata la difficoltà per l'item "d1" è normale, ad ogni modo, i dati di Easiness subito sotto non sono altro che i valori difficoltà moltiplicati per -1 . Questo rende facile identificare in $-0,172$ la difficoltà dell'item. Si osservi che ben il 90% di risposte corrette ha un indice di difficoltà bassissimo ($-2,155$).

Osserviamo l'andamento dei grafici che riportano le ICC per ogni item (figura 5.32) e la TIF della prova (figura 5.33). Le ICC di tutti gli item riassumono quanto descritto poc'anzi circa i livelli di difficoltà e di discriminatività di ogni item. Si può, quindi, osservare che la domanda d20, ad esempio, risulta essere particolarmente difficile in quanto anche soggetti con abilità molto elevate hanno una probabilità pari a 0,7 di rispondere in modo corretto. Le domande più problematiche dal punto di vista del potere discriminante risultano essere la d14 e la d19, in quanto lo scarto nella probabilità di rispondere

```

Results of RM estimation:

Call: RM(X = data)

Conditional log-likelihood: -4597.875
Number of iterations: 12
Number of parameters: 19

Item (Category) Difficulty Parameters (eta): with 0.95 CI:
      Estimate Std. Error lower CI upper CI
d2      0.120      0.095    -0.066    0.305
d3      0.391      0.095     0.205    0.577
d4      0.639      0.096     0.450    0.828
d5      0.372      0.095     0.186    0.558
d6     -1.293      0.114    -1.517   -1.069
d7      0.949      0.100     0.754    1.145
d8      0.213      0.095     0.028    0.398
d9     -2.155      0.149    -2.446   -1.863
d10     0.649      0.097     0.459    0.838
d11     -0.475      0.098    -0.667   -0.283
d12     -0.557      0.099    -0.751   -0.362
d13     0.620      0.096     0.431    0.808
d14     0.908      0.099     0.714    1.102
d15     -0.516      0.099    -0.709   -0.322
d16     -0.836      0.103    -1.039   -0.633
d17     -0.758      0.102    -0.958   -0.558
d18     0.157      0.095    -0.028    0.342
d19     0.466      0.095     0.280    0.653
d20     1.277      0.105     1.071    1.483

Item easiness Parameters (beta) with 0.95 CI:
      Estimate Std. Error lower CI upper CI
beta d1      0.172      0.096    -0.015    0.359
beta d2     -0.120      0.095    -0.305    0.066
beta d3     -0.391      0.095    -0.577   -0.205
beta d4     -0.639      0.096    -0.828   -0.450
beta d5     -0.372      0.095    -0.558   -0.186
beta d6      1.293      0.114     1.069     1.517
beta d7     -0.949      0.100    -1.145   -0.754
beta d8     -0.213      0.095    -0.398   -0.028
beta d9      2.155      0.149     1.863     2.446
beta d10    -0.649      0.097    -0.838   -0.459
beta d11     0.475      0.098     0.283     0.667
beta d12     0.557      0.099     0.362     0.751
beta d13    -0.620      0.096    -0.808   -0.431
beta d14    -0.908      0.099    -1.102   -0.714
beta d15     0.516      0.099     0.322     0.709
beta d16     0.836      0.103     0.633     1.039
beta d17     0.758      0.102     0.558     0.958
beta d18    -0.157      0.095    -0.342   -0.028
beta d19    -0.466      0.095    -0.653   -0.280
beta d20    -1.277      0.105    -1.483   -1.071

```

Figura 5.31: Analisi della difficoltà degli item di un test

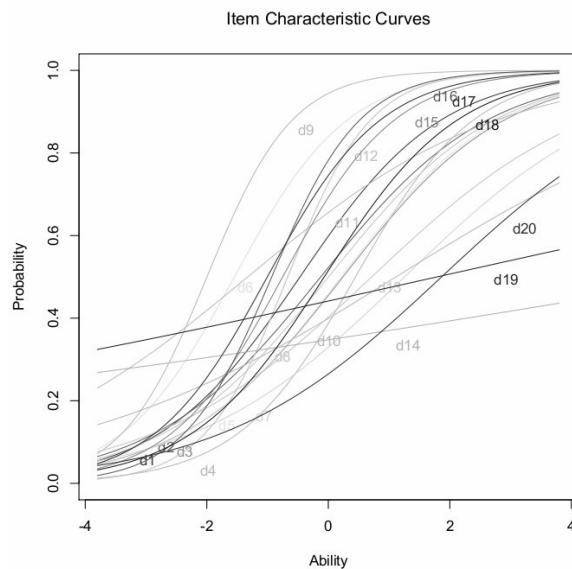


Figura 5.32: ICC degli item della prova

in modo corretto tra soggetti che presentano bassi livelli di abilità e soggetti con elevate abilità è limitato a pochissimi punti percentuali. La domanda più semplice, invece, risulta essere la d9 poiché anche soggetti leggermente al disotto del valore medio di abilità hanno una elevata probabilità di dare una risposta corretta. Dalla lettura dei TIF emerge che il questionario misura tendenzialmente abilità medio basse, pertanto il suo utilizzo in una prova a scopo selettivo necessiterebbe dell'inserimento nella prova di questionari con un livello maggiore di difficoltà e con un più alto potere discriminante. Tuttavia il questionario appare sufficientemente calibrato per cogliere tutti i livelli di abilità possibili.

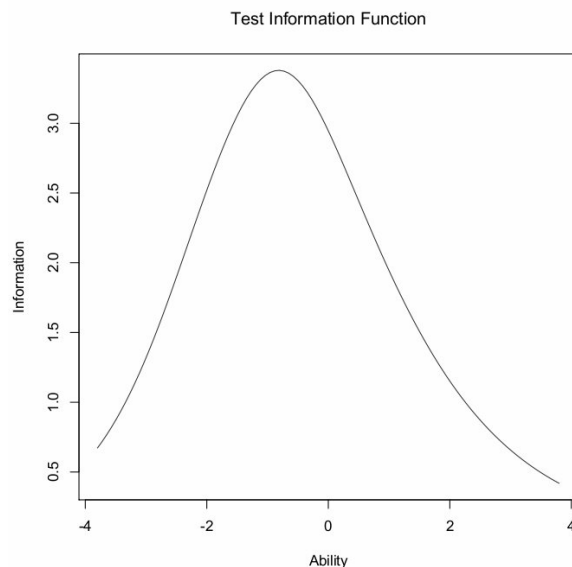


Figura 5.33: TIF degli item della prova

5.3 Classical Test Theory (CTT) e Item Response Theory (IRT): modelli a confronto

L'item analysis classica è la più diffusa in campo educativo perché può essere realizzata con semplici ed intuitivi calcoli. La CTT nasce con l'intento di analizzare la validità e l'attendibilità di un questionario e non di misurare il livello di abilità di coloro che la hanno sostenuta, di conseguenza per la misurazione di tale livello effettua esclusivamente la somma dei punteggi ottenuti.

Una limitazione emersa nell'analisi della CTT, quindi, è da ricercare nella sua incapacità di scindere le caratteristiche della prova dalle caratteristiche del candidato [Henard, 2000]. Già in Hambleton, R. K. & Swaminathan, H. [1985] rileviamo una forte critica alla CTT, in quanto gli autori osservano come, utilizzando la CTT, sia azzardato stabilire se effettivamente un item sia facile o difficile, in quanto questo dipende dalle capacità del campione di esaminandi al quale

è sottoposto un questionario. Al tempo stesso, è difficile stabilire in maniera univoca se un candidato sia più o meno abile in una specifica disciplina, in quanto la misurazione, effettuata tramite i metodi della CTT, è fortemente influenzata dal livello di difficoltà del questionario somministrato. Inoltre nella CTT, la difficoltà degli item e la capacità della persona sono poste su scale diverse [Wright and Stone, 1979]. Il modello IRT supera questi limiti e aiuta il ricercatore a costruire item svincolati dalle specifiche caratteristiche dello studente e soprattutto non influenzati nella costruzione da componenti soggettive che possono inficiarne la validità [Henard, 2000; Wright and Stone, 1979]. La IRT, infatti, come emerso dalla trattazione del modello, colloca la difficoltà dell'item e l'abilità dell'individuo su un'unica scala che è, al tempo stesso, sia "person-free" che "item-free" [Schmidt and Embretson, 2003]. Inoltre, a differenza della CTT, nella IRT, il rapporto tra il costrutto latente e il punteggio vero è non di tipo lineare, poiché la probabilità di rispondere correttamente ad un quesito è descritta da una funzione logistica [Raju *et al.*, 2002].

Se ciò che principalmente distingue la IRT dalla CTT, quindi, è il superamento del concetto di dipendenza tra soggetto e item, tuttavia esistono molte altre differenze: nella IRT l'abilità latente di un soggetto è indipendente dal tipo di test, ossia il punteggio ottenuto da un soggetto non dipende dal particolare set di domande somministratigli. È valida anche l'affermazione contraria, ossia che le caratteristiche di un item non dipendono dal particolare gruppo di soggetti rispondenti. La relazione tra abilità di un soggetto e probabilità di rispondere correttamente ad un item è esplicitata da una funzione di probabilità di forma nota e questo significa che è possibile definire un livello probabilistico per ogni item piuttosto che un livello probabilistico complessivo di tutto il test. La IRT, quindi, concentra la propria attenzione al singolo item che compone un test di valutazione, mentre la seconda pone maggiore attenzione al punteggio globale acquisito nell'intero questionario. L'IRT, quindi, valuta in che modo il soggetto esaminato corrisponde ad ogni singolo item, poiché assume che ogni candidato abbia una specifica abilità per rispondere correttamente ad ogni specifico item. Viene fissata, allora, una scala che rappresenta il livello di abilità (θ) che un candidato può possedere, quindi per ogni livello di abilità si indaga la probabilità di rispondere correttamente all'item

($P(\theta)$). La probabilità di rispondere correttamente ad uno specifico item sarà bassa in corrispondenza di bassa abilità e sarà crescente ad abilità crescenti. Una persona in possesso di una maggiore abilità rispetto ad un'altra dovrebbe avere maggiori probabilità di rispondere correttamente ad ogni domanda che possiede caratteristiche specifiche di difficoltà e discriminatività e, analogamente, se una domanda è più difficile di un'altra significa che per ogni persona la probabilità di rispondere correttamente nella seconda domanda è maggiore.

Le metodologie di stima dei risultati latenti, effettuati applicando i modelli IRT, consentono la trasformazione di indicatori categoriali in misure quantitative e continue [Wright *et al.*, 1989], ricavate come misure probabilistiche, ottenute con metodi inferenziali, che riproducano le loro componenti sistematiche, e quindi tali da prevedere i valori mancanti di uno o più individui su uno o più item [Wright *et al.*, 2000] senza introdurre né informazioni aggiuntive, né modelli causali esterni alla pura logica di valutazione dei risultati [Lovaglio, 2004a].

6 La valutazione nella Didattica Speciale

“Nessuno deve restare alle porte della città.”

(Charles Peguy)

La Didattica Speciale è quella branca della Didattica Generale che si occupa appunto dello studio degli speciali bisogni educativi che, in quanto tali, richiedono strategie metodologiche e didattiche atte a guidare nel suo percorso formativo l'alunno diversamente abile, affinché sia messo in condizione di poter sviluppare al massimo livello tutte le sue potenzialità. Oggi, più che nel passato, ogni insegnante dovrebbe avere conoscenze e competenze mirate ad affrontare gli speciali bisogni educativi dei propri studenti, poiché al di là di coloro che possiedono una specifica certificazione e che quindi usufruiscono di un servizio di sostegno alla persona, in ogni classe sono sempre più frequenti casi di alunni che presentano disturbi del comportamento, deficit dell'attenzione, fragilità emotiva, mancanza di autostima, iperattività o che provengono da realtà socio-culturali di emarginazione e di degrado.

La Didattica Speciale è prima di tutto *didattica dell'inclusione* poiché, nonostante siano passati circa quarant'anni dalla chiusura delle scuole speciali e dall'integrazione degli alunni diversamente abili nella "scuola normale", l'inserimento in classe di un bambino con bisogni educativi speciali necessita sempre di un'opera di sensibilizzazione al problema di docenti, genitori ed alunni. Troppo spesso, infatti, questi bambini sono vissuti come "diversi" e quel che è peggio è il fatto che l'approccio alla diversità viene vissuto con paura, con disagio e non come una risorsa, come una possibilità in più di educare i compagni all'accettazione dell'altro e alla sua inclusione.

Ogni giorno, ogni docente ha a che fare con la diversità che caratterizza tutti i suoi alunni. Quanti alunni uguali esistono in una classe? Nessuno! È per questo motivo che occorre riflettere sul concetto stesso di diversità. Nessun alunno, infatti, è uguale ad un altro, ognuno ha

un substrato sociale, economico e culturale che lo ha formato e che lo accompagna, ognuno ha la sua particolare esperienza di vita, di affetti, di stimoli, di conoscenze, di caratteristiche psico-fisiche, ecc. La nostra fortuna è proprio quella di vivere in un mondo di “diversi” ed è importantissimo educare i giovani al rispetto di questa diversità che ci differenzia e ci caratterizza e, soprattutto, che fa di ognuno di noi una persona speciale. Quando poi in un contesto classe, già di per sé eterogeneo, viene inserito un alunno con speciali bisogni educativi occorre prima di tutto concentrarsi sulla sua specialità, impegnandosi a conoscerla e a comprendere soprattutto quali siano i punti di forza dello studente, poiché è su questi, in una dimensione pro-attiva, che si può improntare un piano di lavoro mirato al superamento delle difficoltà che potrebbe incontrare.

Il considerare la diversità come una “risorsa”, una peculiarità che distingue l’uno dall’altro, non vuole assolutamente, però, rappresentare l’assunzione di una presa di posizione di buonismo o un non voler riconoscere quanto effettivamente siano speciali alcuni bisogni educativi. Tutt’altro, vuole essere una dar voce alla necessità nella scuola di poter contare su un corpo docente preparato ad affrontare qualsiasi tipo di bisogno educativo, indipendentemente dal livello di specialità. Tutti i docenti devono essere in grado di garantire un’adeguata formazione a tutti i loro alunni e a ciascuno di essi, ma purtroppo una specifica competenza in relazione agli speciali bisogni educativi è posseduta, ancora oggi, dai soli docenti di sostegno, ai quali spesso è delegato il compito di occuparsi esclusivamente dell’alunno diversamente abile e che non sono vissuti dai colleghi come risorsa fondamentale per l’acquisizione di nuove competenze. Le competenze in campo psicologico, pedagogico e didattico di un docente di sostegno rappresentano, infatti, una risorsa eccezionale all’interno del team-docenti, che deve essere condivisa se si vogliono garantire interventi formativi di qualità nella scuola di tutti e di ciascuno. Una prospettiva futura della scuola della qualità e dell’innovazione è, infatti, proprio quella di dare a tutti i docenti, non solo a quelli di sostegno, le competenze necessarie ad affrontare gli speciali bisogni educativi.

La Didattica Speciale va oltre la visione del rapporto di insegnamento-apprendimento come relazione di causa-effetto, per la quale si ritiene che alla spiegazione di un concetto da parte dell’insegnante corri-

sponda un apprendimento nell'allievo. Questa visione, fortemente ancorata in una didattica di tipo tradizionale, è stata largamente superata sia nell'ambito della Didattica Generale, sia in quello della Didattica Speciale. Entrambe, infatti, pongono al centro del rapporto insegnante-allievo l'oggetto culturale: in quest'ottica quando si vuole trasmettere un concetto, una conoscenza, un'abilità, o meglio un oggetto culturale, non serve spiegarlo e rispiegarlo più volte fino a quando non viene interiorizzato ma occorre modificare l'oggetto culturale stesso attraverso metodologie e strategie di insegnamento personalizzate; in questo modo è possibile riattraversare lo stesso sapere con direzioni e con scopi diversi. Con questa metodologia, promossa da Wittgenstein [McGinley and Tierney, 1989], e definita "*criss cross landscape*", l'insegnante non procede per definizioni ma per esempi, in tal modo ritorna più volte sugli stessi concetti e "costringe [l'alunno] a percorrere una vasta regione di pensiero in lungo ed in largo ed in tutte le direzioni". Apprendere è questo percorrere e ripercorrere un territorio con un "viaggio" che sarà differente per ogni individuo, poiché la personalizzazione è la trama che chi apprende costruisce intorno al tema da analizzare. Il ruolo del docente è, quindi, quello di stimolare, di dare input, di proporre esempi, di accompagnare; ma non può costruire l'apprendimento e non può trasmettere la conoscenza dato che proprio, riprendendo Wittgenstein, la molteplicità e la connessione che l'alunno compie tra gli esempi diviene la conoscenza stessa.

Il discorso della valutazione, come vedremo nei successivi paragrafi, all'interno della Didattica Speciale è molto complesso e articolato, poiché se è necessario garantire percorsi formativi adeguati all'alunno con bisogni educativi speciali, a maggior ragione occorre valutarne l'efficacia. Il Ministero della Pubblica Istruzione stabilisce che nei confronti degli alunni con minorazioni fisiche o sensoriali non si procede normalmente a valutazioni differenziate, mentre, per gli alunni in situazione di handicap psichico, la valutazione, adeguatamente differenziata, tiene conto degli obiettivi prefissati nel Piano Educativo Individualizzato (PEI). Qualora il PEI abbia individuato per l'alunno diversamente abile obiettivi formativi non riconducibili ai programmi ministeriali e ai Piani di studio previsti per i diversi tipi di scuola, il Consiglio di classe valuta comunque i risultati dell'apprendimento con l'attribuzione di giudizi o di voti relativi esclusivamente allo svol-

gimento del PEI. Tali giudizi o voti hanno valore legale al fine della prosecuzione degli studi e di essi viene fatta menzione in calce alla scheda di valutazione o alla pagella (art. 15 Ordinanza ministeriale 21 maggio 2001, n. 90). Gli alunni valutati in modo differenziato possono partecipare agli esami di qualifica professionale e di licenza di maestro d'arte. Per le prove di esame (art. 318 del Testo Unico - d.lvo 297/1994) sono predisposte per il primo e per il secondo ciclo prove equipolenti e tempi lunghi per l'effettuazione delle prove scritte. Gli alunni con minorazioni fisiche e sensoriali, quindi, secondo il Ministero non necessitano di valutazioni differenziate, contrariamente ai compagni che presentano minorazioni psicofisiche. La questione non è semplice come appare e occorrerà fare un ulteriore passo per comprendere quanto svariati possano essere gli speciali bisogni educativi e quanto complesso possa essere il discorso della valutazione di un percorso formativo personalizzato.

6.1 Gli speciali bisogni educativi

“Io voglio fare come gli altri. Io vorrei andare nei laboratori di chimica e fisica”, con questa citazione Dario Ianes apre il suo discorso sulla Speciale Normalità, mettendo in evidenza il desiderio di un alunno diversamente abile di percepirsi come gli altri: la necessità di sentirsi normale, di frequentare una scuola normale, di fare le cose normali, come ed insieme a tutti gli altri compagni normali. La vera qualità dell'integrazione si costruisce quotidianamente, declinando in vari modi una condizione che lo stesso Ianes definisce di *“speciale normalità*. “La speciale normalità è una condizione di sintesi tra specialità e normalità, che le contiene e le supera entrambe: la normalità si arricchisce di specificità non comuni, di peculiarità, di risposte tecniche particolari; la specialità va ad arricchire le normali prassi, ne penetra le fibre più profonde e le modifica, le rende più inclusive e rispondenti ai bisogni”[Ianes, 2006].

La speciale normalità individua una grande varietà di situazioni con le quali la scuola si trova a confrontarsi giornalmente: alunni normali che possiedono molti tratti di specialità, alunni speciali con i bisogni essenziali della normalità. Troviamo la speciale normalità quando analizziamo cosa sta accadendo in generale sul versante dei

bisogni che presentano gli studenti: nell'eterogeneità che caratterizza le classi troviamo alunni con bisogni educativi speciali dovuti alla presenza di deficit sensoriali o psicofisici, a disturbi della sfera relazionale, a mancanza di autostima, ad una diversa etnia, a vissuti di degrado e se ne potrebbero elencare ancora un'infinità.

Il concetto di normalità può essere declinato in diverse direzioni. Può avere un significato puramente statistico, per indicare la maggior frequenza di un certo evento oppure l'attesa che un certo evento abbia luogo. Così le convenzioni che si stabiliscono tra gli esseri umani si fondano sull'aspettativa che gli individui si comportino nel modo che socialmente è ritenuto normale.

Il *Bisogno Educativo Speciale* è caratterizzato da un coesistere dei concetti di normalità e di specialità. Prima di tutto, infatti, c'è il diritto di tutti i cittadini ad una formazione della persona che garantisca pari dignità sociale ed uguaglianza. È un diritto alla normalità sancito dalla Costituzione che è inalienabile per tutti: è il diritto che ha ogni alunno di sentirsi uguale agli altri ed, al tempo stesso, è il bisogno che tutti abbiamo di uno sviluppo e di una funzionalità il più possibile normale e rispondente alle normali richieste della vita sociale. In questa essenziale normalità, però, troviamo a confrontarci anche con la specialità, la differenza e la peculiarità di alcuni alunni che presentano situazioni deficitarie nella struttura e nelle funzioni corporee, oppure nell'apprendimento, nelle relazioni, in alcuni aspetti psicologici, o ancora, a livello familiare. È sempre la Costituzione a mettere in evidenza che è obbligo della Repubblica *rimuovere gli ostacoli ... che, limitando, di fatto, la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana*.

L'inserimento e l'integrazione di allievi con bisogni educativi speciali all'interno delle classi curricolari sono l'elemento caratterizzante della didattica speciale del sistema scolastico italiano, che sottolinea l'importanza della dimensione sociale dell'apprendimento, perché imparare in contesti e situazioni "normali" fa sentire l'allievo che presenta difficoltà uguale agli altri, e gli offre maggiori possibilità di sviluppo sociale e cognitivo.

Il primo compito di un insegnante nel momento in cui si trova a cominciare una nuova esperienza di insegnamento è sicuramente quello di imparare a conoscere gli alunni che ha davanti, attraverso

un'osservazione mirata e sistemica, poiché prima di programmare un percorso educativo occorre comprendere le peculiarità degli stili di apprendimento, le caratteristiche relazionali, affettive, emotive e conoscere le conoscenze, abilità e competenze possedute alle quali ancorare le nuove conoscenze. Quando un alunno presenta bisogni educativi speciali la fase dell'osservazione è di estrema importanza poiché è necessario cogliere nelle minime sfumature tutte le potenzialità che possano permettere di costruire un percorso formativo valido. Non è sempre facile o immediato individuare in allievi con bisogni speciali, quali sono gli ambiti dei loro possibili interessi, le cose che preferiscono fare, le persone con cui riescono a comunicare più facilmente. Tuttavia riuscire a stabilire un collegamento motivazionale con le attività previste per il loro percorso individualizzato è un elemento chiave. Attivare un buon processo didattico vuol dire riuscire a coinvolgere l'allievo con bisogni speciali in attività che lo gratifichino, e che lo rendano protagonista del proprio apprendimento all'interno del percorso didattico che egli compie insieme a tutti i suoi compagni di classe. Riuscire a creare un contesto sociale accogliente e rassicurante, dove sia possibile per l'allievo stabilire relazioni sociali significative con i propri compagni di classe e con i propri insegnanti, è un elemento fondamentale per la crescita e lo sviluppo delle potenzialità e capacità dell'alunno stesso. Decisivo, in questo contesto, è il ruolo che i compagni di classe possono giocare, diventando di volta in volta tutor del compagno con bisogni speciali e partner dell'insegnante disciplinare in un'esperienza di *cooperative learning*. Un'esperienza didattica di questo tipo è altamente formativa non solo per l'alunno che presenta una disabilità, ma per tutti i compagni che, attraverso un'esperienza concreta di inclusione, maturano un maggior senso di responsabilità e di rispetto per la diversità.

La progettazione del Piano Educativo Individualizzato (PEI) per un alunno diversamente abile deve fondarsi su un lavoro di team che vede coinvolti i docenti curricolari, il docente di sostegno, l'assistente educativa, l'èquipe psico-pedagogica della Asl di appartenenza. La pluralità degli interventi che ruota intorno al bambino deve essere garanzia di un approccio multidisciplinare valido poiché fondato sulla condivisione del percorso formativo. In relazione al livello di gravità del deficit la programmazione individualizzata può prevedere

una semplificazione degli obiettivi programmati per l'intera classe o l'individuazione di obiettivi specifici per il singolo. La diagnosi di disabilità può essere un buon punto di partenza per effettuare indagini circa la tipologia di deficit segnalato, anche se si tratta comunque di linee generali, poiché anche i soggetti che presentano la stessa problematica si differenziano molto per le loro specificità. "L'ICF, International Classification of Functioning, Disability and Health, è la classificazione del funzionamento, della disabilità e della salute, promossa dall'Organizzazione Mondiale della Sanità"[Ianes, 2006]. Questa classificazione completa la classificazione ICD-10 che contiene informazioni sulla diagnosi e sull'eziologia delle diverse patologie. L'ICF, invece, si riferisce alle caratteristiche che presenta il soggetto che vive una specifica situazione patologica. L'ICD-10 e l'ICF usati in modo complementare forniscono un quadro globale della malattia e del funzionamento dell'individuo. L'ICF è strutturato in 4 principali componenti:

- Funzioni corporee
- Strutture corporee
- Attività (in relazione a capacità e performance)
- Fattori ambientali

Il funzionamento e la disabilità sono viste come una complessa interazione tra le condizioni di salute dell'individuo e l'interazione con i fattori ambientali e personali. La classificazione utilizza questi spazi come dinamici e in interazione, non come statici. Siccome la disabilità è data dalle modalità con cui il soggetto interagisce con l'ambiente, l'ICF è applicabile a tutte le persone, anche quelle in perfetta salute. Esso è stato creato appositamente per essere utilizzato a livello internazionale e interculturale con obiettivi molto diversificati, da un utilizzo clinico fino a studi epidemiologici e di politica della salute. Essendo neutrale, il linguaggio utilizzato è stato specificato fin nei minimi dettagli per chiarire al meglio il significato della terminologia utilizzata nel contesto specifico della valutazione del funzionamento. L'accertamento della condizione di disabile viene definito da apposita certificazione rilasciata da una commissione medica collegiale. La certificazione fa riferimento, per l'identificazione della disabilità, ai parametri definiti dalla OMS. Tuttavia l'utilizzo dell'ICF non è ancora diffuso a livello nazionale ma utilizzato solo in alcune regioni in fase

sperimentale, questo determina una disomogeneità nelle certificazioni. Spesso, infatti, si tratta di certificazioni che riportano la sola diagnosi clinica e che non danno ulteriori informazioni circa il funzionamento della disabilità in relazione al caso specifico certificato. È facile leggere ancora certificazioni del tipo “soggetto affetto da trisomia 21; ritardo della sfera cognitiva: medio; ritardo del linguaggio: medio; ritardo psicofisico: medio”. Una certificazione di questo tipo dice ben poco sulle caratteristiche specifiche del soggetto e non esiste in campo formativo una ricetta ad hoc che applicata ad uno specifico deficit conduca a risultati predeterminati. In campo formativo, come abbiamo visto anche sopra, non si può parlare di causa-effetto, tanto meno quando si ha a che fare con bisogni educativi speciali.

In base a quanto previsto dalla Legge finanziaria del 2002 n.289 art. 35 comma 7 e dal decreto applicativo n.185 del 2006, l'individuazione dell'alunno disabile per l'integrazione scolastica avviene attraverso l'accertamento da parte di un Collegio istituito dall'ASL. Di questo fanno parte un neuropsichiatra della Azienda Ospedaliera, uno Psicologo e un Assistente Sociale dell'ASL. Il Collegio si esprime previa presentazione da parte del genitore della domanda con la quale chiede che il proprio figlio venga sottoposto all'accertamento dello stato di disabilità per l'integrazione scolastica. Il certificato deve riportare la diagnosi clinica codificata preferibilmente secondo l'ICD 10 e l'indicazione se trattasi di patologia stabilizzata o progressiva. La relazione clinica deve evidenziare lo stato di gravità della disabilità, il quadro funzionale sintetico del minore con indicazione dei test utilizzati (eventualmente allegando copia dei test stessi) e dei risultati ottenuti, e descrivere le maggiori problematiche nelle aree: cognitiva e neuropsicologica, sensoriale, motorio-prassica, affettivo-relazionale e comportamentale, comunicativa e linguistica, delle autonomie personali e sociali. Secondo le nuove modalità il verbale di accertamento firmato da tutti i componenti il Collegio è rilasciato direttamente al genitore. Successivamente al verbale gli operatori che hanno in carico la situazione predispongono la diagnosi funzionale che, come il verbale di accertamento, verrà rilasciata al genitore. È quindi il familiare a consegnare alla scuola il verbale e la diagnosi funzionale. In questo modo può essere avviato l'iter necessario per garantire l'inserimento, l'integrazione e l'inclusione dell'alunno con speciali bisogni educativi.

Il primo passo da compiere è quello di lavorare in modo dettagliato alla redazione della *Diagnosi Funzionale*, redatta, come stabilito dal DPR 24 febbraio 1994, dall'équipe multidisciplinare (neuropsichiatra, psicologo, terapeuta della riabilitazione, assistente sociale...) di struttura sanitaria pubblica o privata accreditata che ha in carico il minore. La stessa è aggiornata al passaggio di ciclo scolastico ovvero in qualsiasi altro momento vi siano cambiamenti significativi del quadro con conseguente necessità di modifiche alle forme di sostegno alla persona. La Diagnosi Funzionale è uno strumento di estrema importanza per la scuola poiché deve consentire una conoscenza dettagliata della situazione di partenza della persona disabile, con particolare riferimento all'evidenziazione dei livelli di capacità (competenze e abilità possedute), limiti (deficit di competenze e abilità, presenza di problematiche comportamentali, psicologiche e relazionali) e potenzialità (competenze e abilità allo stato latente che possono essere stimulate) presenti nel soggetto. L'équipe multidisciplinare ha il compito di definire gli obiettivi di intervento perseguibili per il superamento dei limiti o l'attivazione delle potenzialità, dando indicazioni per l'applicazione più idonea delle procedure di intervento e delle attività necessarie al perseguimento degli obiettivi stessi.

Alla Diagnosi Funzionale segue la strutturazione del *Profilo Dinamico Funzionale*, redatto dall'unità multidisciplinare che elabora la diagnosi funzionale, dai docenti curricolari e dagli insegnanti specializzati della scuola, con la collaborazione dei familiari dell'alunno, dopo un primo periodo di inserimento scolastico. Il Profilo Dinamico Funzionale indica il prevedibile livello di sviluppo dell'alunno che presenta una situazione di deficit; in esso vengono esplicate le caratteristiche fisiche, psichiche e sociali e affettive della persona e messe in rilievo le difficoltà di apprendimento conseguenti alla situazione di deficit, le possibilità di recupero e le capacità possedute, che primariamente devono essere sostenute, sollecitate e progressivamente rafforzate e sviluppate.

Segue, in fine, la stesura del *Piano Educativo Individualizzato*, il documento nel quale vengono descritti gli interventi integrati ed equilibrati tra di loro, predisposti per l'alunno diversamente abile, per un determinato periodo di tempo, ai fini della realizzazione del diritto all'educazione e all'istruzione. Alla definizione del PEI provvedono

congiuntamente gli operatori delle ASL e, per ciascun grado di scuola, personale insegnante curricolare e di sostegno della scuola, con la partecipazione dell'insegnante operatore psico-pedagogico individuato secondo criteri stabiliti dal Ministero della Pubblica Istruzione e con la collaborazione dei genitori dell'alunno. Il PEI è il progetto che contiene la sintesi coordinata dei 3 progetti:

- educativo-didattico;
- riabilitativo;
- di socializzazione;

nonché le forme di integrazione tra attività scolastiche ed extra-scolastiche (art.5 DPR 24/02/94).

Gli interventi educativi devono, se possibile, essere attinenti alla programmazione educativa prevista per l'intera classe e individualizzati in relazione alle specifiche competenze dell'alunno, prevedendo semplificazioni della programmazione e interventi mirati di recupero e sostegno. Solo nei casi di gravi deficit psicofisici è opportuno programmare un intervento con obiettivi differenziati che devono essere soprattutto centrati sull'acquisizione di comportamenti socialmente accettabili e di un certo livello di autonomia, nonché sul raggiungimento di competenze relazionali che consentano all'alunno diversamente abile una proficua integrazione nella vita sociale della comunità. Il Piano Educativo Individualizzato deve essere indirizzato verso un Progetto di Vita, pertanto deve essere costruito in relazione alle attitudini, agli interessi, alla ricerca di identità e alla consapevolezza. È necessario, prima di tutto, che si aiuti il ragazzo diversamente abile, naturalmente in relazione all'età ed alle proprie potenzialità, a guardare dentro di sé, a verificare i propri interessi, a prendere consapevolezza delle proprie capacità, delle proprie competenze. Il Piano Educativo Individualizzato deve possedere una forte connotazione orientativa e la valutazione del processo formativo dovrà essere funzionale a questa finalità. Fino a qualche anno fa la dimensione orientativa del Piano Educativo Individualizzato inteso come Progetto di Vita era scarsamente considerato. Si era sostanzialmente convinti che fosse impossibile uscire da un percorso obbligato e, in un certo senso, imposto dallo stato stesso di minorazione e si era costretti in una visione valutativa che sottolineava le incapacità e le mancanze di un soggetto, piuttosto che evidenziare i suoi punti di forza. Questa visione, oggi, fortunatamente è stata

messa da parte poiché si è compresa l'importanza di incoraggiare gli alunni che necessitano di bisogni educativi speciali ad effettuare delle scelte personali in relazione alla propria vita. Il ragazzo non può sicuramente compiere questo cammino da solo, ma occorre garantirgli un adeguato sostegno orientativo, che lo accompagni soprattutto nelle fasi di passaggio da un ordine di scuola all'altro, in quanto sono le fasi più delicate dell'integrazione scolastica.

L'attuazione del PEI deve essere sempre costantemente monitorata poiché si tratta di uno strumento aperto che può essere continuamente aggiornato e modificato in relazione alle esigenze emergenti in corso di formazione. Questo comporta una verifica sistematica del processo di insegnamento-apprendimento ed una valutazione dettagliata dei progressi compiuti e delle difficoltà riscontrate.

6.2 Valutare tutti, valutare ciascuno

La valutazione è una fase decisiva del processo di integrazione poiché consente il controllo e il monitoraggio dell'efficacia degli interventi educativi individualizzati. La valutazione è un momento essenziale nel percorso formativo di un alunno con bisogni educativi speciali e deve far parte integrante del PEI, assumendo soprattutto una funzione regolativa all'interno del processo di insegnamento-apprendimento, poiché deve consentire di adeguare il percorso alle esigenze che emergono in itinere.

La valutazione è prima di tutto un diritto che deve essere riconosciuto all'alunno diversamente abile, poiché, soprattutto al termine del ciclo di studi, è giusto che, come i suoi compagni, consegua una certificazione delle competenze acquisite, spendibile per favorire l'ingresso nel mondo del lavoro o per proseguire con gli studi universitari. La C.M. 14.03.08, n. 32 prevede che, per l'esame di stato al termine della scuola secondaria di primo grado, per gli alunni con diagnosi specialistica di dislessia o di altri disturbi specifici dell'apprendimento, sia previsto l'impiego di strumenti compensativi oltre all'assegnazione di maggior tempo per lo svolgimento delle prove e che comunque questi alunni dovranno sostenere tutte le prove scritte. Per gli alunni con disabilità psicofisica, invece, in relazione alla gravità del caso, la valutazione finale viene operata sulla base del piano educativo indivi-

dualizzato, al fine di valutare il processo formativo in rapporto alle potenzialità dei singoli ed ai livelli di apprendimento e di autonomia iniziali. Per l'esame di Stato i docenti preposti al sostegno degli alunni con disabilità partecipano a pieno titolo alle operazioni connesse alla predisposizione e correzione delle prove e alla formulazione del giudizio globale. Gli alunni possono svolgere una o più prove differenziate, in linea con gli interventi educativo-didattici attuati sulla base del Piano Educativo Individualizzato (PEI), secondo le previsioni contenute nell'art. 318 del Testo Unico. Tali prove hanno comunque valore equipollente ai fini della valutazione dell'alunno. Anche in questo caso la commissione può fissare un tempo differenziato per l'effettuazione delle prove. Per quanto riguarda gli alunni con disabilità visiva, è consentito l'utilizzo del supporto digitale della prova nazionale da convertire nelle forme previste, impiegando le strumentazioni in uso (braille, lettura digitale, sintetizzatore vocale). Nel caso di esito negativo delle prove di esame, per gli alunni con disabilità è comunque possibile rilasciare un attestato che certifichi i crediti formativi acquisiti; tale attestato è titolo per l'iscrizione e la frequenza di classi successive, ai soli fini del riconoscimento di crediti formativi, e concorre ad assicurare la frequenza negli istituti di istruzione secondaria superiore. Nel diploma di licenza non deve essere fatta menzione delle prove differenziate sostenute dagli alunni con disabilità. L'Ordinanza Ministeriale 15.03.2007, n.26, relativa all'esame di stato al termine della scuola secondaria di secondo grado, non si differenzia molto da quanto appena citato per la scuola di primo grado. L'Ordinanza stabilisce, infatti, che la commissione d'esame, sulla base della documentazione fornita dal consiglio di classe, relativa alle attività svolte, alle valutazioni effettuate e all'assistenza prevista per l'autonomia e la comunicazione, predispone per gli alunni diversamente abili prove equipollenti a quelle assegnate agli altri candidati, che possono consistere nell'utilizzo di mezzi tecnici o in modalità diverse di esecuzione, o ancora nella scelta di contenuti culturali e professionali differenti. In ogni caso le prove equipollenti devono consentire di verificare che il candidato abbia raggiunto una preparazione culturale e professionale idonea per il rilascio del diploma attestante il superamento dell'esame. Per la predisposizione delle prove d'esame, la commissione d'esame può avvalersi di personale esperto e per il loro svolgimento la stessa

si avvale, se necessario, dei medesimi operatori che hanno seguito l'alunno durante l'anno scolastico. I testi della prima e della seconda prova scritta sono trasmessi dal Ministero anche tradotti in linguaggio braille, ove vi siano candidati non vedenti. Per i candidati che non conoscono il linguaggio braille la Commissione può provvedere alla trascrizione del testo ministeriale su supporto informatico, mediante scanner, autorizzando anche l'utilizzazione di altri ausili idonei, abitualmente utilizzati dal candidato nel corso dell'attività scolastica ordinaria. La Commissione può, inoltre, concedere tempi più lunghi nell'effettuazione delle prove scritte e grafiche e del colloquio, previsti dal comma 3 dell'articolo 16 della legge n.104 del 3/2/1992, ma la dilatazione temporale non può, di norma, comportare un maggior numero di giorni rispetto a quello stabilito dal calendario degli esami. Solo in casi eccezionali, la Commissione, tenuto conto della gravità del deficit, della relazione del consiglio di classe, delle modalità di svolgimento delle prove durante l'anno scolastico, può deliberare lo svolgimento di prove scritte equipollenti in un numero maggiore di giorni. I candidati che hanno seguito un percorso didattico differenziato e sono stati valutati dal consiglio di classe con l'attribuzione di voti e di un credito scolastico relativi unicamente allo svolgimento di tale piano possono sostenere prove differenziate, coerenti con il percorso svolto finalizzate solo al rilascio dell'attestazione di cui all'art.13 del D.P.R. n.323/1998. I testi delle prove scritte sono elaborati dalle commissioni, sulla base della documentazione fornita dal consiglio di classe. Per un discorso di rispetto della persona e di privacy, il riferimento all'effettuazione delle prove differenziate va indicato solo nell'attestazione e non nei quadri affissi all'albo dell'istituto.

Nell'atto valutativo non si può, quindi, prescindere dalla peculiarità e specificità che caratterizzano l'alunno, pertanto, soprattutto nei casi più gravi, non si può pensare di valutare l'alunno in riferimento ad un livello ritenuto standard per tutti gli allievi, poiché in tal modo si negherebbero i bisogni specifici della persona. Il termine di confronto dei progressi di apprendimento del ragazzo diversamente abile non deve essere, quindi, il risultato conseguito dai compagni negli stessi apprendimenti, ma sono il cambiamento delle sue prestazioni comportamentali nel tempo, rispetto al ritmo personale. La valutazione, dunque, deve essere funzionale al Piano Educativo Individualizzato,

deve essere uno strumento che consente agli insegnanti, agli operatori, ai genitori, al ragazzo stesso di monitorare l'andamento del processo educativo, di individuare le finalità e gli obiettivi maggiormente congruenti, di cogliere progressi, stasi, difficoltà nell'integrazione.

Sarebbe riduttivo, quindi, considerare quale funzione della valutazione solo quella sommativa che mira all'accertamento dei risultati, in quanto significherebbe confinare la valutazione all'interno di quello che Petracca [2011] definisce "paradigma della decisione", che utilizza la valutazione esclusivamente per stabilire se un alunno possa essere promosso, se possa conseguire un titolo di studio, se i suoi risultati rientrino nella media, ecc. La valutazione ha anche, e soprattutto, come abbiamo ribadito più volte, una funzione formativa che consiste nel regolamentare il processo di insegnamento-apprendimento. La valutazione, allora, deve riferirsi contemporaneamente al processo e agli obiettivi che l'insegnante a monte ha stabilito: ecco il legame tra valutazione e programmazione, tra valutazione ed azione di insegnamento. In quest'ottica la valutazione diviene il nucleo della relazione insegnamento-apprendimento e assume le connotazioni di valutazione formativa e, al contempo, orientativa, diagnostica, prognostica e continua. La valutazione deve assumere, quindi, la funzione di strumento regolatore dei processi e deve avere un carattere promozionale, per dare all'alunno la percezione esatta dei propri punti forti prima di sottolineare quelli più deboli; di elemento formativo, perché, dando all'alunno la percezione del punto in cui è arrivato, gli consente di capire, all'interno del processo formativo, che cosa deve fare e che cosa deve chiedere ancora alla scuola; di mezzo orientativo, per mettere il ragazzo in condizione di rendersi consapevole dei propri punti di forza e di debolezza e, in relazione ad essi, di acquistare più capacità di scelta e di decisione. Se si garantisce all'alunno diversamente abile, e comunque a tutti gli alunni, questo tipo di valutazione si contribuisce alla costruzione consapevole di una cittadinanza attiva per la realizzazione del proprio progetto di vita.

"Gli handicappati non hanno bisogno di benevolenza- ha spiegato Andrea Canevaro durante il Convegno Internazionale su *La Qualità dell'Integrazione Scolastica*, svoltosi a Rimini nel 2007 - ma non hanno neanche bisogno di scoprire che sono handicappati, perché di questo qualche sospetto già lo avevano; hanno invece bisogno di essere valu-

tati per orientarsi. È un'esigenza speciale tutta loro? Non sembra. E non sembra neanche che di fronte a questo problema gli handicappati vivano una diversità, se si può dire così, diversa dai ragazzi normali. Entrambi hanno la necessità di valorizzare le loro identità e quindi anche le loro diversità".

Questa visione delle finalità della valutazione si inquadra nel campo di quella che viene definita valutazione dinamica, una metodologia valutativa mirata alla conoscenza dei potenziali dell'alunno diversamente abile e dei suoi processi di apprendimento, piuttosto che sulla verifica degli apprendimenti conseguiti. La valutazione dinamica consiste nella valutazione delle strategie del pensiero, della percezione, dell'apprendimento e del problem solving attraverso un processo attivo di insegnamento. In questo processo l'insegnante assume un ruolo da mediatore, poiché guida lo studente alla registrazione sistematica dei progressi conseguiti. La valutazione dinamica riveste un ruolo molto importante per la definizione del Piano Educativo Individualizzato perché consente di valutare le strategie cognitive utilizzate dall'alunno per risolvere un problema, le funzioni cognitive carenti responsabili dell'insuccesso e quanto esse siano modificabili, l'influenza dei fattori emotivi, affettivi e motivazionali nell'apprendimento, le potenzialità di acquisizione di nuove conoscenze attraverso il transfert ad altri contesti delle competenze acquisite. La valutazione di un alunno diversamente abile deve, quindi, essere mirata e sistematica e, al di là della funzione sommativa che gli consente di acquisire una certificazione di competenze, deve documentare tutto il percorso formativo. Sarebbe opportuno, quindi, che i docenti dei diversi ordini di scuola si coordinassero maggiormente al fine di costruire in itinere, durante tutto l'arco di permanenza nella scuola dell'alunno, un curriculum che documenti le attitudini e gli interessi che ha manifestato e le competenze maturate. Il curriculum dovrebbe inoltre prevedere un'area predisposta per l'autovalutazione in modo da consentire allo studente stesso di auto-valutarsi, di conoscersi, di verificare la fattibilità dei suoi progetti.

Un ultimo aspetto che occorre prendere in considerazione, ma assolutamente non meno importante, è quello dell'attribuzione che lo studente diversamente abile dà al suo successo o insuccesso di fronte ad una prova di verifica. Se un bambino che ha fallito un compito,

attribuisce questo fallimento al fatto di non aver studiato adeguatamente o di non aver compreso pienamente la consegna, attribuisce alla frustrazione una causa instabile, che non ha eccessive ripercussioni sui successivi apprendimenti. Se, invece, percepisce questo fallimento come un non essere all'altezza della situazione e, quindi, di non essere in grado di ottenere un risultato migliore, attribuisce alla frustrazione una causa stabile, che influenzerà negativamente i futuri apprendimenti, in quanto lo studente si sentirà incapace di apprendere e di migliorarsi. In questo caso è assolutamente necessario che l'insegnante, piuttosto che evidenziare l'insuccesso, enfatizzi le abilità che l'alunno possiede, al fine di eliminare la possibilità che l'allievo viva il dramma della consapevolezza dell'incapacità di apprendere. Se l'alunno, infatti, arriva a percepirsi come incapace, perderà fiducia e stima di sé e si chiuderà alla possibilità di superare l'ostacolo e di acquisire nuovi apprendimenti. Uno dei compiti della valutazione è anche quello di far acquisire allo studente il senso del limite con un atteggiamento di accettazione positiva, dando un senso di realtà alle difficoltà possedute che non deve essere vissuto come una situazione frustrante. L'accettazione consapevole dei propri limiti, infatti, può incrementare maggiormente la ricerca delle proprie potenzialità e motivare all'elaborazione di un progetto che promuova l'autonomia personale e sociale, un Progetto di Vita rivolto alla promozione di una cittadinanza attiva. L'individuo, infatti, nell'incontro con il proprio limite prova i sentimenti della rassegnazione e della rinuncia, ma a partire da essi può essere supportato alla sperimentazione del sentimento del coraggio e dell'osare. Crescere significa anche divenire lentamente capaci di tollerare i propri limiti, imparando a tollerare la frustrazione e a muoversi nella direzione del migliorarsi. L'impegno ad investire in coraggio e fiducia non riguarda, ovviamente, solo il soggetto diversamente abile, ma coinvolge tutti: dalle figure genitoriali, agli insegnanti, agli specialisti, che condividono la sua personale storia, e più in generale interessa chiunque di noi poiché tutti devono fare i conti con i propri limiti. La disponibilità a riconoscere ed accettare il limite, consente di concretizzare un realistico progetto di vita. Quando le possibilità personali sono limitate, occorre muovere qualche passo in più sul versante sociale, scolastico, sanitario, ambientale, allestendo o potenziando servizi comunitari, formali ed informali, coordinan-

doli in rete, stimolando nuove risorse e offrendo supporti personali e materiali, sempre nell'ottica di considerare il soggetto disabile e la sua famiglia come protagonisti attivi dell'intervento. Ogni individuo possiede, infatti, delle risorse personali che devono essere investite proficuamente all'interno di un progetto personale e professionale. La conoscenza, da parte dell'individuo, delle proprie risorse potenziali è un mezzo essenziale per appropriarsi del proprio avvenire. Ognuno deve, quindi, essere messo in condizione, per investire in coraggio e fiducia, di poter effettuare il bilancio delle proprie competenze, al fine di comprendere pienamente quali siano le proprie risorse, gli interessi, le motivazioni e su queste deve costruire un concreto Progetto di Vita. Anche l'alunno diversamente abile, come già abbiamo anticipato per tutti gli alunni, ha la necessità di una valutazione che miri a verificare non solo ciò che sa, ma ciò che sa fare con ciò che sa, ossia una valutazione fondata su una prestazione reale e adeguata dell'apprendimento. Si tratta, quindi, di coinvolgere il ragazzo diversamente abile in compiti che richiedono di applicare le conoscenze nelle esperienze del mondo reale; l'enfasi è posta sulla riflessione, sulla comprensione e sulla crescita, piuttosto che su risposte fondate sul ricordo di fatti isolati. Una valutazione, per essere realmente autentica, deve proporsi l'obiettivo di valutare il progresso dell'apprendimento nel tempo. Come tale, deve essere continuativa e individualizzata e deve far riferimento a un progetto personalizzato di apprendimento.

6.3 la valorizzazione della persona per mezzo del Computer Adaptive Testing (CAT)

I modelli IRT e la tecnologia hanno consentito l'applicazione di un approccio al testing educativo che è noto come Computerized Adaptive Testing (CAT). Il CAT è uno strumento di misurazione che viene somministrato al computer e che sfrutta la proprietà dell'invarianza per creare un algoritmo in base al quale i soggetti ricevono un test "su misura" per il loro livello di abilità o di tratto: in pratica, il computer si adatta alle capacità del soggetto proponendo item la cui difficoltà va progressivamente a coincidere con il livello del soggetto del soggetto nel costruito.

Come vedremo meglio più avanti, il computer somministra al soggetto uno o due item di prova e, in base alle risposte date dallo studente, stima il presumibile livello di abilità o di tratto del soggetto, selezionando la somministrazione dei successivi item per verificare se il livello stimato sia corretto, aggiustando il tiro ogni volta che il soggetto risponde correttamente o sbaglia, fino a che la stima non diviene sufficientemente stabile e precisa.

Per organizzare somministrazioni di tipo CAT occorre prima di tutto realizzare una *item-bank* in cui siano contenuti un gran numero di item, ognuno con un ben preciso livello di difficoltà, ossia corrispondente ad un certo livello di preparazione dello studente. Ogni item deve possedere un suo "peso specifico" che lo identifichi in relazione al livello di difficoltà. Se, ad esempio, un item ha un peso di determinato peso p_j allora gli studenti con un'abilità uguale o superiore al peso dell'item avranno una probabilità maggior di rispondere correttamente rispetto agli studenti con livelli di abilità inferiore. In relazione alla risposta data dallo studente il sistema stima il livello di abilità e "pesca" la domanda successiva in funzione dell'abilità stimata. Il sistema continuerà con la stessa procedura fino a quando le risposte dello studente si saranno stabilizzate ad uno specifico livello di abilità.

Supponiamo che un docente voglia utilizzare il CAT per la valutazione degli apprendimenti in una classe in cui sia inserito anche un alunno con speciali bisogni educativi. Avrà sicuramente bisogno di una *item-bank* in cui siano contenuti un gran numero di item, ognuno con un ben preciso livello di difficoltà, ossia corrispondente ad un certo livello di preparazione dello studente. Gli item possono variare dai più elementari ai più complessi e devono essere inseriti anche item semplificati in relazione all'abilità dello studente diversamente abile, che devono comunque essere graduati per livello di difficoltà crescente. Supponiamo che ogni singolo computer selezioni una domanda per ogni studente e che all'alunno diversamente abile venga proposta una domanda che abbia una difficoltà pari a 2; se per lo studente questo item è facile avrà più del 50% di probabilità di rispondere correttamente ad esso, in quanto ha evidentemente una preparazione superiore a 2, se invece risulterà difficile avrà una probabilità inferiore al 50% di rispondere correttamente, di conseguenza la sua preparazione è

inferiore a 2, mentre se avrà una probabilità del 50% di rispondere correttamente all'item allora la sua preparazione è uguale a 2. In relazione alla risposta dello studente il computer comincia a stimare il livello di abilità che possiede: "minore di 2" se la risposta è sbagliata, "uguale o maggiore di 2" se la risposta è corretta. Gli item successivi vengono, quindi, selezionati e somministrati in base al loro livello di difficoltà in modo da individuare con più precisione l'effettivo livello di preparazione del soggetto. Il computer procederà in modo adattivo, adattandosi appunto alle risposte del soggetto e mettendolo in condizione di rispondere correttamente in relazione al suo livello di abilità. Quando la risposta sarà errata, il computer ne selezionerà, per la risposta successiva, una più semplice, al contrario tenterà con una più difficile, fino a quando il soggetto comincerà a rispondere sempre correttamente perché il computer avrà individuato il livello di abilità dello studente. Questo procedimento verrà effettuato per tutti gli alunni ai quali viene somministrato il test.

Il CAT permette di ottenere una stima adeguata del livello del soggetto nel costrutto solo se è stato somministrato almeno un numero minimo di item, che a seconda dell'ampiezza concettuale del costrutto può variare da 10 a 20, ovviamente se ogni sfaccettatura del costrutto è stata adeguatamente indagata. Il numero minimo di item, inoltre, consente alla procedura di avere una buona validità: i soggetti con livelli particolarmente alti o bassi nel costrutto, infatti, tenderanno ad avere test più brevi, dato che rispondono correttamente o sbagliano praticamente tutti gli item che vengono loro proposti.

Il CAT ha grandi potenzialità in campo di istruzione sia perché è veloce da somministrare, sia perché è oggettivo e imparziale. Tuttavia affinché funzioni al meglio occorre costruire una item-bank adeguata, ma ciò comporta un lavoro impegnativo e preciso, che non può essere condotto da un solo docente ma dalla collaborazione di docenti differenti. Gli item, infatti, devono innanzitutto coprire adeguatamente il dominio di contenuto, essere formulati nella maniera corretta e devono presentare alternative di risposta che appaiano equi-probabili a chi non conosce la risposta corretta. Tutti gli item della item-bank, inoltre, devono essere validati attraverso una pre-somministrazione ad almeno 500 soggetti, affinché la stima dei parametri degli item (difficoltà, discriminatività, guessing e carelessness) sia attendibile.

Tale stima dei parametri può essere realizzata direttamente dal software, e in base alla proprietà dell'invarianza essa sarà indipendente dal campione utilizzato. Tutto questo processo prende il nome di *calibrazione del test*. Lo scopo è quello di avere item che coprano in modo capillare tutti i possibili livelli di difficoltà degli item sulla scala dei logit, dato che questo corrisponde alla possibilità di poter stimare nel modo più raffinato possibile il livello nel costrutto dei soggetti. È quindi necessario che chi sviluppa gli item sia sufficientemente esperto della procedura di assessment da generare domande che coprano una gamma di difficoltà il più ampia possibile e nel modo più differenziato possibile. Uno dei vantaggi del CAT è l'aggiornamento automatico dei livelli di difficoltà degli item di soggetto in soggetto.

Nell'era delle Tecnologie dell'Informazione e della Comunicazione il CAT ha grandi potenzialità. La selezione adattiva degli item, scelti in base alla stima ottenuta del livello di abilità del soggetto, consente una somministrazione rapida e fornisce un feedback immediato. Tutti gli studenti, indipendentemente dal loro livello di abilità, si trovano in condizione di essere gratificati dalla compilazione del questionario poiché, quando questo si stabilizza al livello dello studente, ognuno si trova in condizione di rispondere correttamente. Per una somministrazione di CAT è necessario, però, che tutti i soggetti abbiano sufficiente familiarità con l'uso del computer.

L'uso del CAT per la valutazione degli apprendimenti in generale e nello specifico anche in presenza di soggetti diversamente abili, consente esperienze di integrazione e dà pari opportunità a tutti gli studenti di vivere in modo gratificante la valutazione. Ovviamente, sarà poi compito dell'insegnante quello di dare un feedback della prova allo studente indicandogli, in relazione all'analisi delle risposte date, quali siano i suoi punti di forza e di debolezza.

Conclusioni e prospettive future

“Le attività di valutazione, pur fondate sulla base dei dati quantitativi assunti con la verifica e la misurazione, coinvolgono una gamma più ampia di operazioni di analisi, di considerazioni e di interpretazioni qualitative dei fattori che hanno prodotto gli esiti registrati con gli strumenti di rilevazione usati.”

(Domenici)

“La via di mezzo è sempre la migliore: ogni eccesso conduce alla rovina.”

(Plauto)

Conclusioni

Nella trattazione di questo lavoro abbiamo cercato di cogliere gli aspetti più significativi di un tema tanto vasto e complesso qual è quello della valutazione. Dopo aver introdotto etimologicamente il termine stesso per comprenderne meglio il significato, abbiamo tracciato il quadro della valutazione nel contesto scolastico italiano, raffrontandolo con quello di altri sistemi educativi europei. Abbiamo, quindi, analizzato le metodologie, le strategie e gli strumenti utilizzati nel campo della valutazione quantitativa e in quello della valutazione qualitativa allo scopo di identificare una metodologia di valutazione più organica e complessa che possa garantire validità ed attendibilità alla valutazione, oggi che la scuola è chiamata ad educare alla complessità e all'imprevedibilità. Le analisi che abbiamo condotto hanno confermato che occorre progettare e riprogettare continuamente iniziative, momenti e finalità, poiché in campo educativo la progettazione di percorsi formativi ha un carattere dinamico che richiede flessibilità e capacità di gestire le interrelazioni, il cambiamento e l'imprevisto. Gli strumenti utilizzati, d'altra parte, e le connesse istruzioni per l'uso, non possono essere precostituiti una volta per sempre, in una sorta di a-priori metodologico, ma vanno selezionati di volta in volta e calibrati sulle specifiche situazioni (luoghi, oggetti, procedure e, soprattutto, persone).

Ci siamo, ancora, soffermati a riflettere sulla valutazione degli alunni con bisogni educativi speciali per scoprire che in realtà ogni alunno, per il fatto di essere una persona unica ed irripetibile, presenta una sua specificità che la valutazione deve valorizzare. È proprio questo, a nostro avviso, il cardine della valutazione: valutare significa attribuire un valore nel senso che la scuola deve valorizzare tutti e ciascuno. Occorre, infatti, dare valore a tutto il processo di insegnamento-apprendimento per valorizzare la formazione della persona. L'attività valutativa dovrà considerarsi, per tanto, come un'attività di pensiero produttivo, poiché deve essere in grado, appunto, di espletare la sua funzione regolativa del processo formativo che consente di gestire il cambiamento e l'imprevisto; come un'attività di pensiero comparativo poiché nasce sempre da un confronto tra ciò che si programma e ciò che si realizza, tra aspettative e risultati, tra un livello di competenze iniziale ed uno finale, tra diversi stili di apprendimento; come un'attività di pensiero critico perché comporta un'accurata analisi, riflessione, progettazione e continua verifica da parte dell'insegnante che deve contestualizzare l'esperienza formativa e da parte dello studente che deve costruire in prima persona la propria crescita culturale e sociale; e ancora come attività di pensiero ermeneutico poiché l'educazione e la formazione non sono avulse dal contesto storico, politico, economico, culturale e sociale. I sistemi scolastici, infatti, sono chiamati ad interpretare la realtà e a progettare interventi finalizzati alla formazione di uomini che siano in grado di vivere la loro esperienza di vita con razionalità, criticità e creatività. Un altro aspetto essenziale, emerso dalla trattazione, è che valutare non significa emettere un giudizio sulla persona. L'insegnante, infatti, è chiamato a valutare l'efficacia del suo intervento educativo attraverso la verifica costante del processo. Questa verifica si basa sulla misurazione degli apprendimenti, sull'osservazione delle modificazioni dei comportamenti, sull'analisi delle competenze, coniugando processi e prodotti, ma sarebbe deontologicamente riprovevole se comportasse un giudicare la persona. Se poi, come tutt'oggi talvolta accade, venisse usata come "arma" per esercitare la propria autorità sugli studenti, denoterebbe, di fatto, un'incapacità ad affermare le proprie competenze professionali. L'autorità non va confusa con l'autoritarismo: contenere, limitare, dare delle regole, definire dei confini non significa reprimere ma, al contrario, fungere da guida rispettando

il desiderio e il bisogno di individualità e di auto-affermazione dell'altro. In questa ottica potremmo affermare che non c'è libertà, non c'è conquista di individualità senza un reale contenimento, senza una guida autorevole che ci insegni a pensare, a fare scelte, ad agire nel rispetto degli altri e di se stessi. La scuola attraverso la valutazione, la co-valutazione e l'autovalutazione promuove gradualmente i processi di identificazione e di differenziazione, che consentono la maturazione della persona nella sua visione olistica e la rendono un soggetto autonomo e libero. Il docente non può favorire questa crescita se impone la sua autorità, ma deve essere invece responsabile e autorevole, poiché solo in questo modo consente allo studente di separarsi, individuarsi e differenziarsi. Gli stessi conflitti circa i ruoli e le responsabilità possono avere un valore costruttivo e formativo, poiché consentono ai ragazzi di confrontarsi e di diventare consapevoli delle differenze, nonché di assumere posizioni attive di responsabilità. Il docente deve saper essere autorevole, non autoritario, e la valutazione è uno degli strumenti che può consentirgli di esprimere la propria autorevolezza, più che l'autorità in sé. I docenti che fanno leva sulla sola autorità, senza dare sostegno alle proprie motivazioni, senza guidare l'alunno a comprendere il significato di un giudizio espresso in una prova di verifica, o che tendono ad imporsi, interrompono l'ascolto, facendo sentire gli studenti giudicati. La valutazione, per diventare un momento esperienziale fondante per la formazione della propria personalità, deve essere condivisa e deve guidare gradualmente lo studente ad auto valutarsi, a conoscersi, a comprendere i propri punti di forza e di debolezza, a trovare le strategie per superare le difficoltà e quelle per migliorarsi, in una presa di coscienza del fatto che ognuno di noi è artefice della costruzione del proprio sapere e del proprio essere.

Il Prof. Bertagna [2006] ha iniziato il suo ultimo libro sul tema della valutazione, *Valutare tutti Valutare ciascuno*, con una citazione tratta dai Racconti chassidici di Milan [1994] filosofo, teologo e pedagogista austriaco, che riporto integralmente di seguito poiché è stata motivo di sollecitazione ad una riflessione profonda sul significato della valutazione. *“L'uomo è come un albero. Se ti metti di fronte ad un albero e lo guardi incessantemente per vedere se cresce e di quanto sia cresciuto, non vedrai nulla. Ma curalo in ogni momento, liberalo dal superfluo e tienilo pulito (...) ed esso, a tempo debito, comincerà a crescere. Lo stesso vale*

anche per l'uomo: l'unica cosa che gli serve è superare lacci e impedimenti, non mancherà di svilupparsi e crescere. Ma è sbagliato esaminarlo in continuazione per scoprire quanto sia cresciuto". Il ruolo dell'educatore, dell'insegnante, del formatore è primariamente quello di mettere ogni discente in condizione di crescere, nel suo processo di formazione, in relazione alle proprie attitudini e, soprattutto, che si coltivi in lui lo stupore, la meraviglia, quell'attimo di sospensione del pensiero che ti lascia rapito di fronte a qualcosa di nuovo e che genera la scintilla, il desiderio di scoprire, di percorrere nuove strade della conoscenza. Il segreto di un apprendimento duraturo e significativo è, quindi, da ricercare nella capacità del docente di rendere il sapere un oggetto di desiderio, di saper motivare l'alunno alla scoperta di nuove conoscenze. Bisogna insegnare agli adolescenti l'arte di ammirare: in ciò consiste, a mio parere, uno dei segreti dell'educazione.

La valutazione degli apprendimenti, pertanto, deve essere vissuta dall'alunno come un momento positivo del processo di insegnamento-apprendimento, evitando sia che nel discente si generino ansie e preoccupazioni, sia che questo momento, cruciale ed indispensabile, venga vissuto come una situazione nella quale si ritiene di dover dimostrare la propria efficienza rispetto ad un apprendimento specifico. È necessario, in quest'ottica, che il docente metta l'alunno in condizione di saper auto valutare il proprio processo di apprendimento. La valutazione non si deve, quindi, limitare ad una verifica di prestazione ma, al contrario, deve essere il momento in cui l'alunno stesso prende coscienza del proprio livello di abilità, conoscenze e competenze maturate. Chi apprende, in questo modo, è reso protagonista del suo stesso processo di apprendimento, poiché è messo in condizione di comprendere il livello di padronanza raggiunto, di conoscere i propri punti di debolezza e, soprattutto, i propri punti di forza e le potenzialità su cui poter far leva per superare le difficoltà incontrate per gestire i propri limiti.

Il sentiero della conoscenza non sempre è facile da percorrere e richiede un impegno intenso e perseverante, quindi è necessario che l'alunno si approcci a questo cammino sentendosi pienamente protagonista del suo processo formativo. Il docente, al tempo stesso, deve assolutamente superare una visione puramente trasmissiva nel suo modo di insegnare e deve comprendere che la formazione, essendo

un processo di insegnamento-apprendimento, rende docenti e discenti co-protagonisti. L'apprendimento, d'altro canto, ci accompagna lungo tutto il corso della vita e molto spesso chi insegna scopre di aver imparato qualcosa di nuovo proprio dai suoi stessi allievi. Il docente non deve riempire la testa dei propri studenti di nozioni e concetti, ma deve guidarli a costruire le proprie conoscenze ed è proprio la valutazione ad avere un ruolo centrale in questo processo. Così si rivolgeva Socrate al suo allievo Teeteto, ben duemilacinquecento anni fa, rendendolo protagonista dei suoi stessi apprendimenti. Il suo insegnamento è ancora attualissimo: *"(...) abbi fiducia in te stesso (...) metticela tutta (...) perché non sei vuoto ma gravido (...) la mia arte maieutica (...) rassomiglia a quella delle levatrici, ma ne differisce in questo, che opera su gli uomini e non su le donne, e provvede alle anime partorienti e non ai corpi. E la più grande capacità sua è, ch'io riesco, per essa, a discernere sicuramente se fantasma e menzogna partorisce l'anima del giovane, oppure se cosa vitale e reale. Poiché questo ho di comune con le levatrici, che anch'io sono sterile: non genero sapienza (...) Io sono dunque, in me, tutt'altro che sapiente, né da me è venuta fuori alcuna sapiente scoperta che sia generazione del mio animo; quelli invece che amano stare con me, se pur da principio appariscano, alcuni di loro, del tutto ignoranti, tutti quanti poi, seguitando a frequentare la mia compagnia, ne ricavano, purché il dio glielo permetta, straordinario profitto (...) ed è chiaro che da me non hanno imparato nulla, bensì proprio e solo da se stessi molte cose e belle hanno trovato e generato; ma d'averli aiutati a generare, questo sì, il merito spetta al dio e a me. (...) se cercherai di diventare gravido di altre cose, dopo di queste, e ci riuscirai, per merito di questa ricerca, sarai pieno di cose migliori"*.

Tutti i sistemi scolastici a livello mondiale stanno affrontando una fase critica dovuta al crollo di tradizionali certezze sulle quali si sono costruiti ed improntati i sistemi stessi. La crisi è dovuta soprattutto al fatto che, in seguito al repentino cambiamento che ha caratterizzato la società dovuto alla globalizzazione, all'innovazione tecnologica, all'incontro di culture diverse, alla messa in discussione di valori fondanti, il mondo della scuola si è trovato disorientato ed ha avvertito la necessità di mettersi nuovamente in gioco al fine di escogitare le metodologie e le strategie più idonee a formare i cittadini del futuro. Alcuni paesi, per reagire a questa situazione, hanno attuato importanti politiche di investimento nel settore educativo, promuovendo la ricerca

al fine di elevare la qualità dell'istruzione. L'Italia, purtroppo, rientra ancora in quel gruppo di paesi che, pur denunciando una situazione difficile della scuola, dovuta a forti squilibri tra costi e benefici e ad una modernizzazione più di facciata che non reale, piuttosto che investire concretamente nell'istruzione rivolge esortazioni moralistiche ad insegnanti ed allievi perché mostrino un maggior impegno.

La valutazione è un'operazione complessa che mette in gioco diverse variabili: la qualità e la quantità delle conoscenze acquisite, la capacità di trasferire le competenze acquisite ad altri contesti per la soluzione di problemi diversi, la corrispondenza tra le aspettative degli insegnanti e le risposte degli allievi, gli aspetti emotivi, affettivi e relazionali, i fattori ambientali, le condizioni socio-culturali e molto altro ancora. Ne consegue che la qualità del giudizio di valutazione non risulta mai, sia in contesti diversi, sia in uno stesso contesto ma con docenti diversi, uniforme e costante. Questa difformità nelle procedure valutative non è da inquadrare solo in una problematica di ordine didattico, legata al fatto che determina valutazioni molto eterogenee tra loro e una non conformità della certificazione di competenze, ma esplicita un problema ancora più complesso, che è quello di una disomogeneità nell'interpretazione complessiva dell'azione formativa. La scuola ha l'obbligo di garantire a tutti e a ciascuno, con un intervento intenzionale e sistematico, partendo dall'orizzonte di esperienze ed interessi, un percorso formativo che consenta di acquisire tutti gli strumenti necessari per divenire un cittadino attivo, responsabile, consapevole e in grado di affrontare l'imprevisto e l'imprevedibile che caratterizzano la società del cambiamento e dell'innovazione. I sistemi scolastici spesso assumono, però, un atteggiamento anacronistico, manifestando resistenza al cambiamento.

Le Nuove Indicazioni per il Curricolo del primo ciclo di istruzione, risultano essere uno strumento utile alla scuola dell'autonomia per muovere i primi passi da protagonista nel campo dell'educazione. Per la prima volta alla scuola è data la possibilità di costruire autonomamente i propri percorsi formativi in relazione alle esigenze del contesto ed ai relativi vincoli e risorse. La valutazione, come momento fondamentale del processo formativo, viene parzialmente svincolata da procedure valutative di tipo top down e diviene lo strumento che consente di costruire un percorso più rispondente alle esigenze

delle diverse realtà locali. Per quanto attiene alla competenza delle scuole circa le nuove forme di valutazione, lo schema di regolamento evidenzia come, in ragione dell'autonomia scolastica, appartiene alle scuole e ai docenti la scelta delle modalità per la valutazione in itinere e per la predisposizione del documento di valutazione. Resta fermo, naturalmente, l'obbligo di rispettare tre elementi fondamentali derivanti dalle norme generali del sistema di istruzione nazionale: la valutazione degli apprendimenti relativi alle discipline previste dai piani di studio; la valutazione del comportamento dell'alunno; l'espressione delle valutazioni, periodiche e finali, con voto in decimi per ciascuna disciplina. Una prima riflessione va fatta sull'affermazione che sostiene che il voto in decimi valorizza la valutazione in funzione del miglioramento dei livelli di apprendimento. Si ritiene che la valutazione, soprattutto nella Scuola Primaria e nella Scuola secondaria di I grado, non possa limitarsi ad una valutazione con sola funzione certificativa poiché è necessario, in relazione alla fascia di età e di sviluppo degli alunni di questo ordine di scuola, affermare la funzione regolativa della valutazione stessa, la sola in grado di consentire, sulla base delle informazioni raccolte in itinere nel processo di insegnamento-apprendimento, un continuo adeguamento delle proposte di formazione alle reali esigenze degli alunni ed ai traguardi programmati. In tal senso deve essere privilegiata una valutazione formativa che prenda in considerazione non solo i risultati raggiunti dal singolo alunno ma l'intero percorso formativo dello stesso, i passi realizzati, la motivazione, l'impegno, le aree di sviluppo prossimale. L'importanza di una valutazione in itinere sta nel consentire di procedere per passi, correggendo o modificando strumenti e modalità di apprendimento sia da parte dei docenti che da parte degli alunni e, al tempo stesso nell'avere una valenza diagnostica in quanto consente di orientare docenti ed alunni verso gli sviluppi successivi del processo di apprendimento. Una valutazione di tipo "ideografico", che consenta un confronto diacronico fra la situazione iniziale e la situazione finale dell'alunno, mirata al recupero dell'originalità del singolo e delle singole situazioni è quella che si ritiene più idonea in questa fase dello sviluppo della persona. Questo tipo di valutazione, infatti, consente di effettuare una riflessione profonda dei risultati conseguiti progressivamente dal singolo studente, di rilevare carenze

da superare ma soprattutto le potenzialità sulle quali ipotizzare gli sviluppi successivi del processo di insegnamento-apprendimento.

La condivisione, con le famiglie e con gli alunni stessi, delle metodologie di valutazione utilizzate dal team docente e delle osservazioni rilevate consente di promuovere nell'alunno una maggior motivazione ad apprendere e una miglior capacità di autovalutazione e, al tempo stesso, promuove un senso di co-protagonismo e co-partecipazione di insegnanti-genitori-alunni al processo di insegnamento-apprendimento. Le polemiche mosse dai docenti manifestano contrarietà alla al sistema di valutazione in scala numerica decimale poiché i docenti ritengono che questa tipologia di valutazione non valorizzi l'alunno e sia penalizzante, soprattutto per gli alunni con maggiori difficoltà di apprendimento, in quanto potrebbe determinare un calo della motivazione ad apprendere e generare forme negative di competizione tra gli alunni. La valutazione in decimi, peraltro, non è assolutamente in sintonia con la personalizzazione dell'azione educativa e formativa. Grande importanza viene, invece, data al giudizio analitico globale di maturazione raggiunto dall'alunno poiché è ritenuta l'unica forma di giudizio che consenta di evidenziare tutto il processo formativo. I docenti a più voci hanno ribadito che la sostituzione del voto al giudizio sintetico, espresso con la scala non sufficiente, sufficiente, buono, distinto, ottimo, sostanzialmente non apporta alcuna innovazione alla valutazione poiché, come abbiamo evidenziato, si tratta comunque di operare con scale ordinali che non sono in grado di quantificare esattamente gli intervalli tra un voto e l'altro, la valutazione su scala, sia essa di tipo numerico che nominale, presuppone inoltre solo una misurazione dei risultati conseguiti e non l'espressione del reale percorso formativo. Si capisce, insomma, che c'è ben altro dietro il voto e che un "6" non è, di per sé, più chiaro, oggettivo e trasparente di un "sufficiente", se non si chiarisce cosa stiamo valutando (conoscenze, abilità, competenze, ...), sulla base di quali criteri (progresso dell'allievo, standard di riferimento, soglie assolute ecc.), utilizzando quali strumenti di verifica (osservazioni, prove tradizionali, prove strutturate, prodotti degli allievi ecc.). E ancora, se la valutazione finale è limitata esclusivamente all'attribuzione di un voto nato dalla media aritmetica dei risultati conseguiti nelle diverse prove di verifica, possiamo affermare che quel "6" conseguito da uno studente che abbia

progressivamente ottenuto le valutazioni parziali "4", "6", e "8", abbia lo stesso significato del compagno che ha ottenuto "8", "6" e "4", o addirittura di chi abbia sempre preso "6"? La valutazione degli allievi è un processo delicato, che interpella innanzi tutto chi la compie, è una ricerca continua di un giusto equilibrio tra promozione, cura, attenzione ai bisogni degli allievi e valorizzazione dei loro impegni, capacità, meriti. Tutto questo sta con fatica dentro un voto e non è certo rappresentato e descritto dalla pura media aritmetica dei voti assegnati in corso d'anno. Chiedere ai docenti di esprimere con un voto il livello di apprendimento, la maturazione globale della persona e la socialità è contrario alla cultura pedagogica della valutazione, che ha prestato sempre di più attenzione ai processi, mettendoli in relazione con gli specifici contesti nei quali si esplicano. La sola valutazione che si può ritenere valida in questa fascia d'età è quella di tipo formativo, finalizzata ad intervenire sui processi e ad eliminare gli ostacoli che possano condizionare l'apprendimento; di tipo globale poiché investe gli aspetti dei comportamenti, degli atteggiamenti, delle relazioni, delle conoscenze e delle competenze; di tipo individualizzato perché ogni alunno è valutato in relazione al proprio personale percorso formativo.

Una piccola riflessione per quanto riguarda l'uso del voto in condotta; non si nega l'importanza di valutare il comportamento dell'alunno ma, se già non è lecito fare medie aritmetiche tra le diverse tipologie di prove di verifica degli apprendimenti e tra discipline che presentano un substrato epistemologico diverso, figuriamoci se un comportamento può far media con un apprendimento. Questa è un'affermazione dell'ex Ministro Gelmini rilasciata in un'intervista: "Il voto in condotta farà media perché sappiamo che l'aumento di episodi di bullismo preoccupa molto genitori e insegnanti. Si torna, dunque, a una scuola del rigore che fa del comportamento un elemento significativo per formare la personalità dei ragazzi". Una scuola punitiva? A parte la già contestata valenza pedagogica della media con altre discipline, quello che sconcerta è che la valutazione del comportamento è inquadrata in una funzione punitiva e non di promozione alla convivenza civile ed alla legalità. La scuola ha primariamente il compito di promuovere negli alunni emozioni, motivazioni, atteggiamenti, conoscenze, relazionalità che, in una visione olistica della persona, andrà a valutare come sviluppo personale e sociale, ossia come comportamento. Nella

valutazione del comportamento devono essere attivati, incentivati ed osservati in ogni singolo alunno aspetti identificativi del suo processo di crescita, quali: la conoscenza e controllo del proprio corpo e delle proprie emozioni; l'accettazione e rispetto di sé, degli altri, delle cose e dell'ambiente in cui vive; il rispetto delle regole convenute; la capacità di ascolto attivo; il desiderio di aiutare i compagni a superare le difficoltà; la capacità di relazionarsi con compagni ed adulti; l'interesse nel valorizzare le potenzialità del gruppo di lavoro, assumendo un ruolo positivo all'interno del gruppo stesso; l'autonomia personale e nello svolgimento del proprio lavoro; la capacità di portare a termine gli impegni assunti o affidati, con senso di responsabilità. La scuola ha il compito di promuovere interventi mirati all'acquisizione di questi comportamenti, senza i quali l'alunno non sarà mai in grado di essere un cittadino attivo. Atti scorretti, fenomeni di bullismo, violazione dei diritti altrui hanno bisogno di interventi di rieducazione e di sensibilizzazione; non è sufficiente bocciare chi ha sbagliato, poiché la bocciatura non garantisce che l'alunno maturi comportamenti diversi e si migliori come persona, al limite può pensare che è opportuno comportarsi meglio per non essere bocciato di nuovo, ma non risolveremo mai il problema se non adottiamo in qualità di educatori tutte le strategie più idonee affinché le regole della convivenza civile siano interiorizzate e condivise, non solo imposte. Persino il carcere ha una funzione educativa e non punitiva, e alla scuola diamo solo strumenti per punire? Il Ministero si dovrebbe preoccupare di dare prima di tutto sostegno adeguato ai docenti affinché questi siano in grado, in stretta collaborazione con le famiglie, di promuovere la socialità, il rispetto, la riflessione, la legalità. Non serve usare il voto per intimare o punire, occorre aiutare i ragazzi a divenire persone migliori, guidandoli a riflettere profondamente sul significato di libertà, di rispetto, di accettazione, di integrazione, di democrazia.

Un ultimo appunto occorre fare sul fatto che la normativa vigente prevede che la certificazione delle competenze con una valutazione espressa in decimi. Le competenze, come abbiamo visto, forniscono un insieme chiaro ed integrato di dimensioni rispetto alle quali possono essere descritte e misurate performance. Un modello di competenze ben progettato, quindi, fornisce i legami fra i comportamenti, le capacità e gli attributi richiesti all'individuo in specifici compiti ri-

chiesti. Per valutare le competenze occorre effettuare una valutazione autentica che consenta di verificare ciò che l'alunno sa fare con ciò che sa, come si avvicina alla risoluzione di problemi concreti della vita reale, in che modo riesce ad affrontare l'imprevisto, se ha maturato un pensiero riflessivo, critico e creativo, se è in grado di condividere le proprie esperienze con altri, confrontandosi e cooperando con il gruppo. A tutto questo, e ad altri aspetti non meno importanti che nel corso della trattazione abbiamo evidenziato come caratteristici della complessità della persona, non si può attribuire un voto in decimi. Dire che un alunno è competente "6", o "8", o "10" non ha alcun significato pedagogico e didattico e non rappresenta un sistema valido di certificazione delle competenze. Ci auguriamo che il Ministero riveda questa posizione, affidandosi alla competenza di esperti del settore, confrontandosi proficuamente con gli altri sistemi educativi, investendo maggiormente nella ricerca e che attuando pratiche di ascolto ai docenti, agli alunni e ai genitori, che ogni giorno vivono da protagonisti la scuola. Nel frattempo la scuola dell'autonomia può sperimentare forme di valutazione che consentano di documentare le competenze possedute dagli alunni attraverso l'uso di rubriche di valutazione. Agendo così, oltre a trovare una soluzione di applicazione immediata per soddisfare le richieste ministeriali, si dà agli alunni la possibilità di acquisire consapevolezza del proprio percorso formativo. Gli alunni, infatti, se si sentiranno coinvolti in prima persona, vivranno esperienze di co-valutazione valide, che consentiranno loro di giungere ad una capacità di autovalutazione, poiché gradualmente comprenderanno il proprio stile di apprendimento e le proprie capacità e scopriranno i propri limiti. In tal modo saranno, infatti, maggiormente motivati a migliorarsi e affronteranno meglio i possibili insuccessi, in quanto loro stessi hanno partecipato a fissare i parametri della valutazione. Con il coinvolgimento e la partecipazione degli alunni si assiste all'emersione delle diversità e alla loro valorizzazione, poiché ognuno diventa consapevole del proprio ruolo e del ruolo degli altri. È utile, inoltre, che la valutazione sia condivisa e discussa con i genitori in tutte le sue fasi, perché in questo modo si costruirà un protocollo di lavoro condiviso scuola-famiglia. Infatti, attraverso la condivisione, sia l'alunno, sia la famiglia sono messi a conoscenza del percorso formativo, dello step in cui si colloca

il proprio figlio nei diversi ambiti di apprendimento e negli aspetti educativi e sociali, dei passi successivi che egli deve compiere e degli interventi mirati di supporto ed aiuto che possono essere necessari per superare difficoltà o per potenziare eccellenze. La competenza si può solo descrivere se non si utilizzano strumenti scientifici di misurazione, in quanto essendo un fenomeno complesso, ricco di elementi taciti oltre che espliciti, ha una strutturazione ologrammatica che abbraccia una moltitudine di dimensioni.

La valutazione nel nostro percorso è andata ad assumere il significato di interpretare e di valorizzare, ha esplicitato la sua funzione di riflessione sul contesto e sugli aspetti specifici che lo caratterizzano, connotandosi come momento essenziale e propulsore dell'attività formativa. Nella relazione interpersonale della valutazione, sia il docente che lo studente scoprono, interpretano e costruiscono significati. Sfuggendo alla logica oggettivistica della misurazione degli apprendimenti, la valutazione assume oggi un carattere formativo complesso, mirato al tempo stesso alla visione globale del processo e all'unicità del singolo. La valutazione diventa, quindi, un processo di ricerca applicata che ha il suo focus nel confronto con le complessità generate da un processo di interazioni e relazioni che si influenzano reciprocamente. In questo senso diventa fonte di apprendimento sia per il docente che per lo studente, poiché determina un continuo rinvio dalla dimensione progettuale a quella più specificatamente valutativa e viceversa, che, come abbiamo visto, devono essere fortemente condivise tra tutti i protagonisti del processo formativo, affinché l'apprendimento sia significativo. Lo sfondo concettuale nel quale si muove la valutazione oggi è, infatti, quello del paradigma della complessità, nel quale, come afferma E. Morin, la valutazione stessa diventa un costrutto complesso in interazione con il fenomeno valutato, finalizzato a rinforzare le dinamiche più che a stabilire un controllo. Per questo motivo appare necessario, in primis, che l'insegnante maturi competenze specifiche in campo valutativo, acquisendo tutte le possibili metodologie e strategie che la ricerca pedagogica e didattica ha, nel correre del tempo, sperimentato; così da riuscire, in un proficuo lavoro di team, a trovare le interconnessioni tra le diverse teorie in modo da strutturare le modalità più idonee a fare della valutazione una concreta esperienza di interpretazione e valorizzazione della complessità. Per non perdere

mai di vista il principio della complessità, che è necessario porre alla base di qualsiasi valido discorso in campo educativo, poiché complesso è l'essere umano, complessa è la società in cui esplica se stesso, complessa è la sua formazione e complesso il compito di valutarla, ci piace concludere questa trattazione ricordando le parole scritte da Morin [2001] ne "I sette saperi necessari all'educazione del futuro", le quali, oltre a dare un breve assaggio dell'idea di complessità, ci rendono consapevoli del fatto che la formazione ed, in seno ad essa, la valutazione devono rispettare e valorizzare questa complessità: *"L'essere umano è esso stesso nel contempo uno e molteplice. Ogni essere, anche il più chiuso nella più banale delle vite, costituisce in se stesso un cosmo. Porta in sé le proprie molteplicità interiori, le proprie personalità virtuali, un'infinità di personaggi chimerici, una poliesistenza nel reale e nell'immaginario, nel sonno e nella veglia, nell'obbedienza e nella trasgressione, nell'ostentato e nel segreto. Porta in sé brulichii larvali in caverne e in abissi insondabili. Ciascuno contiene in sé galassie di sogni e fantasmi, slanci inappagati di desideri e di amori, abissi di infelicità, immensità di glaciale indifferenza, conflagrazioni di astri in fiamme, irruzioni di odio, smarrimenti stupidi, lampi di lucidità, burrasche dementi..."* [1].

Prospettive e sviluppi futuri

I-Value: il prototipo di una piattaforma per la valutazione degli apprendimenti

Per costruire un questionario per la valutazione degli apprendimenti, le funzioni informative dei singoli item e del test più in generale risultano necessarie in quanto consentono di individuare con correttezza gli opportuni item in funzione del loro potere informativo. Nell'esperienza valutativa quotidiana dei docenti, tuttavia, è impensabile che ogni singolo docente possa disporre delle conoscenze teoriche e pratiche utili ad applicare correttamente la IRT. Questa constatazione ha portato alla ricerca di una soluzione possibile che consenta di fornire ogni docente di uno strumento, di uso semplice ed intuitivo, da poter gestire, a livello di utente finale, per strutturare questionari per l'apprendimento validi ed attendibili.

In questo lavoro di ricerca, attualmente, è in fase di definizione un prototipo che dovrebbe consentire, ai docenti che aderiranno liberamente al progetto, di poter costruire questionari di verifica degli apprendimenti su un'apposita piattaforma e-learning e, attraverso un processo di simulazione virtuale della compilazione, effettuare una procedura corretta di calibrazione, basata su modello IRT, che consenta di stimare i parametri degli item, del questionario e dei soggetti. Il costrutto latente che viene misurato dal test concettualmente può assumere valori infiniti, ma nella pratica quotidiana questa scala viene ristretta ad un range più circoscritto, che presenta media 0 e varianza 1. Sulla stessa scala vengono espressi anche i parametri degli item, così da rendere comune per item e soggetti la scala di misura.

Quando si costruisce un test, tuttavia, non si conoscono a-priori né i parametri degli item né la quantità di abilità che il soggetto che risponde al test possiede, di conseguenza, non è automaticamente possibile esprimere sulla stessa scala entrambe questi valori. Il ricercatore, nella prassi, procede ad una pre-somministrazione del questionario ad un campione di almeno 500 soggetti, codifica le risposte fornite dai soggetti agli item, e applica su questa codifica una procedura matematica atta a creare una scala che sia unica per quella combinazione di item. Solo dopo aver ricondotto item e soggetti sulla stessa scala, si possono stimare i valori dei parametri degli item e dei soggetti attraverso il paradigma iterativo, dove la stima dei diversi parametri termina solo quando si ottengono tra un ciclo e l'altro stime stabili e più precise. La procedura di test calibration, o calibrazione del test, viene quindi effettuata per creare una banca di item su cui sono stati stimati i parametri e i livelli di abilità per il campione di calibrazione, così da poter estendere, in virtù del principio di invarianza della misurazione, gli stessi item a nuovi soggetti (per poter stimare i valori di abilità sui parametri degli item che sono noti) oppure il campione di calibrazione a nuovi item (per poter stimare sui valori di θ che sono noti nel campione, i valori dei parametri per i nuovi item). Se la pre-somministrazione e la calibrazione di un test sono largamente diffuse nel campo della ricerca valutativa nazionale ed internazionale, tuttavia risulta pressoché impossibile pensare che un docente, nello svolgimento della sua quotidiana funzione valutativa, applichi procedure così lunghe e complesse. Il docente può essere formato

sull'importanza del rispetto di criteri che rendano valida e attendibile la misurazione di una specifica abilità, ma difficilmente affronta, sia nella sua formazione iniziale sia in servizio, uno studio approfondito di modelli statistici per la valutazione di variabili latenti. D'altra parte è difficilmente ipotizzabile, che possa disporre di un campione elevato e significativo di studenti per calibrare il suo questionario prima di somministrarlo agli studenti della propria classe. Solitamente, quindi, nell'illusione di una valutazione più oggettiva, somministra questionari costruiti in base alla sua esperienza professionale e attribuisce un voto alla prova senza rispettare alcuna procedura corretta di scaling.

La riflessione su queste problematiche, ha portato alla nascita dell'idea di progettare e costruire, in collaborazione con una giovane ricercatrice e un esperto informatico, una piattaforma didattica interattiva da mettere a disposizione dei docenti interessati alla costruzione di questionari di verifica degli apprendimenti. La piattaforma metterà a disposizione dell'utente una serie di strumenti per costruire questionari della sola tipologia a risposta multipla, almeno nella fase iniziale. Tutti gli item inseriti andranno a confluire nella costruzione di un database di domande, classificate per ordine e grado di scuola, area disciplinare in generale ed obiettivi specifici di apprendimento, in modo che con il passare del tempo si possa disporre di una gran quantità di item già validati, in relazione all'analisi delle risposte e dei distrattori, e che non sia necessario per il docente inserire ogni volta nuove domande. Il processore, al termine della costruzione della prova, è pronto per simulare la compilazione virtuale del questionario con un numero crescente di iterazioni, che può essere impostato dall'utente stesso. A seguito della compilazione virtuale, sarà possibile effettuare operazioni di calibrazioni con semplici ed intuitivi comandi che consentiranno all'utente di condurre un'analisi secondo il modello IRT, in particolare utilizzando il Multiple-choice model (MCM). Il sistema restituirà i risultati del fit complessivo del modello, del fit di ogni singolo item e del fit del soggetto, plottando le ICC (Item Characteristic Curve) e la IIC (Item Information curve) del modello utilizzato. Se l'item è costruito in maniera corretta si osserverà che al crescere delle abilità, cresce la probabilità di opzionare la risposta corretta. Sarà possibile inoltre osservare il comportamento delle risposte errate e dei distrattori, in questo caso tendenzialmente si possono riscontrare, per

bassi livelli di abilità, probabilità più alte di contrassegnare le opzioni di risposta sbagliate rispetto a quella corretta. Per consentire ai docenti di interpretare correttamente ed autonomamente i risultati ottenuti nell'analisi dei dati, sarà predisposto uno spazio formativo, nel quale gli stessi docenti potranno consultare materiali esplicativi dei modelli di riferimento, osservare la lettura dei risultati di analisi già condotte, consultare articoli scientifici suggeriti, accedere a cospicui riferimenti bibliografici.

Ringraziamenti

Ogni traguardo, conquista, soddisfazione non è mai il risultato di un'impresa individuale ma vede coinvolte una pluralità di persone che nel progetto hanno creduto e lo hanno, anche inconsapevolmente, supportato... e talvolta anche sopportato!

Prima di tutto mi corre l'obbligo di ringraziare il Professor Tonio Di Battista che ha creduto in me molto prima che riuscissi a crederci io. Mi ha lanciato una proposta che a me sembrava una sfida al di là delle mie concrete possibilità, sostenendomi e spronandomi a raggiungere i miei obiettivi con tenacia e determinazione. Spesso siamo noi i peggiori giudici di noi stessi e le nostre rigidità, i nostri timori, ci impediscono di affrontare nuovi cammini, ma se troviamo sulla nostra strada chi ci dà suggerimenti e ci incoraggia tutto ci sembra più realizzabile. Nella mia esperienza di ricerca ho potuto contare sulla collaborazione di molti ricercatori... ci vorrebbe un libro solo per ringraziarli tutti. Chi si dedica alla ricerca non fa vita facile, il più delle volte non ha orari, non ha certezze sulle prospettive future, vive piccoli successi e grandi delusioni... la passione e la tenacia che accompagnano i ricercatori, dai più giovani ai veterani, trasmettono però una grande energia a chi ha la fortuna di vivere questa esperienza. La ricerca ti consente di provare ripetutamente l'esperienza della meraviglia, dello stupore, della scoperta; ti fa cadere, ti delude, ti smarrisce ma ti incita a provare e riprovare in una continua sfida a te stesso e ai tuoi limiti.

Nella mia esperienza professionale di insegnante e di formatrice, ho avuto modo di apprendere moltissimo dai miei alunni della scuola primaria, dai giovani studenti universitari e dagli adulti in formazione continua. Grazie a tutti loro ho imparato che la lunga strada della conoscenza si percorre a piccoli passi, raggiungendo di volta in volta un nuovo traguardo. E da qui bisogna ripartire, verso nuovi saperi e nuove conquiste, scrutando con curiosità sempre oltre l'orizzonte.

E poi c'è chi, oltre a supportarti, ti sopporta! La famiglia! La famiglia è la più grande fonte di energia sulla quale possa contare l'essere umano. Dall'amore, la cura, l'assistenza che un genitore offre ancor prima che tu venga al mondo, alla capacità di sostenerti, di tenderti la mano, di combattere al tuo fianco affinché tu possa realizzare te stesso, senza mai manifestare stanchezza e scoraggiamento. Il mio babbo e

la mia mamma non sono mai stanchi, o meglio non lasciano mai trasparire la loro stanchezza... sono sempre presenti, anche quando non possono esserci fisicamente... danno consigli ma non si intromettono nelle mie decisioni, consentendomi di essere comunque la protagonista dei miei successi e dei miei errori... camminano al mio fianco, sostenendomi quando vacillo, prendendomi in braccio quando cedo, facendosi da parte discretamente quando ottengo un risultato affinché io possa viverlo come esclusivamente mio... sono in me fisicamente ed emotivamente in ogni istante... sono la concretizzazione della forza dell'amore!

Mio marito e i miei figli rappresentano quella quotidianità rassicurante che mi dà serenità e voglia di vivere con intensità ogni momento. Un abbraccio, una parola, un litigio, ogni gesto è un'opportunità di confronto e di crescita continua. In famiglia si scoprono e si sfidano i propri limiti e al tempo stesso si valorizzano e si accrescono i pregi che ognuno possiede. Quando si diventa genitori si comprende pienamente la potenza del sentimento di amore con tutte quelle sfumature che ti fanno vivere i più svariati sentimenti di gioia come di tristezza, di speranza come di timore... non c'è sentimento che possa conoscere l'uomo che sia più ricco e variegato dell'amore per i propri figli, che raccolga in sé tutti gli stati d'animo e le sensazioni che l'uomo può provare. Grazie Francesca e Samuele, perché con voi cresco, apprendo e rinasco ogni giorno.

La vita, le mie scelte, il caso, la fortuna... qualunque sia stata la causa ho avuto l'opportunità di avere al mio fianco un compagno di avventura insostituibile, mio marito, che mi dà quotidianamente il sostegno e la libertà di vivere pienamente ogni attimo ed ogni occasione. Con te no ho bisogno di dire parole, a noi basta il silenzio di uno sguardo, di un tocco, di un respiro...

La mia allegria e il mio ottimismo si rinnovano ogni giorno grazie al calore delle moltissime persone che amo e che sono sempre al mio fianco: mia sorella, Lellino, i miei cognati, i miei suoceri, i miei nipoti, i miei amici di sempre, la mia grande e magnifica famiglia.

*Grazie
Cristiana*

Bibliografia

- A Scherbaum, C., Finlinson, S., Barden, K., and Tamanini, K. (2006). Applications of item response theory to measurement issues in leadership research. *The Leadership Quarterly*, **17**(4), 366–386.
- Ackerman, T. A., Gierl, M. J., and Walker, C. M. (2005). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice*, **22**(3), 37–51.
- Adams, R. J., Wilson, M., and Wang, W.-c. (1997a). The multidimensional random coefficients multinomial logit model. *Applied psychological measurement*, **21**(1), 1–23.
- Adams, R. J., Wilson, M., and Wang, W.-c. (1997b). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*, **22**(1), 47–76.
- Anastasi, A., Urbina, S., et al. (1997). *Psychological testing*. Prentice Hall Upper Saddle River, NJ.
- Bain, A. (1859). *The emotions and the will*. Longman, Green, Longman, Roberts, and Green.
- Baker, F. B. and Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*, volume 176. CRC.
- Barbaranelli, C. e Natali, E. (2005). *I test psicologici: teorie e modelli psicometrici*. Carocci, Roma.
- Barton, M. A. and Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model.
- Becker, G. S. (1964). *Human Capital*. Columbia University Press for the National Bureau of Economic Research, New York.

- Belok, M. V., Dellow, D. A., McCarthy, J. M., Neufeldt, H. G., Duimstra, E., Bronars Jr, J. C., Johanningmeier, E., Calabro, H., Erickson, R., Franklin, A., *et al.* (1979). Book review section 1. *Educational Studies*, 10(2), 201–222.
- Bertagna, G. (2006). *Pensiero manuale. La scommessa di un sistema educativo di istruzione e di formazione di pari dignità*. Rubbettino Editore.
- Binet, A. (2006). *Introduction à la psychologie expérimentale: 1894*. Editions L'Harmattan.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*.
- Birnbaum, I. M. (1969). Prior-list organization in part-whole free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 8(6), 836–838.
- Bloom, B. S. (1990). *Tassonomia degli obiettivi educativi*. Giunti e Lisciani Editori, Firenze.
- Bloom, B. S. e. a. (1956). Taxonomy of educational objectives. *New York: David McKay*, 19, 56.
- Bond, T. and Fox, C. (2007). *Applying the rasch model*, (2nd). *Fundamental measurement in the human sciences*.
- Bottani, N. (2011). Concorso per dirigente, non si scherza sui questionari. *Tuttoscuola*. <http://www.tuttoscuola.com/cgi-local/disp.cgi?ID=26489>.
- Bottani, N. e Cenerini, A. (2003). *Una pagella per la scuola*. Centro studi Erickson, Trento.
- Bottani, N. e Tuijnman, A. (1994). *Indicatori internazionali dell'educazione: struttura, sviluppo e interpretazione*, in OCSE, *Valutare l'insegnamento*. Armado Editore, Roma.
- Bottani, R. (2002). *Insegnanti al timone*. Il Mulino, Bologna.
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 1904-1920, 3(3), 296–322.

- Campbell, D. T., Stanley, J. C., and Gage, N. L. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin Boston.
- Cantrell, C.E. (1999). *Item response theory: understanding the one-parameter Rasch model*, volume 5. Thompson Ed., Stamford.
- Card, D. (2003). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, **69**(5), 1127–1160.
- Ceccatelli, C. and Di Battista, T. (2011a). Analysis of the correlation among self-esteem, effective learning and disorder of relationship between peers: an application of the lisrel model in the primary school. Ranchi - India. Rinpas.
- Ceccatelli, C. and Di Battista, T. (2011b). *The Relationship between the Self-Esteem and the Effective Learning: A Case Study in Primary School*, chapter 21, pages 357–370. Nova Science Publishers, handbook on psychology of self-esteem edition.
- Ceccatelli, C., Di Battista, T., Marianacci, R., and Tateo, A. (2010a). Lisrel model for a confirmatory analysis: relationship between low self-worth level and victim of bullying. *Procedia Social Behavioral Sciences*, **9**, 1612–1616.
- Ceccatelli, C., Marianacci, R., and Tateo, A. (2010b). Lisrel model for a statistical analysis of bullying: prevention and education. In *Atti di convegno MTISD*.
- Ceccatelli, C., Di Battista, T., Marianacci, R., and Di Zio, S. (2011). Career paths in a gender perspective: An application of multilevel irt models. In *International Meeting of the Classification and Data Analysis Group*.
- Ceccatelli, C., Di Battista, T., and Fortuna, F. (2012a). Burnout, learning and self-esteem at school: an empirical study. In *The 46th Scientific Meeting of the Italian Statistical Society*.
- Ceccatelli, C., Di Battista, T., Fortuna, F., and Maturo, F. (2012b). Giochi statistici ed innovazione didattica. In *Innovazione didattica e formazione degli insegnanti. Il ruolo dell'Università e della Mathesis*.

- Ceccatelli, C., Di Battista, T., and Mignani, S. (2012c). Motivazione, apprendimento e valutazione nella scuola secondaria di ii grado: il caso delle olimpiadi nazionali di statistica. *Induzioni: demografia, probabilità, statistica a scuola*, pages 31–66.
- Ceccatelli, C., Di Battista, T., and Fortuna, F., M. F. (2013). L'item response theory come strumento di valutazione delle eccellenze nella scuola. *Science & Philosophy*.
- Ceccatelli, C., Di Battista, T., Fortuna, F., and Maturo, F. (in press). Best practices to improve the learning of statistics - the case of the national olympics of statistics in italy. *Procedia - Social and behavioral science*.
- Checchi, D. (2003). Scelte di scolarizzazione ed effetti sul mercato del lavoro. *Lucifora, C.(a cura di), Mercato, occupazione e salari: la ricerca sul lavoro in Italia, Mondatori, Milano*.
- Chiandotto, B. (2004). Sulla misura della qualità della formazione universitaria. *Studi e note di economia*, 3, 2004.
- Chiandotto, B. e Bacci, S. (2005). Un modello multilivello per l'analisi della durata degli studi universitari. *Modelli statistici per l'analisi della transizione università-lavoro*, pages 63–86.
- Chiorri, C. (2011). *Teoria e tecnica psicometrica - Costruire un test psicologico*. McGraw-Hill.
- Cipollone, P., Montanaro, P., and Sestito, P. (2010). L'istruzione. *Il Mezzogiorno e la politica economica dell'Italia*, page 77.
- Collins, R. (1992). *Teorie sociologiche*. Il Mulino, Bologna.
- Cortina, J. M. (1993). What is coefficient alpha? an examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Crivellari, C. (2004). *Professori nella scuola di massa. Dalla crisi del ruolo alla formazione universitaria*. Armando Editore.
- Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. ERIC.

- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**(3), 297–334.
- Cronin Jr, J. J. and Taylor, S. A. (1992). Measuring service quality: a reexamination and extension. *The Journal of Marketing*, pages 55–68.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.
- de Lalande, J. J. L. F. (1790). *Voyage en Italie, contenant l'histoire & les anecdotes les plus singulieres de l'Italie, & description; les usages, le gouvernement, le commerce, la littérature, les arts, l'histoire naturelle, & les antiquités*. Tome Troisième, Genève.
- De Landsheere, G. (1973). *Elementi di docimologia. Valutazione continua ed esami*. La Nuova Italia, Firenze.
- Decroly, O. (1929). *La fonction de globalisation et laenseignement*. Maurice Lamertin.
- Di Battista, T. (2012). *Metodi e tecniche della valutazione: un approccio statistico*. Franco Angeli, Milano.
- di Piero Cipollone, P. M. and Sestito, P. (2010). Misure di valore aggiunto per le scuole superiori italiane: i problemi esistenti e alcune prime evidenze.
- Díaz, A. M. M. and Cepedab, E. (2010). Synthesizing the ability in multidimensional item response theory models. *Revista Colombiana de Estadística*, **33**(1), 127–147.
- Domenici, G. (1993). *Manuale della valutazione scolastica*. Editori Laterza, Bari.
- Donoghue, J. R. (1994). An empirical examination of the irt information of polytomously scored reading items under the generalized partial credit model. *Journal of Educational Measurement*, **31**(4), 295–311.
- Edgar, M. (2001). I sette saperi necessari all'educazione del futuro. *Milano, Raffaello Cortina*.

- Edgeworth, F. V. (1888). The statistics of examinations. *Journal of the Royal Statistics Society*, pages 599–635.
- Eisinga, R., Grotenhuis, M. t., and Pelzer, B. (2012). The reliability of a two-item scale: Pearson, cronbach or spearman-brown? *International journal of public health*, pages 1–6.
- Espinosa, M. P. and Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, **54**(5), 415–425.
- Ferguson, G. A. (1942). Item selection by the constant process. *Psychometrika*, **7**(1), 19–29.
- Ferrario, E. (1995). La valutazione dei sistemi scolastici: gli indicatori di qualità. *Dirigenti Scuola*, **5**, 13.
- Foray, D. (2006). *L'economia della conoscenza*. Il Mulino, Bologna.
- Fox, J. and Glas, C. A. W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, **68**(2), 169–191.
- Fox, J. P. (2004). Applications of multilevel irt modeling. *School Effectiveness and School Improvement*, **15**(3–4), 261–280.
- Fox, J. P. (2005). Multilevel irt using dichotomous and polytomous response data. *British Journal of Mathematical and Statistical Psychology*, **58**(1), 145–172.
- Fox, J. P. and Glas, C. A. W. (2001). Bayesian estimation of a multilevel irt model using gibbs sampling. *Psychometrika*, **66**(2), 271–288.
- Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice*, **7**(2), 33–38.
- Gattullo, M. (1968). *Didattica e docimologia: Misurazione e valutazione nella scuola*. Armando Editore, Roma.
- George, D. e Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. Allyn & Bacon, Boston.

- Gori, E., Sanarico, M., and Plazzi, G. (2005). La valutazione e la misurazione nelle scienze sociali: oggettività specifica, statistiche sufficienti e modello di rasch. *Non profit*, 3, 605–644.
- Gori, E. e Vittadini, G. (1999). *La valutazione dell'efficienza ed efficacia dei servizi alla persona. Qualità e valutazione nei servizi di pubblica utilità*. ETAS.
- Guilford, J. (1950). *Fundamental statistics in psychology and education* mcgraw-hill book company. New York, pages 347–348.
- Guttman, L. (1949). *The basis for scalogram analysis*. Bobbs-Merrill, College Division.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. *Setting performance standards: Concepts, methods, and perspectives*, pages 89–116.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer Nijhoff, Boston.
- Henard, D. (2000). Item response theory. *Reading and understanding more multivariate statistic*, pages 67–98.
- Hudson, B. (a cura di) (1975). *Introduzione alle tecniche di valutazione*. Zanichelli, Bologna.
- Ianes, D. (2006). *La speciale normalità: strategie di integrazione e inclusione per le disabilità dei bisogni educativi speciali*. Edizioni Erickson.
- Istruzione, M. P. (2007). *Indicazioni per il curricolo per la scuola dell'infanzia e per il primo ciclo d'istruzione*.
- Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. *Journal of Educational Measurement*, 38(1), 79–93.
- Kaplan, M. and Saccuzzo, P. (1997). *Psychological testing*, phoenix: Color corporation.
- Kuder, G. F. and Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160.

- Laugier, H. and Weinberg, D. (1935). Étude comparée des notes d'examens des étudiants et des étudiantes en sciences. *Le Travail humain*, 3(1), 62–81.
- Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh. Section A. Mathematical and Physical Sciences*, 61(03), 273–287.
- Lévy, P. (1994). *L'intelligence collective: pour une anthropologie du cyberspace*. La Découverte Paris.
- Levy, P. (1996). *L'intelligenza collettiva. Per un'antropologia del cyberspazio*. Feltrinelli, Milano.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Lord, F. (1952). A theory of test scores. *Psychometric monographs*.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. ERIC.
- Lord, F. M. (2005). Standard errors of measurement at different ability levels. *Journal of Educational Measurement*, 21(3), 239–243.
- Lord, F. M., Novick, M. R., and Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
- Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Publishing Company, Massachusetts.
- Lovaglio, P. G. (2004a). Investimento in capitale umano e disuguaglianze sociali.
- Lovaglio, P. G. (2004b). La stima di risultati clinici con la rasch analysis. *Statistica*, 64, 127–144.
- Lucisano, P. e Salerni, A. (2003). *Metodologia della ricerca in educazione e Metodologia della ricerca in educazione e formazione*. Carocci, Roma.

- Maier, K. (2001). A rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*, **26**(3), 307–330.
- Mair, P. and Hatzinger, R. (2007). *Extended Rasch modeling: The eRm package for the application of IRT models in R*. Department of Statistics and Mathematics, University of Economics and Business, Vienna.
- Margiotta, U. (1999). *L'insegnante di qualità*. Armando Editore, Roma.
- Mariani, A. M. (1991). *Valutare gli insegnanti*. La Scuola.
- Marzano, R.J (2006). *Classroom Assessment and Grading That Work*. ASCD, Alexandria, VA.
- Mason, L. (1996). *Valutare a scuola: prodotti, processi, contesti dell'apprendimento*. CLEUP.
- Mcdonald, I. (1999). *Remote paging in a single address space operating system supporting quality of service*. Citeseer.
- McKinley, W. and Tierney, R. J. (1989). Traversing the topical landscape reading and writing as ways of knowing. *Written Communication*, **6**(3), 243–269.
- McKinley, R. and Mills, C. (1989). Item response theory: Advances in achievement and attitude measurement. *Advances in social science methodology*, **1**, 71–135.
- McTighe, J. and Ferrara, S. (1996). Performance-based assessment in the classroom: A planning framework. *A handbook for student performance assessment in an era of restructuring*, pages 1–9.
- Mialaret, G. (1977). *Ciencias de la educación*. Oikos-tau.
- Milan, G. (1994). *Educare all'incontro: la pedagogia di Martin Buber*, volume 96. Città Nuova.
- Misiti, R. (1954). Presentazione del reattivo “progressive matrices”(adulti e fanciulli) di jc raven. *Ricerca Scient., Rome*, **24**, 1052–1058.

- Mislevy, R. J. (1985). Estimation of latent group effects. *Journal of the American Statistical Association*, **80**(392), 993–997.
- Morin, E. (2001). I sette saperi.
- Morin, E. (2005). *Educare per l'era planetaria. Il pensiero complesso come metodo di apprendimento*. Armando Editore, Roma.
- Morin, E., Ciurana, É.-R., Motta, D. R., Morin, E., and Morin, E. (2005). *Educare per l'era planetaria: il pensiero complesso come metodo di apprendimento*. Armando.
- Mosier, C. I. (1940a). A modification of the method of successive intervals. *Psychometrika*, **5**(2), 101–107.
- Mosier, C. I. (1940b). Psychophysics and mental test theory: Fundamental postulates and elementary theorems. *Psychological Review*, **47**(4), 355.
- Mosier, C. I. (1941). A psychometric study of meaning. *The journal of social psychology*, **13**(1), 123–140.
- Nunnally, J. C. (1978). *Psychometric Theory*. McGraw-Hill, New York.
- Nunnally, J.C. (1976). *Misurazione e valutazione nella scuola*. OS, Firenze.
- O'Connell, A. A. and McCoach, D. B. (2008). Multilevel measurement modeling. *Multilevel modeling of educational data*, pages 345–388.
- OCSE (1998). *Esami delle politiche nazionali dell'istruzione*. Armando Editore, Roma.
- Ostini, R. and Nering, M. L. (2005). *Polytomous item response theory models*. Sage Publications, Incorporated.
- Pellerey, M. (2004). Le competenze individuali e il portfolio. *La Nuova Italia*.
- Petracca, C. (2000). *Progettare per competenze*. Elmedi, Milano.
- Petracca, C. (2011). Approccio per competenze nella scuola. *Dalle Indicazioni al curriculum*, page 52.

- Piéron, H. (1922). V. étude psychotechnique de quelques tests d'aptitude. *L'année psychologique*, **23**(1), 144–175.
- Plake, B. S. and Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment. *Handbook of classroom assessment*, pages 53–68.
- Platone, T. (1999). traduzione e note di m. VALGIMIGLI, introduzione e note aggiornate di AM IOPPOLO, Laterza, Roma-Bari.
- Polanyi, M. (1979). *L'economia della conoscenza*. Armando Editore, Roma.
- Raju, N. S., Laffitte, L. J., and Byrne, B. M. (2002). Measurement equivalence: a comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, **87**(3), 517.
- Rasch, G. (1960). Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests.
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics*, **4**(3), 207–230.
- Ritter, N. (2010). Understanding a widely misunderstood statistic: Cronbach's alpha. *Online Submission*.
- Rogers, C. (2012). *On becoming a person: A therapist's view of psychotherapy*. Download iTunes eBook.
- Ruch, G. M. (1923). The achievement quotient technique. *Journal of Educational Psychology*, **14**(6), 334.
- Sampson, R. J. and Raudenbush, S. W. (1999). Systematic social observation of public spaces: a new look at disorder in urban neighborhoods. *American Journal of Sociology*, **105**(3), 603–651.
- Schizzerotto, A. e Barone, C. (2006). *Sociologia dell'istruzione*. Il Mulino, Bologna.

- Schmidt, K. M. and Embretson, S. E. (2003). Item response theory and measuring abilities. *Handbook of psychology*.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, 8(4), 350.
- Sears, L. (2007). Edward lee thorndike (1874-1949): A look at his contributions to learning and reading. *Shaping the Reading Field: The Impact of Early Reading Pioneers, Scientific Research, and Progressive Ideas*, page 119.
- Sgalambro, L. (2002). Le competenze e gli indicatori di valutazione. Monasta, A.(a cura di). *Organizzazione del sapere, discipline e competenze*. Carocci, Roma, pages 115–144.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, pages 201–292.
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 1904–1920, 3(3), 271–295.
- Spencer, L. M. and Spencer, S. M. (2008). *Competence at Work models for superior performance*. John Wiley & Sons.
- Starch, D. and Elliott, E. C. (1912). Reliability of the grading of high-school work in english. *The School Review*, 20(7), 442–457.
- Swartz, E. and Bakari, R. (2005). Development of the teaching in urban schools scale. *Teaching and Teacher Education*, 21(7), 829–841.
- Taylor, B. M. and Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly*, pages 134–146.
- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, pages 43–64.
- Tessaro, F. (1997). *La valutazione dei processi formativi*. Armando Editore, Roma.
- Tessaro, F. (2002). *Metodologia e didattica dell'insegnamento secondario*. Armando Editore, Roma.

- Thorndike, E. L. (1918). The seventeenth yearbook of the national society for the study of education.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review; Psychological Review*, **34**(4), 273.
- Thurstone, L. L. (1959). The measurement of values.
- van der Linden, W. J. and Veldkamp, B. P. (2004). Constraining item exposure in computerized adaptive testing with shadow tests. *Journal of Educational and Behavioral Statistics*, **29**(3), 273–291.
- Varisco, B. M. (2004). *Portfolio: valutare gli apprendimenti e le competenze*. Carocci, Roma.
- Vertecchi, B. (1988). *Manuale della valutazione: analisi degli apprendimenti*. Editori riuniti.
- Vertecchi, B. (1993). *Decisione didattica e valutazione*. La Nuova Italia, Firenze.
- Vertecchi, B. (2003). *Manuale della valutazione. Analisi degli apprendimenti e dei contesti*, volume 4. Franco Angeli.
- Visalberghi, A., Maragliano, R., and Vertecchi, B. (1978). *Pedagogia e scienze dell'educazione*. Mondadori, Milano.
- Wiggins, G. and McTighe, J. (1998). Understanding understanding. Wiggins, G. and McTighe J., *Understanding by Design. Association for Supervision and Curriculum Development*, pages 38–67.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. Jossey-Bass.
- Williams, N. J. e Beretvas, S. N. (2006). Dif identification using hglm for polytomous items. *Applied Psychological Measurement*, **30**(1), 22–42.
- Wood, W. B. (2009). Innovations in teaching undergraduate biology and why we need them. *Annual Review of Cell and Developmental Biology*, **25**, 93–112.

- Wright, B., Linacre, J., *et al.* (1989). Observations are always ordinal; measurements, however, must be interval. *Archives of physical medicine and rehabilitation*, **70**(12), 857.
- Wright, B. D. and Masters, G. N. (1982). *Rating Scale Analysis*. *Rasch Measurement*. ERIC.
- Wright, B. D. and Stone, M. H. (1979). *Best Test Design*. *Rasch Measurement*. MESA Press, Chicago.
- Wright, B. D., Mok, M., *et al.* (2000). Rasch models overview. *Journal of applied measurement*, **1**(1), 83–106.
- Wright, E. (1982). Modern psychoanalytic criticism. *Modern Literary Theory*, **131**, 147–71.
- Zeithaml, V. A., Berry, L., and Parasuraman, A. (1991). Understanding customer expectations of service. *Sloan Management Review*, **32**(3), 42.
- Zinbarg, R. E. *e. a.* (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for ω . *Applied Psychological Measurement*, **30**(2), 121–144.