

Un po' di storia della valutazione scolastica: letture e riflessioni¹

(a cura di *Guido Benvenuto* e *Andrea Giacomantonio*)

Il volume, in formato ipertestuale, può essere navigato utilizzando l'indice che segue. Per scorrere e aprire il collegamento usare il comando CTRL+clic sulla pagina specifica

Indice

PREMESSA	V
PARTE 1. LA DOCIMOLOGIA: DALLA CRITICA AGLI ESAMI ALL'ANALISI DEI PRINCIPI DELLA VALUTAZIONE	1
1. NASCITA DELLA DOCIMOLOGIA (PIÉRON, H., EXAMENS ET DOCIMOLOGIE)	7
2. STEREOTIPI ED EFFETTI DI ALONE (DE LANDSHEERE, G., ÉVALUATION CONTINUE ET EXAMENS. PRECIS DE DOCIMOLOGIE).....	12
3. I LIMITI DELLE PROVE [...] DI ACCERTAMENTO DELLE CONOSCENZE (DOMENICI, G., MANUALE DELLA VALUTAZIONE SCOLASTICA).....	15
4. IDÒLA (EFFETTI DISTORSIVI) NELLA VALUTAZIONE INTERNA (BENVENUTO, G., GLI ESAMI SONO MATURI?).....	20
5. MISURAZIONE E VALUTAZIONE (VISALBERGHI, A., MISURAZIONE E VALUTAZIONE NEL PROCESSO EDUCATIVO).....	24
6. QUANDO POSSIAMO FIDARCI DELLA VALUTAZIONE (VERTECCHI, B., ORIGINI E SVILUPPI DELLA DOCIMOLOGIA).....	29
7. LA VALUTAZIONE IN DIECI FOTOGRAMMI (GIOVANNINI, M. L., LA VALUTAZIONE)	32
PARTE 2. APPROCCI E FUNZIONI DELLA VALUTAZIONE	41

¹ Questo materiale riprende integralmente un materiale ad uso di studio per i corsi di Docimologia (Università Sapienza di Roma, Corso di Laurea in Scienze dell'Educazione e della Formazione) messo a punto e curato da Guido Benvenuto e Andrea Giacomantonio, dal titolo *La valutazione scolastica: letture e riflessioni: Un'antologia di testi sulle teorie della valutazione*, stampato da Nuova Cultura, Roma, 2008.

IV

8. APPROCCI ALLA VALUTAZIONE (STAME, N., TRE APPROCCI PRINCIPALI ALLA VALUTAZIONE: DISTINGUERE E COMBINARE)	51
9. LE FUNZIONI DEL CONTROLLO SCOLASTICO NEL PROCESSO EDUCATIVO (GATTULLO, M., DIDATTICA E DOCIMOLOGIA. MISURAZIONE E VALUTAZIONE NELLA SCUOLA)	60
10. FUNZIONI DELLA VALUTAZIONE (VERTECCHI, B., MANUALE DELLA VALUTAZIONE. ANALISI DEGLI APPRENDIMENTI)	69
11. LA VALUTAZIONE CONTESTUALIZZATA: L'ALTERNATIVA AI TEST STANDARDIZZATI (GARDNER, H., L'EDUCAZIONE DELLE INTELLIGENZE MULTIPLE)	78

PARTE 3. GLI STRUMENTI DEL CONTROLLO SCOLASTICO: LA TECNICA DEL TESTING

87

12. UNA TIPOLOGIA DELL'OSSERVAZIONE (POSTIC, M., DE KETELE, J. M., OSSERVARE LE SITUAZIONI EDUCATIVE)	100
13. LA COSTRUZIONE DI UN "PROFILO" (BONCORI, G., PROFILO, IN ENCICLOPEDIA PEDAGOGICA)	114
14. STRUMENTI DI MISURAZIONE E PROVE OGGETTIVE: DEFINIZIONI E TIPOLOGIE (MELCHIORI, R., ADAS. IL LABORATORIO DELLA VALUTAZIONE)	119
15. UN QUADRO DI SINTESI DEGLI STRUMENTI DI VERIFICA SCOLASTICA (BENVENUTO, G., METTERE I VOTI A SCUOLA. INTRODUZIONE ALLA DOCIMOLOGIA)	124
16. LA COSTRUZIONE DELLE PROVE OGGETTIVE (DOMENICI, G., MANUALE DELLA VALUTAZIONE SCOLASTICA)	129
17. FORMULARE I QUESITI (ESEMPI E ALCUNE REGOLE PER LA COSTRUZIONE)(BENVENUTO, G., LOPRIORE, L., COSTRUZIONE E SOMMINISTRAZIONE DEGLI STRUMENTI)	135
18. CORREZIONE DELLE PROVE OGGETTIVE DI PROFITTO E ANALISI DELLE PRESTAZIONI (VERTECCHI, B., DECISIONE DIDATTICA E VALUTAZIONE)	144
19. UNA DIVERSA NOZIONE DI "SUFFICIENZA" (VERTECCHI, B., DECISIONE DIDATTICA E VALUTAZIONE)	147

Premessa

La raccolta di testi e documenti che compone il volume è stata pensata con un duplice scopo: “leggere” insieme agli studenti dei corsi universitari testi “originali” (anche se in alcuni casi siamo di fronte a traduzioni da varie lingue) di ricerche e studi condotti in ambito di teoria e prassi della valutazione nell’ultimo secolo; “documentare” alcune tappe della riflessione docimologica e di teoria della valutazione.

Sempre più spesso all’università gli studenti di corsi di laurea o di formazione quali quelli per la specializzazione all’insegnamento sono coinvolti in attività didattiche condotte attraverso lezioni dialogiche e momenti laboratoriali. Lo studio collettivo di testi e la lettura autonoma di approfondimento sono azioni didattiche spesso secondarie, maggiormente sacrificate per il basso numero di ore per modulo e il poco tempo a disposizione per uno studio individuale.

Un’antologia, quindi, per ragionare sull’evoluzione dei concetti, delle metodologie, sulle finalità e funzioni in campo valutativo è un tentativo, forse controcorrente, per offrire materiali e spunti alla riflessione autonoma e collettiva.

Sono stati individuati tre filoni antologici, per introdurre e proporre le diverse letture.

Il primo ambito riguarda gli studi e le tematiche docimologiche, ovvero attinenti alla critica e all’analisi delle valutazioni scolastiche. Da quando ci si è interrogati sulle modalità di verifica degli apprendimenti scolastici, sulle prassi valutative e quindi sull’eventuale incidenza della soggettività e sulle distorsioni presenti nelle “misurazioni” scolastiche, si è costantemente registrata un diffusa preoccupazione da parte dei docenti e degli studenti per la complessità dell’atto valutativo e le connaturate forme di arbitrarietà che esso incontra.

Il secondo filone antologico sviluppa le finalità e funzioni della valutazione. Si indagano qui i modelli impliciti ed espliciti che guidano i docenti nell’operare comparazioni, misure e giudizi valutativi.

Il terzo e ultimo filone, dopo aver illustrato i diversi formati di prove per la verifica degli apprendimenti, si sofferma su un particolare formato valutativo che negli ultimi cinquan-

ta anni ha fatto levare scudi e alzare barricate ideologiche: la valutazione oggettiva, ovvero il testing. Non c'è docente che non si sia schierato, più o meno apertamente, in questo dibattito. Accanto ai fautori e difensori dei test, che consapevoli dei limiti che tali strumenti presentano, ne sottolineano la loro funzione "democratica", vale a dire "la stessa prova per tutti", si pongono i sinceri oppositori, che, in nome di una necessaria apertura e dialogicità nella valutazione, rimandano a formati valutativi meno strutturati e coercitivi possibili.

I tre filoni pongono ogni docente e operatore scolastico di fronte a quesiti e livelli interpretativi diversi. I testi presentati e introdotti nelle diverse sezioni si offrono al tempo stesso come "punti di vista" che in distinti momenti della storia della scuola sono stati avanzati da autorevoli studiosi e come "nodi problematici" della teoria/prassi valutativa.

I testi propongono spesso differenti modelli valutativi e distinte concezioni della scuola, spaziando dagli anni '60 del secolo scorso ai giorni nostri, con una preponderanza di contributi prodotti negli anni '80-'90.

Ci auguriamo che dal loro accostamento e dalla loro lettura possano scaturire il giusto stimolo alla riflessione critica, l'approfondimento delle tematiche e la curiosità di allargare lo sguardo con la lettura integrale delle fonti riportate.

Guido Benvenuto e Andrea Giacomantonio

Parte 1. La docimologia: dalla critica agli esami all'analisi dei principi della valutazione

a cura di Guido Benvenuto²

L'etimologia del termine docimologia, proposto nei primi anni '30 da Henri Piéron, è rintracciabile in alcuni vocaboli greci che rimandano al contesto dell'esaminare inteso in senso ampio. Da un lato abbiamo infatti il significato di prova o di esame (δοκιμασία/dokimasía)³, dall'altro, con il termine λογία da λογος (logos: discorso, ma anche riflessione scientifica) si aggiunge lo spessore di studio e contesto di riflessione sull'esaminare.

L'accezione iniziale per docimologia è quindi quella di *studio destinato alla critica e al miglioramento delle votazioni scolastiche* (Pieron, 1929, 1934). Nei primi studi il termine aveva, a ben vedere, un "carattere negativo criticando i sistemi di votazione e dimostrando la mancanza di fedeltà e di validità agli esami". Nel tempo il campo semantico si allarga, includendo l'analisi dei comportamenti di chi esamina, prevalentemente l'insegnante, e di chi è esaminato, sostanzialmente lo studente. Nel 1971 De Landsheere definisce la docimologia *scienza che ha per oggetto lo studio sistematico degli esami, in particolare dei sistemi di votazione e del comportamento degli esaminatori e degli esaminati*.

Nel suo uso più contemporaneo il termine ha ampliato ulteriormente lo spettro, confondendosi con lo stesso processo valutativo: *lo studio delle prove di esame, delle verifiche scolastiche e dei sistemi di valutazione in educazione* (cfr. De Ketele, 1982).

La docimologia, quindi, da un iniziale carattere di critica dei voti assume una valenza sempre più costruttiva, in quanto *riflessione o risposta scientifica per contrastare la personalizzazione nella valutazione scolastica* (v. soggettività/soggettivismo), attraverso la proposta di "metodi e tecniche di valutazione più obiettivi o, quanto

² Parte di questa presentazione è estratta da: Benvenuto, G. (2003), *Mettere i voti a scuola. Introduzione alla docimologia*, Carocci, Roma, pp. 29-39.

³ Le indicazioni lessicali concernenti il termine "dokimasía" sono tratte dal *Dizionario greco-italiano* di Rocci L. (1939), Ed. Dante Alighieri, Roma. Altri termini con la stessa radice: δοκιμή (dokimé: provare) e δοκιμάζω (dokimázo: esaminare, provare), δοκιμαστής (dokimastés: esaminatore), δοκιμαστικός (dokimastikós: atto ad/per esaminare). Alcuni riferimenti storici ci aiutano a contestualizzare le indicazioni lessicali (cfr. S.Battaglia, *Grande Dizionario nario della lingua italiana*, Unione tipografica editrice torinese, 1961). Nell'antica Grecia la prova di esame (dokimasía) riguardava ad esempio il cittadino eletto per voto o scelto a sorteggio per ricoprire una magistratura cittadina. La prova, in base al diritto pubblico ateniese, serviva a verificare se il cittadino possedesse i requisiti richiesti. Ma altre prove riguardavano i giovani al momento di entrare nell'efebato, o i cavalieri per controllare se avevano i mezzi richiesti dalla legge.

meno, più rigorosi, e mettendo a punto i mezzi per rendere le classificazioni confrontabili, in maniera da assicurare una maggiore giustizia scolastica" (De Landsheere, 1971).

1.1 La prima accezione del termine docimologia è quindi: studio delle prove di esame

L'esigenza di studiare e correggere alcune disfunzioni nelle valutazioni di esame nel '900 nasce dai sensibili cambiamenti nei sistemi scolastici che si svilupparono in Europa ad inizio del secolo. Le tematiche docimologiche sono strettamente collegate all'aumento della richiesta di formazione, al crescere della scuola di massa che, come vedremo, dagli anni '60 in poi incomincia a richiedere forme di valutazione più oggettive e diverse modalità di accertamento del rendimento scolastico.

La problematica della valutazione in contesti di esame era talmente avvertita in ambito psicologico e pedagogico da essere presente in due delle 15 tesi della "scuola attiva" che Claparède enunciò al Congresso di igiene mentale nel giugno del 1922 a Parigi. Nel suo Rapporto, che presenta la concezione funzionale dell'educazione, leggiamo:

"13) Le riforme qui preconizzate saranno possibili solo se il sistema degli esami verrà profondamente trasformato. La necessità dell'esame spinge molti insegnanti, loro malgrado, a sovraccaricare la memoria più che a sviluppare l'intelligenza. Salvo, forse, per un minimo di cognizioni indispensabili, gli esami dovrebbero essere soppressi, e sostituiti da una valutazione data in base ai lavori individuali fatti durante l'anno, oppure per mezzo di studi adeguati. 14) La psicologia sperimentale è in grado di fornire alla pedagogia pratica dei metodi adatti al controllo del valore dei metodi didattici e del rendimento scolastico. Essa ci fornisce anche dei metodi per la valutazione mentale (tests mentali)".

Il dibattito sull'uso o rifiuto dei test a scuola nasce proprio dalla consapevolezza che molte verifiche scolastiche non presentano sufficiente rispetto della validità e affidabilità nell'esaminare. E sono proprio questi temi, centrali per la nascita e sviluppo degli studi docimologici, a far sentire l'esigenza di studi e ricerche sperimentali, intraprese nei primi anni del Novecento in molte nazioni europee.

Piéron pubblica nel 1936 (cfr. Pieron, 1963) i risultati della commissione francese sull'analisi rigorosa delle modalità di effettuazione degli esami terminali della scuola secondaria, il *baccalauréat*. In particolare si studiarono gli scarti⁴ che i vari correttori presentavano nell'attribuzione di votazioni alle stesse prove. Si confrontarono le votazioni fornite da 30 insegnanti, già commissari di esami, suddivisi in 6 diversi

⁴ Se chiediamo a due correttori di assegnare voti ad un numero di n prove, per ciascuna prova potremo avere lo stesso voto o un voto diverso. Se il voto è lo stesso, lo scarto è zero, se il voto è diverso lo scarto è dato dalla differenza fra i due voti. La media di questi scarti ci dà lo scarto medio.

gruppi di esaminatori; ogni esaminatore corresse campioni di 100 elaborati scritti (composizione francese, versione latina, composizione inglese, matematica, dissertazione filosofica e fisica), assegnando un giudizio (valutazione) con voto su scala in ventesimi. Confrontando le votazioni fornite dai singoli insegnanti si dimostrò che lo scarto medio tra i correttori era molto elevato (4 punti per inglese, matematica e fisica, 7 punti per francese e filosofia) e che lo scarto massimo (12/13 punti) rivelava l'inesistenza di criteri uniformemente condivisi e applicati nella valutazione.

In sintesi, quando correggiamo si presentano diversi elementi di distorsione:

- ciascun valutatore è sensibile ad alcune caratteristiche delle prove, mentre tende a trascurarne altre;
- i punteggi attribuiti dai vari correttori davano luogo a distribuzioni diverse (ognuno interpretava a suo modo una scala di voti);
- non v'era concordanza tra correttori;
- per ottenere un voto vero (livello della prova al di là delle differenze) il numero dei correttori sarebbe dovuto essere altissimo (13 per la matematica, 127 per la dissertazione filosofica!).

Altri studi furono sviluppati negli anni a seguire, e non solo in Francia, a evidenziare l'interesse generale per la problematica sollevata (v. Pieron, 1963, Visalberghi, 1955, Gattullo, 1968, De Landscheere, 1971, Vertecchi, 1991).

Per quanto riguarda l'Italia gli studi docimologici prendono l'avvio alla fine degli anni '50.

Sono Visalberghi e Calonghi i primi ad aprire il dibattito e a sensibilizzare sulle tematiche docimologiche. Nel '54 Calonghi pubblica su *Orientamenti Pedagogici* "L'insufficienza e soggettività delle valutazioni scolastiche degli insegnanti"; nel '55 Visalberghi con *Misurazione e valutazione nel processo educativo* apre un dibattito sulle funzioni e utilizzazioni del *testing* a scuola. Nel 1958, in occasione del Congresso internazionale di psicologia applicata si tiene poi il "Simposio di docimologia". L'ampiezza della bibliografia sulle tematiche docimologiche negli anni '50 dimostra il nascente interesse per la problematica. I primi studi e ricerche si rivolgono soprattutto all'individuazione di metodologie e interventi per la riduzione della soggettività nella valutazione scolastica e alla riflessione sui voti⁵.

Il testo di Visalberghi (1955) è centrale per questo ampio dibattito. In esso, infatti, si discute la distinzione tra i termini "misurazione" e "valutazione", a tutt'oggi essenziale per distinguere le fasi e le funzioni nei diversi momenti di verifica e valutazione in contesti formativi. L'invito non è a schierarsi per misure quantitative o

⁵ Il celebre testo che in campo psicometrico affronta le tematiche della misurazione di determinate variabili in campo di valutazione scolastica è *Psychometric Methods* di Guilford (1954). Tra i vari interventi in campo più pedagogico nel dibattito sviluppatosi negli anni '50 citiamo: Calonghi (1954b, 1954c, 1955, 1956, 1958), Corallo (1955), Lepore (1957), Pedrazzi (1955), Pisani (1956), Visalberghi (1954).

forme di valutazioni più qualitative, ma a distinguere tra fasi che permettono di quantificare maggiormente (la misurazione) e fasi che precedono o seguono la quantificazione e/o la raccolta mirata di informazioni. In questi casi si parla di atti valutativi (valutazione), cioè atti che richiedono un giudizio. Che tra i due termini, misurazione e valutazione, ci sia una stretta interdipendenza è fuor di dubbio: «nonché sopprimere la *valutazione*, la *misurazione* nasce dalla valutazione e nella valutazione confluisce» (ibidem p.18). La questione è che per la gran parte delle verifiche, e il *testing* ne è solo un esempio, la distinzione va accuratamente mantenuta. Nella preparazione di test gli elementi qualitativi sono innegabili, ad esempio nella scelta degli elementi da sottoporre a verifica, nella composizione delle domande e nella valutazione dei risultati, ma anche quelli quantitativi risultano manifesti, ad esempio nella somministrazione, correzione e possibile analisi statistica. Ecco allora che «la valutazione non interviene soltanto *dopo* l'impiego dei *tests*, interviene anche *prima*, cioè nella fase del loro apprestamento» (ibidem p.17).

Il discorso andrebbe declinato per la gran parte delle forme di verifica, tradizionali o più alternative. Le fasi di misurazione e di valutazioni sono fasi di uno stesso processo che si conclude con la formulazione di giudizi o con l'attribuzione di voti. Chi intende i due termini come antitetici o contrapposti confonde le fasi del processo di valutazione, rilevazione e analisi, con i possibili momenti nei quali esso si sviluppa. Già il termine "misura", ricordava Visalberghi, ha due significati principali: a) operazione di conteggio o confronto; b) abito di equilibrio e di discrezione.

La riflessione aperta da Visalberghi spinge da un lato a migliorare le competenze nella costruzione e uso di test, nel senso ovviamente di test ben fatti, intelligenti e capaci di rilevare i processi di apprendimento, ma dall'altro a costruire giudizi più obiettivi possibili: «Noi dobbiamo domandarci non già quali tipi di prove siano più «calde» ed umane, ma quali tipi di prove servano meglio ai fini di un giudizio impegnativo e motivato, ricco di conseguenze pratiche importanti, e che deve perciò essere giusto ed obiettivo quant'è possibile» (ibidem p.16).

Il testo di Visalberghi (1955) sulla misurazione e valutazione nel contesto educativo fornisce quindi una prima riflessione sull'esigenza di ricorrere a diverse forme di verifica nella scuola. Studi di Calonghi (vedi bib.) e successivamente di Gattullo (1968) manifestano l'ampliarsi del dibattito e l'approfondimento dei temi specifici. Dalla docimologia come problema scolastico per i singoli docenti e studenti si inizia a discutere della necessità di "servizi nazionali": "Non esiste un servizio regolare di docimologia in Italia, e quindi tutta la nostra poca sperimentazione riesce alquanto approssimata"⁶. Andreani Dentici (1969) allarga il quadro toccando le problematiche della misurazione dell'abilità mentale e del rendimento scolastico e offrendo sia una critica alle prove tradizionali sia metodologie e strategie alternative per la misurazione di specifiche abilità nella scuola.

⁶ Cfr. Visalberghi, A., *Istruzione programmata*, in Commissione Nazionale italiana per l'Unesco, *Le nuove tecniche d'apprendimento*, Atti del Convegno di Bologna 6-9 aprile 1967, Ente Autonomo per la Fiera di Bologna, Bologna, 1969, p.214.

Nel 1971 De Landsheere con la pubblicazione del celebre testo *Évaluation continue et examens. Précis de docimologie* riprende pienamente il termine « docimologia » che definisce « una scienza che ha per oggetto lo studio sistematico degli esami, in particolare dei sistemi di votazione, e del comportamento degli esaminatori e degli esaminati ». L'intero lavoro parte dall'esigenza di contenere la soggettività e altre disfunzioni presenti nello svolgimento degli esami per dotare i docenti di strumenti che affianchino l'osservazione continua.

1.2 Disfunzioni e distorsioni nell'attribuzione di valutazioni a scuola

I principali “nemici” della docimologia sono senza dubbio le distorsioni valutative a scuola. Gli “effetti” perversi che si presentano tutte le volte che ci si trova ad attribuire una votazione a prove non strutturate quali ad esempio quelle di scrittura o di tipo orale, vedi l'interrogazione, sono chiari a tutti i docenti e studenti di ogni ordine e grado, in Italia e altrove. Si può dire che le fonti di distorsione di giudizio in cui si incorre in questa fase sono più o meno un luogo comune della scuola.

Molti i testi che si sono occupati di passare in rassegna le forme della distorsione valutativa a scuola. Tra le letture che presentiamo in questa sezione le prime sono dedicate alle distorsioni e “effetti” dovuti ad eccesso di soggettività nella valutazione scolastica.

Dopo una prima lettura “storica” (Lettura 1.: *Nascita della docimologia*), sulla nascita della docimologia, tratta da uno degli ultimi testi di Pieron, fondatore e primo diffusore del termine, si passa alle letture scelte per discutere il tema delle distorsioni nella valutazione: tra i primi a denunciare il problema si presenta De Landsheere (Lettura 2.: *Stereotipi ed effetti alone*), tra i primi a discutere la problematica in Italia vi è Domenici (Lettura 3.: *I limiti delle prove di accertamento delle conoscenze*) e si chiude il tema con un quadro sinottico (Lettura 4.: *Benvenuto, 2003: Idòla (effetti distorsivi) nella valutazione interna*).

Le letture che seguono, in questa prima sezione, allargano invece il quadro per individuare e discutere le dimensioni su cui la docimologia ha investito in termini di studi e di ricerche sul campo. Il testo che apre la riflessione, sia terminologia, sia teorica, è quello nel quale, negli anni '50, Visalberghi dibatte la necessaria differenziazione tra un livello di accertamento, di misurazione e uno di giudizio, di valutazione (Lettura 5.: *Misurazione e valutazione*). Seguono due riflessioni: la prima di Vertecchi, che discute sulle condizioni che rendono “accettabile” la valutazione, perché condotta con serietà e rigore (Lettura 6.: *Quando possiamo fidarci della valutazione*); la seconda di Giovannini, che illustra le coordinate generali della valutazione, inquadrando il tema della affidabilità all'interno di una cornice più ampia di funzioni e tempi della valutazione (Lettura 7.: *La valutazione in dieci fotogrammi*).

Bibliografia

- Andreani Dentici, O. (1968), *Abilità mentale e rendimento scolastico*, Firenze, La Nuova Italia.
- Benvenuto, G. (2003), *Mettere i voti a scuola. Introduzione alla docimologia*, Carocci, Roma.
- Calonghi, L. (1954b), "Difficoltà scolastiche e prove diagnostiche", *Orientamenti Pedagogici*, I, 1, pp.66-69.
- Calonghi, L. (1954c), "Le prove oggettive (per un controllo più oggettivo del rendimento scolastico)", *Orientamenti Pedagogici*, I, 5, pp.476-490.
- Calonghi, L. (1955), "Prove e scale oggettive, tests ed esami", *Orientamenti Pedagogici*, II, 2, pp.165-170.
- Calonghi, L. (1956), *Test ed esperimenti*, PAS, Torino.
- Calonghi, L. (1958), "I voti dicono e non dicono", *Orientamenti pedagogici*, 2.
- Calonghi, L. (1961), *Sussidi per la valutazione scolastica*, PAS, Zürich.
- Corallo, G. (1955), "Valutazione del profitto", in *Idee e fatti delle scuole d'America*, Hermes, Salerno.
- de Ketele, J.M, (1982), *La docimologie*, Cabay, Louvain-La-Neuve.
- De Landsheere, G. (1971), *Évaluation continue et examens. Précis de docimologie*, Éditions Labor, Bruxelles, (tr. it., *Elementi di docimologia. Valutazione continua ed esami*, La Nuova Italia, Firenze, 1973).
- Gattullo, M. (1968), *Didattica e docimologia*, Armando, Roma.
- Guildford, J.P. (1954), *Psychometric Methods*, McGraw-Hill, New York.
- Lepore, G. (1957), "Valutazioni psicometriche e rendimento scolastico", *La Scuola Secondaria*, 13.
- Pedrazzi, L. (1955), "Verso nuove tecniche di valutazione scolastica", *Il Mulino*, 40, pp. 99-133.
- Pisani, V. (1956), "Il problema psico-pedagogico delle valutazioni scolastiche", *Puer*, n.3.
- Piéron, H. (1929), "La technique des examens et la nécessité d'une docimologie", *L'enseignement scientifique*, II, 17, pp. 193-199.
- Piéron, H. (1934), "Étude critique de la valeur sélective du certificat d'étude et comparaison de cet examen avec une épreuve par tests", *Études docimologiques sur le perfectionnement des examens et des concours*, Publications de Travail Humaine, 3.
- Piéron, H. (1963), *Examens et docimologie*, Presses Universitaires de France, Paris, (tr. It. *Esami e docimologia*, Armando Armando, Roma, 1965).
- Vertecchi, B. (1991), *Origini e sviluppi della docimologia*, Giunti e Lisciani, Teramo.
- Visalberghi, A. (1954), "Socialità nella scuola e nuove tecniche di accertamento del profitto", *Rivista di Psicologia sociale*, 3-4, pp.77-82.
- Visalberghi, A. (1955), *Misurazione e valutazione nel processo formativo*, Edizioni di Comunità, Milano.

1. **Nascita della docimologia (Piéron, H., *Examens et docimologie*)⁷**

Primi riconoscimenti, in Francia.

Nell'opera *Essais sur l'enseignement en général et sur celui des mathématiques en particulier*, pubblicata nel 1805 da un certo F. Lacroix (*L'Enseignement mathématique*, 25 giugno 1937), sono state notate queste assennate riflessioni:

Giacché il sapere autentico in matematica non è costituito dallo sforzo di memoria il quale, più che dilatarle, restringe le facoltà, del tutto ingiustificata è la prassi d'un esame orale e mnemonico per appurare la capacità dei giovani che si dedicano allo studio delle scienze. Per questo è accaduto che gli uomini più colti hanno concordemente ed onestamente riconosciuto le loro perplessità circa una loro promozione, ove fossero stati sottoposti ad un siffatto esame. Tale asserzione è stata fatta, con la modestia che in misura così eminente lo contraddistingue, dal Lagrange stesso, in una delle sue lezioni al Politecnico... Come allora si può equamente pretendere dai discepoli ciò che non si oserebbe esigere dal maestro? Si ignora il tempo che si fa loro perdere nel rivedere o nel ripetere incessantemente la materia d'esame, al solo scopo di tenerli in esercizio ed in condizione di rispondere su tutto ciò che hanno imparato? Si è disposti a credere che il disgusto, inevitabile conseguenza di tanta monotonia, circoscriva il progresso dei giovani al limite ove finisce la loro prova o non li induca addirittura ad eliminare sollecitamente dal cervello quelle medesime nozioni apprese con tanta fatica per fame ostentazione un sol giorno?

... Molti di essi, talora sotto la guida degli stessi maestri, studiano l'umore, le consuetudini degli esaminatori, mossi dall'unico scopo di cercare il modo per abbreviare e mitigare la prova che devono subire.

Siffatte critiche precorrevano di poco l'istituzione napoleonica del 1808 circa gli esami universitari ed i diplomi richiesti ai docenti ai vari livelli, abilitazione magistrale, *baccalauréat*, *licence*⁸, laurea. Da allora, non è ignoto ad alcuno il pullulare di

⁷ Tratto da: Piéron, H. (1963), *Examens et docimologie*, Presses Universitaires de France, Paris, 1963; tr. it. *Esami e docimologia*, Armando Armando, Roma, 1965, pp. 9-13; Conclusioni pp. 183-184.

⁸ Si chiama *licence* un titolo conferito dalle facoltà universitarie, il quale sanziona gli studi superiori seguiti da studenti provvisti di *baccalauréat* o titolo equipollente. La *licence* è titolo di studio necessario per gli aspiranti al *doctorat*. Ne esiste un particolare tipo, chiamata *licence d'enseignement*, richiesta a chi voglia partecipare ai concorsi per l'insegnamento secondario. I *licenciés*, dopo almeno 2 anni di studi e la discussione d'una tesi, possono sostenere un esame (*doctorat*), che può essere considerato alla stregua del nostro esame di laurea. (*N.d.T.*).

esami e concorsi e, considerato l'incessante aumento di tempo e di sforzi richiesti, vien fatto di domandarsi in che misura il loro rendimento li giustifichi.

Durante la frequenza scolastica è emerso che la preoccupazione degli esami tendeva a far assumere come scopo, nell'insegnamento, ciò che doveva essere unicamente un mezzo per controllarne l'efficacia. In una conferenza alla Società di Pedagogia nel luglio 1961, il prof. Zamansky, preside della Facoltà di Scienze di Parigi, si dichiarava colpito dal fatto che l'insegnamento di secondo grado sembrava imperniarsi sull'idea fissa che le famiglie, non appena il loro rampollo compie i 10 anni, sono già ossessionate dal concorso d'ammissione al quale dovrà presentarsi a distanza di 8 anni. Da quel momento si impostano gli studi così da assicurare il buon esito al concorso.

D'altro canto, l'illimitata fiducia attribuita agli esaminatori ed alle commissioni, investiti d'autorità sovrana, non sempre è apparsa esente da pericoli, specie durante le prove orali, e molti esaminatori scrupolosi non hanno fatto mistero d'una preoccupazione talvolta angosciata di fronte alle loro responsabilità e, al tempo stessa, del loro stupore nel vedere con quale leggerezza taluni colleghi esplicassero le loro funzioni. Tale fu il caso del prof. Mauriac, preside della Facoltà di Medicina a Bordeaux, nel suo *Esame di coscienza* d'un professore, pubblicato nella "Presse médicale" del 30 gennaio 1937: trovatosi un giorno, così egli narra, nell'aula in cui si tenevano gli esami di *baccalauréat*

due esaminatori su cinque mi hanno stupito col loro modo di fare: non appena sedutosi davanti a loro, il candidato assumeva la fisionomia dell'imputato: quindi voce tagliente, sorriso beffardo, facile ironia, frasario sconcertante...

Per converso, un'estrema benevolenza induce talora un esaminatore sensibile, quando il candidato sia in difficoltà, a prendere su di sé e domande e risposte.

La funzione del caso negli esiti favorevoli o sfavorevoli è spesso accettata come fatto ineluttabile, ma suscita altresì veementi proteste, come quella del giovane Raymond Poincaré, studente di filosofia, il quale così si esprime circa il *baccalauréat*:

È una prova scarsamente morale. Per quanta si lavori per cinque mesi dall'alba al tramonto, la sola fortuna ha stabilito il nostro destino.

La ricerca Piéron-Laugier, 1922

Complessivamente però i docenti hanno serbato illimitata fiducia nella validità delle selezioni. Di essa ho trovato molteplici prove allorquando, agli albori del secolo, fui del novero di coloro che si sforzarono di costituire una psicotecnica scientifica-

mente valida e praticamente efficace. Nell'elaborazione di test atti ad individuare talune attitudini e ad istituire pronostici futuri negli scolari, si calcolavano correlazioni con gli esiti favorevoli negli esami scolastici. Orbene, per i docenti essi fornivano inconcussi criteri di validità e, in caso di discordanze, si consideravano i test come privi di significato. Era quindi normale porsi il problema dello stesso criterio, di sottoporlo direttamente ad uno studio critico, di esaminare gli esami stessi. Così appunto, insieme con mia moglie e con l'amico Henri Laugier, intrapresi la prima ricerca di docimologia⁹, riguardante l'esame di licenza elementare nella sessione estiva del 1922.

Una batteria di 6 test concernenti diverse capacità (permutazione di lettere, formazione di vocaboli, frasi assurde, analogie, ricordo immediato di parole, percezione visiva di mutamenti) fu somministrata a 117 scolari che sostenevano l'esame e si rilevarono le valutazioni attribuite in sede di scrutinio finale, nonché i voti riportati nelle diverse prove d'esame, suddividendoli in 3 gruppi: A (storia e geografia, recitazione, ortografia) riguardante soprattutto acquisizioni mnemoniche; B (composizione, lettura espressiva, aritmetica) riferentesi a capacità intellettuali; C (disegno, calligrafia, canto, ginnastica, cura del quaderno) legato a qualità diverse o capacità extra-intellettuali.

Era legittimo attendersi una correlazione debole o trascurabile fra il terzo gruppo e gli altri due, ed un po' più forte tra il secondo ed il primo. Orbene, tutte le correlazioni furono di valore medio, quasi uguale (0,461 di C con A e 0,406 con B, 0,429 di B con A). Sembrava insomma esistere un fattore generale atto a contraddistinguere un buon candidato, con un certo effetto di «alone» negli esaminatori, notissimo effetto d'iperestimazione o d'ipoestimazione generale dei buoni e dei cattivi.

La valutazione di scrutinio finale ha avuto con quella in sede d'esame una correlazione media (0,460 e 0,515) in due delle tre scuole dalle quali provenivano i candidati, ed una più forte nella terza.

Con l'insieme dei test di capacità, la correlazione delle prove del gruppo C fu decisamente negativa ($- 3,80$); positive invece, ma debolissime, coi gruppi A e B (0,228 e 0,144).

La correlazione di ciascuno dei test col risultato conseguito all'esame fu praticamente nulla, in particolare per i due che maggiormente dipendevano dall'intelligenza (analogie e frasi assurde).

La conclusione di quella prima indagine docimologica era stata la seguente:

La licenza elementare, esame scolastico, può fornire indicazioni su un tratto, non certamente trascurabile, dell'individualità infantile: l'attitudine scolastica, fun-

⁹ Ho proposto questo termine movendo dai vocaboli greci concernenti gli esami: *dokimè* e *dokimasia* (prova d'esame), *dokimàzo* (esaminare), *dokimastès* (esaminatore), *dokimasikòs* (atto ad esaminare).

zione complessa di tratti di carattere (buona volontà, docilità) e di capacità d'acquisizione e d'istruzione. Si tratta però sostanzialmente d'un dato alquanto povero e del tutto insufficiente. Ove si voglia praticare un orientamento razionale ed avvalersi di utili selezioni in questa o quella direzione, quando ad esempio si accede agli studi dell'ordine medio-superiore, è pacifico che non ci si può assolutamente limitare ad un siffatto esame di tipo tradizionale e neppure si può ad esso attribuire valore decisamente eliminatorio.

Non era stato affrontato il problema della fedeltà dei voti, in caso di ripetizione dell'esame o di cambiamento d'esaminatore.

[...]

CONCLUSIONE

Le ricerche docimologiche, intraprese da una quarantina d'anni a questa parte, esigono una sistematica prosecuzione. Sin d'ora possiamo stabilire alcuni dati essenziali.

I - Siccome negli esami tradizionali il caso ha una parte eccessiva, è necessario studiare misure capaci d'attenuarla per quanto è possibile:

1. *preparare al loro compito gli esaminatori e fornir loro cognizioni elementari di statistica.*
2. *attuare sistematicamente nelle commissioni una coordinazione nelle scale di valori.*
3. *per le prove scritte non implicanti dati che possano essere facilmente resi oggettivi, generalizzare la prassi d'una duplice correzione, con determinazione del voto definitivo previa discussione.*
4. *usare metodi d'indagine per valutare obiettivamente un'estensione delle cognizioni ed un livello di capacità acquisite senza timore di ricorrere a test standardizzati.*
5. *dissociare sempre il controllo delle formazioni educative dalla prognosi delle attitudini, confuso dal baccalaurèat nel suo svolgimento classico.*

II - In una necessaria preoccupazione d'adattamento, si determini la parte ed il valore degli esami e dei concorsi.

1. *Precisare negli esami il livello che si può esigere, fondandosi su una percentuale approssimativa delle ammissioni che si discosti sempre dal 50%.*
2. *In materia d'esami e di concorsi che aprano la via verso studi superiori, si inseriscano prove di prognosi circa le forme d'attitudini intellettuali richieste dagli studi in parola.*

III - Come politica educativa generale, si cerchi per quanto possibile di sostituire alle selezioni per esami un orientamento continuo fondato su una conoscenza profonda di ogni personalità, ottenuta col concorso di metodi obiettivi.

IV - Considerato il sempre più rapido aumento delle cognizioni umane, per fronteggiare il quale non basta un allungamento della scuola dell'obbligo:

1. *si rinunzi ad un enciclopedismo di base, diventato inaccessibile e si eviti un sovraccarico d'acquisizioni superflue, a pro' d'un solido minimum strettamente indispensabile.*
2. *si studi in tutti i campi la possibilità d'inserire una continua preoccupazione educativa, con un costante incitamento a non interrompere l'attività mentale.*
3. *si prepari durante la scuola dell'obbligo un uso integrale di tutti gli strumenti culturali che la civiltà moderna offre agli uomini e ci si preoccupi di suscitare curiosità ed interesse senza schiacciare la memoria sotto il peso di sovraccarichi transitori a causa d'esami mal concepiti che rischiano di recare come conseguenza il disgusto per la cultura.*

2. Stereotipi ed effetti di alone (de Landsheere, G., *Évaluation continue et examens. Précis de docimologie*)¹⁰

Nella situazione scolastica abituale, l'insegnante conosce ciascuno dei suoi alunni e può quindi dosare, sfumare le sue classificazioni in funzione di un effetto desiderato: da una parte si incoraggia sopravvalutando il lavoro, mentre altrove si dà prova di una eccezionale severità per dare uno choc che si spera salutare. In questi casi, l'insegnante agisce deliberatamente, in piena coscienza.

Nei fenomeni di stereotipia e di alone le cose vanno in tutt'altro modo.

(a) Stereotipia

Per *stereotipia*, intendiamo una inalterabilità più o meno dichiarata, che si instaura nel giudizio formulato nei confronti dell'allievo.

La stereotipia deriva da una contaminazione dei risultati. Un lavoro mediocre predispone a pensare che lo sarà anche il secondo; se questo si verifica, la tendenza a dare un voto mediocre al terzo lavoro si accresce ancora, e così di seguito. Nel professore sovraccarico di compiti da correggere, la deformazione si produce tanto più facilmente.

Si avrebbe torto a credere che la stereotipia influenzi solamente le valutazioni a base dichiaratamente soggettiva (dissertazione, composizione di storia e, in generale, risposte che implicano un giudizio di valore). Essa colpisce anche esercizi «oggettivi» come il dettato ortografico. L'esperienza seguente lo dimostra.

Un professore di lingua materna fa regolarmente dettati. Ben presto egli conosce gli alunni che abitualmente riescono meglio e quelli che riescono meno bene in questo esercizio. Se si determina la frequenza degli errori «dimenticati», non colti dal correttore, constatiamo che le dimenticanze in favore degli alunni bravi sono significativamente più elevate che quelle verso gli alunni meno preparati. Nel primo caso, l'insegnante si aspetta di non trovare errori; nel secondo, li "aspetta al varco"¹¹.

La situazione può diventare gravissima, se interviene una mancanza di coscienza professionale. Ci era stato segnalato (nell'ambito dell'insegnamento secondario) un caso di stereotipia dichiarata di cui un alunno era vittima per le versioni latine. A titolo di verifica, abbiamo fatto fare gli altri compiti successivamente da un altro a-

¹⁰ Tratto da: De Landsheere, G. (1971), *Évaluation continue et examens. Précis de docimologie*, Éditions Labor, Bruxelles; tr. it. *Elementi di docimologia. Valutazione continua ed esami*, Firenze, La Nuova Italia, 1973, pp. 30-32 e pp. 41-43.

¹¹ Vedere a questo proposito M. Zillig, *Beliebte und unbeliebte Volksschulerinnen*, in «Arch. für die ges. Psychologie», n. 12, 1934, p. 32, cit. da E. Hohn, *Der schlechte Schuler*, München, Piper, 1967, p. 105.

lunno della classe, da un alunno dello stesso livello didattico, classificato ottimo in un'altra scuola, poi da un laureato in filologia classica: il voto non è cambiato di un mezzo punto su venti.

Abbiamo ripetuto questo genere di esperienze in altre occasioni. Quanti genitori, inoltre, a volte molto competenti nell'ambito della materia della quale facevano occasionalmente il compito dei loro bambini, hanno provato qualche delusione o meraviglia quando sono venuti a conoscenza del «loro» voto!

Tuttavia sarebbe ingiusto dedurne che tutti gli insegnanti accusati in tal senso manchino di coscienza e di responsabilità. Sono i metodi di valutazione più che le persone che bisogna mettere in discussione.

(b) Effetto di alone

L'*effetto di alone* presenta un dichiarato carattere passionale. Spesso, si sopravvalutano le risposte di un alunno di bel portamento, dallo sguardo franco, dalla dizione gradevole. Tuttavia non bisogna generalizzare. Conosciamo professori che, per anti-conformismo, favoriscono una certa sciatteria o bizzarria che non sempre sono di buon gusto.

Anche la grafia può influenzare il correttore, sia per motivi di leggibilità, sia per motivi chiaramente affettivi. Gli specialisti della pubblicità sanno da molto tempo che il modo di presentare il messaggio esercita un'influenza considerevole sul suo effetto.

[...]

Altre Critiche

Molte altre critiche vengono rivolte agli esami. Il rapporto della commissione consultiva sugli esami di insegnamento secondario in Gran Bretagna, depositato fin dal 1911, ce ne fornisce una lista. Noi l'allungheremo un po', pur avendo l'impressione di non essere affatto esaurienti.

Per l'alunno:

- dedica troppa energia a ripetere le idee degli altri anziché a sviluppare la sua creatività;
- è premiato per apprendimenti spesso effimeri (si è dimostrato che fino all'80% delle conoscenze, soprattutto nozionistiche, apprese per l'esame, sono scomparse quindici mesi dopo);
- impara l'obbedienza passiva alle ingiunzioni;
- la sua attitudine ad esprimersi prende talvolta il sopravvento sul contenuto (mostreremo più tardi quanto gli handicap socio-culturali si rivelano soprattutto nell'ambito dell'espressione verbale);

- spesso rinuncia ad esprimere un giudizio personale, per piegarsi alle idee del professore che valuterà l'esame;
- ciò che impara può essere viziato da uno spirito di competizione che rasenta a volte la piaggeria;
- impara a speculare sulla fortuna: nell'esame tradizionale il piccolo numero di domande riflette idiosincrasie del professore, apre la via al gioco dei suggerimenti;
- gli esami tradizionali, la loro lunga preparazione e il periodo di affaticamento che ne segue, accorciano considerevolmente l'effettivo anno scolastico;
- gli esami tradizionali impediscono il lavoro di gruppo, ed esaltano invece il valore della performance individuale, origine d'egoismo;
- la minaccia d'insuccesso che essi fanno pesare — spesso raddoppiata dalla paura di un provvedimento familiare — incita alla frode. Il successo di coloro che frodano costituisce un cattivo esempio per gli alunni rimasti onesti;
- gli esami danno una idea falsa del lavoro adulto dove, più l'argomento è complesso e difficile, più si ricorre a consigli e a opere di consultazione.

Per il professore:

- insegna in funzione dell'esame, perfino delle esigenze particolari dei membri di una commissione esterna. Vedremo che in buone condizioni, lo spirito dell'insegnamento e dell'esame deve essere unico. Qui si allude ad esami estranei agli obiettivi educativi;
- è soffocato nella sua metodologia, se lo spirito dell'esame è estraneo agli obiettivi pedagogici che gli sembrano essenziali. Come praticare una pedagogia della scoperta, dell'esplorazione personale — processi lenti, ma generatori di apprendimenti profondi — se l'esame imposto è di carattere enciclopedico?
- schiavo del programma, non lascia procedere gli alunni secondo il loro ritmo e quindi digerire la materia;
- tende a dare troppa importanza alle attitudini e alle nozioni utili all'esame.

Su questa base, si è battezzato effetto di contraccolpo il fenomeno per il quale i professori modificano il metodo ed il contenuto del loro insegnamento in funzione dell'evoluzione degli esami imposti dall'esterno.

Importanti aspetti educativi possono così essere trascurati e spesso lo sono stati.

Riassumendo, gli esami tradizionali presentano generalmente gravi difetti di costruzione. La loro validità è discutibile. La valutazione dei lavori è, da parte sua, aggravata da grossi difetti. Inoltre, gli esami possono nuocere alla salute fisica e mentale degli alunni.

Aggiungeremo infine che, in alcuni casi, i professori svolgono il loro corso in funzione di un esame e non il contrario [...].

3. I limiti delle prove [...] di accertamento delle conoscenze (Domenici, G., *Manuale della valutazione scolastica*)¹²

Quei limiti strutturali producono *effetti* tanto più perniciosi quanto più alterano surrettiziamente il grado di affidabilità della rilevazione. Tra i più diffusi, e più studiati in letteratura perché particolarmente pericolosi, vi sono gli effetti di *alone*, di *contrasto*, di *stereotipia*.

Altri ancora, quali *l'effetto della profezia che si autoverifica* o *effetto Pigmalione* e quello della *distribuzione forzata dei risultati formativi*, pur dipendenti dalla strumentazione valutativa impiegata, sono prodotti soprattutto dal diffusissimo modello didattico che incardina le sue principali attività [...] nella «trasmissione delle conoscenze e delle informazioni» e nel «controllo degli apprendimenti conseguiti dagli allievi» (in questo caso la valutazione assolve più una funzione fiscale che formativa [...]).

Effetto di alone. Consiste nell'alterazione del giudizio riferito a una specifica prestazione in forza dell'influenza esercitata dai precedenti giudizi. Questi si espandono ben oltre il loro spazio originario, proprio come l'alone di una macchia, fino a caratterizzare, in modo impertinente, anche gli elementi di strutturazione della nuova valutazione. Accade così che l'apprezzamento di una prova finisca con il non rispecchiare la qualità reale di quella e solo di quella specifica prestazione presa in esame.

Dal punto di vista di un allievo che ha sempre avuto voti o giudizi positivi la cosa potrebbe risultare persino piacevole, anche se ingiusta; ma da quello di chi per una qualche ragione è stato giudicato negativamente, l'effetto risulterebbe doppiamente ingiusto e frustrante. Ben lo sa chi, per rimontare posizioni inizialmente sfavorevoli, ha dovuto dar prova di possedere le qualità richieste molte e molte volte prima di un loro «oggettivo» riconoscimento. In altri termini, si è innavvertitamente disposti a trovare una qualsiasi giustificazione alle *défaillances* di chi è andato frequentemente bene nel passato, così come si sospende il giudizio (magari mettendoci accanto un bel punto interrogativo) su prestazioni in sé positive di coloro per i quali si sono registrati in precedenza risultati giudicati negativi.

L'effetto di alone si manifesta anche nella pratica del calcolo della media aritmetica dei voti (o dei giudizi equivalenti) assegnati nell'arco dei trimestri o dei quadrimestri. Pur prescindendo, come vedremo più avanti, dall'infondatezza di un simile calcolo quando si impieghino scale di misura ordinali, quali sono ap-

¹² Tratto da: Domenici, G. (1993), *Manuale della valutazione scolastica*, Roma-Bari, Laterza, pp.42-46.

punto i voti, se dopo una eventuale sequenza di «quattro» l'allievo mostri di padroneggiare tutte le conoscenze fino a quel momento promosse, sarebbe un non senso attribuire come voto sintetico quello risultante dalla media di tutti i voti. Significherebbe quanto meno inquinare il giudizio sulle prestazioni cognitive con elementi a esso estranei quale ad esempio la costanza nell'applicazione (per altri versi importantissima) e simili.

Effetto di contrasto Si può verificare soprattutto durante le interrogazioni orali, ma anche con la correzione dei compiti scritti a stimolo e risposta aperti. Consiste, come l'effetto di alone, in una soprastima o sottostima di una determinata prova, ma questa volta rispetto a standard di prestazioni ideali del docente, o, più frequentemente, rispetto a precedenti, contestuali o immediatamente successive prove di altri allievi. Se ad esempio uno studente espone in modo brillante una certa questione e un altro in modo quasi stentato, magari solo dal punto di vista linguistico, in assenza di misurazioni oggettive del possesso delle abilità sottoposte a controllo, si è portati, quasi inesorabilmente, a sottostimare la prestazione del secondo rispetto a quella del primo o viceversa a soprastimare la prima rispetto alla seconda. I giudizi che ne conseguono risentiranno più del contrasto immediatamente percepito tra le due esposizioni che dall'analisi del possesso dei requisiti sottoposti davvero a controllo.

Effetto di stereotipia Consiste nella scarsa alterabilità dell'opinione che ci si fa di un allievo, dal punto di vista cognitivo o non cognitivo, così che ogni giudizio potrebbe risentire positivamente o negativamente della generalizzazione dell'opinione originaria. Per molti versi possono a questo assimilarsi i cosiddetti *effetti di pregiudizio, di empatia* e simili.

Certamente risultano del tutto diverse e assai spesso accettabili sul piano valutativo talune prove scritte quali esercizi, problemi e simili, almeno quando non costringano a seguire meccanicamente una sorta di algoritmo risolutivo della questione, ma offrano sovrabbondanza di dati rispetto a quelli strettamente necessari per svolgerli. Ma anche di questi strumenti si parlerà nel capitolo delle prove semistrutturate.

Qui e ora, è utile ribadire che molta parte delle cosiddette prove tradizionali (sono in realtà quelle più massicciamente usate ancora oggi), a causa della loro struttura e delle finalità d'uso, non consentono la rilevazione di ciascuna e di tutte le abilità che si vorrebbero verificare, costringendo in questo modo ad assumere iniziative e attività successive alla verifica non sempre rispondenti alle necessità di apprendimento di ciascuno e di tutti gli allievi, e quindi incapaci di far correggere, eventualmente, approcci didattici, scelte di mezzi e risorse, canali comunicativi di fatto rivelatisi poco pertinenti ed efficaci e/o a consolidare scelte mostratesi invece particolarmente felici e produttive.

Le considerazioni appena fatte non devono indurre a comportamenti di pregiudiziale ripulsa di quelle prove, quanto piuttosto di consapevole conoscenza dei loro limiti metrologici intrinseci, perchè come vedremo, cambiandone le finalità valutative specifiche e buona parte della loro struttura, possono trovar posto, anche se di necessità assieme ad altri strumenti, nell'armamentario valutativo che gli insegnanti dovrebbero impiegare.

Per concludere la breve carrellata sugli «effetti» prodotti dalla strumentazione valutativa diffusamente usata e dalle funzioni di verifica prevalentemente svolte nel fare scuola quotidiano, risulta utile illustrarne ancora due, anche se, come si è detto, essi derivano primariamente e ancor più dalla tipologia dell'organizzazione della didattica in auge ancora oggi.

Effetto della distribuzione forzata dei risultati Consiste nell'accettazione, non sempre implicita da parte di molti docenti, dell'ipotesi secondo cui gli esiti della formazione non possano non rispecchiare l'andamento della curva normale di molti fenomeni naturali, o degli errori accidentali delle misurazioni, ovvero non possano che distribuirsi secondo la famosa curva a campana di Gauss. Così che, necessariamente, al termine di un anno scolastico, di un ciclo o di un corso di studi solo un 20 per cento circa degli allievi raggiungerebbe, a livelli ritenuti ottimali, gli obiettivi generali della formazione, un altro 20 per cento si situerebbe ai livelli più bassi (andando magari ad alimentare il *drop-out*), mentre l'altro 60 per cento circa si attesterebbe su livelli di apprendimento considerati, a seconda dei punti di vista, più o meno sufficienti, oppure del tutto mediocri, come con buona approssimazione viene illustrato dalla fig. 3.1.

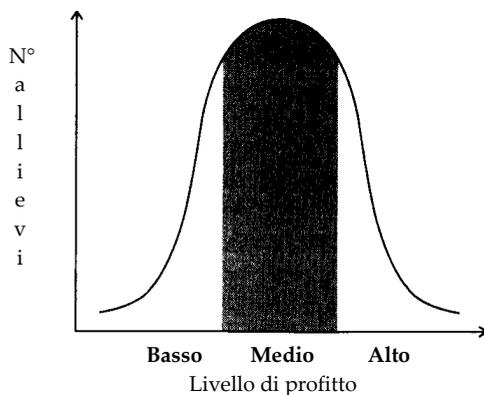


Fig. 3.1. La distribuzione «forzata» dei risultati scolastici.

L'ineluttabilità di una simile distribuzione, ritenuta tale perché puntualmente si verifica al termine di ogni anno o ciclo formativo in quelle situazioni in cui

l'organizzazione dei processi formativi è di tipo tradizionale, in realtà rappresenta una sorta di costrizione dei risultati dovuta alla mancata valorizzazione di molte delle funzioni specialistiche della didattica e della verifica. Alla sua base vi è l'interpretazione degli allievi come soggetti appartenenti a una popolazione che possiede caratteristiche distribuite casualmente, e che pertanto nei loro comportamenti cognitivi non potranno discostarsi da una qualsiasi «popolazione» di eventi casuali. Sappiamo invece che, proprio in forza della ormai massiccia esposizione prescolastica alla televisione (che garantisce almeno una omogeneità linguistica nazionale di fatto inesistente solo una decina di anni fa) e dei risultati che diffusamente si registrano almeno dopo i primi due tre anni di esperienza scolastica, la popolazione degli studenti, quanto meno in rapporto a certe precondizioni dell'apprendimento scolastico, non è più di tipo casuale: nonostante talune differenze interindividuali, molti elementi di conoscenza accomunano a tal punto i suoi componenti da non poterli considerare come appartenenti a un universo del tutto indifferenziato.

Se gli esiti formativi risultano poi differenziati, evidentemente la scuola, anziché operare costruendo sulla base comune e valorizzando le differenze, avrà invece favorito l'ampliamento delle diversità attraverso una piatta azione didattica. La scuola appare in tal modo non già una struttura intenzionalmente volta alla modificazione significativa delle conoscenze degli allievi, ma semplicemente come apparato casuale di promozione casuale degli apprendimenti.

Se anche si considerassero gli allievi una popolazione casuale, magari in rapporto alla diversità degli interessi e delle eventuali attitudini e delle motivazioni e dei valori presente all'inizio dell'esperienza scolastica (peraltro un po' meno plausibile oggi di quanto non lo fosse negli anni Sessanta, quando questa critica venne mossa dai teorici del *Mastery learning*), la scuola si comporterebbe - per fare un parallelo con la medicina - come uno strano ospedale dove, nonostante le accorte diagnosi e le molteplici terapie, gli ammalati vengono dimessi nelle stesse condizioni, se non peggiori, di quelle d'ingresso.

Effetto Pigmalione È l'effetto edipico della profezia che si autoverifica. Non nel banale senso magico per cui basti di per sé una certa previsione perché essa si attui, quanto piuttosto in quello per cui a certe predizioni del successo o dell'insuccesso scolastico di ciascun allievo, fatte dai docenti sulla scorta di taluni elementi informativi disponibili, conseguono comportamenti verbali e non verbali spesso inconsapevolmente coerenti con quelle previsioni. Queste finiscono così con l'orientare e indirizzare non solo gli atteggiamenti del docente, ma anche e persino quelli dell'allievo, fino a strutturare e consolidare le condizioni che di fatto facilitano la concretizzazione degli esiti delle loro ormai coincidenti aspettative.

Ovviamente la questione ha senso con riferimento ai grandi numeri, poiché è sempre possibile e per fortuna, che almeno in taluni allievi essere consapevoli della

«disistima» degli altri catalizzi invece di inibire una reattività positiva. Data l'età degli allievi è però difficile che il non ricevere fiducia e gratificazioni connesse alla valorizzazione delle competenze possedute, l'essere bloccati in iniziative magari goffe ma finalizzate all'autovalorizzazione, ecc., innesti una sfida a lungo termine per mostrare di essere diversi da quello che soprattutto i docenti fanno loro credere di essere.

4. Idòla (effetti distorsivi) nella valutazione interna (Benvenuto, G., *Gli esami sono maturi?*)¹³

La scuola, almeno così come l'abbiamo intesa fino ad oggi, con soggetti (docenti e studenti) che valutano altri soggetti (fondamentalmente studenti) è probabilmente tra i principali contesti ad alta presenza di distorsioni valutative.

Le distorsioni valutative ci ricordano da vicino gli idòla baconiani¹⁴ che assediavano la mente umana sì da rendere difficile l'accesso alla verità, così come gli effetti distorsivi nella valutazione si frappongono ad un rigoroso e condiviso accertamento degli apprendimenti raggiunti.

Come Bacone indicò, la conoscenza deve prevedere una solida fase di raccolta di informazioni e dati tale da permettere una costruzione rigorosa di elementi di conoscenza, di giudizi conoscitivi potremmo dire per il nostro discorso sulla valutazione scolastica. E questo procedere conoscitivo, dalla raccolta dei singoli elementi conoscitivi alla delineazione di un'attenta valutazione, di un mirato profilo conoscitivo è un procedere induttivo.

È proprio l'induzione che, oltre a fornire una chiave costruttivistica dei giudizi, previene una gran parte delle distorsioni che provengono da abusi soggettivistici, cioè dall'avere delle idee e visioni personali (e a volte preconette) che influenzano eccessivamente, dalle quali deduciamo i giudizi senza tener conto dell'effettivo riscontro della realtà esperienziale. Il procedere induttivo, oltre a fornire il senso di progressiva costruzione del giudizio valutativo, indica una forma di conoscenza che non ipotizza la necessaria presenza di una realtà oggettiva. Certo per costruire un giudizio, una qualsiasi valutazione come quelle che costantemente formuliamo a scuola, abbiamo bisogno di procedere con logica e rigore riconoscibile. Ecco la necessità di smascherare gli idòla della valutazione, per poi procedere a forme di rilevamento condivise e alla costruzione di giudizi più legati alla realtà fattuale che a quella soggettiva.

¹³ Tratto da: Benvenuto, G. (2007), *Gli esami sono maturi? Critiche e prospettive docimologiche per gli esami di stato*, Anicia, Roma, pp. 45-48.

¹⁴ Francesco Bacone (Londra, 1561-1626) nel suo progetto di *Instauratio Magna* critica la cultura tradizionale e propone una riforma del sapere. Nel *Novum Organum*, unica parte scritta del suo progetto, Bacone prima polemizza e critica la cultura tradizionale (*pars destruens*) e poi delinea un metodo di ricerca (*pars construens*). Le critiche sono quattro, e le presenta come i vizi fondamentali che impediscono agli uomini di costruire un nuovo sapere, e sono: gli idoli della tribù (fondati nella natura umana), idoli della spelonca (propri dell'individuo), idoli del mercato (legati al linguaggio), idoli del teatro (derivanti dalle varie filosofie). La liberazione dai fantasmi (idòla) è il presupposto per conoscere la natura nella sua concretezza. Il metodo di ricerca è l'induzione, procedimenti sistematico che "consiste nel trarre e far sorgere assiomi dall'esperienza".

La principale famiglia di distorsioni è sicuramente legata alla soggettività di chi valuta. Troppa letteratura (narrativa e filmica), oltre all'esperienza diretta, ci aiuta a ricordare quel docente che ha "piegato" la valutazione da un'altra parte rispetto alla nostra opinione, o quel momento nel quale elementi di discriminazione o di parzialità sono stati manifesti.

Il modo più semplice per eliminare l'interferenza della soggettività nella valutazione è quello di predisporre forme di verifica che non prevedano la valutazione in diretta a carico di uno o più soggetti valutatori, ma con forme di valutazione indirette, quali le prove strutturate. Altra possibilità, ma assai meno proponibile almeno in larga scala nella scuola, è quella del confronto intersoggettivo, cioè compiuto da più valutatori sempre in base a criteri prestabiliti o concordabili.

Questi ragionamenti portano nella scuola all'uso sempre più frequente di prove strutturate quali i test, a misurazioni cioè che si presentano formalmente più oggettive. Il principio docimologico qui sotteso è molto semplice: nelle prove strutturate massimo è l'accordo sulla predefinizione delle risposte, e infatti esse permettono una correzione automatizzata, nelle prove con diverso grado di strutturazione si deve invece ricorrere all'accordo nella valutazione da parte di più esaminatori (studenti o docenti).

Ma le prove strutturate non risolvono il problema della valutazione scolastica. Eludiamo alcune distorsioni ricorrendo ad esse ma abbiamo bisogno di procedere a forme di verifica per quelle abilità che non possiamo saggiare in esclusiva modalità strutturata. I test introducono una semplificazione nelle procedure, un notevole risparmio di tempo e di sforzi nella correzione che spesso è a dispetto della profondità di analisi e della valutazione di abilità cognitive di livello superiore. E non stupisce quindi che docenti, soprattutto di area scientifica e tecnologica, riconoscono la superiorità che prove di diverso tipo hanno nella valutazione della complessità degli apprendimenti, ma si rivolgono con maggior frequenza a quelle strutturate, per mancanza di tempi e necessità di sintesi (cfr. Ben-Chaim, Zoller, 1997).

La soluzione [...] sta nell'impegnare strumenti di verifica diversificati nei differenti tempi della didattica e nel preparare i docenti a specifiche metodologie per l'accordo nella valutazione. I test, forme di valutazione strutturate, aiutano a individuare i punti di forza e debolezza in senso comparativo e in riferimento a standard riconosciuti e condivisi largamente. Le altre forme di valutazione scolastica, quali quelle semi-strutturate, permettono valutazioni maggiormente analitiche e invitano lo studente a impegnarsi su piani di produzione delle conoscenze e conseguente interpretazione e riflessione critica. Nel caso di prove semi-strutturate, quali ad esempio le prove di scrittura, si dovrebbe procedere con metodologie che garantiscano una maggiore trasparenza e affidabilità nella correzione. Ad esempio si potrebbe procedere alla revisione delle prove valutate equivalenti, per riequilibrare eventuali disparità, oppure valutare alcune prove a campione (o proprio quelle che hanno richiesto maggior concentrazione nella formulazione del giudizio) dopo alcuni giorni dalla prima valutazione. Nel caso di prove orali, si potrebbe lavorare con

coppie di studenti, variando la composizione nei diversi giorni, oppure far utilizzare agli studenti stessi specifiche griglie di correzione di determinati elementi discussi e concordati con gli studenti stessi. Soprattutto poi nelle verifiche che assumono una forte caratterizzazione misurativa e selettiva, quali ad esempio le "interrogazioni quadrimestrali" o all'interno di esami che gli effetti indesiderati, le distorsioni nella valutazione, presenti nella gran parte delle valutazioni, devono essere contrastati cercando di controllare l'affidabilità delle correzioni e ricorrendo a eventuali correttivi.

Certo alcune forme di distorsione sono meno gravi di altre. Il danno sembra maggiore tanto più si presentano nelle fasi di misurazione o selezione. Una cosa è fornire un giudizio parziale o impreciso in fase di discussione didattica, altra cosa è utilizzarlo per stilare una graduatoria di merito per il passaggio di corso. Questa distinzione tra diversi momenti valutativi ci costringe a dare dei pesi alle diverse forme di distorsione: nel caso di una dialettica processuale siamo disposti anche ad accettare alcune forme distorsive, mentre nel caso di selezioni dobbiamo assolutamente evitarle per un generale discorso di equità e giustizia valutativa.

Ed è in questa prospettiva che passeremo quindi rapidamente in rassegna alcuni delle principali distorsioni, o deformazioni arbitrarie nella valutazione, ripartendole per famiglie di problemi (cfr.: De Landsheere, 1971; Domenici, 1993; Giovannini, 1994; Mason, 1996; Benvenuto, 2003). L'obiettivo non è semplicemente quello di campionarle, ma di individuare quei solidi motivi che spingono all'uso più razionale di forme di verifica, a procedure meno discutibili dal punto di vista dell'attribuzione di voti, o a metodologie che permettano il contenimento di quelle distorsioni.

Bibliografia

- Benvenuto, G. (2003), *Mettere i "voti" a scuola. Alcune questioni docimologiche*, Carocci, Roma.
- Ben-Chaim, D. and Zoller, U., (1997), "Examination-Type Preferences of Secondary School Students and Their Teachers in the Science Disciplines", *Instructional Science*, 25, 347-367.
- De Landsheere, G. (1971), *Évaluation continue et examens. Précis de docimologie*, Éditions Labor, Bruxelles, (tr. tt., *Elementi di docimologia. Valutazione continua ed esami*, La Nuova Italia, Firenze, 1973).
- Domenici G. (1993), *Manuale della valutazione scolastica*, Laterza, Bari.
- Giovannini, M.L. (1994), *Valutazione sotto esame*, Ethel editoriale Giorgio Mondatori, Milano.
- Mason, L. (1996), *Valutare a scuola. Prodotti, processi, contesti dell'apprendimento*, Cleup, Padova.

Idòla (distorsioni e forzature) nella valutazione

Stimolo/strumento	Docente/valutatore	Situazione
Interferenze connesse alla rilevazione o alle procedure valutative	Disparità di chi valuta legate a tratti comportamentali e aspettative	Interpretazioni parziali o forzate connesse al contesto
<p>1) <i>Contiguità (prossimità) domande</i>: problema degli strumenti di rilevazione dovuto alla vicinanza di alcune domande. La contiguità può interferire e influenzare le modalità di risposta ma anche l'interpretazione del valutatore</p>	<p>5) <i>Indulgenza/severità</i>: generosità e sopravvalutazione (effetto chioccia: sono tutti miei pulcini) o all'opposto eccessivo rigore e sottovalutazione (effetto XXX: sono tutte capre). Questi comportamenti valutativi spesso sconfinano nell'errore sistematico</p>	<p>11) <i>Contraccolpo</i> (valutazione orientata/finalizzata): si selezionano gli elementi di verifica intermedia in funzione degli esami/verifiche finali</p>
<p>2) <i>Successione/contrasto</i> (contiguità domande e sovra-sotto stima per contrasto)</p>	<p>6) <i>Pigmalione (aspettative)</i>: adeguamento degli studenti alle aspettative dei docenti o rilevazione da parte del docente dei soli risultati attesi</p>	<p>12) <i>Maturazione</i>: il presunto profitto dello studente è in effetti dovuto alla sua crescita psico-fisica</p>
<p>3) <i>Contiguità prove</i>: quando le prove sono troppo ravvicinate e di natura simile l'interferenza può provocare o una demotivazione o un naturale miglioramento</p>	<p>7) <i>Pregiudizio</i>: In base ad aspettative si manipolano/selezionano prove (stimoli/domande)</p>	<p>13) <i>Interferenza del contesto</i> (classe studenti, genitori)</p>
<p>4) <i>Tempi e modalità di correzione</i></p>	<p>8) <i>Stereotipia</i> (forte incidenza di precedenti giudizi)</p>	<p>14) <i>Contagio</i> (conoscenza delle valutazioni degli altri colleghi)</p>
	<p>9) <i>Alone</i>: tendenza del valutatore ad accentuare alcune caratteristiche e a generalizzare di conseguenza la valutazione</p>	
	<p>10) <i>Distribuzione forzata dei risultati</i> (assimilazione alla curva normale)</p>	

5. Misurazione e valutazione (Visalberghi, A., *Misurazione e valutazione nel processo educativo*)¹⁵

I. Misurazione e valutazione

La parola «misura» ha due significati principali, che non sono scollegati come taluni aspetti della vita moderna potrebbero far temere. Non c'è nessuna ragione di fondo per cui la «misura» intesa come operazione di conteggio o confronto non debba accompagnarsi con la «misura» intesa come abito di equilibrio e discrezione. Si potrebbero fare, è vero, sottili analisi circa l'origine classica dei due significati ed il loro uso rinascimentale, ma non crediamo che i risultati sarebbero in contrasto con la semplice osservazione di buon senso che l'abito stesso del misurare, implicando l'attitudine a vedere un più ed un meno dove il giudizio affrettato scorge qualità assolute è esso stesso un abito di riflessività, di moderazione e di prudenza. Come si spiega allora che la tendenza contemporanea a «misurare » tutto si accompagni spesso con atteggiamenti completamente opposti? Non pretendo di risolvere in questa sede un così vasto problema, per quanto evidentemente in pochi casi esso può apparire tanto pertinente come quanto la «misura» investe il campo dei valori più propriamente umani – intelligenza, carattere, cultura -; e li investe, si lamenta da molti, presentandosi nella più deplorabile unilateralità di significato.

Mi limiterò a congetture che forse vale per la «misura» ciò che vale per la scienza in generale: la cosa più grave è di volersi fermare a metà strada, dimenticando o non volendo ricordare che essa sola, come la famosa lancia d'Achille, può guarire i mali che produce. Fermarsi a metà strada significa, di solito, tentare uno sfruttamento affrettato di risultati provvisori e parziali, tentarlo in base a vecchi abiti che non sono quelli che la scienza stessa, coraggiosamente perseguita, fa maturare, ma appartengono invece all'oscuro passato.

Così John Dewey vedeva nell'uso distorto dei *tests* di intelligenza, con cui taluno pretendeva d'istituire una sorta di classismo o castalismo aggiornato, «un segno della influenza che ancora esercitano su di noi gli ordinamenti feudali»¹⁶. Del resto, non furono certo gli psicologi che idearono ed elaborarono i *tests* d'intelligenza a prospettare o incoraggiarne impieghi siffatti; da essi ci vennero piuttosto seri

¹⁵ Tratto da: Visalberghi, A. (1955), *Misurazione e valutazione nel processo educativo*, Edizioni di Comunità, Milano, pp.11-18.

¹⁶ «Individuality, Equality and Superiority», in *The New Republic* del 13 dicembre 1922.

ammonimenti a non usarli come giocattoli per pericolosi trastulli. Oggi la misurazione dell'intelligenza è diventata, per sua interna legge di sviluppo, una disciplina talmente complessa e criticamente consapevole della selva di interrogativi in cui si muove, che difficilmente verrebbe in mente a qualcuno che ne sia minimamente informato, di farne un uso così sciocco come quello di dividere per suo mezzo le persone in intelligenti e stupide e di determinare così la loro destinazione sociale. In pochi individui ho trovato un senso della «misura» così sviluppato quanto nei pochi veri specialisti in psicomelia che ho avuto la fortuna di conoscere.

Credo che all'incirca lo stesso discorso valga per l'altro tipo fondamentale di *testing* quello del profitto o del rendimento scolastico. L'analogia tra *tests* psicologici e *tests* di profitto è meno significativa di quel che si crede: questi ultimi sono strumenti pedagogici («strumenti di lavoro rigorosamente pedagogici» li qualifica il Decroly) il cui apprestamento e il cui impiego vanno riservati, di norma, agli educatori forniti della necessaria competenza. Il loro uso da parte degli psicologi non è meno ricco di pericoli dell'uso dei *tests* psicologici da parte degli educatori. Tuttavia, anche a prescindere dalle somiglianze esteriori che sono parecchie, c'è il fatto innegabile che in ambedue i casi si tratta di «misurazioni» applicate a operazioni mentali. Una tale pretesa di «misurare lo spirito» appare a taluni talmente irraguardosa per lo «spirito» stesso da destinare senz'altro all'insuccesso qualunque tentativo di mostrar loro che queste pratiche diaboliche non solo non sono poi tanto diaboliche, cioè non negano il mondo di valori in cui essi hanno fede (purchè si tratti di un effettivo mondo di valori, e non di un bagaglio di rugginose superstizioni), ma addirittura sono suscettibili di cooperare validamente all'avvento del regno dello spirito! Il *testing* culturale, non meno di quello psicologico, costituisce un ausilio a comprendere meglio il singolo individuo, la persona umana nella sua originalità peculiare; può certo essere distorto ad altri fini, contro la volontà ed i suggerimenti degli studiosi che se ne occupano, ma se un simile timore autorizza al rifiuto, perchè non rifiutare in blocco tutte le tecniche su cui si fonda il mondo moderno?

Nel caso dei *tests* o prove oggettive e di profitto¹⁷ un primo equivoco da

¹⁷ Usiamo l'espressione «prove oggettive di profitto» come corrispondente italiana dell'espressione inglese «objective achievement tests» il termine copre dunque tutta la vasta gamma di possibili *tests* rivolti a saggiare il profitto piuttosto che l'intelligenza o altro, purchè possiedano il carattere dell'«oggettività» (siano cioè suscettibili di venir corretti in modo identico da qualsiasi correttore sufficientemente abile - ma in casi estremi anche da una macchina -, di aver quindi assegnato un punteggio con criteri esattamente definiti e di eliminare perciò completamente la cosiddetta «equazione personale» del correttore). Distingueremo però fra «prove oggettive (di profitto) di classe» («class-room» o «informal objective (achievement) tests») e «prove oggettive (di profitto) tipo» o «prove tipo» («standardized objective (achievement) tests »): le prime sono prove oggettive più o meno «alla buona che gli insegnanti preparano per la loro classe o le loro classi, le seconde sono le

dissipare è quello che nasce dalla credenza che, con il loro impiego, si sostituisca la «misurazione» alla «valutazione». Ammesso che il testing sia senz'altro «misurazione» nel senso pieno del termine (ciò è vero soltanto per le prove oggettive tipo, corredate da norme precise per l'assegnazione del punteggio), rimane tuttavia assolutamente infondata la credenza che esso elimini ogni giudizio valutativo. Vale in realtà l'opposto: la pratica del testing mette in luce tali e tanti problemi di autentica valutazione da obbligare ad un impegno in questa direzione quale raramente vien dato di trovare in altri casi.

L'equivoco nasce in primo luogo dalla comoda illusione che i metodi tradizionali di accertamento del profitto non comportino aspetti tecnici e quantitativi, ma siano puri atti «originali», giudizi distintamente qualitativi, penetranti «intuizioni» della realtà che ci sta di fronte. Per fortuna, le cose stanno in modo diverso: *l'arte di esaminare* è, come tutte le arti, una sintesi di tecnica e di atti originali d'intuizione. La tecnica, o le tecniche, che un esame orale o scritto comporta, sono di assai varia natura, e provengono in parte dalla tradizione, in parte sono desunte dalla osservazione e dall'esperienza. Che io sappia, nessuno ne ha mai tentato un'analisi esauriente. Accanto ad una quantità di accorgimenti psicologici, spesso molto sottili, intervengono varie considerazioni quantitative (circa l'importanza ed il numero delle nozioni saggiate, e circa il grado di conoscenza ed assimilazione delle stesse) in parte rispecchiate dagli stessi verbali d'esame, quando ci sono; infine c'è il problema dell'assegnazione del voto (la funzione intermedia del giudizio è del tutto apparente), ed è un problema

prove oggettive approntate in vista di un impiego molto più largo, e che prima di essere utilizzate per scopi impegnativi hanno subito tutta una serie di controlli, in gran parte di natura statistica (soprattutto per eliminare i quesiti poco o niente affatto discriminativi), ed infine sono state somministrate, sempre in via sperimentale ad un campione opportunamente scelto e sufficientemente ampio, al fine di determinare, sempre statisticamente, quale sia la loro difficoltà media, quale la dispersione tipica del punteggio, e spesso di determinare anche la corrispondenza fra la scala dei punteggi «grezzi» e una graduatoria o classifica tipica, desunta dai risultati complessivi ottenuti dal campione («centilaggio», «decilaggio», «tetronaggio», ecc.).

Raymond Buyse ed altri autori vorrebbero riservare il termine «test» solo alle «prove tipo» e l'espressione «épreuves objectives» la restringono alle sole «prove oggettive di classe», ma non ci sembra convenzione che possa plausibilmente adottarsi: non vediamo infatti la ragione di discostarci talmente dall'uso anglosassone, e in particolare di attribuire un significato antonomastico così ristretto al vocabolo «test» che continua ad avere in inglese l'accezione amplissima di «prova» dei più vari generi. Si noti poi che in italiano il continuo impiego del termine inglese «test» riesce assai meno tollerabile che in francese, tant'è vero che anche in psicologia lo si sostituisce spesso (con «reattivo»); non precludiamoci un'analogia possibilità in pedagogia, anche se, per un bel po', di «prove tipo» non avremo da parlare troppo, per la buona ragione che di adatte all'impiego nelle scuole italiane non ce ne sono nè se ne potranno preparare tanto presto.

veramente grosso, come sanno tutti gli esaminatori coscienziosi, anche per il fatto che, secondo che spiegherò più avanti, alla nostra classificazione in decimi non può venir assegnato nessun significato univoco. In tutti gli esaminatori c'è la tendenza ad interpretare la scala da 1 a 10 come una «scala d'intervallo», e lo dimostra l'uso dei mezzi punti, dei «più» e dei «meno», il che, ai nostri fini presenti, ci sembra sufficiente a concludere che un aspetto tecnico ed in parte «misurativo» è presente, in forme più o meno spiegate, anche nei sistemi tradizionali ad esame. L'analisi minuta (l'analisi, per esempio, della registrazione stenografica di un'interrogazione orale) ci dimostrerebbe che le considerazioni di questo tipo sono strettamente fuse con atti di vera e propria valutazione non riconducibili a schemi tecnici di sorta. C'è chi vede in questa stretta fusione il segno della superiorità inattaccabile dei sistemi tradizionali di valutazione rispetto alla «meccanicità» ed «aridità» delle prove oggettive. Quest'opinione è per un verso fondata, ma per un altro nasce anch'essa da un equivoco. Statuisce essa infatti il valore di quello che retoricamente si chiama «contatto d'anime», ma che in termini più semplici e più validi può dirsi la superiore letizia del commercio spirituale (o, deweyanamente, della «transazione») fra persone pensanti, di fronte ad altri modi più esteriori e più poveri di rapporto interpersonali. La prima è senza dubbio una delle forme più alte, anche se forse la più difficile a realizzarsi sul serio, che può assumere l'opera educativa. L'equivoco sta in ciò, che le superiori virtù del dialogo hanno, in se stesse, poco a che fare con quelle che sono le finalità degli esami ed in genere delle prove di accertamento del profitto. Noi dobbiamo domandarci non già quali tipi di prove siano più «calde» ed umane, ma quali tipi di prove servano meglio ai fini di un giudizio impegnativo e motivato ricco di conseguenze pratiche importanti. E che deve perciò essere giusto ed obiettivo quant'è possibile.

Ma forse l'espressione «ai fini di un giudizio» non risulta chiara in connessione con l'impiego dei *tests* o prove oggettive. Molti pensano che i *tests* giudichino da soli, cioè bastino da soli a spedire i singoli individui verso destini diversi, suppergiù come le pesatrici automatiche degli uffici postali destinano a diversi scomparti le lettere e i plichi di diverso peso. Ora è vero che i *tests* possono funzionare in questo modo, in qualche caso. Per esempio, l'ammissione ad un qualche corso di un qualche *College* statunitense può essere subordinata al conseguimento di un punteggio minimo in un *test* sui fondamenti della materia trattata. Ma sono usi molto rari, e del resto giustificabili, tanto più che gli insufficienti possono ricorrere, di solito, a prove d'appello anche d'altro genere, e che gli approvati non si tatueranno il voto sulla fronte, ma dovranno subito cominciare a dar prova in vari altri modi delle loro capacità. Nessun esame veramente importante si decide con un *test* o con una media automatica fra diversi *tests*, neppure negli Stati Uniti, ch'è la nazione dove il *testing* del profitto è più diffuso. Ci sono sempre altre prove, compiti scritti d'altro genere, e di solito anche un colloquio - per non dire della varietà di disparati elementi presi a volta a volta in considerazione, non esclusi i rapporti sulle qualità sociali o sportive del candidato. L'atto decisivo è dunque un giudizio, una valutazione [...], in cui i risultati dei *tests*

entrano come dati di fatto molto importanti, ma non esclusivi.

Ma la valutazione non interviene soltanto *dopo* l'impiego dei *tests*, interviene anche *prima*, cioè nella fase del loro apprestamento. Si tratti di semplici prove oggettive di classe o di prove oggettive tipo, tutto il loro valore (più esattamente, la loro «validità») deriva dal modo in cui si è valutata l'importanza delle nozioni da richiedere, degli abiti intellettuali da saggiare. Per le prove tipo, non ci si accontenta molte volte del giudizio diretto dei competenti, ma si tenta di determinarne la «validità» studiando la loro correlazione con *criteri* opportunamente scelti. Si tratta allora di valutare questi criteri. In certi casi, come vedremo, i criteri sono costituiti da medie fra i risultati di un gran numero di prove scritte di tipo comune, scelte con speciale accuratezza e corrette con eccezionale attenzione ed impegno. Le capacità valutative di decine di specialisti servono da fondamento alla validità del *test*. Ma il *test* come tale non valuta più, misura. Esso rappresenta un momento intermedio, di pura misurazione, inserito fra serie complesse di atti valutativi. Ciò che nelle prove tradizionali è fuso assieme, e talvolta confuso, nella pratica del *testing* viene accuratamente distinto. Nonché sopprimere la *valutazione*, la *misurazione* nasce dalla valutazione e nella valutazione confluisce. Essa mantiene però una provvisoria, ma nettissima autonomia, che le permette di raggiungere, come vedremo, una precisione ed un'attendibilità altrimenti impossibili.

6. Quando possiamo fidarci della valutazione (Vertecchi, B., Origini e sviluppi della docimologia)¹⁸

I risultati della ricerca sugli esami di *baccalauréat* misero in evidenza una grande diversità nei criteri di giudizio da parte dei commissari. Non si poteva nutrire se non una scarsa fiducia nell'esito degli esami, se a seconda del commissario che avrebbe espresso il giudizio un allievo poteva vedere il proprio impegno coronato da un successo o da un insuccesso. Eppure alla ricerca furono chiamati a partecipare insegnanti esperti, ciascuno dei quali era, come singolo, degno di fiducia, e senza dubbio non privo di qualità personali.

Come spiegare la contraddizione che veniva a determinarsi?

Come era possibile che insegnanti esperti potessero formulare giudizi così diversi fra loro?

Il fatto è che la ricerca sul *baccalauréat* fu effettuata in una fase dello sviluppo storico dei sistemi scolastici dei paesi industrializzati che possiamo considerare di svolta. Si continuava ad operare secondo criteri tradizionali, ma stavano mutando profondamente le condizioni di intervento. Le qualità personali degli insegnanti non potevano più essere considerate una condizione sufficiente ad assicurare l'adeguatezza nello svolgimento delle varie funzioni connesse alla pratica della formazione scolastica. Si sarebbe visto con evidenza sempre maggiore nei decenni successivi che le qualità personali costituiscono un prerequisito necessario, ma non sufficiente per definire un profilo professionale capace di far fronte alle nuove esigenze della formazione scolastica. Stava infatti cambiando il modo di considerare la qualità del prodotto della scuola. In un contesto elitario, quando agiscono meccanismi capaci di limitare severamente l'accesso all'istruzione, si considera positiva la qualità della formazione se i risultati di un numero modesto di allievi si presentano come decisamente positivi.

Ma quando il contesto cambia, e la scuola vede un deciso ampliamento della propria base sociale di reclutamento, anche il criterio secondo cui giudicare la qualità della formazione deve essere ridefinito. I risultati molto buoni conseguiti da una minoranza di allievi costituiscono sempre un dato positivo, ma il giudizio che si esprime sulla capacità del sistema di soddisfare le esigenze della formazione sarà piuttosto influenzato dai risultati ottenuti dalla maggioranza.

Il criterio col quale si giudicano i risultati scolastici non è, a ben vedere, diverso da quello che siamo abituati a considerare per esprimere giudizi che riguardano l'a-

¹⁸ Tratto da: Vertecchi, B. (1991), *Origini e sviluppi della docimologia*, Teramo, Giunti e Lisciani, pp. 40-44.

deguatezza di ciò che vien fatto in altri settori della vita sociale. Se dobbiamo esprimere un giudizio che riguarda il servizio sanitario o l'amministrazione della giustizia non ci basterà sapere se qualche malato ha riacquisito la salute o se qualche malfattore è incorso nelle sanzioni della legge, ma vorremo essere rassicurati sul fatto che quanti hanno bisogno di cure le ricevano e che ogni cittadino onesto non debba temere che altri turbino impunemente le condizioni della convivenza civile. Non ci comportiamo diversamente se dobbiamo esprimere un giudizio sulla produzione, invece che di servizi, di beni fisici. Se dobbiamo acquistare un'automobile, non sarà tanto importante sapere se il costruttore sia in grado di produrre saltuariamente un bolide da 300 all'ora, ma vorremo essere rassicurati sul fatto che il motore funzioni correttamente per un congruo numero di chilometri, che la carrozzeria non si copra di ruggine dopo qualche giorno di esposizione alle intemperie ecc. In breve, sia che si tratti di servizi, sia che si tratti di beni fisici, il nostro giudizio viene influenzato dalle caratteristiche che osserviamo nella generalità dei casi.

È come dire che la qualità di un prodotto si stabilisce sulla base di uno *standard*, e non del presentarsi accidentale di aspetti positivi.

Ciò vale anche per il prodotto dell'attività formativa: la qualità del servizio scolastico sarà positiva se la generalità di quanti ne fruiscono consegue i risultati desiderati. Questa condizione non contrasta certamente con la possibilità da parte di alcuni allievi di conseguire risultati eccezionalmente positivi. Interessa infatti che sia raggiunto il livello che abbiamo definito come *standard*: ciò che va oltre è sempre apprezzabile.

Quando si opera su grandi quantità, assicurare uno *standard* positivo è indubbiamente difficile. Occorre definire con grande esattezza le caratteristiche che si intendono osservare nel prodotto, e mantenere il processo in condizioni di controllo costante. Quando il processo, com'è nel caso della formazione, coinvolge l'attività di un gran numero di persone, occorre contenere la dispersione che deriva dalle differenze nelle qualità individuali, negli stili di lavoro, negli atteggiamenti. Un repertorio di competenze professionali che si riferisce ad una attività complessa è positivamente apprezzabile quando realizza un equilibrio tra la necessità di consentire l'espressione delle doti personali e quella di assicurare il raggiungimento dello *standard* che si è definito per il prodotto di una determinata attività.

Assicurare la qualità della formazione nelle società contemporanee richiede una trasformazione del contesto che consenta una razionale organizzazione del lavoro. La condizione preliminare è che si disponga di informazioni capaci di offrire descrizioni realistiche circa lo svolgersi delle attività. Proprio questa condizione ha fatto apparire l'inadeguatezza delle pratiche tradizionali della valutazione: gli insegnanti maturavano propri convincimenti ed esprimevano giudizi secondo criteri che non avevano bisogno di rendere espliciti verso l'esterno.

Le loro qualità personali assicuravano anche la qualità del giudizio. Ciascuno poteva organizzare il proprio lavoro nel modo che riteneva più opportuno, senza bisogno di rendere le proprie scelte compatibili con quelle degli altri.

Lo sviluppo della scolarizzazione ha richiesto revisioni nella organizzazione del lavoro scolastico che hanno progressivamente spostato la responsabilità delle scelte dai singoli alla struttura entro cui operano.

L'opinione pubblica ha incominciato ad esprimere propri giudizi sulla qualità della formazione non in riferimento al lavoro svolto da questo o quell'insegnante, ma integrando i dati di cui dispone sull'insieme del sistema scolastico. La correttezza di tali giudizi, così come l'adeguatezza delle scelte che hanno incidenza sulla formazione richiedono informazioni analitiche, formulate in modo da poter essere comunicate senza eccessive deformazioni. Si comprende quindi perché modi di valutazione affidati a criteri impliciti non apparissero più accettabili e perché la ricerca si sia orientata alla definizione di nuove condizioni di accettabilità degli elementi (in primo luogo la misura delle prestazioni) sui quali si sarebbe fondata l'espressione dei giudizi di valutazione.

Tali condizioni sono fondamentalmente due:

- la prima è che si sappia con esattezza qual è l'aspetto che una determinata procedura (per esempio una misurazione) tende a rilevare e che la procedura stessa sia in grado di evidenziare tale aspetto senza che ad esso se ne sovrappongano altri al punto da renderlo indistinguibile, e quindi da non consentire di separare l'informazione che si riferisce ad esso da quella che dovrebbe essere riferita agli altri aspetti. Occorre inoltre che i dati di cui si viene in possesso siano confrontabili con altri ad essi omogenei (per esempio, se si tratta di misurazioni di grandezze fisiche, a quelli di strumenti che fungano da riferimento). Queste condizioni definiscono la *validità* delle informazioni;
- la seconda condizione è che le informazioni siano rilevate secondo criteri accuratamente definiti, ricorrendo ad una strumentazione che dia luogo a dati uniformi se utilizzata in condizioni uniformi, che rilevi dati uguali se l'uso è iterato su oggetti uguali, che sia insensibile al variare di chi rileva il dato. Questa seconda condizione si definisce attendibilità.

Una parte considerevole della ricerca docimologica si è rivolta alla precisazione delle condizioni formali e delle soluzioni tecniche capaci di assicurare livelli accettabili di validità e di attendibilità alle informazioni sulle quali si fonda l'attività valutativa.

7. La valutazione in dieci fotogrammi (Giovannini, M. L., *La valutazione*)¹⁹

1.1 Primo: rendimento dell'alunno e altro...

La valutazione punta i suoi riflettori sull'alunno, ma anche sui percorsi formativi.

Le caratteristiche della scuola "a nuovo indirizzo", tesa a perseguire l'uguaglianza nella diversità, richiedono il superamento di un modello valutativo del "passato" nel quale la valutazione non soltanto mira a controllare il rendimento dell'alunno, ma fa prevalentemente ricadere su di lui la responsabilità dei risultati non soddisfacenti. Il vecchio modello, infatti, considera determinanti per la riuscita scolastica le variabili di tipo attitudinale e di tipo moralistico (le "doti" naturali, l'impegno, la volontà ecc.) con scarso riferimento all'adeguatezza delle modalità e delle procedure adottate nell'insegnamento. In contesti didattici nei quali si voglia mettere in atto un reale diritto allo studio e realizzare effettive opportunità rispetto ai bisogni cognitivi e socioaffettivi degli allievi, diventa necessario rivedere le funzioni della valutazione e allargarne l'"oggetto" oltre la prestazione dell'allievo. Per promuovere una capacità di studio sempre più autonoma, nell'attività didattica viene infatti rivolta particolare attenzione anche ai processi cognitivi e a quelli metacognitivi dell'alunno (e dunque al tipo di strategie attraverso cui l'alunno arriva a quei risultati, alle modalità utilizzate per organizzare le conoscenze, alle ipotesi e alle giustificazioni che vengono esplicitate, alla capacità di variare strategia quando questa non porta alla soluzione, nonché ai processi motivazionali strettamente interrelati ai processi e strategici cognitivi e metacognitivi). La valutazione, pertanto, è diretta non solo a «verificare se certe conoscenze e competenze sono state acquisite, ma anche a diagnosticare il livello di significatività e di utilizzabilità che esse hanno raggiunto, il livello di coscienza della validità e significatività dell'organizzazione interna delle conoscenze, la capacità di autoregolazione delle strategie di loro acquisizione e applicazione». Tali informazioni sono infatti indispensabili per realizzare interventi formativi adeguati ai singoli allievi. Oltre che al singolo, è importante fare riferimento alla situazione del gruppo (classe, interclasse o gruppo più allargato), non tanto per utilizzare un criterio valutativo basato sul confronto con l'andamento medio, quanto per riflettere sulla effettiva praticabilità della proposta formativa e sugli effetti da essa prodotti.

Nella valutazione rientrano inoltre le molteplici variabili del contesto didattico e organizzativo predisposte per favorire la costruzione delle conoscenze e delle abilità nei destinatari/protagonisti dell'apprendimento; questo al fine di poter individuare i

¹⁹ Tratto da: Giovannini, M.L. (1955), *La valutazione*, Mondadori, Milano, pp. 12-22.

punti forti e quelli deboli e di poter riflettere sulla loro funzionalità ed efficacia rispetto agli apprendimenti degli allievi, con eventuali conseguenti azioni di miglioramento.

In una scuola che intenda caratterizzarsi per l'intenzionalità formativa, per il rispetto-valorizzazione delle diversità culturali e cognitive, per la continuità, per la flessibilità organizzativa e didattica e per lo stile collegiale dei docenti, vanno riservati alla valutazione momenti specifici di incontro dei docenti della classe e del plesso: ciò non soltanto per una riflessione sull'andamento degli apprendimenti degli allievi, ma anche per pronunciarsi sulla congruenza e sulla validità delle scelte via via effettuate rispetto ad alcuni principi base. Non è certamente da trascurare il problema della difficoltà di individuare un criterio di riuscita in relazione a tali caratteristiche connotative. Crediamo comunque sia già molto importante che gli insegnanti, alla luce dei "segni di riconoscimento" della didattica propria di una scuola a nuovo indirizzo, si interrogino su quali siano per loro gli indicatori di un insegnamento efficace e procedano a un controllo empirico delle loro posizioni. Non solo nella valutazione degli allievi, ma anche in quella più estesa dei percorsi formativi compiuta da un gruppo di docenti è da rifiutare un uso della valutazione che induca a demotivare, a discriminare, a classificare, a sollecitare la competitività o a categorizzare. La valutazione infatti trova la sua ragion d'essere e la sua forza nell'esigenza di un "servizio scuola" trasparente e connotato da una tensione al cambiamento innovativo, non certo in quella di classificare gli studenti e gli insegnanti in bravi e non bravi.

1.2 Secondo: un impegno collegiale

La valutazione costituisce un importante ambito di riflessione e di confronto tra i docenti.

L'impegno collegiale nel processo di valutazione della qualità dei percorsi formativi non esclude affatto l'importanza di una dimensione individuale; essa però deve sapersi collegare a quella collettiva. In tale confronto rientrano pure gli usi, i tempi e i modi della valutazione.

Una tale pratica aiuta non soltanto a ripensare all'efficacia dei contesti predisposti, delle procedure e delle modalità messe in atto, ma anche a rendere esplicite le concezioni dell'insegnamento e dell'apprendimento che di fatto si privilegiano e, soprattutto, induce a confrontare le proprie idee e concezioni relative allo sviluppo delle capacità dei singoli studenti, concezioni che ovviamente stanno alla base dei giudizi che si formulano. Il confronto dovrebbe mirare, innanzitutto, a individuare delle modalità esplicite e condivise di rappresentazione dell'intervento formativo e della valutazione per realizzare tra i docenti del plesso un effettivo "patto didattico". Riuscire a superare la cosiddetta "cultura dell'individualismo a oltranza" o una "partecipazione collegiale costretta"² non è facile. Viene richiesta infatti una riflessione sulle proprie "teorie soggettive", derivate non solo dalla formazione ma dai valori culturali e sociali, dall'esperienza e dall'affettività di ciascuno. Sono necessarie inoltre la disponibilità all'autocritica e alla revisione delle proprie posizioni, uno scambio costruttivo con gli altri e la volontà di risolvere insieme

i problemi. Aspetti questi che contrastano con le diversità di attese e di coinvolgimento, e in parte anche con ciò che gli insegnanti dichiarano di praticare ("fare tutto il possibile dentro la loro classe" e "doversela sbrogliare" continuamente alla meglio di fronte ai problemi che la classe impone). Il confronto può invece costituire non solo un aiuto al superamento di molte difficoltà individuali, ma anche un elemento irrinunciabile per una gestione collegiale dei percorsi formativi. Quest'ultima dovrebbe lasciare ai soggetti coinvolti lo spazio per potere verificare nel contesto in cui operano le proprie e altrui posizioni, nell'ottica di una scuola intesa come luogo di apprendimento non soltanto per gli alunni, ma per tutti coloro che hanno compiti formativi.

1.3 Terzo: uno zoom sulla situazione iniziale

La valutazione iniziale si caratterizza per la funzione prognostica nei confronti dei percorsi didattici progettati.

Espressioni quali valutazione iniziale e in itinere sono state create per evidenziare la necessità di non collocare il momento valutativo soltanto a percorso terminato; tuttavia esse vengono utilizzate solitamente in riferimento al controllo delle acquisizioni possedute dagli allievi e non nel senso più ampio qui privilegiato. La valutazione iniziale, oltre a un riferimento temporale, assume un significato logico rispetto alla intenzionalità dell'intervento formativo. Occorre infatti che il team di classe e di plesso si interroghi, in relazione alle "caratteristiche" di ciascun allievo, sulla rispondenza della proposta formativa a un'utilizzazione articolata delle attrezzature, dei sussidi e dei materiali didattici, a un'alternanza di situazioni e di attività in un'ottica di continuità dell'itinerario didattico. La ricognizione delle caratteristiche degli alunni dovrebbe consentire di conoscere i loro interessi e i loro atteggiamenti, le modalità di acquisizione del sapere da loro privilegiate e in particolare le loro rappresentazioni spontanee, in quanto consentono di identificare le difficoltà di acquisizione (cioè gli ambiti di conoscenza più problematici) e «le zone di sviluppo prossimale rispetto a cui muoversi con l'intervento formativo»³.

Sulla base della progressiva raccolta di tali informazioni tramite prove mirate e osservazioni in contesti diversificati, gli insegnanti possono progressivamente adattare la proposta formativa alle esigenze degli alunni in modo da realizzare un avanzamento positivo di ciascuno. Il giudizio prognostico si riferisce proprio alla pertinenza del percorso formativo e alla "adeguatezza" dei contesti didattici che si intendono predisporre (non dovrebbe essere riferito all'allievo né creare pregiudizi nei suoi confronti). Una simile valutazione si contrappone dunque a quel tipo di valutazione iniziale che finisce per diventare una sorta di «predicazione che tende ad autorealizzarsi, per effetto dei mutamenti che provoca a livello degli atteggiamenti di insegnanti ed allievi»⁴. La valutazione iniziale consente di fare una prima verifica delle caratteristiche dei percorsi formativi progettati, in particolare della flessibilità delle strategie didattiche e organizzative, della continuità formativa, della congruenza rispetto alle conoscenze precedenti dell'alunno e alla loro organizzazione (nonché della funzionalità delle procedure,

utilizzate per la loro rilevazione).

1.4 Quarto: soste in itinere

La valutazione intermedia aiuta a decidere le soluzioni concrete più opportune ai fini di un'adeguata regolazione del percorso formativo.

Pronunciarsi a percorso già effettuato, seppure a ragion veduta, sulla sua scarsa flessibilità ed efficacia, sarebbe in realtà troppo tardi e inutile. E pertanto auspicabile che vengano fatti dei bilanci intermedi sulla "adeguatezza" e sulla funzionalità di quanto realizzato, sull'effettivo miglioramento conseguito rispetto a quanto previsto. Questo aiuta a vagliare le scelte e le opportunità, e a cogliere in modo sistematico l'evoluzione del processo formativo.

Fare il punto, per esempio, sulle modalità di funzionamento dei rapporti con l'extrascuola e con le famiglie, dei raccordi tra le varie classi, del lavoro collegiale, sull'effettivo uso flessibile degli spazi e dei tempi, sull'articolazione delle attività, sull'uso di media diversi, sulla coerenza tra contenuti e metodologie didattiche adottate, sull'evoluzione delle conoscenze e abilità scolastiche, consente di individuare gli aspetti sui quali si può agire e sui quali è più opportuno convogliare le energie.

La riflessione sull'andamento delle competenze degli alunni viene effettuata dagli insegnanti a vari livelli: dal singolo alla classe, a classi parallele, non escluso il confronto, là dove ci sia la possibilità, con riferimenti anche esterni (altre scuole). Scopo della valutazione intermedia è infatti quello di aiutare gli insegnanti a ripensare all'efficacia delle metodologie didattiche adottate (per rendere il percorso formativo più rispondente alle esigenze dei singoli) e di sollecitare ciascun alunno a riflettere sul proprio percorso di apprendimento, sulla propria "concezione" dell'imparare, sulle strategie adottate e sulle difficoltà incontrate (in modo da condurlo a una responsabilità sempre maggiore nella gestione del proprio apprendimento). L'informazione che deve essere fornita all'alunno non è quella relativa a dove egli si situa rispetto agli altri, quanto piuttosto quella che si riferisce ai traguardi da raggiungere, ai progressi realizzati e ai modi che hanno favorito l'acquisizione. È opportuno, a tal fine, che le pratiche valutative non solo riducano gli aspetti comparativi ma coinvolgano direttamente lo studente nel giudizio sulla qualità delle sue acquisizioni e prestazioni. La precisazione che il successo o l'insuccesso possono essere riferiti, prima ancora che a lui, all'organizzazione delle attività didattiche consente di eliminare, o almeno di contenere, la competitività tra gli alunni e i vari sotterfugi cui molti di loro ricorrono in una situazione didattica tradizionale per nascondere ciò che non sanno e/o non hanno capito.

1.5 Quinto: un bilancio complessivo

La valutazione finale comporta un consuntivo sui percorsi formativi attuati.

Non solo i dati valutativi dei bilanci che si effettuano durante il percorso, ma anche quelli di un bilancio complessivo sono utilizzati dai vari attori in direzione di una accumulazione di informazioni, nella prospettiva di contribuire a predisporre successivamente degli interventi capaci di assicurare a tutti le migliori condizioni di riuscita. Essi mirano a verificare l'effettivo raggiungimento dei traguardi formativi da parte degli allievi e la "tenuta formativa" delle condizioni di efficacia previste (collegialità, continuità, flessibilità ecc.).

Anche il riferimento a parametri e a *standard* esterni viene utilizzato, nella situazione di un bilancio complessivo, non a scopo classificatorio, ma per esplorare le scelte compiute e la loro efficacia.

L'analisi delle informazioni disponibili viene effettuata a vari livelli per giungere fino a quello del collegio dei docenti, per una loro ricaduta sulle successive decisioni formative e sulle scelte organizzative.

1.6 Sesto: un processo sistematico

Le valutazioni informali attuate quotidianamente richiedono il sostegno di modalità sistematiche e ben definite.

Nella vita di tutti i giorni, le valutazioni che esprimiamo sono sovente il frutto di reazioni affettive spontanee, di cui può riuscirci difficile dare una spiegazione razionale. Per esempio, a volte capita di esprimere mentalmente un giudizio negativo nei confronti di una persona per il solo fatto che indossa un abito che a noi non piace; oppure di apprezzare molto un paesaggio, da altri nemmeno preso in considerazione. In tali situazioni è possibile esprimere le proprie impressioni per influenzare gli altri o anche soltanto per comunicare. Si tratta, tuttavia, di impressioni generali, almeno nei casi in cui non ci sia consapevolezza degli elementi e delle basi su cui si fondano le valutazioni espresse.

Nel compimento di un sorpasso, per esempio, si valuta la velocità e la distanza, l'ingombro dei veicoli e la larghezza della strada per decidere in merito; tuttavia i giudizi sono basati su percezioni personali peraltro diverse in un altro sorpasso e le informazioni su cui si fonda la scelta si confondono con la scelta stessa.

Le valutazioni espresse in ambito formativo, riferite alla validità del progetto da realizzare, alla funzionalità degli itinerari effettuati e pure ai livelli di abilità raggiunti dagli alunni, dovrebbero, per contro, superare quanto più possibile tali caratteristiche "impressionistiche". Richiedono pertanto che il giudizio sia opportunamente sorretto da informazioni, dati, misure attendibili e pertinenti rispetto a ciò che si vuole valutare, in modo che la loro ricaduta sull'attività didattica successiva sia utile.

In senso stretto la valutazione consiste in un'operazione di apprezzamento, di attribuzione di un valore, spesso in riferimento a un ideale o a una norma. Occorre però considerare che, da un punto di vista strutturale, la valutazione consiste in un processo che si articola in varie operazioni (individuazione del che cosa valutare, raccolta delle

informazioni necessarie, elaborazione ed espressione del giudizio). Tali operazioni, nei fatti, vengono compiute sempre, ma la convinzione che la valutazione si risolva in un solo atto fa trascurare non soltanto la corretta conduzione di tutte le altre operazioni ma anche la loro funzione propria e autonoma⁵.

1.7 Settimo: indicatori non ambigui

La determinazione di ciò che si intende valutare avviene tramite la scelta e la definizione di indicatori non ambigui.

La prima operazione del processo valutativo è l'identificazione dell'"oggetto" della valutazione, sia che ci si riferisca agli apprendimenti degli allievi, sia alle caratteristiche dei percorsi formativi. In alcuni casi tale identificazione è semplice, in altri è assai più complessa. Un accorgimento importante, per non privilegiare gli aspetti su cui è più semplice reperire le informazioni, è la messa in atto di tutte le possibili condizioni affinché l'aspetto/situazione da valutare sia preso in considerazione e sia espresso tramite indicatori definibili e identificabili.

Tra gli interrogativi che è opportuno porsi prevalgono i seguenti: gli indicatori scelti si riferiscono a ciò che si vuole valutare? Qual è il legame tra ciò che si vuole valutare e le scelte che stanno a monte? Quali i criteri che hanno indotto a scegliere quegli indicatori? Le risposte a tali quesiti costituiscono non solo un'occasione di confronto, ma una prima condizione dell'affidabilità del processo valutativo.

1.8 Ottavo: una base affidabile e pertinente

Le misure/informazioni raccolte devono essere pertinenti e adeguate.

La valutazione tradizionale è caratterizzata soprattutto dal sovrapporsi della fase di rilevazione delle informazioni/misurazione a quella della valutazione in senso stretto. Pertanto gli elementi di giudizio vengono confusi con gli elementi descrittivi e spesso addirittura li sostituiscono.

La *misurazione* consiste nell'insieme delle operazioni che si conclude con l'attribuzione di numeri a oggetti, eventi, o a qualche loro qualità, secondo precise regole, in modo da quantificare i risultati della rilevazione relativi all'oggetto misurato. Essa non va confusa con la fase propriamente valutativa. Anche nel caso in cui si ricorra a una rilevazione di tipo qualitativo, occorre per quanto possibile distinguere le due fasi.

La correttezza della fase relativa alla definizione degli indicatori costituisce una condizione necessaria ma non sufficiente in rapporto alla correttezza della fase relativa alla rilevazione dei dati. Questi potrebbero essere:

- non pertinenti, in quanto non relativi all'oggetto della valutazione;
- non validi, in quanto raccolti tramite una modalità che non garantisce che siano quelli che si vorrebbero effettivamente raccogliere;
- non affidabili, in quanto sono state compiute delle scorrettezze nella rilevazione.

Occorre pertanto, nella fase di raccolta delle informazioni, far sì che esse

siano numerose, ma anche utilizzare gli adeguati accorgimenti per minimizzare l'effetto delle differenti fonti di errori.

Quanto più numerose e corrette risulteranno le informazioni raccolte, tanto più il giudizio sarà fondato e permetterà di assumere le decisioni educative più opportune. Occorre infatti tenere presente che il tipo di strumento o di procedura utilizzato o il diverso genere di contesto possono far variare le informazioni raccolte; pertanto è opportuno utilizzare differenti modalità, salvaguardandone però la congruenza e la validità rispetto al tipo di informazioni che si ricercano.

Le informazioni rilevate possono essere oggetto di immediato giudizio oppure essere accumulate per successivi e più ampi giudizi.

1.9 Nono: individuazione dei punti forti e deboli

I giudizi vanno formulati in modo chiaro e in rapporto a criteri trasparenti.

Le decisioni che vengono assunte nelle svolte principali del percorso formativo, spesso frutto di negoziazioni tra i partecipanti, risulteranno tanto più trasparenti quanto meno generica sarà la formulazione del giudizio conclusivo su cui si basano.

Purtroppo, per effetto di giudizi generici e non sufficientemente "supportati", può risultare difficile argomentare le decisioni che si vogliono assumere. A volte, inoltre, i giudizi relativi alle acquisizioni degli allievi tendono a investire l'allievo e non la prestazione da lui fornita ("è un alunno disattento", "è pigro"...), e sono completamente disgiunti dal contesto didattico nel quale sono stati manifestati i comportamenti tanto criticati. Si tende infatti ad accentuare il determinismo dei tratti di personalità e delle componenti extrascolastiche, anziché quello della scuola e degli interventi didattici. Quanto all'articolazione dei giudizi, non ci si può limitare a esprimere una soddisfazione/insoddisfazione generica nei confronti del risultato dell'intervento formativo; è opportuno invece essere in grado di indicare il criterio di accettabilità e precisare in che cosa tale risultato corrisponda alle attese e in che cosa si discosti. Questo è importante non soltanto per i risultati delle valutazioni che devono orientare i successivi interventi, ma anche per la valutazione di bilancio. Si pensi per esempio ai giudizi estesi che caratterizzavano le schede di valutazione, che spesso dicevano e non dicevano, sicché, in ultima analisi, non potevano certo costituire un valido strumento di informazione e di aiuto per i genitori e gli alunni, e neppure per la continuità formativa.

1.10 Decimo: un archivio dinamico

Viene data molta importanza alla documentazione, quale modalità in grado di permettere una ricostruzione sistematica dei percorsi formativi.

La documentazione qui considerata non è quella prevista dalla normativa ministeriale e neppure quella costituita dai diari o "quadernoni" nei quali vengono riportati appunti e annotazioni su aspetti particolari della giornata o sulle

modalità di attuazione del lavoro in classe. Questi ultimi infatti costituiscono una documentazione fruibile individualmente e non partecipabile e comunicabile agli altri.

Si fa qui riferimento soprattutto a un archivio nazionale, al momento inesistente, disponibile per chiunque avesse interesse a consultarlo, contenente informazioni sui risultati scolastici in termini di conoscenze e abilità apprese e sui percorsi, metodologie e organizzazione che li hanno prodotti. La disponibilità di un tale archivio consentirebbe inoltre di confrontare i dati di una classe con quelli di gruppi più allargati appartenenti alla stessa popolazione studentesca. È la dimensione collettiva su ampia scala, infatti, che permette di identificare le incoerenze e la praticabilità dei progetti curricolari, e di aggiustare, per esempio, gli stessi programmi scolastici al fine di renderli più adeguati all'età e ai saperi degli studenti. Ciò presuppone però la disponibilità di "banche dati" relative a prove, a esiti formativi che non siano direttamente legati a specifici itinerari didattici ma che tuttavia potrebbero costituire un valido parametro di riferimento per i docenti, utile anche per riflettere sull'andamento degli effetti degli itinerari realizzati. La documentazione costituisce anche un elemento che accresce la trasparenza verso l'esterno della qualità della scuola e del progetto educativo del circolo/istituto, e consente ai cittadini la possibilità di fornire un valido apporto per un miglioramento degli interventi formativi.

Note

1. M. Pellerey, *La valutazione diagnostica dei processi cognitivi e metacognitivi*, in AA.VV., *La valutazione nella scuola media*, «Studi e Documenti degli Annali della Pubblica Istruzione», 1993, pp. 278-294.
2. A. Hargreaves, M. Fullan, *Understanding Teacher Development*, Teachers College Press, 1991.
3. L.S. Vygotskij, *Storia dello sviluppo delle funzioni psichiche superiori e altri scritti*, tr. it., Firenze, Giunti-Barbera, 1974, e anche dello stesso autore *Pensiero e linguaggio*, a cura di L. Mecacci, Bari, Laterza, 1990.
4. B. Vertecchi, *Decisione didattica e valutazione*, Firenze, La Nuova Italia, 1993, p. 88.
5. Al fine di evitare la sovrapposizione delle operazioni della misurazione con quelle della valutazione vera e propria, Gattullo ha proposto di utilizzare il termine *controllo* quale complesso delle operazioni; cfr. in particolare M. Gattullo, *Didattica e Docimologia. Misurazione e valutazione nella scuola*, Roma, Armando, 1967.

Parte 2. Approcci e funzioni della valutazione

di Andrea Giacomantonio

Il concetto di *forma* è variamente presente e soggetto a diverse, a volte opposte, definizioni nella storia del pensiero occidentale²⁰. Il postulato fondamentale della *Gestalttheorie* (teoria della forma) è che il tutto, vale a dire la



Fig. 1 – Figura ambigua

forma, è più della somma delle parti, l'unità delle quali è garantita da precise leggi di organizzazione (somiglianza, prossimità, simmetria, chiusura, continuità di direzione ecc.). Il modo più semplice per esemplificare l'approccio della psicologia della forma è riflettere sulla figura ambigua rappresentata di fianco. In essa tendiamo a cogliere alternativamente (mai contemporaneamente) due forme: un calice o due volti l'uno di fronte all'altro. Tale possibilità dipende sia dall'uso sapiente fatto dall'artista delle leggi di organizzazione

del campo percettivo, sia dalla nostra predisposizione ad applicare queste leggi per interpretare i dati sensoriali. Nel caso presente l'emergere delle due forme dipende dal colore che poniamo alternativamente sullo sfondo o in primo piano²¹.

Nel secolo scorso il concetto di *forma* è stato utilizzato per indicare "unità" diverse, le cui leggi hanno un potere connettivo variabile. Si pensi, da un lato, alle *forme di vita*, di cui parla Ludwig Wittgenstein per designare l'insieme delle attività (tra le quali un ruolo privilegiato è ricoperto dal linguaggio) che unisce i diversi membri di una comunità o di una cultura²²; dall'altro, al modo con cui Thomas Kuhn usa metaforicamente lo stesso concetto per spiegare la natura dei paradigmi scientifici (approssimativamente immaginabili come la visione storicamente determinata che una comunità scientifica condivide di una disciplina e dell'intero universo) e le difficoltà che incontra un ricercatore quando assumere alternativamente il punto di vista di due paradigmi in competizione²³. A prescindere dai legami tra le due posizioni teoriche (quella di Wittgenstein e quella di Kuhn), nelle prossime pagine il concetto di

²⁰ Per una sintetica descrizione del ruolo del concetto di forma nella storia del pensiero occidentale cfr. Vuillemin, J. (1979), *Forma*, in *Enciclopedia*, vol. VI, Torino, Einaudi, pp. 314-323.

²¹ Per una presentazione storica della psicologia della Gestalt cfr. Sambin (1980), M., *La psicologia della Gestalt*, in Legrenzi, P. (a cura di) (1980), *Storia della psicologia*, Bologna, il Mulino, pp. 111-145; Mecacci, L. (1994), *Storia della psicologia del Novecento*, Roma-Bari, Laterza, pp. 47-94.

²² Cfr. Wittgenstein, L. (1953), *Philosophische Untersuchungen*, Oxford, Basil Blackwell; trad. it. *Ricerche filosofiche*, Torino Einaudi, 1967, v., p. es., § 19, § 23, § 241, § 325 e *passim*.

²³ Cfr. Kuhn, T. S. (1970²), *The Structure of Scientific Revolution*, Chicago, The University of Chicago; trad. it. *La struttura delle rivoluzioni scientifiche*, Torino, Einaudi, 1969², pp. 139-165.

forma verrà utilizzato in entrambe le accezioni; in particolare nella seconda quando si farà riferimento ai modelli valutativi, nella prima laddove gli orizzonti del discorso si ampliano sino a considerare le relazioni, per esempio, tra sistemi sociali, sistemi scolastici e didattica.

In campo pedagogico-didattico sono identificabili forme (paradigmi o modelli) differenti, spesso coesistenti e almeno parzialmente sovrapponibili sul piano teorico e metodologico. Una siffatta coesistenza comporta effetti non sempre positivi sia per il decisore politico sia soprattutto per chi opera sul campo. Si pensi, ad esempio, alla difficoltà di definire rigorosamente la missione della scuola all'interno della società presente e alla tensione che ciò comporta tra curricula espliciti ed impliciti; all'impossibilità di identificare con certezza i contenuti dell'insegnamento ed i relativi metodi; alla proliferazione di procedure e strumenti valutativi. Un epifenomeno della situazione descritta consiste nella diffusa attribuzione di significati diversi agli stessi termini, usuale tanto tra i ricercatori quanto nella classe docente, che spinge Benedetto Vertecchi a parlare di *indeterminatezza linguistica* e Maria Lucia Giovannini addirittura di *giungla terminologica*²⁴.

La stessa proliferazione di forme o paradigmi si rileva in campo valutativo. Nel testo di Nicoletta Stame (lettura 8), ad esempio, si individuano tre approcci che, sebbene differiscano sul piano teorico, coesistono e spesso interagiscono sul piano applicativo. L'approccio *pragmatista – della qualità* e quello *costruttivista-sociale* solo di recente – ovvero con la diffusione dalla fine degli anni '80 dei sistemi di auto- ed etero-valutazione di istituto – hanno iniziato ad esercitare un'influenza sulle pratiche valutative in uso nel sistema scolastico italiano. Ma, ahinoi!, l'uso sconsigliato di questi modelli (in particolare del primo) fatto dalla Commissione di valutazione istituita nel 2001 dal Ministro Moratti, ne ha promosso una percezione fallace e distorta. È difficile oggi trovare un docente che percepisca correttamente i vantaggi ed i problemi associati all'uso di questi strumenti.

Diversamente l'approccio positivista-sperimentale è conosciuto meglio, almeno sul piano teorico. I Decreti Delegati del '74, infatti, hanno promosso e dato l'avvio a diverse (quasi-)sperimentazioni²⁵. Non tutte, certamente, hanno seguito un disegno sperimentale rigoroso, ma non mancano esperienze esemplari²⁶. La posizione del

²⁴ Vertecchi, B. (1993), *Decisione didattica e valutazione*, Scandicci (FI), La Nuova Italia, pp. 8-9; Giovannini, M. L. (1994), *Valutazione sotto esame*, Milano, Ethel-G. Mondatori, p. 48.

²⁵ Campbell D.T., Stanley J.C. (1963), *Experimental and Quasi-experimental Designs for Research*, Boston, Houghton Mifflin Company.

²⁶ Tra i più noti e perseveranti divulgatori dell'approccio positivista-sperimentale alla valutazione si annoverano Aldo Visalberghi e Maria Corda Costa. In merito cfr., a puro titolo esemplificativo, Visalberghi, A. (a cura di) (1964), *Educazione e condizionamento sociale*, Bari, Laterza; Corda Costa, M., Vertecchi, B., Visalberghi, A. (1975), *Orientamenti per la sperimentazione didattica*, Torino, Loescher. Per comprendere meglio il contributo fornito dai due autori alla promozione del metodo

Ministero della Pubblica Istruzione nei confronti di questo approccio è sempre stata ambigua. Nonostante pregevoli iniziative di formazione della classe docente e di promozione di ricerche, molte sperimentazioni istituzionali – soprattutto quando si trattava di valutare determinate politiche dell'educazione – sono state svolte in assenza di un disegno sperimentale (si veda, per esempio, quella relativa alla sperimentazione nella scuola primaria delle Indicazioni nazionali elaborate sotto il ministro Moratti²⁷). Assumendo che la situazione italiana sia comparabile con quella francese, una spiegazione plausibile di questo fenomeno è stata fornita da Louis Legrand: «La situazione più arcaica, e che disgraziatamente tende in Francia a riproporsi senza fine, è quella in cui l'avvio delle innovazioni non è accompagnato da alcuna sperimentazione né da alcun ricorso alla documentazione scientifica disponibile [...]. Le carenze di tale procedura sono evidenti. Nessuna preoccupazione di valutazione oggettiva dei comportamenti e dei rendimenti»²⁸. Un *modus operandi* che risulta vantaggioso esclusivamente per i soggetti promotori della riforma che, in questo modo, evitano di sottoporre il proprio operato al giudizio della comunità scientifica²⁹.

Per quanto riguarda la valutazione degli apprendimenti, nella scuola italiana si assiste alla convivenza di due forme (o modelli) fondamentali di valutazione, troppo spesso non sufficientemente distinte: la forma classica e quella docimologia, riconducibile al modello positivista sperimentale. Per descriverle è opportuna una sommaria ricostruzione storica, in cui le pratiche valutative vengano riesaminate nei contesti in cui di volta in volta si inseriscono, vale a dire in determinate configurazioni (forme) assunte dalla didattica e dalle relazioni tra scuola e società³⁰.

Nei paesi industrializzati sino alla prima guerra mondiale i sistemi scolastici avevano prevalentemente una struttura duale, fondata su base classista, che permetteva alla maggior parte dei giovani di acquisire le competenze necessarie per l'ingresso nel mercato del lavoro e poche altre nozioni di carattere civico e morale; mentre riservava ad un'élite, quella che avrebbe in seguito composto la classe dirigente della società, l'accesso agli ordini di studi superiori. Alla base di questa rela-

sperimentale in didattica, però, è necessario considerare lo spazio che, sotto la loro direzione, le riviste *Scuola e città* e *La ricerca* hanno dedicato alla diffusione dei risultati di ricerche sperimentali. Gli eredi di questa "scuola" si concentrano oggi intorno alla rivista *Cadmo*, diretta da Benedetto Vertecchi.

²⁷ *Strumenti per la Riforma. Progetto R.I.So.R.S.E. Rapporto Nazionale*, numero monografico di *Annali della Pubblica Istruzione*, 3-4, 2004.

²⁸ Legrand, L. (2004), *Politiche dell'educazione*, Roma, Anicia, pp. 49-50.

²⁹ *Ibidem*, pp. 50-51.

³⁰ Per la relazione tra scuola e società, letta attraverso il concetto di uguaglianza delle opportunità educative, cfr. Husén, T. (1972), *Social Background and Educational Career*, OCSE, Parigi; trad. it. *Provenienza sociale e carriera scolastica*, Torino, Loescher, 1974, pp. 19-35.

zione tra scuola e società, denominata *conservatrice* da Husén, vi è la convinzione che ogni essere umano ha ricevuto (per volere divino e, secondo alcuni, coerentemente all'origine sociale) una specifica dotazione di capacità che determina il livello di istruzione cui ognuno può e deve accedere.

Dopo il primo conflitto mondiale, in particolar modo negli anni '40 e '50, prevale una forma *liberale* di relazione tra scuola e società, per cui l'educazione formale rappresenta lo strumento con il quale ogni individuo deve conseguire lo status sociale corrispondente alla sua dotazione intellettuale innata³¹. Il compito principale dello Stato, quindi, consiste nel rimuovere gli ostacoli che impediscono la frequenza scolastica, assicurando una generalizzata uguaglianza delle opportunità.

In entrambi i contesti storici la didattica assume la forma detta *tradizionale*, che consiste nello svolgimento di un'attività di insegnamento uniforme (nella quale, in altri termini, gli stimoli forniti, i metodi utilizzati, il materiale offerto ecc. non sono individualizzati), che non considera l'influenza sul processo di apprendimento delle caratteristiche (cognitive, motivazionali ecc.) che gli allievi presentano all'inizio dell'intervento didattico³². L'eventuale programmazione dell'attività di insegnamento è centrata sui contenuti e al discente è lasciata la responsabilità del successo scolastico. La valutazione, svolta in modo estemporaneo e intuitivo attraverso compiti in classe e interrogazioni, ha sostanzialmente la funzione di selezionare gli allievi conformi ai parametri legati al possesso del sapere, alle tipologie culturali e ai sistemi di valore richiesti o imposti dalle classi dominanti. L'esito prevalente consiste nella riproduzione delle stratificazioni sociali preesistenti e nell'esclusione delle classi subalterne dall'esercizio del diritto all'istruzione. Inoltre, la necessità di assicurare l'adesione individuale al sistema di valori della classe sociale egemone determina una sistematica contaminazione delle attività valutative con giudizi di valore, il che rafforza la tendenza all'espulsione dal sistema scolastico di chi proviene dai gruppi non egemonici³³.

Il mutamento delle condizioni sociali e del mercato del lavoro, come già si intravedeva negli anni '60, favoriscono l'accesso ai sistemi formativi di sempre maggiori percentuali di giovani³⁴. La necessità di assicurare loro sia competenze funzionali

³¹ La nostra stessa costituzione (art. 34) si ispira ad un'impostazione simile della questione. Cfr. Ragazzini, D. (1987), *Dall'educazione democratica alla riforma della scuola*, Napoli, Liguori, pp. 68-105.

³² Per una definizione delle *variabili di soglia* cfr. Lastrucci, E. (2007²), *Progettare, programmare e valutare l'attività formativa*, Roma, Anicia, pp. 126-127.

³³ Vertecchi, B. (1986²), *La verifica del prodotto scolastico*, in Visalberghi, A. (1986²), *Pedagogia e scienze dell'educazione*, Milano, Mondadori, pp. 224-225.

³⁴ Nel 2000 solo il 3,9% dei giovani compresi tra i 15 ed i 24 anni non possiede alcun titolo o solo quello elementare. Cfr. Isfol (2001), *Federalismo e politiche del lavoro. Rapporto annuale 2001*, Milano, Franco Angeli, pp. 350-351.

alla crescita economica e civile dello Stato, sia la capacità di custodirne il perenne rinnovo, rendono inaccettabili i costi economico-sociali che comporta la forma tradizionale di didattica in termini di esclusione dal circuito formativo e di dispersione del talento. Nelle pagine iniziali del brano di Vertecchi riprodotto di seguito è riportata una critica al concetto tradizionale di profitto, la cui validità, è bene sottolineare, non può essere considerata in assoluto o astrattamente dal contesto socio-economico di riferimento: l'adeguatezza, infatti, della forma tradizionale della didattica a determinate configurazioni sociali è indiscutibile, lo è meno nell'attuale scenario storico.

Per assicurare che ogni studente consegua il successo formativo è necessario valorizzare le risorse individuali e conseguentemente *individualizzare* gli obiettivi ed i percorsi formativi³⁵. Nei paesi industrializzati la risposta a questa esigenza ha assunto sin dalla metà degli anni '60 una forma *tecnologica*, sviluppando le prime proposte, dal carattere determinista e riduttivo, in strategie di insegnamento-apprendimento aperte e flessibili, quali il *mastery learning*³⁶. Esso prevede: 1) la strutturazione delle sequenze di apprendimento in moduli articolati in unità didattiche³⁷; 2) un'attività valutativa al termine di ognuna di queste ultime 3) lo svolgimento periodico di attività di recupero o compensative qualora si sia rilevato il mancato raggiungimento degli obiettivi previsti; 4) la cura delle relazioni con l'insegnante ed

³⁵ Tale necessità è riconosciuta in modo generale e spesso trasversale alle diverse convinzioni politiche. A puro titolo esemplificativo cfr. Vertecchi, B., La Torre, M., Nardi, E. (1994), *Valutazione analogica e istruzione individualizzata*, Scandicci (FI), La Nuova Italia, e Tenuta, U. (2001), *Il piano dell'offerta formativa. Moduli e unità didattiche. La programmazione nella scuola dell'autonomia*, Roma, Anicia. Per una rassegna degli studi riguardanti l'individualizzazione dei percorsi di apprendimento cfr. Orlandi, M. (con una nota di Vertecchi, B.) (1994), "Il dibattito sull'individualizzazione in Italia dal 1970 al 1993 attraverso le riviste", *Cadmo*, 5-6, pp. 147-163.

³⁶ I testi classici che individuano gli estremi del processo evolutivo che ha consentito alle *tecnologie dell'istruzione* di trasformarsi in *tecnologie educative* sono, da un lato, quello di Cram, D. (1961), *Explaining «Teaching Machines» and Programming*, San Francisco, Fearon Publishers; trad. it. «Macchine per insegnare» e istruzione programmata, Firenze, La Nuova Italia, 1967; dall'altro quello di Block, J. H. (a cura di) (1971), *Mastery Learning, Theory and Practice*, New York, Holt, Rinehart and Wiston; trad. it. *Mastery learning. Procedimenti scientifici di educazione individualizzata*, Torino, Loescher, 1971. Una breve ricostruzione storica dell'uso delle tecnologie in campo educativo è in Maragliano, R., Vertecchi, B. (1986²), *Mezzi didattici e tecnologie dell'insegnamento*, in Visalberghi, A. (1986²), *Pedagogia e scienze dell'educazione*, cit., pp. 193-222 e in Domenici, G. (1994), *Dalla metodologia chiusa alle nuove strategie didattiche*, in Vertecchi, B. (a cura di) (1994), *La didattica. Interpretazioni e strutture*, Napoli, Tecnodid, pp. 49-71. La figura 2 è un adattamento di quella presente in Vertecchi, B. (1999), *L'Archivio Docimologico per l'Autovalutazione delle Scuole. Cos'è, come si usa*, Milano, Franco Angeli, p. 40.

³⁷ Per quanto riguarda la didattica modulare, oltre al testo già citato di Tenuta, cfr. Domenici, G. (1998), *Manuale dell'orientamento e della didattica modulare*, Roma-Bari, Laterza.

i colleghi di corso; 5) la percezione da parte dell'allievo dei traguardi formativi come espressione delle proprie esigenze³⁸.

In queste forme didattiche la valutazione, come emerge chiaramente dalla figura, assume un carattere strategico.

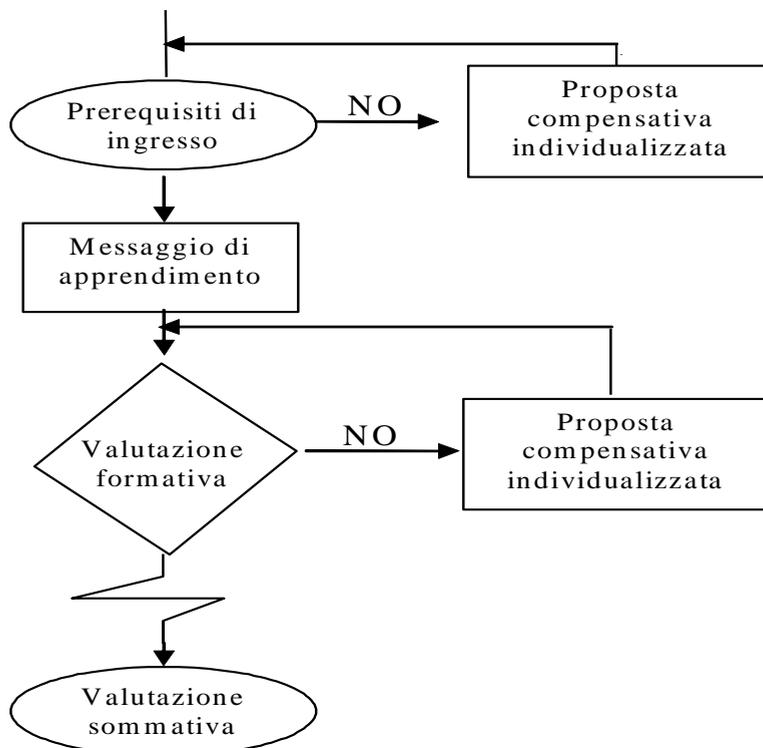


Fig. 2 – *Mastery learning*

Esso è ben rappresentato nella seconda parte del saggio di Vertecchi che riportiamo (lettura 10), dove in rapporto al tempo si distinguono tre tipi di attività valutative che svolgono ben otto funzioni. In merito a queste ultime è necessaria solo una pre-

³⁸ Le presentazioni sintetiche del *mastery learning* sono numerosissime. Tra queste si consiglia Visalberghi, A. (1974), «*Mastery Learning*: nuove strategie didattiche per rendere omogeneo il profitto, in Corda Costa, M. (a cura di) (1974), *Metodologia e psicologia*, Torino, Loescher, pp. 130-136. Per la necessità di integrare tale strategia di insegnamento-apprendimento con attività meno strutturate cfr. Visalberghi, A. (1988), *Insegnare ed apprendere. Un approccio evolutivo*, Scandicci (FI), La Nuova Italia, pp. 111-122. Per una critica sintetica e acuta alla *pedagogia per obiettivi* e, quindi, al *mastery learning* cfr. Calidoni, P. (1993), «Genesi, significato, limiti e prospettive della cosiddetta «pedagogia per obiettivi»», *Annali della Pubblica Istruzione*, 4, pp. 382-389.

cisazione. La valutazione (con funzione) diagnostica ha lo scopo di accertare il profitto degli studenti per decidere se essi hanno bisogno di svolgere un'attività compensativa prima di procedere con l'unità didattica prevista dal programma. Sia la valutazione dei prerequisiti sia quella formativa, quindi, svolgono una funzione diagnostica dato che entrambe, sebbene in momenti diversi dell'*iter* formativo, hanno lo scopo di accertare se gli allievi posseggono le conoscenze necessarie per iniziare o proseguire il percorso educativo o se è opportuno proporre interventi di recupero. La stessa relazione lega la valutazione prognostica, da un lato, con quella predittiva, dall'altro, con gli esami e i giudizi di orientamento previsti alla fine di un corso di istruzione o formazione³⁹.

La logica sottostante a questa forma di valutazione è espressa con estrema chiarezza da Mario Gattullo in un testo che rappresenta oramai un classico imperituro della docimologia: *Didattica e docimologia. Misurazione e valutazione nella scuola*⁴⁰ (lettura 9). In esso il complesso delle operazioni da svolgere per valutare il profitto di uno studente è denominato *controllo scolastico*. Quest'ultimo rappresenta uno dei quattro *momenti* essenziali di ogni processo educativo. Il primo consiste nella definizione degli *scopi* che si intendono conseguire; essi orientano la scelta degli *strumenti* e dell'*organizzazione* dell'intervento didattico, il cui fine ultimo ovviamente è di permettere all'allievo di raggiungere i traguardi precedentemente definiti. Il compito del controllo di conseguenza è di accertare in modo rigoroso se tale conseguimento è avvenuto ed in quale misura. I risultati che esso deve produrre, quindi, sono due: il primo consiste nell'esatta *misurazione* del profitto degli studenti; sulla base del quale bisogna comprendere se esiste uno scarto tra i risultati attesi (gli scopi del processo educativo) e quelli raggiunti (rilevati con la stessa misurazione). Questo atto rappresenta la *valutazione* in senso stretto.

La necessità della loro comparazione implica, inoltre, che scopi e risultati raggiunti siano espressi nello stesso linguaggio. Per misurare con precisione le conoscenze acquisite dagli allievi bisogna necessariamente riferirsi a elementi osservabili e misurabili, ossia al loro *comportamento*. Ne consegue logicamente che gli scopi educativi riguardano modificazioni comportamentali che rappresentano specularmente l'oggetto del controllo scolastico. È, però, evidenza indiscutibile che i fini del sistema d'istruzione hanno orizzonti vastissimi. Lo stesso Visalberghi, per esempio, identifica il compito della scuola con la formazione del cittadino, del produttore e del consumatore⁴¹. Tali scopi sono raggiungibili solo alla conclusione di un percorso di istruzione di un certo respiro, che dovrebbe coincidere con gli anni necessari per

³⁹ La relazione tra le funzioni della valutazione è spiegata molto bene in Melchiori, R. (2002), *Adas. Il laboratorio della valutazione*, vol. 1, Milano, Franco Angeli, pp. 12-15.

⁴⁰ Roma, Armando, 1968.

⁴¹ Cfr. Visalberghi, A. (1988), *Insegnare ed apprendere. Un approccio evolutivo*, cit., pp. 100-103.

l'espletamento dell'obbligo scolastico. Pertanto, mentre conviene articolare le *mete educative* in *finalità* che riguardano una o più discipline, è assolutamente necessario un processo di *operazionalizzazione*⁴² capace di tradurre queste ultime in *obiettivi comportamentali*⁴³. Le *tassonomie degli obiettivi*, come vedremo nel prossimo capitolo, rappresentano un valido strumento di ausilio per lo svolgimento di queste operazioni⁴⁴.

Il punto debole della proposta di Gattullo è probabilmente identificabile proprio nella sua definizione di valutazione, che consistendo nella creazione di una graduatoria degli studenti in ragione del loro profitto lascia inevase due esigenze fondamentali: quella di verificare se ogni singolo discente ha conseguito tutti gli obiettivi dell'attività didattica e quella di avere informazioni di ritorno riguardanti l'efficacia e l'efficienza della organizzazione approntata e degli strumenti utilizzati.

Nel primo caso la soluzione è molto semplice.

Descrittori dell'apprendimento	Allievi					% delle lacune
	1	2	3	...	n	
1. Sa tradurre in cifre un numero espresso in lettere	0	/	/	0	/	
a) intero	/	/	/	/	0	
b) decimale	0	/	/	0	/	
2. Sa sottrarre numeri	/	/	/	/	0	
a) interi	0	/	/	0	/	
b) decimali	/	/	/	/	0	
c) interi e decimali	0	/	/	0	/	
...	/	/	/	/	0	
m	/	0	0	/	/	
Osservazioni:						

Fig. 3 – Esempio di lista di descrittori dell'apprendimento

AmMESSO che gli obiettivi operazionalizzati equivalgano a *descrittori dell'apprendimento* e posta una relazione tra questi ultimi e gli *item* che compongono uno strumento di misurazione, è possibile rappresentare in una tabella a doppia

⁴² Tra i classici si consiglia Mager, R. F. (1976), *Goal Analysis*, Belmont, Fearon Publishers-Lear Siegler, Education division; trad. it. *L'analisi degli obiettivi*, Teramo, Giunti & Lisciani, 1976.

⁴³ Cfr. Lastrucci, E., *Op. cit.*, pp. 212-216.

⁴⁴ Anche in questo caso la letteratura in merito è molto vasta. A scopo introduttivo cfr. Vandeveld, L., Vander Elst, P. (1977), *Peut-on préciser les objectifs en éducation? Illustration de deux modèles*, Bruxelles, Editions Labor; trad. it. *Obiettivi educativi e modelli didattici* in B. S. Bloom e J. P. Guilford, Roma, Armando, 1977.

entrata, come quella che segue, tutte le conoscenze possedute dagli studenti (nella figura 3 il segno / indica l'obiettivo raggiunto e O quello non conseguito), con in-dubbio vantaggio per la valutazione formativa⁴⁵.

L'impossibilità dell'approccio docimologico alla valutazione scolastica di fornire indicazione circa gli elementi che non hanno funzionato come previsto all'interno di un processo d'insegnamento che ha registrato un insuccesso, è una diretta conseguenza in campo pedagogico-didattico della tesi di Duhem-Quine⁴⁶. Il programma didattico, infatti, può essere considerato alla stregua di un insieme di ipotesi che legano i livelli di soglia degli studenti, gli obiettivi, gli strumenti, l'organizzazione dell'intervento ed una serie astrattamente non determinabile di variabili di contesto, e che viene corroborato dal successo scolastico e falsificato dall'insuccesso. In quest'ultimo caso il riscontro empirico non indica quale delle ipotesi è sbagliata. Potremmo aver misurato scorrettamente i livelli di soglia, definito male gli obiettivi o aver commesso errori nell'organizzazione dell'intervento; ma non ci è dato di sapere esattamente cosa bisogna revisionare e migliorare. È, quindi, necessario come riconosce anche Gattullo⁴⁷, attingere ad altre fonti le informazioni che il controllo scolastico non fornisce autonomamente.

Una soluzione parziale al problema è proposta da Melchiori, che adotta un modello di rappresentazione dei progetti formativi che scardina le usuali logiche lineari a favore di soluzioni reticolari⁴⁸. In tale modello tutti gli elementi componenti sono definiti in relazione l'un dell'altro; in altri termini vengono esplicitate tutte le connessioni esistenti tra obiettivi, attività, risorse e sistema di valutazione. In questo modo se un intervento non ha successo è possibile restringere l'area di ricerca degli elementi che non hanno funzionato. Ponendo, ad esempio, che un'iniziativa formativa voglia conseguire solo due obiettivi e che la maggior parte degli studenti non ha raggiunto uno dei due, si deduce che l'errore è probabilmente da ricercare solo nelle attività o in uno degli altri elementi connessi a tale traguardo.

Il testo con cui si conclude questa seconda parte dell'antologia è di Howard Gardner e si intitola *La valutazione contestualizzata: l'alternativa ai test standardizzati* (lettura 11). Si tratta di una critica serrata alla docimologia, all'applicazione del modello positivista-sperimentale alla valutazione scolastica. Le argomentazioni addotte dall'Autore non ci sembrano generalizzabili: nel nostro Paese, per esempio, la tecni-

⁴⁵ Cfr. Domenici, G. (1981), *Descrittori dell'apprendimento*, Teramo, Giunti & Lisciani. L'argomentazione è riportata più brevemente in Idem (2001²), *Manuale della valutazione scolastica*, Roma-Bari, Laterza, pp. 190-195.

⁴⁶ Per un'introduzione storica al problema cfr. Gillies, D., Girello, G. (1995), *La filosofia della scienza nel XX secolo*, Roma-Bari, Laterza, pp. 121-142.

⁴⁷ Cfr. Gattullo, M. (1968), *Op. cit.*, p. 16.

⁴⁸ Cfr. Melchiori, R. (2000), *Progetto SEPI. Linee guida per l'autoanalisi di istituto e il miglioramento della qualità del servizio scolastico*, SPE-PSN-PTI, 28 febbraio, pp. 17-20.

ca del *testing* non è diffusa come negli Stati Uniti, dove – e questa ci sembra l'intima essenza delle critiche mosse da Gardner – l'aspetto tecnico prevale sul fine, gli strumenti di verifica sulla funzione didattica.

Tali critiche, tuttavia, rimangono di indiscutibile interesse avvertendoci della necessità di rendere semplice e naturale il processo di valutazione, di utilizzare strumenti diversi al variare dell'oggetto sottoposto a controllo, di elaborare sistemi di valutazione capaci di descrivere il complesso profilo di uno studente, di conservare un certo equilibrio tra esigenze didattiche e metrologiche. Sul secondo e sul terzo punto di questo elenco torneremo nella terza parte di questo lavoro; in questa sede ci sembra opportuna una breve riflessione su un elemento implicito nelle argomentazioni di Gardner, ossia sul rapporto tra valutazione di prodotto e di processo.

La forma docimologica della valutazione scolastica misura il conseguimento da parte degli allievi di un certo risultato, ovvero presta attenzione al prodotto del processo di apprendimento. In realtà, da un punto di vista didattico, potrebbe essere assai utile conoscere e pilotare i processi di apprendimento. Cercheremo di spiegarne la ragione molto brevemente. Supponiamo che le conoscenze di uno studente siano rappresentabili attraverso una mappa concettuale, ovvero attraverso un insieme di concetti e di operazioni che li pongono in relazione. In questa prospettiva l'acquisizione di un nuovo concetto comporta una riconfigurazione più o meno significativa della mappa concettuale. Questo processo, da un lato, richiede tempo, dall'altro, è causa della parziale e completa ridefinizione dei concetti e delle capacità operative già possedute.

Con queste premesse si comprende intuitivamente la ragione per cui per un docente (e per lo stesso discente) è importante controllare e pilotare il processo di apprendimento: si tratta di sapere come evolve la forma della mappa, di individuare quali strategie cognitive (motivazionali e volitive) ha attivato lo studente e di orientarlo conseguentemente. In altre parole, si tratta di facilitare i processi di apprendimento⁴⁹.

⁴⁹ Per la rappresentazione della conoscenza individuale attraverso le mappe concettuali cfr. Pellerrey, M. (1994²), *Progettazione didattica. Metodi di programmazione educativa scolastica*, Torino, SEI; Novak, J. D., Gowin, D. B. (1984), *Learning how to learn*, New York, NY, Cambridge University Press; tr. it. *Imparando ad imparare*, SEI, Torino, 1989. Per l'importanza della valutazione di processo cfr. Lichtner, M. (2004), *Valutare l'apprendimento: teorie e metodi*, Milano, Franco Angeli.

8. *Approcci alla valutazione (Stame, N., Tre approcci principali alla valutazione: distinguere e combinare)*⁵⁰

1. L'approccio "positivista-sperimentale"⁵¹

L'approccio "positivista-sperimentale" è nato negli USA contemporaneamente al varo dei primi programmi di "Guerra alla povertà". Ad esso hanno portato contributi fondamentali dai primi autori legati al *Bureau of Applied Social Research* (Hyman, 1967; Suchman, 1967; la prima Carol Weiss 1972)⁵², ai classici della sperimentazione (Campbell 1999)⁵³, così come autori più eclettici come Rossi e Freeman (1999)⁵⁴ e Chen (1990)⁵⁵.

Esso è stato concepito per programmi volti ad affrontare gravi e ricorrenti problemi sociali (disoccupazione, degrado urbano, criminalità, ecc.). Secondo una logica da azione razionale, si pensa che i programmi siano articolati in obiettivi da raggiungere, mezzi tramite cui raggiungerli, e risultati attesi, e che si collochino in un ciclo politico che va da decisione ad implementazione a valutazione a nuova decisione.

Il compito della valutazione consiste nel verificare e misurare se gli obiettivi sono stati raggiunti: gli obiettivi sono l'elemento rispetto al quale avviene il confronto con il risultato ottenuto. Questo comporta riuscire a definire quali sono gli obiettivi

⁵⁰ Tratto da: Stame N. (2001), *Tre approcci principali alla valutazione: distinguere e combinare*, in Palumbo, M. (2001), *Il processo di valutazione. Decidere, programmare, valutare*, Franco Angeli, Milano, pp. 21-46.

⁵¹ Uso il termine "positivista" non in senso spregiativo, come qualcuno ha pensato, ma perché questo è il nucleo teorico da cui ha tratto origine. Aggiungo il termine "sperimentale" perché in questo approccio la metodologia utilizzata è qualificante. Come si vedrà in seguito, altri usano nomi diversi, che colgono aspetti particolari di quella concezione: oggettivista, razionale, ecc.

⁵² Hyman, H. H. (1967), *Evaluating Social Action Programm*, in Lazarsfeld, P. F. (a cura di) (1967), *The Uses of Sociology*, New York, Basic Book; Suchman E. A. (1967), *Evaluative Research: Principles and Practice in Public Service and Social Action Programs*, New York, Russel sage Foundation; Weiss, C. (1972), *Evaluation Researches*, Englewood Cliffs, NJ, Prentice Hall [n.d.c.]

⁵³ Per una bibliografia sul filone della sperimentazione sociale negli anni '70 cfr. Stame (1998: 45-49). Per gli sviluppi nei decenni successivi cfr. Manski e Garfinkel (1992). [I testi citati in questa nota sono: Stame, N. (1998), *L'esperienza della valutazione*, Roma, SEAM; Mansky, C., Garfinkel, I. (1992), *Evaluating Welfare and Training Programs*, Cambridge, MA, Harvard University Press; quello citato nel testo è: Campbell D., con Russo J. (1999), *Social Experimentation*, Thousand Oaks, CA, Sage - n.d.c.].

⁵⁴ Il volume *Evaluation: a Systematic Approach* di Rossi e Freeman, la cui prima edizione è del 1980, è giunto alla sua sesta edizione nel 1999 (co-autore M. Lipsey) [Thousand Oaks, CA, Sage - n.d.c.]. Col procedere delle edizioni successive il testo si è via via ampliato ed ha tenuto sempre più conto degli altri approcci.

⁵⁵ Chen, H. T. (1990), *Theory-Driven Evaluation*, Newbury Park, CA, Sage [n.d.c.]

del programma (cosa sempre difficile), per poter individuare le variabili da studiare e gli indicatori rispetto ai quali effettuare delle misurazioni. Per far fronte alle ovvie difficoltà di aver a che fare con obiettivi poco chiari che sono il frutto di compromessi politici, si è poi introdotta la distinzione tra scopi (*goals*) e obiettivi (*objectives*) (Rossi, Freeman e Lipsey 1999: 94)⁵⁶: mentre i primi esprimono delle aspirazioni non quantificabili, sono questi ultimi (formulati con indicatori quantitativi che misurano lo stato di una variabile⁵⁷) che divengono la meta da raggiungere.

Poiché il programma è formulato come un'ipotesi di un cambiamento desiderato, la valutazione tende a verificare se tale cambiamento si è verificato, e se ciò dipende veramente dal programma e non da altra concausa: lo scopo della valutazione è verificare l'efficacia del programma nel conseguire un obiettivo che dovrebbe risolvere un problema, e quindi la generalizzabilità di un tale intervento in altre situazioni simili. Per tali motivi, si ritiene più adatto il metodo sperimentale (o quasi-sperimentale, Campbell 1969)⁵⁸, nel quale si mette a confronto un gruppo sperimentale sottoposto al programma con un gruppo equivalente che non riceve il programma (gruppo di controllo): in questo modo, le differenze osservate tra i due gruppi dovrebbero indicare le responsabilità del programma nel realizzare il risultato desiderato.

All'inizio, con le grandi sperimentazioni sociali degli anni '70, si è trattato di trovare un metodo per valutare programmi sperimentali, in cui il disegno sperimentale di valutazione è coevo al varo dell'azione. In questo caso ci si preoccupa di confrontare la situazione post-intervento con quella pre-intervento. I modelli impiegati erano prevalentemente di tipo quasi-sperimentale, e consistevano nell'osservazione prima/dopo di diversi gruppi equivalenti cui veniva somministrato l'*input*, ma senza gruppo di controllo. Successivamente si sono sviluppati metodi sperimentali basati sul confronto tra situazione con/senza intervento, ottenuto tramite osservazioni nel corso dell'attuazione di un intervento⁵⁹.

Le tecniche di ricerca usate sono tutte quelle quantitative: dal questionario ai modelli econometrici, ossia tutte quelle che consentono entrambi i tipi di confronto. Si tratta di un metodo di difficile applicazione, sia per le difficoltà di trovare dati sulla situazione pre-intervento o di controllare gli effetti di altre possibili cause, sia per la difficoltà stessa di osservare un gruppo (tanto sperimentale che di controllo) al di fuori del suo contesto.

Nel caso di un confronto *pre/post* intervento il disegno valutativo viene impostato all'inizio, e la valutazione viene fatta *ex post*; in quello di un confronto con/senza

⁵⁶ Rossi, P. H., Freeman, H. E., Lipsey M. W. (1999⁶), *Evaluation: a Systematic Approach*, Thousand Oaks, CA, Sage [n.d.c.].

⁵⁷ Invece di programmare la eliminazione, o anche solo la riduzione, della disoccupazione, ci si può prefiggere di "creare un tot di posti di lavoro", di "diminuire la disoccupazione di tanti punti percentuali", di "portare il livello di disoccupazione alla media europea".

⁵⁸ Campbell, D. (1969), "Reforms as Experiments", *American Psychologist*, 4 [n.d.c.]

⁵⁹ Rebecca Maynard (2000) definisce questi sperimentalisti della seconda generazione come quelli "dal volto umano" [Il testo citato in questa nota è: Maynard, R. A. (2000), "Sociologi, economisti, psicologi ... o semplicemente buoni valutatori? Lezioni dall'esperienza statunitense", *Rassegna Italiana di Valutazione*, 17-18 - n.d.c.].

intervento viene fatta a intervalli regolari. I valutatori sono esterni al programma, e si ritiene che pertanto siano obiettivi e non ne influenzino l'andamento. Coerentemente con l'idea di teoria del programma presente in questi autori, non si indaga il processo di attuazione: secondo la nota distinzione di Scriven (cfr. Stame 1998: 55)⁶⁰ non si fa una valutazione "costruttiva" (*formative*), per decidere cosa sta funzionando o meno ed eventualmente modificare il programma insieme agli amministratori, ma si attende di vedere i risultati del programma per permettere al decisore di stabilire se esso merita di essere proseguito o interrotto (valutazione riepilogativa, o *summative*). Questa formulazione estremistica iniziale è stata poi mitigata con lo sviluppo di molte tecniche di monitoraggio e di misurazione intermedia dei risultati, sulla scia del lavoro iniziato da Rossi e Freeman con la loro idea di "valutazione cucita su misura dei programmi" (1999: 164)⁶¹.

A questo approccio corrisponde un'idea di utilizzazione "strumentale" della valutazione: la valutazione dovrebbe fornire *input* di informazione che i politici trasformino in decisioni.

Il limite principale di questo approccio è che esso può dimostrare che qualcosa si modifica dopo che è stato introdotto un *input*, o perché è presente una variabile, ma non è in grado di dire perché ciò avviene. Alcuni autori sostengono espressamente che ciò non è necessario. Rispetto a questo, dal suo interno si sono levate voci critiche della centralità della metodologia, a favore di una centralità della teoria, per concentrarsi sulla ricerca del perché il programma dovrebbe funzionare come previsto: è l'approccio della "valutazione guidata dalla teoria" (*theory-driven evaluation*) sviluppato da Chen (1990)⁶², il quale sostiene che compito della valutazione è chiarire la teoria alla base del funzionamento del programma, individuare ipotesi forti che possano essere testate, per escludere le ipotesi rivali e far uscire dalla incertezza sulle cause che inficerebbe tutto l'approccio. Come vedremo in seguito, questo tema è al centro del dibattito attuale, e ha trovato risposte significative negli sviluppi che più avanti considero come forme di "contaminazione tra approcci".

Altro limite dell'approccio "positivista-sperimentale" è che si osserva solo ciò che ci si è predisposti ad osservare, non si è attrezzati per scoprire e capire effetti inattesi. Ciò è una conseguenza stessa del disegno di valutazione, e della sottovalutazione di tutto quello che ha a che vedere con il processo di implementazione di un programma, con il comportamento della burocrazia, ecc. Inoltre, i suoi risultati si possono conoscere solo dopo, anche molto dopo, la fine del programma, mentre i politici vorrebbero poter sapere subito come sta andando, dato che spesso il loro nome è legato alla riuscita del programma.

Tuttavia, nonostante questi ed altri limiti di cui è stato tacciato (gli obiettivi non sono chiari, le informazioni non sono sufficienti, è difficile poter generalizzare un successo, i decisori non usano "razionalmente" la valutazione), questo approccio continua ad essere ritenuto quello più adatto, tanto nei programmi di investimenti

⁶⁰ Stame, N. (1998), *L'esperienza della valutazione*, cit. [n.d.c.].

⁶¹ Rossi, P. H., Freeman, H. E., Lipsey M. W. (1999⁶), *Evaluation: a Systematic Approach*, cit. [n.d.c.].

⁶² Chen, H. T. (1990), *Theory-Driven Evaluation*, cit. [n.d.c.].

pubblici (in cui è più facile confrontare obiettivi e risultati) quanto nei programmi sociali, come quelli di “riduzione della povertà” finanziati dal *welfare state* negli USA e di sviluppo sostenibile finanziati dalle agenzie mondiali di cooperazione allo sviluppo, o in quelli europei per la coesione sociale, che sono formulati in termini di *input* (voci di spesa; assi-misure-azioni) e di *output* (risultati). La presa intellettuale di questo approccio, che continua ad essere visto come il metodo valutativo per eccellenza, si può far risiedere tanto nel desiderio dei politici di avere risposte chiare, quanto nella persistente influenza del paradigma scientifico che lo sottende presso consulenti e centri pubblici di ricerca.

2.2. L'approccio “pragmatista – della qualità”

L'approccio che chiamiamo “pragmatista - della qualità” riunisce due scuole per tanti aspetti diverse, ma che hanno in comune la pietra di paragone: giudicare un programma in base ad un'idea di valore, che può essere un concetto di merito detenuto da attori interni al programma e/o dai valutatori, o uno standard definito all'esterno del programma.

Esso è nato in ambiente pragmatista, come opposizione intellettuale al positivismo, e prende spunto dalla disputa sui valori che per il pragmatismo, contrariamente al positivismo, sono centrali in ogni attività di ricerca, e che devono tanto più esserlo in una ricerca come quella valutativa che ha la sua ragion d'essere nel giudicare. Secondo Scriven, fondatore dell'approccio (egli lo chiama in vario modo: “prospettivista”, “transdisciplinare”), il valutatore non deve farsi influenzare dagli obiettivi del programma, e la sua deve essere una “*goal free evaluation*”; mettendo il programma e i suoi obiettivi sopra tutto - come nell'approccio “*goal oriented*” dei positivisti - ci si fornirebbe invece un alibi per una pretesa neutralità valoriale del valutatore. Quest'ultimo deve dare un giudizio sul programma in base alle proprie competenze e ai valori, che sono “oggettivi” e propri alla situazione ed al programma.

Per Scriven (1992)⁶³ il giudizio di valore si scompone in due aspetti, che vanno usati entrambi: *merit*, o valore intrinseco ad un'attività, e offerto secondo lo standard di “qualità” di quell'attività (ad esempio, un corso di francese, che ha valore perché l'insegnante di francese insegna bene il francese); *worth*, o valore estrinseco e quindi che incontra i bisogni dei destinatari, nel contesto in cui si svolge (nell'esempio, per quanto l'insegnante di francese sia bravo il corso non ha valore in una situazione in cui si preferisce l'insegnamento dello spagnolo). Questi giudizi vanno utilizzati entro la “logica del valutare”⁶⁴, che - per Scriven - si compone delle

⁶³ Scriven, M. (1992⁴), *Evaluation Thesaurus*, Newbury Park, Sage Publications.

⁶⁴ Scriven sostiene che questo modo di procedere assomiglia molto alla valutazione dei prodotti che fanno le associazioni dei consumatori. Per un aggiornamento del dibattito sulla “logica del valutare” si veda Fournier (1995), che contiene anche un contributo dello stesso Scriven [Il testo citato in nota è: Fournier, D. M. (a cura di) (1995), *Reasoning in Evaluation: Inferential Links and Leaps*, New Directions for Evaluation, n. 68, San Francisco, Jossey-Bass - n.d.c.].

fasi di: stabilire criteri di merito e standard, misurare la *performance* dei singoli programmi, dare un punteggio, ordinare più programmi in una graduatoria, sintetizzare i risultati in un giudizio finale di valore.

Qui dunque l'elemento di confronto non è più ciò che un singolo programma può voler raggiungere in una data situazione (obiettivi), ma ciò a cui tutti i programmi simili dovrebbero mirare su una certa scala di merito se vogliono essere considerati di buona qualità. Si tratta di una differenza di fondo, che viene spesso sottovalutata considerando che anche gli obiettivi potrebbero essere formulati come *standard*, ma senza tener conto del diverso ruolo assunto dal valutatore nel giudicare un programma rispetto ad altri simili⁶⁵.

Alcuni di questi aspetti possono essere rinvenuti in tutti quei modi di valutare il rendimento di istituzioni o di servizi che si ispirano alle teorie manageriali del *Total Quality Management*, e che sono entrati nella gestione dei servizi pubblici (socio-assistenziali, formativi, ecc.). In effetti, essi si adattano bene a valutare il rendimento di prestazioni che non rientrano in programmi *ad hoc*, ma che si svolgono regolarmente perché rispondono a bisogni ricorrenti e stabili dei cittadini, e per le quali bisogna stabilire se sono offerte in modo adeguato, secondo criteri di efficacia ed efficienza. La qualità diventa quindi la proprietà positiva da rilevare: e poiché è concetto astratto, si tenta di scomporlo in vari aspetti, per cui si cercano degli indicatori, rispetto ai quali stabilire degli standard di raggiungimento. Ad es., per quanto riguarda la salute, Donabedian (1990)⁶⁶ suggerisce di distinguere le tre grandi categorie che compongono un servizio (struttura, processo, esito), rispetto alle quali individua i criteri di adeguatezza, imparzialità, continuità, ecc., che dovranno essere analizzati con indicatori specifici, e su cui occorrerà stabilire gli standard. Si tratta di un'operazione molto complessa, e soggetta a molti condizionamenti politici. Come si fissano gli standard? Ossia, rispetto a cosa ci si confronta: un *benchmark* (che è una soglia elevata, che potrebbe sempre essere aumentata, come nei sistemi di Miglioramento Continuo della Qualità), un'altra situazione simile, la propria situazione nel periodo precedente (cfr. Wholey 1983: 155 e segg.)⁶⁷? Sono mete da perseguire o soglie minime sotto cui non si può scendere? Chi li fissa: l'autorità politica, le amministrazioni, gli esperti, la concertazione tra le parti sociali? Sono nazionali o stabiliti localmente dalle amministrazioni⁶⁸? Come si vede, la fissazione degli standard può agire come stimolo al miglioramento, così come può consentire di adattarsi su strategie difensive.

⁶⁵ Basti pensare che in un famoso contributo Scriven (1976) ha criticato il metodo sperimentale contrapponendogli il proprio *modus operandi* [Il testo citato in nota è: Scriven, M. (1976), "Maximizing the Power of Causal Investigations: the Modus Operandi Methods", *Evaluation Studies Review Annual*, 1 - n.d.c.].

⁶⁶ Donebian, A. (1989), *La qualità dell'assistenza sanitaria*, Roma, Nuova Italia Scientifica [n.d.c.].

⁶⁷ Wholey, J. (1983), *Evaluation and Effective Public Management*, Boston, Little and Brown [n.d.c.].

⁶⁸ Questa problematica è stata al centro della discussione sulla attuazione delle Carte dei Servizi in Sanità. Cfr. Lo Schiavo (1999) per un interessante confronto tra Italia e UK. [Il testo citato in nota è: Lo Schiavo, L. (1999), "Qualità dei servizi pubblici e valutazione", *Il nuovo governo locale*, numero speciale dedicato alla Carta dei servizi - n.d.c.]

Se la fissazione degli standard è attività più legata alla programmazione, la valutazione ha comunque un ampio campo di azione nel misurare le *performance* rispetto agli standard, e nel dare giudizi rispetto agli scostamenti. Per queste attività ci si serve soprattutto di modelli di “giudizi degli esperti” (*Delphi groups*, NGT), che raramente coinvolgono gli *stakeholders*, se non per la valutazione della soddisfazione degli utenti. Ciò ha il pregio di sembrare semplice e fattibile; ma ha anche il difetto della rigidità, dato che i modi di prestare un servizio e le esigenze degli utenti cambiano in continuazione, e quindi si rende necessario trovare meccanismi di aggiustamento continuo degli standard. Rispetto a questo problema, tuttavia, si può dire che ci siano stati recentemente sviluppi molto significativi nei metodi di indagine, sia dal punto di vista delle tecniche per assegnare le preferenze (analisi multi-criteri, analisi delle preferenze: cfr. Taralli 1998⁶⁹), che da quello del coinvolgimento di “esperti” laici e “del terreno” (operatori, beneficiari) a fianco degli esperti professionali (cfr. Bezzi 2001)⁷⁰.

Questo approccio è la base su cui sono nati i sistemi di “*performance review*” che si adattano alle politiche del *New Public Management* (Gore 1993; Pollitt 1995)⁷¹. Ad esso sono stati portati contributi fondamentali sia da autori che si sono occupati del settore pubblico in generale, come Wholey (1989)⁷², sia da autori impegnati in settori particolari come la sanità o la formazione, ed è quello cui si ispirano le nostre politiche di riforma del settore pubblico (Azzone e Dente 1999)⁷³.

Esso è uno strumento nelle mani del *management* pubblico, che vuole sapere quanto sta producendo il suo servizio (la propria efficienza nel fornire un dato servizio, il *value for money* di cui parlano gli inglesi) o nelle mani di un’amministrazione centrale, che vuole confrontare servizi simili offerti da strutture diverse. E nel sistema sempre più sviluppato dell’*outsourcing* è visto come mezzo per confrontare le prestazioni esterne rispetto ai requisiti richiesti dall’amministrazione pubblica. Per questi motivi, si può dire che qui si prevede un uso “strumentale” della valutazione, per consentire decisioni amministrative.

Le ragioni della sua grande diffusione fanno tuttavia capire alcuni dei problemi principali che esso sta attualmente incontrando. Nei recenti sviluppi che abbiamo appena accennato l’approccio della qualità si è venuto identificando con la pratica della certificazione, e con l’applicazione anche ai servizi pubblici del sistema ISO 9000, che era nato con lo scopo di standardizzare la produzione di merci e di servizi privati. In tal modo si è venuti a modificare radicalmente la concezione iniziale del valutatore: da colui che dà giudizi in base alla propria capacità lo si è ridotto a chi

⁶⁹ Taralli, S. (1998), *La scala delle priorità obbligate: una teoria per rilevare le gerarchie di valore*, in Bezzi, C., Palumbo, M. (a cura di), *Strategie di valutazione*, Perugia, Arnaud-Gramma [n.d.c.].

⁷⁰ Bezzi, C. (2001), *Il disegno della ricerca valutativa*, Perugia, Giada [n.d.c.].

⁷¹ Gore, A. (1993), *Creating a Government that Works Better and Costs Less. Report of the National Performance Review*, New York, Random House; Pollitt, C. (1995), “Justification by Work or by Faith? Evaluating the New Public Management”, *Evaluation*, 2 [n.d.c.].

⁷² Wholey, J. (1989), *Improving Government Performance*, San Francisco, Jossey Bass [n.d.c.].

⁷³ Azzone, G., Dente, B. (a cura di) (1989), *Valutare per governare. Il nuovo sistema dei controlli nelle Pubbliche Amministrazioni*, Milano, ETAS [n.d.c.]

certifica il raggiungimento di standard e assicura l'osservanza di procedure, entrambi fissati da entità esterne. Secondo lo stesso Scriven (1982)⁷⁴, coloro che parlano della qualità dei servizi riferendosi a standard, norme ISO, accreditamenti ecc., si riferiscono solo al *merit*, e si limitano a fare una *formative evaluation* per il *management* (che assomiglia ad una valutazione *goal oriented*), e non riescono a fare anche una valutazione orientata al consumatore, che dovrebbe essere *goal free*, e basata su un giudizio di *worth*.

2.3. L'approccio "costruttivista - del processo sociale"

L'approccio che abbiamo chiamato "costruttivista — del processo sociale" si rifà a una serie di modelli. Da una parte, la discussione epistemologica sui paradigmi avviata con la "quarta generazione della valutazione" di Guba e Lincoln (1992)⁷⁵; dall'altra, i contributi attenti alla valutazione come attività che si riferisce ad un processo sociale e si svolge in un contesto politico ogni volta mutevole (Cronbach *et al.* 1980)⁷⁶ con cui i valutatori interagiscono (la *responsive evaluation* di Stake 1980)⁷⁷; inoltre, tutti quei modelli che insistono sulla finalità della valutazione di sviluppare le capacità degli utilizzatori, siano essi *stakeholders* o beneficiari, come *l'empowerment evaluation* (Fetterman *et al.* 1996)⁷⁸, la "valutazione orientata all'utilizzatore" (Patton 1986)⁷⁹; infine, consideriamo dentro questo approccio anche una serie di contributi provenienti al mondo della *policy analysis*, che tendono a mettere in luce il processo attraverso cui si attuano i programmi, ciò che succede di atteso o di inatteso. Al di là delle ovvie differenze, tutti questi modelli hanno in comune un'attenzione al contributo dei vari attori, e a cosa un programma diventa mentre viene attuato, molto più che a come è stato disegnato.

Esso viene applicato non solo - come è ovvio - quando si ha a che fare con situazioni nuove e con programmi pilota⁸⁰, ma anche perché si pensa che non esistano altri criteri di successo che quelli che vengono definiti all'interno della situazione specifica. Ciò significa che - per quanto un programma sia sempre avviato per otte-

⁷⁴ Scriven, M. (1982), *Evaluation Thesaurus*, Beverly Hills, Ca, Sage [n.d.c.].

⁷⁵ Guba, E., Lincoln, Y. (1989), *Fourth Generation Evaluation*, Newbury Park, Ca, Sage [n.d.c.].

⁷⁶ Cronbach, J. L. *et al.* (1980), *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*, San Francisco, Jossey-bass [n.d.c.].

⁷⁷ Stake, R. (1980), *Program Evaluation, Particularity Responsive Evaluation*, in Dockrell, W. B., Hamilton, D. (a cura di) (1980), *Rethinking Educational Research*, London, Hodder and Stoughton [n.d.c.].

⁷⁸ Fatterman, D. L., Kaftarian, S. J., Wandersman A. (1996), *Empowerment Evaluation. Knowledge and Tools for Self-Assessment & Accountability*, Thousand Oaks, Sage [n.d.c.].

⁷⁹ Patton, M. (1986), *Utilization-focused Evaluation*, Newbury Park, CA, Sage [n.d.c.].

⁸⁰ Nei programmi pilota dovrebbe usarsi un metodo esplorativo (Schön *et al.* 1984: 27) e non sperimentale (cfr. Stame 1998: 51) [I due testi citati in nota sono: Schön, D., Drake, W. D., Miller, R. I. (1984), "Social, Experimentation as Reflection-in-Action: Community Level Nutrition Intervention Revisited", *Knowledge*, 1; Stame, N. (1998), *L'esperienza della valutazione*, cit. - n.d.c.].

nere un cambiamento - la pietra di paragone, ossia la definizione di cosa sia un successo, è un concetto che si forma in un momento successivo al varo del programma, e non precedente: è quando si vede come affronta i problemi, lo si confronta con altre situazioni e si sente l'opinione dei vari *stakeholders* che si può formulare un giudizio su cosa funziona bene o male, su cosa possa essere considerato un successo, anche se ciò dovesse avvenire in base ad effetti inattesi (Hirschman 1988; Tandler 1992)⁸¹. La realtà può riservare molte più sorprese di quanto sia previsto nello schema di un programma.

L'idea di fondo è stata avanzata nei testi classici di Cronbach *et al.* (1980)⁸² e Stake (1980)⁸³. La valutazione deve tener conto del fatto che ogni volta che si attua un programma esso muta a contatto con il contesto: i problemi sentiti dai vari *stakeholders* saranno diversi, e le conclusioni raggiunte per un programma non potranno essere generalizzate ad altre situazioni in cui vengono attuati programmi simili. Ciò non significa tuttavia che non si possano trarre delle lezioni da applicare anche altrove.

Da una parte, un programma va analizzato nel suo contesto, sociale e istituzionale. Per fare ciò ci si servirà di analisi di casi, di interviste in profondità, ma non solo: potranno essere utili anche strumenti di indagine quantitativa se colgono un aspetto che gli attori ritengono centrale. La cosa importante è che le parti siano coinvolte in una valutazione partecipata⁸⁴, che il processo sia seguito nelle sue varie fasi da valutatori che interagiscono con gli *stakeholders*, e che in presenza di interessi e posizioni in conflitto i valutatori possano assumere ruoli che vengono chiamati in letteratura con nomi che vanno da "arbitri", a "negoziatori", a "facilitatori"⁸⁵.

Dall'altra parte, il valutatore dovrà fare appello alla propria esperienza per confrontare il caso analizzato con altri, più o meno simili. Dal confronto nasceranno giudizi, e lezioni che potranno poi essere reinterpretate in altri contesti.

Questo approccio è finalizzato ad un uso "conoscitivo" della valutazione, sia nel senso che esso tende a comprendere la situazione attraverso le interpretazioni che

⁸¹ Hirschman, A. O. (1975), "Political Economics" e possibilismo, in Idem (1988), *Come complicare l'economia*, a cura di Meldolesi, L., Bologna, il Mulino; Tandler, J. (1992), *Progetti ed effetti*, a cura di Stame, N., Napoli, Liguori [n.d.c.].

⁸² Cronbach, J. L. *et al.* (1980), *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*, cit. [n.d.c.].

⁸³ Stake, R. (1980), *Program Evaluation, Particular Responsive Evaluation*, cit. [n.d.c.].

⁸⁴ Su questo tema la letteratura è copiosa. Rimando per una buona sintesi a Whitmore (1998) e Jackson e Kassam (1998) [I due testi citati in nota sono: Whitmore, E. (a cura di) (1998), *Understanding and Practicing Participatory Evaluation*, New Directions for Evaluation, n. 80, San Francisco, Ca, Jossey Bass; Jackson, E., Kassam, Y. (1998), *Knowledge Shared: Participatory Evaluation in Development Cooperation*, West Hartford, CT, Kumarian Press - n.d.c.].

⁸⁵ Qui si può notare una differenza forte con gli altri due approcci. In particolare Scriven ha accusato Patton di non distinguere tra consulenza (in cui si stipula un contratto con un cliente, il quale si impegna a seguire i consigli del consulente) e valutazione (in cui il valutatore si limita a dare un giudizio sulla qualità di un programma/prodotto); al che Patton ha reagito dicendo che per essere utile all'utilizzatore il valutatore si deve mettere nei panni di quest'ultimo, «pensare e parlare con le sue parole».

ne danno gli attori, a definire i problemi, a chiarire ciò che si può ottenere coi programmi; sia nel senso del potenziamento delle capacità dei beneficiari di utilizzare il programma per meglio gestire le proprie attività e favorire un processo di apprendimento.

9. Le funzioni del controllo scolastico nel processo educativo (Gattullo, M., *Didattica e docimologia. Misurazione e valutazione nella scuola*)⁸⁶

1. I COMPITI GENERALI DEL CONTROLLO SCOLASTICO

Da un punto di vista formale è possibile distinguere, nel processo educativo, quali ne siano le direzioni e i contenuti effettivi, quattro momenti: vale a dire, gli *scopi* che in funzione del processo stesso s'intendono realizzare, gli *strumenti*, che in relazione a quegli scopi sono utilizzati, l'*organizzazione* concreta secondo cui è articolato l'uso di quegli strumenti, e finalmente il *controllo*, inteso ad accertare se e fino a qual punto gli scopi proposti sono stati raggiunti mediante l'utilizzazione degli strumenti prescelti, e del modo in cui questi ultimi sono stati utilizzati.

a) Cosa bisogna intendere col concetto di *scopo* del processo educativo? Il significato di esso non è univoco, né in relazione alle direzioni e ai contenuti effettivi che esso può assumere, né in funzione dell'importanza che gli scopi possono avere nello stesso processo educativo, né in rapporto alla giustificazione culturale della loro presenza. In relazione ai contenuti e alle direzioni del processo educativo, si può parlare, tra gli altri, di scopi morali, sociali, religiosi, intellettuali, professionali, estetici, i quali tra loro possono essere in rapporti di accordo (totale o parziale), di continuità, di indifferenza o di opposizione. In funzione dell'importanza degli scopi nel processo educativo, se ne può certo compiere una distinzione generale, che li può classificare in finali e immediati, subordinando i secondi ai primi. In rapporto alla giustificazione culturale della presenza, si può parlare di scopi presenti per richieste ed esigenze sia di carattere sociale, sia di carattere psicologico, sia di carattere più propriamente pedagogico, senza peraltro escludere che tali esigenze e richieste possano reciprocamente escludersi o accordarsi. Quel che più qui conta rilevare, comunque, sta nel fatto che, indipendentemente dalla problematicità del concetto di scopo educativo, riconducibile alle ragioni di cui prima s'è detto, esso costituisce momento formalmente essenziale del processo educativo.

b) La scelta degli *strumenti*, da utilizzare per la realizzazione degli scopi del processo educativo, è evidentemente in funzione primaria degli scopi medesimi, nel senso di una loro subordinazione a questi ultimi: sarebbe assurda la scelta d'uno strumento che impedisse di fatto il raggiungimento degli scopi prefissi, o fosse loro indifferente. Gli strumenti educativi - tra cui soprattutto sono da vedere i contenuti

⁸⁶ Tratto da: Gattullo M. (1968), *Didattica e docimologia. Misurazione e valutazione nella scuola*, Roma, Armando, pp. 9-17, 31-34, 42-48, 51.

di cultura, di cui il soggetto educativo si deve impadronire, e i sussidi didattici, utilizzati principalmente nella vita scolastica - sono tuttavia prescelti anche in funzione di mediazioni di carattere teorico, da identificare nelle dottrine pedagogiche e didattiche, a loro volta giustificate in modo esclusivo o prevalente dalla tradizione educativa o da ragioni di carattere psicologico.

c) L'*organizzazione* degli strumenti educativi è da vedere a diversi livelli, di differente complessità: a un livello più semplice sono da vedere, per esempio, l'organizzazione della lezione da parte dell'insegnante, i rapporti tra una lezione e un'altra, il legame tra attività dell'insegnante e attività degli scolari: tutte forme d'organizzazione del lavoro educativo riconducibili, in modo esclusivo o prevalente, alla tradizione educativa o a esplicite teorie pedagogiche o didattiche; a un livello più complesso sarebbe da vedere, per esempio, l'organizzazione del «curriculum» degli studi d'un qualsiasi tipo di scuola, le scelte riguardanti il quale sarebbero anch'esse riconducibili, in modo esclusivo o prevalente, alla tradizione educativa o a teorie pedagogiche o didattiche presenti in modo più o meno esplicito.

d) Nel quadro finora prospettato ha funzione e significato il *controllo* educativo: quel ch'esso deve compiere si identifica nell'accettamento se gli strumenti, utilizzati secondo l'organizzazione educativa prescelta, hanno permesso, e fino a qual punto, di raggiungere gli scopi proposti.

e) Qualche esempio potrà chiarire meglio il discorso che finora s'è condotto. Uno degli *scopi* dell'istruzione nella scuola dell'obbligo, dunque riferibile a tutti i soggetti educativi, risiede nell'apprendimento delle capacità di leggere e scrivere: tale scopo può essere inteso principalmente come legato alla direzione intellettuale del processo educativo di tale tipo di scuola; può essere inteso come immediato, se si riflette che il suo raggiungimento è strumentale in relazione a scopi di più complesso significato culturale; le giustificazioni della sua presenza sono poi da ricondurre soprattutto a ragioni di carattere sociale (l'opportunità che tutti sappiano leggere e scrivere; il diritto di ciascuno a possedere tali indispensabili strumenti culturali). Gli strumenti che l'insegnante può utilizzare per raggiungere un tale scopo possono esser diversi (sillabari, alfabetieri, lavagne, ecc.), ma essi lo saranno però volta a volta in modo diverso a seconda dell'*organizzazione educativa* del lavoro didattico ch'egli ha prescelto (prima di tutto, quella del metodo: il quale esige, a sua volta, un diverso tipo di utilizzazione dell'esperienza e degli interessi degli scolari); e ambedue i momenti formali di cui ora s'è detto, strumenti e loro organizzazione educativa, possono esser giustificati in funzione dell'esperienza dell'insegnante o di precise scelte culturali ch'egli deliberatamente ha compiuto. La funzione del *controllo* che l'insegnante compie, infine, nei modi ch'egli riterrà più opportuni (per esempio: dettati, brevi composizioni, prove oggettive di ortografia e di grammatica, letture silenziose o ad alta voce) è quella appunto di vedere se gli scolari sanno leggere e

scrivere, di accertare i diversi livelli della realizzazione di tali capacità, e di giudicare accettabili o no i livelli che è stato possibile accertare [...].

f) Se quanto finora s'è detto ha potuto chiarire in modo esplicito la *funzione* del controllo nel processo educativo, ha pure aggiunto, quantunque in modo implicito, che tale funzione vi è *indispensabile*. L'intera struttura «scopi - strumenti - organizzazione» sarebbe monca se il controllo non permettesse di accertare il raggiungimento dei primi, e il contributo degli altri ad esso; d'altro canto, un controllo a sè, privo di rapporti con tale struttura, sarebbe evidentemente privo di significato. Ma va pure detto fin d'ora in modo esplicito che l'identificazione dei momenti, che formalmente possono dirsi costitutivi del processo educativo, non ne esclude una gerarchia, il cui vertice dovrebbe esser visto negli scopi da raggiungere, e i cui altri momenti costitutivi sarebbero con funzione subordinata rispetto ad essi: gli strumenti e la loro organizzazione educativa, da una parte, col compito di permettere il raggiungimento degli scopi, e il controllo da un'altra, col compito di accertare, prima di tutto, se il raggiungimento degli scopi è stato o no compiuto. La subordinazione del controllo agli scopi da raggiungere implica dunque l'assurdità d'un processo educativo organizzato, in tutto o in parte, non in funzione degli scopi ma dei controlli, quasi che gli esami o i concorsi fossero e mete educative significative da raggiungere e da superare; ma un'assurdità analoga sarebbe ancora da vedere in una subordinazione del controllo agli strumenti educativi, da identificare nella maggior parte dei casi coi contenuti culturali, e dunque nella riduzione di esso al semplice accertamento della misura in cui tali contenuti sono o no assimilati da parte degli scolari.

2. L'OGGETTO DEL CONTROLLO SCOLASTICO

a) Il discorso sulla funzione del controllo scolastico nel processo educativo, e sui rapporti tra esso e gli altri momenti che costituiscono lo stesso processo, ne può già chiarire in modo sufficiente l'oggetto, che è da identificare con gli scopi stessi del particolare processo educativo, di cui il controllo medesimo fa parte [...].

c) [...] Quali ne siano le direzioni effettive, gli scopi del processo educativo implicano sempre, se sono realizzati, modificazioni più o meno grandi della personalità degli scolari; tali trasformazioni tanto più possono dirsi rispondenti alle finalità prefissate, quanto più in modo strutturato e definitivo esse hanno luogo. Da un tal punto di vista, oggetto del controllo scolastico può esser definita la presenza, nella personalità del soggetto educativo, delle trasformazioni se possibile strutturate e definite, che sono scopi del processo educativo.

D'altro canto, le trasformazioni della personalità degli scolari possono esser controllate fino a un certo punto: quel ch'è possibile accertare è limitato a una sfera molto ristretta del loro comportamento; è per esempio molto problematico effettuare indagini riguardanti il mondo affettivo della personalità, o le motivazioni profonde

di certi suoi atteggiamenti e comportamenti. Appunto per tali motivi, appare necessario restringere l'ambito delle competenze del controllo scolastico, definendone l'oggetto, con precisione maggiore, come la presenza, nel soggetto educativo, delle trasformazioni, se possibile strutturate e definitive, solo del suo comportamento obbiettivamente controllabile, che sono scopi del processo educativo.

Qualche esempio potrà chiarire quel che ora s'è detto. Se scopo dell'istruzione primaria è tra l'altro l'acquisizione, da parte degli scolari, delle capacità elementari di leggere e scrivere, è pure da dire che tale acquisizione può dirsi veramente realizzata soltanto a patto della sua strutturazione definitiva nel comportamento degli scolari: non potrebbe dirsi effettivamente raggiunto uno scopo, come quello che s'è detto, se dopo la conclusione della scuola dell'obbligo quelle capacità elementari scomparissero o anche solo scemassero. Un tale scopo, peraltro, è pressoché tutto traducibile in termini di comportamento obbiettivamente controllabile, dal momento che le abilità che con esso si identificano possono essere sollecitate in tale direzione.

[...]

4. I RAPPORTI TRA CONTROLLO SCOLASTICO E SCOPI DEL PROCESSO EDUCATIVO

[...] *b*) Si supponga compiuto il controllo scolastico relativo a una determinata fase del processo educativo, e si suppongano i risultati da esso dati del tutto insoddisfacenti: l'insegnante dovrà certamente chiedersi quali siano i motivi che hanno potuto portare a quei risultati. Le possibili risposte teoriche sono diverse, e legate ai diversi momenti del processo educativo; in relazione agli obbiettivi di questo, esse ne possono prospettare o l'effettiva irraggiungibilità (gli scolari non erano maturi per conseguirli), o l'inadeguatezza della formulazione (per cui essi non furono previsti fin da principio, ma solo nel momento conclusivo, in modo preciso e funzionale), o il mancato perseguimento sistematico (per cui gli scopi furono di fatto tralasciati durante lo sviluppo del processo educativo). Non è certo il controllo educativo da solo a poter dare una risposta adeguata a tali interrogativi; esso può tuttavia contribuire affinché quella sia il più adeguata possibile, dopo aver fatto sorgere i problemi che ne hanno sollecitato la ricerca.

5. I RAPPORTI TRA CONTROLLO SCOLASTICO E STRUMENTI E ORGANIZZAZIONE DEL PROCESSO EDUCATIVO

a) Il controllo compiuto a conclusione del processo educativo, o d'una sua fase, proprio perché ha la funzione di accertare raggiunti o no (e nel primo caso, a quali diversi livelli) gli scopi del processo medesimo, può dare indicazioni circa l'efficacia o l'insufficienza degli strumenti educativi utilizzati. Nel caso in cui gli scopi educativi fossero stati realizzati in modo soddisfacente, il giudizio sui mezzi adoperati dovrebbe esser senz'altro positivo; al contrario, nel caso opposto, gli strumenti potrebbero esser messi in discussione, e la responsabilità del mancato raggiungimento degli obiettivi proposti potrebbe esser attribuita alla loro insufficienza, o alla loro estraneità rispetto al processo educativo in cui sono stati utilizzati, o alle eccessive difficoltà, ad affrontare le quali la loro presenza potrebbe aver costretti gli scolari.

b) Indicazioni e suggerimenti analoghi, circa l'efficacia dell'organizzazione degli strumenti del processo educativo, potrebbero esser forniti dai risultati del controllo scolastico, tanto nel caso in cui tali risultati fossero soddisfacenti (l'organizzazione prescelta sarebbe così da giudicare funzionale), quanto nel caso contrario (quell'organizzazione potrebbe allora esser giudicata insufficiente, o troppo macchinosa, o implicante la presenza di obiettivi estranei a quelli che realmente si sarebbero voluti perseguire).

[...]

MISURAZIONE E VALUTAZIONE NEL CONTROLLO SCOLASTICO

1. LE TRE FASI DEL CONTROLLO SCOLASTICO

In qualunque modo esso sia condotto, e quale ne sia il livello della funzionalità, il controllo scolastico si articola strutturalmente in tre momenti, i quali sono da identificare nella scelta degli obiettivi da sottoporre ad accertamento, nelle operazioni da compiere per accertare presenti gli obiettivi prescelti, e nel giudizio da dare sui risultati dell'accertamento operato.

a) La scelta dell'oggetto da sottoporre ad accertamento costituisce il primo passo da compiere per effettuare il controllo scolastico. Essa può identificarsi con quella relativa agli obiettivi del processo educativo, e di fatto si identifica con essa, o le è collegata, nel caso in cui i controlli sono effettivamente subordinati a quegli obiettivi; ma anche quando ciò non accade, vale a dire, quando i controlli sono compiuti in modo più o meno arbitrario, ciò che essi intendono accertare è oggetto preventivo d'una scelta, che precede le operazioni concrete dell'accertamento. Anche nella più

semplice interrogazione orale, quando appunto le domande sono per così dire inventate sul momento, e spesso difficilmente ne potrebbe esser dimostrata la connessione con gli scopi finali del processo educativo, la formulazione di esse è preceduta da una scelta, che può esser giustificata con le ragioni più diverse, dalle più valide (l'effettiva intenzione di chieder qualcosa di significativo per la preparazione dello scolaro) alle meno valide (la stanchezza di formulare sempre le stesse domande, e perciò la volontà di cambiare argomento; oppure, l'ispirazione cercata e trovata nel libro di testo).

b) Scelto l'oggetto del controllo, si presenta l'esigenza di accertarne il raggiungimento da parte dello scolaro: e ciò vien compiuto con le modalità prescelte (che sono prove scritte od orali, nel caso del controllo degli apprendimenti intellettuali; prove diverse, oppure l'osservazione diretta, nel caso dell'accertamento della condotta). Tale accertamento può portare sia a constatare presenti o assenti quelle trasformazioni della personalità degli scolari, ch'erano da identificare con gli obbiettivi del processo educativo, sia a constatare livelli diversi della presenza di quelle stesse trasformazioni. Nell'interrogazione orale ciò può accadere, per esempio, nell'accertare a quante domande il candidato ha dato risposta soddisfacente, a quante egli ha dato risposte parziali, e a quante non ha saputo rispondere.

c) Il risultato dell'accertamento viene successivamente interpretato, e nei suoi confronti vien formulato un giudizio, che può essere di sufficienza o no, oppure può aver gradi diversi, sia al di sopra del livello di sufficienza, sia al di sotto di esso. Il voto, comunque esso venga espresso, rappresenta appunto il risultato conclusivo del controllo, in quanto esso riassume in un giudizio complessivo quello che l'interrogazione orale, nell'esempio prima riportato, è riuscita ad accertare della preparazione dello scolaro.

d) Da quanto finora s'è detto dovrebbe apparire chiaro che le tre s'identificano, la prima in un atto di valutazione, che può esser discusso, e sull'oggetto del quale può anche non sussistere accordo; la seconda in un atto d'accertamento, nei confronti dei risultati del quale, una volta che ne fosse stato stabilito l'obbiettivo, non dovrebbero poter sussistere contestazioni; la terza in un atto di valutazione, i criteri del quale, seppur discutibili, dovrebbero essere obbiettivamente determinabili.

[...]

3. ACCERTAMENTO O MISURAZIONE?

S'è finora intenzionalmente utilizzato, per indicare il secondo momento del controllo scolastico, il termine accertamento, e di esso s'è detto soltanto che può limitar-

si a constatare raggiunti o no gli obiettivi del processo educativo, oppure che può giungere anche a individuare gradi diversi, sia dell'avvenuto sia del mancato raggiungimento degli obiettivi prefissi. Non s'è voluto parlare di misurazione, nonostante il fatto che la maggior parte degli accertamenti del controllo educativo ne pretenda le caratteristiche, a causa del fatto che è lecito ritenerla possibile solo con le opportune garanzie. Il discorso che segue cercherà in breve di chiarirle.

a) La definizione generale del concetto di misurazione può esser data dicendo che essa consiste in un processo, il cui risultato è l'attribuzione di un numero agli oggetti, o a qualche loro qualità; vale a dire, con altre parole, l'assegnazione di numeri agli oggetti, dopo aver constatato, sulla base del confronto tra due classi - quella degli oggetti (o delle loro qualità), e quella dei numeri appunto che un'assegnazione del genere è possibile; nel processo educativo, la misurazione implica l'assegnazione dei numeri agli individui, o ai gruppi, nei confronti dei quali sono state misurate le trasformazioni corrispondenti agli scopi educativi.

[...]

6. I CRITERI DELLA VALUTAZIONE

I criteri, in base ai quali hanno luogo interpretazione e giudizio valutativo dei risultati dell'accertamento compiuto, possono essere diversi; si cercherà qui di seguito di indicarne le caratteristiche tipiche distintive.

a) Un primo modo di giudicare l'apprendimento o la condotta, che furono misurati, può esser identificato con l'utilizzazione *di criteri per così dire assoluti*: nel qual caso, per esempio, si può dare il voto massimo solo se mancano del tutto gli errori, o se ve ne sono soltanto in una certa misura; o si dà un giudizio d'insufficienza solo per la constatazione d'un certo tipo di risposta, o per il fatto che a una certa domanda non è stata data risposta; o ancora, si toglie un punto (o una porzione di punto) per ogni errore accertato. Il criterio è in questo caso indicato come assoluto, proprio per il fatto che il giudizio è dato in modo del tutto indipendente da confronti o riferimenti coi risultati delle misurazioni compiute in relazione ad altri soggetti.

b) Un secondo modo di giudicare il rendimento scolastico e la condotta è da individuare nel riferimento, che precede il giudizio, al rendimento o alla condotta del gruppo, ristretto o largo, di cui il soggetto fa parte. Il riferimento, di cui si dice, può esser compiuto in modo approssimativo (per esempio, in una serie di esami orali universitari, il voto che vien dato ai candidati che in una giornata seguono il primo può esser il frutto di confronti con quelli che l'hanno preceduto), o in modo preciso (per esempio, in una prova oggettiva, si calcola il rendimento medio del gruppo, e si

attribuiscono, con particolari criteri, che qui non è il caso di anticipare, voti superiori a quello medio ai rendimenti superiori alla media, e voti inferiori a quello medio ai rendimenti inferiori al punteggio - grezzo - medio). Anche del gruppo, cui si fa riferimento, si può dire in modo diverso: il riferimento può esser compiuto nei confronti del gruppo ristretto (la classe, o i candidati all'esame d'una giornata), di cui il candidato fa parte, o nei confronti dell'intera popolazione, o d'un suo campione rappresentativo (come accadrebbe quando si volessero costruire prove oggettive standardizzate), di cui pure il soggetto fa parte. Com'è ovvio, quanto più preciso sarà stato il riferimento, e quanto più rappresentativo sarà stato il gruppo, tanto più attendibili saranno i giudizi sui risultati della misurazione.

c) Un terzo modo di giudicare, spesso preso in considerazione, s'identifica col riferimento alle possibilità, o alle capacità del soggetto, per cui si può per esempio affermare che uno scolaro non ha reso quanto avrebbe potuto, o che avrebbe potuto fare di più; che un altro scolaro ha dato quanto poteva, e nulla di più; e che un terzo, infine, ha reso più di quanto teoricamente ci si poteva attendere ragionevolmente da lui. Invece che alle possibilità (intellettuali) del soggetto si può far riferimento alla situazione in cui egli si trovava all'inizio del processo educativo, per cui si può per esempio affermare che uno scolaro ha compiuto progressi relativamente maggiori o minori d'un altro, nonostante il fatto che i loro rendimenti siano differenti, da un punto di vista per così dire quantitativo.

Tabella 1SCHEMA RIASSUNTIVO DELLE FASI DEL PROCESSO
DEL CONTROLLO SCOLASTICO

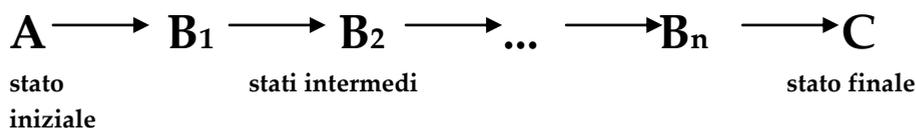
- I. DETERMINAZIONE DELL'OGGETTO DELLA MISURAZIONE (opera di valutazione).
- II. PROCESSO DELLA MISURAZIONE che prevede:
- A) L'utilizzazione di uno *strumento* con la funzione di:
 - 1. *stimolare*⁸⁷ (nel caso in cui le situazioni fossero artificiali);
 - 2. *registrare*⁸⁵ comportamenti degli scolari.
 - B) L'intervento dell'esaminatore col compito di *leggere* (analiticamente e sinteticamente) i comportamenti registrati.
- III. VALUTAZIONE DEI RISULTATI REGISTRATI E LETTI, la quale:
- A) Può implicare la presa in considerazione:
 - 1. del rendimento del gruppo (ristretto o rappresentativo) di cui il soggetto fa parte
e/o
 - 2. delle capacità del soggetto o del suo rendimento iniziale.
 - B) Implica l'utilizzazione d'un sistema di simboli (o di giudizi) entro una gamma prestabilita.

⁸⁷ Le funzioni di stimolazione e di registrazione possono esser compiute direttamente dall'esaminatore.

10. Funzioni della valutazione (Vertecchi, B., Manuale della valutazione. Analisi degli apprendimenti)⁸⁸

La critica del concetto tradizionale di profitto costituisce [...] una premessa necessaria per impostare correttamente un'analisi delle funzioni della valutazione. Possiamo considerare il processo di formazione come l'attività che partendo da una condizione iniziale A consente il raggiungimento di una condizione terminale C, passando attraverso una serie di condizioni intermedie B₁, B₂,... B_n. In ciascun momento del processo il vantaggio è rappresentato dalla differenza (che si presume positiva) fra la condizione raggiunta e quella iniziale (perciò da B₁ - A fino a C - A). In una logica valutativa di tipo tradizionale ci si accontenta di rilevare, allievo per allievo, la differenza: se almeno parte degli allievi raggiunge la condizione C, tale risultato autorizza non solo l'espressione di un giudizio positivo su di essi, ma anche di sostenere che l'insegnamento è stato validamente impartito. Il limite di questo modo di argomentare è l'astrattezza. Si presume infatti che la condizione A sia generalizzata fra gli allievi, e ben distinta dalla condizione C: ciò era probabilmente vero quando la popolazione scolastica era molto più ristretta, ma non è certamente vero oggi, perché l'estrema disomogeneità delle caratteristiche degli allievi fa sì che si debba considerare la condizione A come propria di ciascuno di essi. In altre parole, la condizione A non può più essere considerata neanche una semplificazione descrittiva di uno stato iniziale della popolazione scolastica. È possibile, per esempio, che per una parte limitata di allievi la condizione A si avvicini o si sovrapponga alla condizione C: in questo caso il profitto sarebbe solo apparente, e quindi non sarebbe giustificato un giudizio positivo su di esso.

Fig. 18. *Articolazione del processo di istruzione*



È chiaro che il giudizio resta positivo da un punto di vista formale, ma nulla può ricavarsi da tale giudizio circa la qualità del lavoro didattico che è stato svolto. Del resto, è sempre possibile pensare che nella scuola qualche risultato positivo sia comunque raggiunto, anche se gli interventi sono del tutto inadeguati. Ma tali risultati sono da attribuire all'effetto di circostanze casuali: partendo da essi non è possibile affermare né che l'azione didattica è stata efficace, né che non lo è stata.

⁸⁸ Tratto da: Vertecchi, B., *Manuale della valutazione. Analisi degli apprendimenti*, Roma, Editori Riuniti, 1984, pp. 82-89 e 126-127.

L'interpretazione del profitto cambia se, invece di centrare l'attenzione sul risultato ottenuto dal singolo allievo, cerchiamo di stabilire quale sia stato il risultato collettivo. Diremo allora che la condizione C deve essere accertata sull'intera popolazione di allievi. Per poter affermare che l'attività formativa ha conseguito un risultato positivo, occorrerà che non si riscontrino difformità di rilievo nelle prestazioni individuali rispetto a C; in altri termini la condizione C (che indica quale traguardo si persegue nella formazione) sarà descrittiva del risultato effettivamente conseguito dalla grande maggioranza degli allievi. Ne deriva che da un punto di vista didattico hanno maggior valore dimostrativo i risultati negativi (che in ogni caso stanno ad indicare inadeguatezze nelle scelte didattiche) di quelli positivi (che in qualche misura possono sempre essere attribuiti all'incidenza di fattori non scolastici).

Abbiamo affermato che non basta considerare il vantaggio che l'attività formativa determina per gli allievi, ma occorre tener conto anche di quello che ricade sulla capacità professionale degli insegnanti. E ancora troppo radicata la convinzione che la valutazione non coinvolga i docenti se non come titolari della capacità di giudicare gli allievi. Ciò equivale a sostenere che l'attività didattica è adeguata per definizione, e perciò non soggetta a verifica. Anche questa posizione poteva avere una sua legittimità quando le dimensioni del sistema scolastico erano molto più contenute delle attuali, perché l'insegnamento si presentava come un dato invariabile, nei confronti del quale gli allievi erano tenuti soggettivamente ad adattarsi (e ciò avveniva nei due modi «classici» della lezione privata per chi poteva permetterselo e dell'abbandono della scuola per chi era privo di mezzi). In una scuola che si rivolge a tutta la popolazione, l'insegnamento deve essere invece considerato un dato variabile, le cui caratteristiche si modificano a seconda delle circostanze. L'onere dell'adattamento non può più essere fatto ricadere sugli allievi, ma deve essere assunto dalla scuola. Pertanto l'adeguatezza dell'attività didattica costituisce un oggetto di giudizio, certamente non per fini di controllo gerarchico o burocratico, ma perché è necessario stabilire se attraverso l'esperienza è stato conseguito un *profitto*, e cioè se fra la condizione A (iniziale) e quella C (terminale) si riscontra un accrescimento della capacità professionale di affrontare situazioni nuove e complesse. C'è infine un aspetto nell'interpretazione del profitto che coinvolge le scuole nel loro insieme, in quanto strutture che erogano un servizio di rilevante importanza sul piano sociale. Non ha senso infatti pronunciare giudizi sull'attività delle scuole, e denunciare la qualità scadente dei risultati che si ottengono, se contemporaneamente non si insiste sulla necessità di una verifica obiettiva del prodotto della formazione, che tenga conto, oltre che del livello degli apprendimenti conseguiti dagli allievi, dell'acquisizione da parte delle scuole di nuove capacità di intervento. Se si trascura questa dimensione del problema valutativo, si rinuncia implicitamente ad affermare una idea di progresso nella didattica. Finora il servizio offerto dalle scuole è consistito nella riproposizione di attività conformate su un modello consueto, senza considerare che attraverso l'accumulazione delle competenze è possibile disporre di nuove risorse per l'educazione. È solo un modo di aggirare il problema porre l'enfasi sul

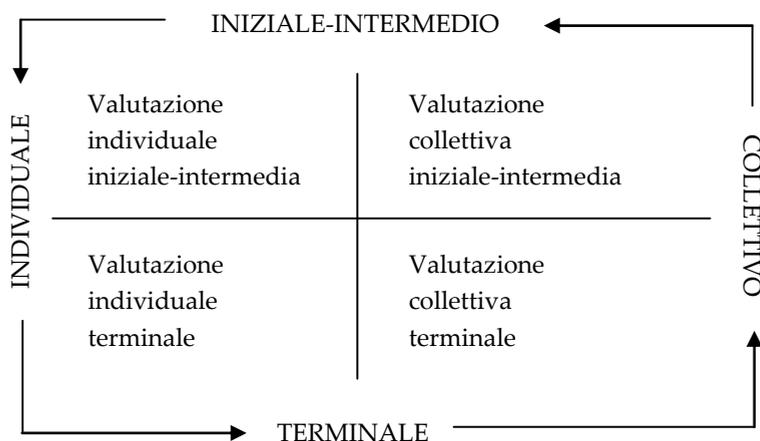
valore dell'esperienza individuale, che si affinerrebbe attraverso il tempo: in questo modo si può giustificare la competenza di un precettore privato, ma non quella di un operatore che si colloca all'interno di una struttura pubblica. L'esperienza assume un effettivo valore conoscitivo solo se la si sottopone a verifica: bisogna quindi valutare l'esperienza per estenderne e generalizzarne le indicazioni, ma anche per individuarne i limiti, e precisare nuove linee di intervento. C'è insomma un profitto delle scuole che consiste nell'incremento della capacità di praticare la formazione.

Per una valutazione che corrisponda ad una interpretazione comprensiva del concetto di profitto occorre tener conto dello sviluppo attraverso il *tempo* del processo formativo e dell'oggetto sul quale si esprime il giudizio. Distingueremo il tempo, e cioè la fase dell'attività di formazione in cui interviene la verifica, in *iniziale, intermedio, terminale*; per ciò che riguarda l'oggetto del giudizio, distingueremo se si tratta di accertare gli apprendimenti dei singoli allievi (valutazione *individuale*) o aspetti d'insieme della formazione che coinvolgono il funzionamento della scuola e l'organizzazione della didattica (valutazione *collettiva*). Incrociando tempi e oggetti della valutazione, è possibile distinguerne le quattro *funzioni* fondamentali (vedi la fig. 19).

La prima funzione della valutazione è collettiva e iniziale-intermedia, e coincide sostanzialmente con l'attività di giudizio che presiede all'assunzione di scelte impegnative circa l'attività da svolgere. Questa funzione coincide largamente con quanto comunemente si indica come *programmazione didattica*. È possibile infatti definire quest'ultima come un giudizio preventivo circa l'adeguatezza delle soluzioni da praticare rispetto agli intenti che ci si propone di raggiungere. Le scuole per programmare hanno bisogno di acquisire *informazione* (debbono cioè disporre di dati analitici relativi alle caratteristiche della popolazione scolastica e alla disponibilità di risorse fisiche e professionali); tale informazione definisce lo stato di partenza dell'attività formativa. È necessario poi precisare quali traguardi si intendano raggiungere, e cioè definire lo stato finale. Quando si disponga di questi due riferimenti resta da precisare l'itinerario didattico, tenendo conto delle risorse di cui si dispone, o che è possibile acquisire (il tempo disponibile, i materiali didattici, le prove di valutazione, le competenze professionali ecc.). Tutto ciò richiede un forte impegno valutativo: i dati iniziali costituiscono il supporto per esprimere un giudizio realistico circa la validità di questa o quella scelta didattica. L'espressione di un giudizio siffatto è tanto più impegnativa quando la scelta didattica si presenta differenziata, per poter corrispondere alle diverse esigenze degli allievi. L'accuratezza della programmazione è una condizione necessaria se si vuole che i risultati (ossia il profitto) siano soddisfacenti non solo per alcuni, ma per tutti, o quasi tutti, gli allievi. C'è anche un'altra ragione per la quale è possibile qualificare la programmazione didattica come un'attività valutativa. Se non si vuole che programmare sia solo una occasione rituale per esporre intenzioni poco impegnative, occorre precisare in che modo sarà verificato il lavoro svolto, e quindi definire il *criterio* sulla base del quale

verranno apprezzati i risultati. Un modo per dare concretezza alla programmazione consiste nel mettere a punto le prove che verranno utilizzate per la verifica finale.

Fig. 19. Scansione temporale e oggetti dell'attività valutativa



La seconda funzione della valutazione è individuale e iniziale-intermedia. Ha lo scopo di fornire una informazione continua e dettagliata circa il modo in cui i singoli allievi accedono ad una procedura di apprendimento e quindi procedono attraverso di essa. La disponibilità di tale informazione è indispensabile se si vogliono assumere decisioni didattiche tempestive, per corrispondere alle necessità di ciascun allievo differenziando la proposta formativa (per questa ragione questa funzione della valutazione si dice *formativa*). Per rendersi conto della rilevanza di questa funzione della valutazione, basta riflettere al fatto che troppo spesso gli insegnanti dispongono di una informazione relativa all'apprendimento quando una determinata attività è già stata conclusa. Se la verifica si effettua *dopo* lo svolgimento di una attività didattica, essa non può avere altro scopo che di registrare se qualcosa è avvenuto o non è avvenuto, senza che si determinino le condizioni per utilizzare i dati acquisiti per migliorare, attraverso rettifiche opportune delle modalità di intervento, la qualità dei risultati scolastici. La valutazione formativa interviene invece *durante* i processi di apprendimento, ed ha lo scopo di accertare in modo analitico quali abilità ciascun allievo stia acquisendo, rispetto a quali incontri difficoltà: gli insegnanti possono quindi attivare tempestivamente quegli interventi compensativi che appaiono più opportuni.

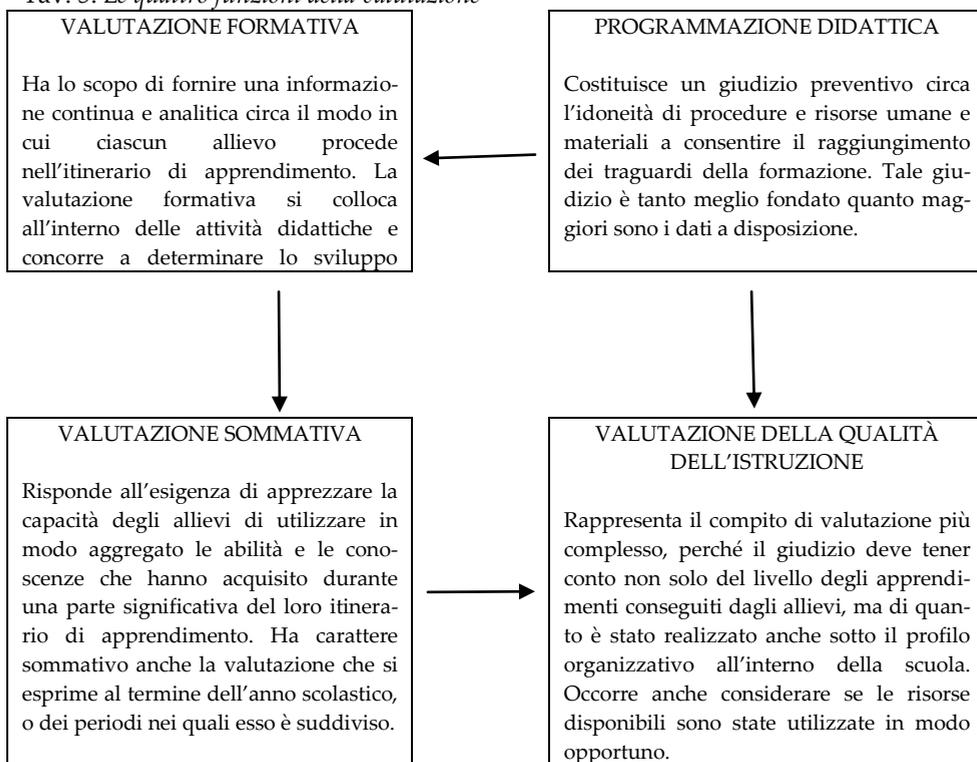
La terza funzione della valutazione tende a consentire l'espressione di un giudizio complessivo sull'apprendimento conseguito dagli allievi. Si tratta quindi di una valutazione terminale e individuale, volta a trarre le somme del lavoro svolto (si definisce per questo valutazione *sommativa*). Nella prassi scolastica alcuni momenti di valutazione sommativa sono previsti istituzionalmente: è il caso delle valutazioni

trimestrali o quadrimestrali, o delle prove di esame. Sarebbe tuttavia limitativo ritenere che una serie di scadenze scandite dal calendario esaurisca la necessità di disporre di valutazioni sommative. Occorre invece prevedere che tali valutazioni intervengano secondo un criterio di opportunità didattica, stabilito in sede di programmazione. Da questo punto di vista è opportuna una valutazione sommativa ogni volta che il lavoro svolto consenta di individuare un complesso organico di abilità e di conoscenze, che costituisca una significativa fase di avvicinamento rispetto al traguardo terminale. A differenza della valutazione formativa, che si colloca con intenti analitici all'interno del processo di apprendimento, la valutazione sommativa verifica la capacità di organizzare funzionalmente le conoscenze acquisite. Infine, la quarta funzione della valutazione è collettiva e terminale. Ha lo scopo di verificare in quale misura siano stati raggiunti i traguardi definiti in sede di programmazione, quale è stata l'efficacia delle scelte effettuate, se i mezzi didattici impiegati erano idonei allo scopo, quale incremento di competenze e di capacità professionali sia stato conseguito. In breve, questa quarta funzione della valutazione costituisce l'occasione in cui si effettua un bilancio consuntivo dell'attività svolta: è come dire che la scuola valuta se stessa. Per questa ragione, si può parlare di valutazione della *qualità dell'istruzione*. Per cogliere la rilevanza della valutazione della qualità dell'istruzione occorre ricordare che la formazione scolastica è una attività di tipo sequenziale: ogni corso o anno di studio non costituisce qualcosa di separato, ma si collega alle attività che lo hanno preceduto o che lo seguiranno. Il carattere sequenziale della formazione costituisce un criterio per interpretare il curriculum degli allievi, ma anche l'organizzazione del lavoro nelle scuole. Dal punto di vista degli allievi, la formazione costituisce un *continuo*, nel quale quanto segue richiama le abilità che sono state acquisite in precedenza. Per le scuole, è necessario che le scelte che riguardano le attività da effettuare abbiano come termine di riferimento le esperienze già compiute: la valutazione della qualità dell'istruzione è dunque la premessa per le successive programmazioni. Se si trascura questa funzione della valutazione, si riduce l'attività formativa alla casualità delle scelte effettuate volta per volta, o ad una stanca ripetizione di schemi didattici logorati. Il collegamento tra valutazione della qualità dell'istruzione e programmazione didattica consente di interpretare la valutazione come un flusso circolare di giudizi che investono ogni aspetto del processo formativo, sia nelle sue fasi di progettazione e organizzazione, sia in quelle di realizzazione (vedi tav. 3).

Ciascuna funzione della valutazione presenta modalità specifiche di espressione del giudizio. La programmazione didattica è il frutto di un confronto al quale debbono partecipare tutti i docenti che operano in una determinata scuola; scelte più determinate vanno quindi effettuate all'interno dei consigli di interclasse nella scuola elementare e di classe in quella secondaria. Spetta a ciascun insegnante specificare ulteriormente criteri e tecniche di intervento, perché la didattica non abbia a risentire di condizionamenti determinati dalle circostanze, e che si risolvono in improvvisazioni incoerenti con gli intenti che si perseguono. La valutazione formativa

può essere gestita parte dal docente (qualche volta dai docenti) che conduce una certa attività didattica, parte dal singolo allievo (in questo secondo caso si realizza una forma di *autovalutazione*). La valutazione sommativa compete ai singoli docenti per quanto riguarda le rispettive discipline, a più docenti se si tratta di attività interdisciplinari, al consiglio di classe quando occorre effettuare un consuntivo del livello di competenza raggiunto dagli allievi. Infine la valutazione della qualità dell'istruzione costituisce uno degli impegni più significativi non solo per i colleghi dei docenti, ma anche per l'insieme degli organi collegiali a governo della scuola, per il confronto che in essi è possibile stabilire con i rappresentanti dei genitori e delle forze sociali.

Tav. 5. *Le quattro funzioni della valutazione*



[...]

È possibile [...], all'interno delle funzioni già considerate [...], distinguere ulteriormente:

1. se la *valutazione iniziale* ha lo scopo di effettuare una ricognizione delle conoscenze che l'allievo possiede all'inizio di un determinato itinerario di studi, per esempio all'inizio della scuola media, si parla di *valutazione d'ingresso*. Più specifico è il significato dell'espressione *valutazione dei prerequisiti*, in quanto si riferisce al possesso da parte dell'allievo di capacità strumentali e operative e di conoscenze necessarie per inserirsi positivamente in una determinata procedura di apprendimento (così, se l'oggetto dell'attività di insegnamento-apprendimento è costituito dallo studio dei criteri di uguaglianza dei triangoli, bisogna che l'allievo già possieda alcune conoscenze di geometria piana, utilizzi correttamente una serie di parole, sappia tracciare disegni di vari tipi di triangoli e riconoscerli ecc.). Infine, se l'attività valutativa iniziale tende a stimare quale potrà essere la difficoltà che ciascun allievo dovrà superare nel corso del processo di apprendimento al fine di predisporre percorsi individualizzati, la soluzione potrà essere costituita dal ricorso a prove analogiche;

2. abbiamo già visto che una interpretazione impegnativa della *valutazione intermedia* è costituita dalla *valutazione formativa*. Essa è finalizzata ad offrire la possibilità di una immediata compensazione di eventuali difficoltà che un allievo abbia incontrato nell'apprendimento. Sia la valutazione iniziale, sia quella intermedia hanno lo scopo di consentire una diagnosi delle conoscenze che un allievo possiede: si parla quindi di *valutazione diagnostica*;

3. la *valutazione finale* interviene, come è evidente, a compimento di una attività più o meno lunga di formazione. L'espressione *valutazione sommativa* è appropriata per indicare la valutazione finale quando si tratta di apprezzare i risultati che si sono conseguiti attraverso itinerari didattici di tipo individualizzato, nei quali si è posta una specifica attenzione nel sostenere le esigenze dei singoli allievi;

4. quando la verifica tende ad anticipare il risultato che un allievo potrà conseguire seguendo un determinato itinerario di studi, si usa l'espressione *valutazione predittiva*. Occasioni di valutazione predittiva (ma anche di discriminazione sociale) erano i vecchi *esami di ammissione* alla scuola media. Anche se attenuato, conservano un certo carattere predittivo gli *esami di maturità*, che sono insieme terminali di un livello e di ammissione al livello successivo (questo carattere era più evidente quando la maturità costituiva il titolo finale dei licei, mentre per gli altri tipi di scuola secondaria superiore il diploma finale era variamente denominato). In qualche misura sono predittivi anche i *giudizi di orientamento*: si tratta infatti di anticipare, sulla base degli elementi di valutazione di cui si dispone, quale sarà il successo negli studi che un allievo potrà conseguire scegliendo questo o quell'altro indirizzo per proseguire nel suo percorso scolastico, o per avviarsi ad attività di lavoro. Trattandosi di una valu-

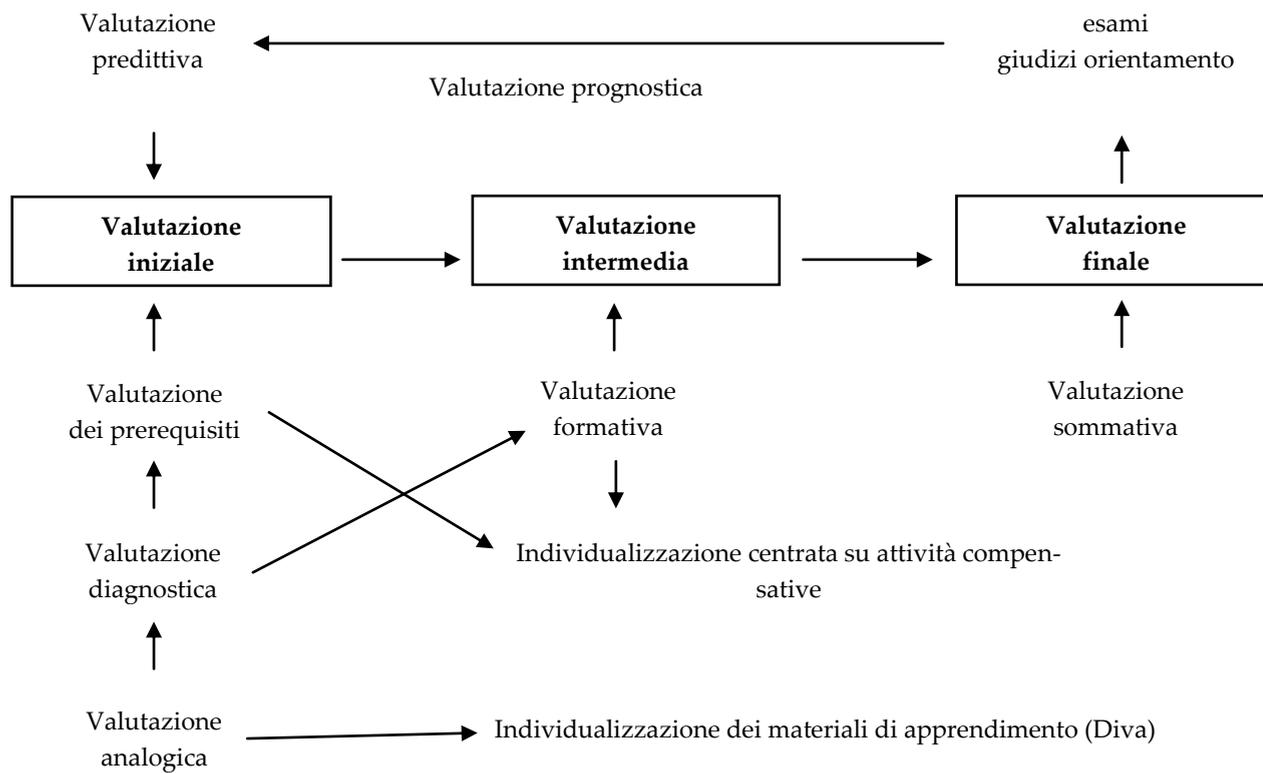


Fig. 25. *Aspetti della valutazione nel processo di insegnamento-apprendimento*

tazione che anticipa un certo corso di eventi (talvolta concorre anche a determinarlo) si può parlare di *valutazione prognostica*.

Non è infrequente, nel dibattito corrente sulla valutazione, assistere a contrapposizioni artificiali tra una funzione e l'altra della valutazione. Per esempio, la valutazione formativa viene opposta a quella sommativa, con l'argomento che la prima sarebbe «buona» perché non discriminante, e l'altra evidentemente «cattiva» perché dà luogo ad apprezzamenti differenziati sui livelli di abilità raggiunti dagli allievi. Tutto ciò non ha senso. I vari aspetti della valutazione sono ugualmente necessari, perché rispondono ad esigenze diverse della didattica (si veda la fig. 25). Semmai il problema è quello di qualificare la pratica della valutazione, perché si possa disporre di informazioni correttamente utilizzabili ai fini della conduzione del lavoro didattico.

Altrettanto ingiustificata è la contrapposizione tra *etero-valutazione* (cioè la valutazione espressa dagli insegnanti) ed *auto-valutazione* (ossia la valutazione che gli allievi sono *in grado* di esprimere direttamente sulle abilità che hanno acquisito). È errato ritenere che la prima sia espressione di una impostazione autoritaria del rapporto educativo, mentre la seconda starebbe a testimoniare dell'apertura e della disponibilità che anima gli insegnanti: entro un quadro attentamente programmato, si può fare in modo che gli allievi apprezzino direttamente la qualità delle prestazioni che sono in grado di fornire (per esempio, in situazioni di valutazione formativa), ma ciò non elimina la necessità di una valutazione che sia il risultato di apprezzamenti effettuati dagli insegnanti.

11. La valutazione contestualizzata: l'alternativa ai test standardizzati (Gardner, H., *L'educazione delle intelligenze multiple*)⁸⁹

Ecco una scena familiare quasi ovunque, oggi, negli Stati Uniti: diverse centinaia di studenti sfilano in una grande aula d'esame. Siedono nervosi, in attesa che vengano distribuiti i plichi sigillati. All'ora stabilita ricevono i fascicoli con qualche parola di spiegazione, e il test formale ha inizio. Nell'aula non si sente volare una mosca, mentre, ciascuno seduto al suo banco, gli studenti si danno da fare con le loro matite ad annerire i cerchietti di cui sono disseminati i fogli delle risposte. Qualche ora dopo, il test si conclude, e i fascicoli vengono ritirati; di lì ad alcune settimane, a casa di ogni studente, e a tutti i *colleges* ai quali essi hanno fatto domanda d'ammissione, viene inviata una scheda recante una serie di punteggi. Il risultato di un test eseguito in mezza giornata diventa così un potente fattore determinante nella decisione del futuro di ogni studente.

Ecco ora una scena ugualmente familiare, in moltissime società preindustriali di tutti i tempi: un ragazzino di dieci-undici anni entra nella casa di un uomo, maestro di una professione, di un'arte o di un mestiere. Al principio, il fanciullo deve eseguire compiti umili, aiutando il maestro a prepararsi per il suo lavoro o a pulire la bottega alla fine della giornata. Durante questa fase iniziale, il fanciullo ha l'opportunità di osservare il maestro al lavoro, mentre quello può studiare il giovane per scoprirne doti particolari o gravi lacune. Un mese dopo l'altro, l'apprendista penetra lentamente nella prassi del lavoro del maestro. Dopo aver preso parte agli aspetti più periferici della sua arte o della sua professione, egli acquisisce infine familiarità con l'intera gamma delle prestazioni che essa comporta. Orientato dai dettami della tradizione, ma anche tenendo conto delle particolari capacità e della motivazione del giovane, il maestro guida il suo cammino attraverso le varie tappe che lo porteranno dallo stadio iniziale di principiante, a quello di lavorante finito. Infine, dopo diversi anni di addestramento così guidato, il giovane sarà pronto a intraprendere in proprio il mestiere - l'arte - che ha appreso.

Sebbene entrambe le scene che ho descritto siano idealizzate, dovrebbero tuttavia essere facilmente riconoscibili da chiunque sia interessato alla valutazione e alla preparazione dei giovani. In effetti, le due situazioni rappresentano due estremi: il primo modello, quello del "test formale", è una forma di valutazione oggettiva e decontestualizzata, che può essere adottata e messa in pratica su vasta scala con una certa garanzia di ottenere risultati riproducibili. Il secondo modello, quello dell'apprendistato, viene realizzato quasi interamente all'interno di un contesto naturale, comprendente i particolari aspetti di un mestiere. La valutazione si basa su

⁸⁹ Tratto da H.Gardner: *L'educazione delle intelligenze multiple*, Milano, Anabasi, 1993 (tr. it. 1995), pp. 131-133 e 147-153.

una precedente analisi delle capacità necessarie in un particolare mestiere, ma può anche essere influenzata da fattori soggettivi, ivi comprese le opinioni personali del maestro sul suo apprendista, i suoi rapporti con altri maestri, o la sua esigenza di altri tipi di servizi.

Dovrebbe essere chiaro che queste due forme di valutazione intendono soddisfare esigenze diverse. Gli apprendistati avevano senso quando, nell'ambiente urbano, la pratica diretta dei vari mestieri era la principale forma di impiego dei giovani. I test formali sono invece un mezzo, tipico dei nostri tempi, per confrontare le prestazioni di migliaia di studenti che hanno ricevuto un'istruzione scolastica. Tuttavia, queste forme di valutazione non si limitano ai due contesti originari descritti in precedenza. Nonostante la società cinese fosse di natura prevalentemente agraria, per più di duemila anni essa si servì dei test formali per selezionare i funzionari del governo imperiale. E, per lo stesso motivo, in numerose forme d'arte, pratiche atletiche e aree di ricerca scientifica (Polanyi, 1958)⁹⁰, gli apprendistati e le forme di valutazione contestualizzata ad essi legate continuano a essere usati anche nella nostra società altamente industrializzata. Perciò, la scelta dei "test formali" in contrapposizione all'"apprendistato" non è imposta esclusivamente dal momento storico o dai principali mezzi di produzione di una società. In particolare, nella nostra, sarebbe possibile utilizzare il metodo dell'apprendistato in misura maggiore di quanto non si faccia. La maggior parte degli osservatori contemporanei, me incluso, non rimpiange la fine del sistema dell'apprendistato obbligatorio, con i suoi frequenti eccessi e il suo sfacciato sessismo; sotto diversi punti di vista, infatti, i test formali contemporanei costituiscono una forma di valutazione più equa e più facilmente giustificabile. Ciò nonostante, alcuni aspetti del modello dell'apprendistato sono coerenti con le nostre attuali conoscenze sull'apprendimento e sul modo in cui valutare al meglio le prestazioni individuali.

La nostra società si è dimostrata eccessivamente incline ad abbracciare la filosofia dei test formali; io sostengo che alcuni aspetti del modello di apprendimento e di valutazione tipico dell'apprendistato - apprendimento che io definisco "contestualizzato" - potrebbero essere proficuamente reintrodotti nel nostro sistema educativo (vedi Collins, Brown e Newman, 1989)⁹¹. Dopo un'esposizione delle origini storiche dei test standardizzati e della concezione monodimensionale dell'attività mentale spesso implicita in tali metodi di valutazione, suggerirò l'esigenza di una concezione della mente e dell'apprendimento di gran lunga più ampia di quella che ha informato le posizioni precedenti.

Il mio obiettivo è quello di immaginare forme di educazione e modalità di valu-

⁹⁰ Polanyi, M. (1958), *Personal Knowledge*, Chicago, University of Chicago Press; tr. it., *La conoscenza personale*, Milano, Rusconi libri, 1990 [n.d.c.].

⁹¹ Collins, A., Brown, J. S., Newman, S. E. (1989), *Cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics*, in Resnick, L. (a cura di), *Cognition and instruction: Issues and agendas*, Hillsdale, N. J., Lawrence Erlbaum [n.d.c.].

tazione che abbiano solide radici nell'attuale conoscenza scientifica e che possano contribuire al raggiungimento di obiettivi pedagogici illuminati. Nella seconda metà del capitolo descriverò brevemente la natura di una società che adottasse questo sistema di valutazione.

[...]

Aspetti generali di un nuovo approccio alla valutazione

Se oggi dovessimo tornare indietro per tracciare un nuovo approccio alla valutazione, potremmo cercare di incorporarvi i seguenti aspetti principali.

PORRE L'ACCENTO SULLA VALUTAZIONE, NON SUL TEST

La propensione per i test, in America, si è spinta troppo oltre. Sebbene, limitatamente a certi scopi, alcuni test siano effettivamente utili, l'industria legata a questi strumenti ha preso il volo in modo irragionevole. Molti di coloro che cercano di capire la base teorica o concettuale della loro validità sono rimasti delusi. Sembra che molti test siano stati ideati per creare - non per soddisfare - un'esigenza. Personalmente nutro sentimenti ambivalenti nei confronti di tali strumenti, mentre non posso dire altrettanto riguardo alla valutazione. Secondo me, è giusto che le persone colte, e quelle che sono loro affidate, siano impegnate in una regolare attività di riflessione sui propri obiettivi, sui vari mezzi per raggiungerli, sul proprio successo - o insuccesso - relativamente ad essi, e sulle implicazioni della valutazione ai fini della riformulazione tanto degli obiettivi, quanto delle procedure.

La valutazione può essere definita come l'ottenimento di informazioni sulle capacità e le potenzialità degli individui, con il duplice obiettivo di fornire a loro un'utile retroazione, e alla comunità circostante dati preziosi. Ciò che distingue la valutazione, così intesa, dai test, è in primo luogo il fatto che la prima favorisce tecniche che ricavano informazioni sull'individuo nel corso delle sue consuete prestazioni; in secondo luogo, essa diffida degli strumenti formali, somministrati in un ambiente neutrale decontestualizzato.

Secondo me, psicologi ed educatori incaricati della valutazione dovrebbero rendere più facile questa stima (vedi Cross & Angelo, 1988)⁹². Dovremmo ideare metodi e misure che ci aiutino a operare valutazioni regolari, sistematiche e utili. In qualche caso, questo lavoro potrebbe anche tradursi nella produzione di "test formali". In moltissimi altri casi però, mi aspetterei un esito diverso.

⁹² Cross, K. P., Angelo, T. (1988), *Classroom assessment techniques: A handbook for faculty*, Ann Harbor, National Center for Research to Improve Postsecondary Teaching and Learning (NCRITL) [n.d.c.].

LA VALUTAZIONE COME PROCESSO SEMPLICE, NATURALE E PROGRAMMABILE

Invece di essere imposta sporadicamente "dall'esterno" durante l'anno scolastico, la valutazione dovrebbe divenire essa stessa parte del naturale ambiente di apprendimento, ed essere effettuata il più possibile "al volo", come parte del naturale impegno di un individuo in una situazione di apprendimento. Probabilmente, all'inizio la valutazione dovrebbe essere introdotta in modo esplicito, ma dopo un po', gran parte di essa avrebbe luogo spontaneamente sia da parte degli allievi che dell'insegnante, con una scarsa necessità di riconoscimento esplicito o di definizione dall'una o dall'altra parte.

A questo proposito, è interessante fare qualche considerazione sulla valutazione delle capacità cognitive dell'esperto. Da un lato, raramente è necessario che l'esperto sia valutato da altri, salvo nel caso di situazioni competitive: si assume infatti che egli sia in grado di condurre la propria attività con pochi controlli esterni. Tuttavia, è anche vero che l'esperto è costantemente impegnato in un processo di autovalutazione - una valutazione spontanea, che ha luogo quasi senza riflessione cosciente, nel corso del suo lavoro. Quando cominciai a scrivere articoli per riviste specializzate dipendevo moltissimo dalle critiche, precise e dettagliate, che mi venivano mosse da docenti ed editors, ora, gran parte della necessaria operazione di valutazione avviene nella mia mente a un livello preconscious, quando siedo alla scrivania scarabocchiando, mentre batto a macchina una prima versione del lavoro, o durante la correzione di una sua precedente stesura.

Quando la valutazione entra gradualmente a far parte del paesaggio educativo, non deve più essere distinta dal resto delle attività svolte in classe. Proprio come avviene in un buon apprendistato, insegnanti e studenti sono costantemente impegnati in un processo di valutazione. Non c'è più nemmeno bisogno di "insegnare per la valutazione", perché essa è ubiquitaria: l'esigenza di disporre di test formali, a questo punto, potrebbe atrofizzarsi del tutto.

VALIDITÀ ECOLOGICA

Un problema che si ripresenta con moltissimi test formali è quello della loro validità, e cioè della loro correlazione con qualche criterio (Messick, 1988)⁹³. Come abbiamo visto, i test per la creatività non sono più molto usati, proprio perché la loro validità non è mai stata adeguatamente dimostrata. Il valore predittivo dei test d'intelligenza e dei test per stimare le attitudini scolastiche è stato spesso messo in dubbio in considerazione della loro scarsa utilità nel prevedere le prestazioni individuali oltre il successivo anno scolastico.

⁹³ Messick, S. (1988), *Validity*, in Linn, R. (1988³), *Educational Measurement*, New York, Macmillian [n.d.c.].

Tornando al nostro esempio dell'apprendistato, non sarebbe logico mettere in dubbio la validità dei giudizi del maestro. Egli è così intimamente associato al suo apprendista che probabilmente è in grado di prevedere i suoi comportamenti con un elevato grado di accuratezza. Quando tale predizione non ha luogo in modo attendibile, invece, ecco sorgere dei problemi. Credo che le attuali forme di valutazione si siano spinte troppo lontano: quando gli individui vengono valutati in situazioni che ricordano più da vicino "le reali condizioni di lavoro", infatti, è possibile fare predizioni molto migliori su quelle che saranno le loro future prestazioni. È strano che moltissimi scolari americani passino ore su un unico esercizio (il test formale) quando si pensa che, una volta terminata la scuola, ben pochi di loro - forse nessuno - si imbattono più in uno strumento simile.

STRUMENTI "APERTI" ALL' INTELLIGENZA

Come abbiamo visto, moltissimi test risentono di un forte pregiudizio a favore di due tipi di intelligenza, quella linguistica e quella logico-matematica. Gli individui che hanno la fortuna di possedere questa combinazione particolare di facoltà probabilmente daranno buone prestazioni nella maggior parte dei test formali, pur non essendo particolarmente esperti nell'ambito conoscitivo effettivamente sotto esame. Per lo stesso motivo, individui che hanno difficoltà nell'intelligenza linguistica, o in quella logico-matematica, o anche in entrambe, possono dare prestazioni scadenti nei test formali anche quando l'oggetto della valutazione è un'altra attività - e questo semplicemente perché non riescono ad acquisire la necessaria padronanza del particolare formato di moltissimi strumenti standardizzati.

La soluzione del problema - più facile a dirsi che non a mettersi in pratica - è quella di ideare strumenti "aperti" che sondino l'intelligenza in modo diretto osservandola all'opera, invece di procedere operando una digressione, e passando attraverso le facoltà logiche e linguistiche. L'intelligenza spaziale, ad esempio, può essere valutata chiedendo all'individuo di orientarsi in un territorio che non gli è familiare; l'intelligenza corporea osservando il modo e la facilità con cui egli impara e ricorda una nuova danza o un esercizio fisico; l'intelligenza interpersonale può essere valutata osservando l'individuo in una contrattazione, oppure il modo in cui egli escogita una via d'uscita in una complessa riunione di lavoro. Questi esempi, così alla buona, ci mostrano il modo di ideare misure più "aperte all'intelligenza", sebbene non necessariamente applicabili nel laboratorio di psicologia o nell'aula dove solitamente si somministrano i test formali.

IMPIEGO DI MISURE MULTIPLE

Nell'educazione, poche cose sono più deleterie del trarre ampie implicazioni educative dal punteggio ottenuto in un unico test, ad esempio nella Wechsler Intelligence Scale for Children. Perfino i test di intelligenza contengono sottotest, e bisognerebbe almeno tener conto della "dispersione" e delle strategie per accostarsi a particolari items (Kaplan, 1983)⁹⁴.

Ancora più auspicabile, è che ci si rivolga a una gamma di misure ideate specificamente per attingere a diversi aspetti della capacità in questione. Si considerino, ad esempio, gli standard di ammissione a un programma per bambini dotati. Nel nostro paese, il 75% almeno dei programmi ammette semplicemente sulla base del QI: 129 e sei fuori, 131 e ce l'hai fatta. È proprio un sistema poco felice! Non ho alcuna obiezione sull'uso del QI, ma perché non prestare anche la dovuta attenzione alle cose già prodotte dal bambino, ai suoi obiettivi e al suo desiderio di un programma, alle sue prestazioni durante un periodo di apprendimento insieme ad altri bambini "dotati", e ad altre misure ugualmente non invadenti? Spesso ho la sensazione che si potrebbero compiere enormi progressi educativi, se solo, invece di apparire sullo sfondo di un unico grafico monodimensionale, il segretario del dipartimento dell'educazione si presentasse alle telecamere mostrando alle proprie spalle numerosi grafici, ciascuno riferito a un aspetto differente dell'apprendimento e della produttività intellettuale.

SENSIBILITÀ ALLE DIFFERENZE, AL LIVELLO DI SVILUPPO E ALLE FORME
DI EXPERTISE INDIVIDUALI

I programmi di valutazione che ignorano le grandi differenze fra gli individui e fra i loro livelli di sviluppo e di expertise sono sempre più anacronistici. In linea di principio, i test formali potrebbero essere adattati in modo da tener conto di queste variazioni documentate. Tuttavia, ciò richiederebbe la sospensione di alcuni degli assunti chiave dei test standardizzati, quali ad esempio l'uniformità degli individui per certi aspetti fondamentali, e la propensione verso strumenti di valutazione economicamente efficienti.

Le differenze individuali dovrebbero anche essere al centro dell'attenzione nel corso della preparazione degli insegnanti e degli specialisti della valutazione. Chi è investito della responsabilità di valutare i giovani dev'essere addestrato formalmente a compiere tali distinzioni; non ci si può aspettare che gli insegnanti arrivino a costruirsi da soli le loro tassonomie empiriche delle differenze individuali. Tale introduzione dovrebbe aver luogo nei corsi di pedagogia o durante il tirocinio. Una volta che queste distinzioni vengono mostrate agli insegnanti, esse divengono qual-

⁹⁴ Kaplan, E. (1983), *Process and achievement revisited*, in Wapner, S., Kaplan, B. (a cura di), *Towards a holistic development psychology*, Hillsdale, N. J., Lawrence Erlbaum [n.d.c.].

cosa di vivo e reale purché si dia loro l'opportunità di osservare bambini con profili intellettuali diversi e di interagire con essi.

Ecco che diviene allora possibile tener conto di tali differenze in modo implicito. I buoni insegnanti, sia che lavorino con una seconda elementare, sia che insegnino pianoforte ai piccolissimi o i fondamenti teorici del disegno sperimentale ai laureati, hanno sempre saputo che con diversi tipi di studenti sono necessari approcci diversi. Questa sensibilità alle differenze individuali può diventare parte della competenza dell'insegnante e può essere utilizzata nel corso dell'insegnamento, come pure in occasione della valutazione. Sarebbe anche possibile - forse ottimale - che gli insegnanti maturassero la propria intuizione delle differenze individuali servendosi di occasioni di valutazione che dessero spazio alla riflessione, messe a punto tenendo presente il particolare ambito al quale sono dirette.

IMPIEGO DI MATERIALI INTRINSECAMENTE CAPACI DI SUSCITARE INTERESSE E MOTIVAZIONE

Uno degli aspetti più contestabili dei test formali, sebbene raramente notato, è l'intrinseca opacità dei materiali impiegati. Quando mai qualcuno trova un test - o una sua parte - stimolante? Probabilmente solo la legislazione sulla "trasparenza" ha consentito che i soggetti sottoposti a test mettessero in discussione le soluzioni, ed è stato solo allora che il dibattito sui singoli items è diventato accessibile a chiunque fosse interessato ad esso.

Invece, non dev'essere così. Un buono strumento di valutazione può essere un momento di apprendimento. Fatto ancora più importante, è assolutamente auspicabile che la valutazione abbia luogo contestualmente, quando gli studenti stanno lavorando a problemi, progetti o prodotti che li impegnano davvero, che stimolano il loro interesse e li motivano a far bene. Può darsi che esercizi di questo tipo non siano facili da ideare come i test a scelta multipla, ma è di gran lunga più probabile che essi stimolino l'intero repertorio delle capacità di uno studente e diano informazioni potenzialmente utili per consigliarlo e per trovargli un'adeguata collocazione.

APPLICAZIONE DELLA VALUTAZIONE NELL'INTERESSE DELLO STUDENTE

Un aspetto ugualmente deprecabile dei test formali è il loro uso dei punteggi. Così classificati, gli individui constatano a quale percentile appartengono, e traggono conclusioni sul proprio merito scolastico, se non addirittura sul proprio merito complessivo. Secondo me, gli psicologi passano troppo tempo a classificare gli individui e troppo poco ad aiutarli: la valutazione dovrebbe avere invece lo scopo primario di aiutare gli studenti. Spetta allo specialista della valutazione fornire allo studente una retroazione che possa essergli utile nel presente: spetta dunque a lui identificare doti e debolezze, dargli suggerimenti su che cosa studiare o su che cosa lavorare, indicargli quali abitudini sono produttive e quali no, come pure che cosa

possa aspettarsi dalle future valutazioni, e così via. È importante soprattutto che parte di tale retroazione abbia la forma di suggerimenti concreti, che indichino allo studente i punti di forza relativi sui quali basarsi, indipendentemente dal livello in cui egli si colloca rispetto a un gruppo di riferimento.

Muniti dei risultati della ricerca sulla cognizione e lo sviluppo umano, e alla luce di questi auspici per un nuovo approccio alla valutazione, dovrebbe essere possibile cominciare a ideare programmi più adeguati di quelli attualmente disponibili. Senza avere l'ambizione, forse eccessiva, di creare una "nuova alternativa ai test formali", negli ultimi anni i miei colleghi ed io, del Project Zero di Harvard, ci siamo impegnati in una serie di progetti che mettono in primo piano i nuovi approcci alla valutazione. Nella prima parte di questo libro ho descritto i nostri numerosi sforzi tesi a valutare contestualmente le doti intellettuali degli studenti. Qui, invece, cercherò di inquadrare questi tentativi nel contesto più ampio della valutazione scolastica.

Parte 3. Gli strumenti del controllo scolastico: la tecnica del testing

a cura di Andrea Giacomantonio

L'obiettivo principale di questa parte è di presentare le procedure per utilizzare la tecnica del *testing* al fine di misurare e valutare il profitto degli studenti. In questa prospettiva iniziare con un lungo paragrafo riguardante l'osservazione rappresenta ovviamente una provocazione. I motivi che giustificano tale scelta sono diversi.

In primo luogo l'impostazione del lavoro di Marcel Postic e Jean Marie De Ketele esemplifica chiaramente l'*approccio critico operativo*, che, almeno a livello ipotetico, costituisce uno strumento metodologico capace di ridurre l'indeterminatezza linguistica che affligge la didattica, in particolare se essa assume la forma tecnologica di cui si è parlato precedentemente. L'operazionismo nasce all'inizio del XIX secolo, ad opera di Percy W. Bridgman, con un preciso programma di ricerca: definire in modo univoco il significato dei concetti teorici della fisica moderna, caratterizzata da un largo uso del formalismo matematico, facendo corrispondere ad ognuno di

essi determinati insiemi di operazioni fisiche o mentali⁹⁵. In altre parole i «nostri concetti devono essere descritti in termini di operazioni fisiche [o *carta e matita*] effettive, cioè, che possono effettivamente essere eseguite»⁹⁶. La nostra ipotesi è che tanto il controllo scolastico, quanto le attività di progettazione e programmazione, possono essere definite e descritte sul piano metodologico in termini operazionali, senza cedere a tentazioni positiviste o neopositiviste⁹⁷.

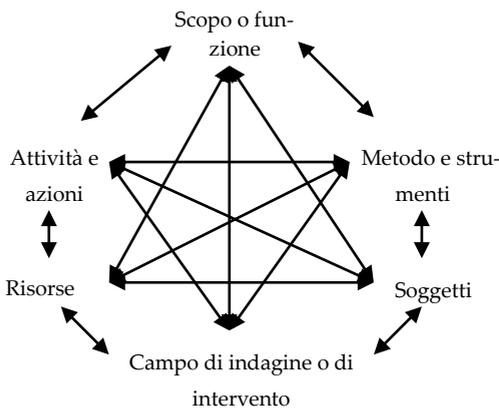


Fig. 3 – Modello definizione operativa

⁹⁵ Cfr. Bridgman, P. W. (1927), *The logic of modern physics*, New York, The Macmillan Company; trad. it. *La logica della fisica moderna*, Torino, Boringhieri, 1965.

⁹⁶ Bridgman, P. W. (1934), *Seconda reazione di un fisico alla teoria degli insiemi*, in Idem (1969), *La critica operativa della scienza*, a cura di Bruno Cermignani, Torino, Boringhieri, pp. 39-40.

⁹⁷ Ciò è possibile assumendo una prospettiva critica che considera il "linguaggio operativo" da utilizzare in ambito didattico alla stregua di un *gioco linguistico* praticato da una determinata comunità secondo regole rigorose autodeterminate. Per il concetto di *gioco linguistico* cfr. Wittgenstein, L. (1953), *Philosophische Untersuchungen*, e la lettura che ne fornisce Kenny, A. J. P. (1973), *Wittgenstein*, Allen Lane The Penguin Press, London; trad. it. *Wittgenstein*, Torino, Boringhieri, 1984.

Premesso, infatti, che un'operazione è un'«attività regolata»⁹⁸, ovvero, in termini pragmatisti, «un corso d'azione intrapreso tenendo conto di certe condizioni di fatto e anticipando mentalmente certi risultati intermedi e finali»⁹⁹, è ipotizzabile che essa sia definibile, come fanno Postic e De Ketele nel caso particolare dell'osservazione pedagogica¹⁰⁰, esplicitando i sei elementi riportati in figura. Definire il controllo scolastico in termini operazionali, quindi, vuol dire esplicitare le operazioni in cui consistono la *misurazione* e/o la *valutazione*, quando, perché e da chi ognuna di esse viene svolta ecc. Utilizzare un "linguaggio operativo" in sede di elaborazione di un progetto o di un sistema di valutazione implica pertanto uno sforzo notevole di chiarificazione di ciò che si vuole fare, che si intende realizzare; costringe ad abbandonare formule generiche tristemente diffuse per cui, nelle ultime righe dei progetti didattici, si risolve il problema della valutazione dichiarando semplicemente che l'apprendimento degli studenti sarà valutato tramite strumenti "scientifici"; favorisce la comunicazione all'interno delle équipes di insegnanti e tra queste e i soggetti che non appartengono al mondo scolastico, per mezzo di un linguaggio che si riferisce ad attività concrete e non ha conseguentemente bisogno di ricorrere a "tecnicismi" più o meno scientificamente giustificati.

È importante sottolineare che i legami tra gli elementi che concorrono alla definizione operativa della valutazione, hanno un carattere *transazionale*. La concezione newtoniana della scienza ci ha abituato a considerare il mondo come composto da «entità stabili e comunque prefissate rispetto [...] all'osservazione e all'indagine»¹⁰¹ che interagiscono tra loro. Con l'introduzione del concetto di transazione, la nostra prospettiva cambia radicalmente. In un celebre esempio John Dewey spiega tale concetto paragonandolo ad una transazione commerciale: essa «[...] fa di uno dei partecipanti un compratore e dell'altro un venditore. Nessuno dei due è un compratore o un venditore se non *in* una transazione e *a causa di* una transazione in cui l'uno e l'altro siano impegnati. E questo non è tutto: certe particolari cose *diventano* dei beni o dei servizi perché sono impegnate in certe transazioni. Non esiste alcuna transazione commerciale senza delle cose che siano, sempre soltanto in una transazione e a causa di essa, beni, utilità o servizi. Inoltre a causa dello

pp. 185-205. Per la relazione tra scienza e giochi linguistici Phillips, D. L. (1977), *Wittgenstein and the Scientific Knowledge. A Sociological Perspective*, London, The Macmillan Press; trad. it. *Wittgenstein e la conoscenza scientifica. Un approccio sociologico*, Bologna, il Mulino, 1981, p. 140; e Mattei, F. (2002), *Il linguaggio scientifico tra scienza e didattica*, in Idem (2002), *Scienza religione filosofia. Intersezioni pedagogiche*, Roma, Anicia, pp. 15-88. Per un'introduzione ai programmi di ricerca del positivismo e del neopositivismo cfr. Oldroyd, D. (1986), *The Arch of Knowledge. An Introductory Study of the History of the Philosophy and Methodology of Science*, New York and London, Methuen; trad. it. *Storia della filosofia della scienza*, Milano, Il Saggiatore, 1989.

⁹⁸ Bridgman, P. W. (1934), *Seconda reazione di un fisico alla teoria degli insiemi*, cit., p. 151.

⁹⁹ Visalberghi, A. (1958), *Esperienza e valutazione*, Torino, Taylor, p. 182.

¹⁰⁰ Un'impostazione simile è anche in Boncori, G. (1994) *Guida all'osservazione pedagogica*, Brescia, La Scuola, pp. 21-22.

¹⁰¹ Visalberghi, A. (1958), *Esperienza e valutazione*, cit., p. 6.

scambio o del trasferimento, entrambe le *parti* (nome idiomatico per *partecipanti*) subiscono un mutamento; e i beni subiscono, per lo meno, un mutamento di *luogo* in conseguenza del quale essi acquistano o perdono certe relazioni o «capacità» connettive, rispetto a quelle che possedevano prima»¹⁰². Nell'ambito della definizione operativa della valutazione, quindi, tutti gli elementi considerati si definiscono reciprocamente, ovvero nessuno di essi esiste a prescindere dagli altri. Poniamo che l'operazione *x* consista 1) nella misurazione (attività) 2) svolta al termine di un'unità didattica con una prova oggettiva (metodo e strumenti), 3) del profitto (campo d'indagine), 4) degli studenti, da parte dell'insegnante interessato (soggetti), 5) in un'ora di tempo (risorsa temporale), 6) per valutare se qualche allievo non ha conseguito gli obiettivi previsti e necessita, quindi di un'attività di recupero (funzione). Ciò che sosteniamo è che se viene modificato anche uno solo degli elementi che definiscono in termini operazionali tale attività, gli altri devono assumere una forma più o meno diversa. Nel caso specifico se viene modificata la funzione e, pertanto, si desidera misurare le conoscenze acquisite dai discenti in una certa disciplina alla fine di un anno scolastico, i cambiamenti sono radicali: un'ora non è più sufficiente, è opinabile che lo sia una sola prova oggettiva ecc. Si comprende più facilmente qual è l'intensità del legame tra gli elementi che costituiscono un sistema di valutazione, se si esamina l'effetto esercitato dal cambiamento di funzione sugli strumenti di valutazione aperti (o non strutturati)¹⁰³. Se la funzione dell'atto valutativo è formativa, un insegnante può utilizzare l'interrogazione dialogica o il tema tradizionale oppure, in maniera più informale, può valorizzare le riflessioni ed i commenti degli studenti. Quando la funzione diviene sommativa, invece, a questi strumenti sono preferibili l'interrogazione e il colloquio pluridisciplinare¹⁰⁴.

Le conseguenze concrete di tale posizione sono diverse e in questa sede ne considereremo solo una a titolo esemplificativo. È diffusa l'abitudine di cercare prove oggettive per la misurazione del profitto presso enti di ricerca o docenti che ne hanno già fatto uso. In entrambi i casi lo scopo per cui è stato costruito lo strumento di rilevazione è differente da quello che ha motivato la ricerca. Nel primo riguarda la valutazione comparativa di risultati conseguiti da diversi sistemi di istruzione; nel secondo caso lo scopo è relativo all'accertamento del conseguimento da parte di *certi* alunni degli obiettivi di un *determinato* programma di insegnamento. In una prospettiva transazionale ciò comporta che gli strumenti di rilevazione utilizzati a tali scopi sono differenti tra loro e non possiedono per lo stesso motivo le caratteristiche per soddisfare le esigenze di valutazione di altri interventi didattici. Lo stesso

¹⁰² Dewey, J., Bentley, A. F. (1946), *Knowing and known*, Boston (Mass.), The Bacon Press; trad. it. *Conoscenza e transazione*, Firenze, La Nuova Italia, 1974, pp. 311-312.

¹⁰³ Cfr., in merito, il testo riportato nelle pagine seguenti di Benvenuto, G., *Un quadro di sintesi degli strumenti di verifica scolastica*.

¹⁰⁴ È opportuno sottolineare come, secondo Benvenuto, il cambiamento di funzione esercita sugli strumenti di valutazione semi-strutturati un effetto più intenso di quello descritto. Cfr. *Ibidem*.

ragionamento può estendersi alle prove che spesso appaiono sui libri di testo adottati dalle scuole al termine delle unità didattiche. In generale, pertanto, l'adozione di un test elaborato in un altro contesto, anche se validato, non può mai essere compiuta senza sottoporre lo stesso strumento ad una approfondita revisione, che lo renda compatibile con le caratteristiche del curriculum svolto e degli studenti cui deve essere somministrato.

Le conseguenze legate all'adozione di una prospettiva transazionale investono anche il tema dell'oggettività e quello connesso del ruolo del soggetto valutatore. Nel saggio che apre il capitolo, Postic e De Ketele distinguono tra l'osservatore indipendente e quello partecipante (lettura 12). Nel caso specifico del controllo scolastico il valutatore assume solitamente il punto di vista dell'"osservatore indipendente" e si tende a distinguerlo nettamente dal soggetto valutato; l'isolamento della variabile di disturbo rappresentata dall'equazione personale del valutatore è stata addirittura una delle ragioni che hanno promosso la nascita e la crescita della docimologia¹⁰⁵. Ma alcuni sviluppi della filosofia della scienza non ci permettono di liquidare il problema così semplicemente. Si pensi, ad esempio, al significato che il *principio di indeterminazione* di Werner Heisenberg riveste per le scienze sociali. Il fisico dimostra che l'introduzione a livello subatomico degli strumenti di osservazione altera lo stato del sistema studiato, rendendo impossibile la determinazione simultanea dei valori che assumono due proprietà della stessa particella: non si possono, per esempio, misurare contemporaneamente lo *spin* e la velocità di un elettrone. Il principio di Heisenberg segna, quindi, la morte del demone di Pierre-Simon de Laplace¹⁰⁶, ovvero del determinismo e dei postulati ad esso collegati, primo tra tutti quello riguardante la distinzione tra osservatore e oggetto osservato, tra ricercatore e oggetto di studio¹⁰⁷. Ne esce valorizzato l'approccio transazionale per cui la distinzione tra insegnante-valutatore, studente-valutato e profitto-oggetto di valutazione non è preesistente all'atto stesso della valutazione. Sul versante metodologico ciò comporta che ogni docente deve 1) sforzarsi di controllare tutte le variabili implicate in un processo valutativo (da quelle psicologiche a quelle logistiche e docimologiche) (modulo descrittivo del criterio transazionale); 2) considerare se stesso, i propri strumenti concettuali, quelli operativi, le proprie preferenze, come variabili socialmente e storicamente determinate che indirizzano e orientano gli atti valutati-

¹⁰⁵ Cfr. E. Lastrucci (1991), *Funzioni e aspetti della valutazione*, in Pietro Lucisano, E. Lastrucci (a cura di) (1991), *Apprendere ad insegnare per insegnare ad apprendere*, Roma, Borla, pp. 61-66.

¹⁰⁶ Secondo il noto positivista la storia e il futuro non avrebbero più segreti per un demone dotato di un'eccezionale capacità computazionale, cui sia data la possibilità di conoscere lo stato di tutte le particelle dell'Universo in un certo momento.

¹⁰⁷ Per un'introduzione ai problemi epistemologici connessi con le recenti scoperte della fisica e della biologia cfr. Cini, M. (1994), *Un paradiso perduto. Dall'universo delle leggi naturali al mondo dei processi evolutivi*, Milano, Feltrinelli.

vi (modulo critico del criterio transazionale)¹⁰⁸. Già nel 1955 Visalberghi evidenziava come ogni forma di accertamento delle conoscenze è preceduta e seguita da valutazioni e decisioni che influenzano l'uso delle stesse prove oggettive¹⁰⁹. Si pensi, per esempio, ad un fatto elementare che spesso passa inosservato: per valutare le conoscenze degli studenti non possiamo aspettare che esse si manifestino "naturalmente", ma siamo costretti a ricorrere ai test, ovvero, come afferma Melchiori (lettura 14) nel brano di seguito riportato, a stimoli artificiali¹¹⁰; ne consegue che la stessa valutazione del profitto è inevitabilmente segnata da un "peccato originale": essa, infatti, risente dello strumento artificiale con cui le conoscenze sono state misurate. Si rifletta allo stesso modo sul carattere convenzionale sia dei criteri che, secondo Domenico, orientano la costruzione di una prova, sia delle regole alle quali per Benvenuto bisogna attenersi per l'elaborazione dei quesiti. Ognuno di questi elementi è carico di "soggettività" o fa riferimento a teorie valide all'interno di precise coordinate storico-geografiche.

Per quanto riguarda l'oggettività, la definizione tecnica più comune è quella presentata da Benvenuto¹¹¹, per cui essa si configura come il grado di accordo esistente tra i giudizi espressi da più valutatori sulla stessa prestazione; nel caso di un test, essendo le risposte alle varie domande prestabilite e conseguentemente la correzione univoca, l'accordo deve essere necessariamente massimo. In altri termini, con tali strumenti di misurazione sembra possibile isolare completamente l'equazione personale dell'insegnante-valutatore. Ciò, però, è vero solo parzialmente, in particolare per quanto riguarda il campo limitato della correzione delle prove somministrate e della relativa assegnazione dei punteggi. Ma cosa dire in merito alle valutazioni e alle decisioni che precedono e seguono le operazioni di misurazione del profitto tramite la tecnica del *testing*? Si pensi, per esempio, all'individuazione dei contenuti da sottoporre ad accertamento. Per quanto riguarda le prove *standardizzate* trovare l'accordo su questo argomento a volte richiede anni, come nel caso della prima vasta indagine comparativa internazionale promossa dall'*International Association for the evaluation of the Educational Achievement* (IEA): essa riguardava sei discipline d'insegnamento (da cui il nome *six subject survey*) e per elaborare e tradurre nelle

¹⁰⁸ Visalberghi, A. (1958), *Esperienza e valutazione*, cit., pp. 8-9. Un esempio dell'influenza delle strutture concettuali dell'insegnante sulla valutazione è rappresentato da quanto affermato precedentemente a proposito dei processi valutativi tipici della didattica tradizionale.

¹⁰⁹ Visalberghi, A. (1955), *Misurazione e valutazione nel processo educativo*, Milano, Edizioni di Comunità.

¹¹⁰ È opportuno sottolineare come il testo di Melchiori sia particolarmente rilevante per le convenzioni linguistiche e per le distinzioni che introduce. Esso, infatti, definisce cos'è una prova oggettiva, un quesito, un elemento di prova ecc.; distingue le prove in base alle caratteristiche formali dello stimolo e della risposta, al grado di standardizzazione, alla funzione didattica che assolvono ecc.

¹¹¹ Benvenuto, G. (2003), *Mettere i voti a scuola. Introduzione alla docimologia*, Roma, Carocci, pp. 110-112.

lingue dei ventuno paesi partecipanti tutte le prove ed i questionari sono stati necessari ben quattro anni¹¹². Per le prove *informali* o di classe¹¹³ la situazione, se possibile, è ancora più delicata. Capita spesso, infatti, che l'oggettività e la "scientificità" del controllo scolastico siano semplicisticamente identificate con l'uso di strumenti di misurazione composti da domande chiuse¹¹⁴. L'impegno principale di chi sostiene questa posizione si risolve nella mera costruzione del test, che spesso, da un lato, non rileva se sono stati conseguiti gli obiettivi previsti dall'attività di insegnamento, dall'altro, pone quesiti i cui risultati sono difficilmente trattabili statisticamente o teoricamente poco significativi. Le stesse osservazioni sono valide anche nel caso degli strumenti tradizionali, sebbene in quest'ambito la situazione sia ancor più grave.

Coerentemente con l'ottica transazionale e operativa che abbiamo adottato, riteniamo che l'unica strada percorribile per assicurare l'oggettività del controllo scolastico sia quella di renderlo 1) organico al processo di insegnamento-apprendimento (come, nella parte seconda, propongono sia Vertecchi, identificando provocatoriamente la valutazione collettiva-iniziale con la programmazione, sia Gattullo, insistendo sulla relazione tra gli scopi educativi e lo stesso controllo); 2) rigoroso con l'adozione condivisa dalla comunità degli insegnanti di precise teorie, procedure e strumenti (non necessariamente coincidenti con quelle proposte in questa sede). Tale posizione ci sembra condivisa dallo stesso Vertecchi quando afferma che se «[...] una prova [o un qualsiasi strumento di misurazione] è stata predisposta partendo da una accurata definizione degli obiettivi e preoccupandosi di assicurare le necessarie caratteristiche di validità e attendibilità, possiamo ritenere che la verifica risultante dalla somministrazione della prova stessa sarà obiettiva: in valutazione, l'*obiettività* non va intesa nel senso di una assoluta quanto improbabile corrispondenza del dato al fatto, ma in quello di un superamento della intuitività, della estemporaneità, della casualità degli apprezzamenti che riguardano la prestazione degli allievi»¹¹⁵.

Le considerazioni relative all'oggettività ci consentono di riflettere, nell'ottica della responsabilità, sulla relazione tra singolo docente, équipe di insegnanti e valutazione. Il controllo scolastico, come si è detto, assume una certa forma in ragione delle categorie concettuali di carattere storico-sociale possedute dall'insegnante-valutatore. Per essere più precisi potremmo dire che la visione del mondo, della

¹¹² Per una sintetica presentazione dei principali risultati delle ricerche IEA svolte in Italia cfr. Laeng, M. (1992), *Pedagogia sperimentale*, Scandicci (FI), La Nuova Italia, pp. 202-240. In merito un utile percorso bibliografico è stato redatto da Lucisano, P., Siniscalco, M. T. (1994), *Rassegna bibliografica delle ricerche IEA*, in «Cadmo», a. II, n. 5-6, dicembre, pp. 164-186.

¹¹³ Tre tipologie per classificare le prove oggettive sono presentate da Melchiori nel brano riportato.

¹¹⁴ Per una tipologia degli strumenti di misurazione si rimanda allo stesso brano di Melchiori e al primo di Benvenuto – intitolato *Un quadro di sintesi degli strumenti di verifica scolastica*.

¹¹⁵ Vertecchi, B. (1984), *Manuale della valutazione. Analisi degli apprendimenti*, cit., pp. 102-103.

scuola e della sua missione sociale, le forme assuntive, la teorie (psicologiche, sociologiche, pedagogico-didattiche, docimologiche ecc.), gli strumenti posseduti e condivisi (ad un livello variabile di consapevolezza) dalla comunità dei ricercatori e degli insegnanti producono una certa forma di valutazione. Dobbiamo, quindi, riconoscere che le azioni di ogni docente, che ogni atto valutativo non sono mai analizzabili come fatti in sé, in termini astratti o decontestualizzati; ma partecipano del flusso delle esperienze di vita complessive di un singolo e di una comunità. A livello di microsistema ne consegue che sul piano logico-epistemologico la responsabilità del controllo scolastico non può essere attribuita al singolo docente, ma al Collegio dei Docenti e ai consigli di (inter-)classe. Sul piano pratico ciò è indispensabile, come è sottolineato da Vertecchi (nella parte seconda), soprattutto se non ci si limita a misurare esclusivamente il profitto degli studenti, ma si tiene presente l'insieme delle dimensioni e delle variabili che consente di delineare e aggiornare un profilo per ogni allievo; ovvero se il controllo scolastico assume una forma *integrata* o polidimensionale.

Quest'ultimo carattere della valutazione contribuisce a spiegare perché abbiamo aperto il capitolo con il saggio di Postic e De Ketele (lettura 12). L'integrazione, infatti, prevede l'uso di strumenti di misurazione diversi in ragione degli obiettivi di apprendimento previsti, degli scopi della rilevazione, delle caratteristiche peculiari degli studenti, della situazione ecc. Questa tesi è sostenuta e tradotta sul piano operativo da Benvenuto¹¹⁶, che esplicita le associazioni esistenti tra le funzioni, i tempi e gli strumenti di valutazione (lettura 15). Come si è visto, inoltre, in questa prospettiva è fondamentale che il controllo si avvalga dei contributi di diversi professionisti con profili e competenze differenti: prima di tutto quelli dei docenti che compongono un consiglio di classe o interclasse, di seguito quelli eventuali di un psicologo, di un pedagogista ecc. D'altro canto, al fine di fornire in tempo reale le informazioni essenziali per la regolazione degli interventi didattici, è altresì necessario che il controllo svolga, nei momenti opportuni, tutte le sue funzioni: dalla misurazione dei prerequisiti all'orientamento. Infine, esso deve prendere in considerazione diversi oggetti: per assicurare il successo formativo, ovvero il conseguimento delle mete educative, infatti, non ci si può limitare alla misurazione del solo profitto, ma bisogna anche rilevare e intervenire sulle caratteristiche degli allievi che influenzano in modo determinante i processi di apprendimento.

Per questo motivo Lastrucci propone di elaborare all'inizio del percorso scolastico un *profilo complessivo* del discente, che richiede la realizzazione di un insieme di operazioni denominato *valutazione diagnostica*¹¹⁷. Il suo obiettivo è di rilevare ciò che gli allievi già sanno e quanto potenzialmente sono in grado di apprendere. Le informazioni da rilevare, perciò, non riguardano soltanto le conoscenze pregresse, ma

¹¹⁶ Si veda il primo contributo dell'Autore presentato in questa parte: *Un quadro di sintesi degli strumenti di verifica scolastica*.

¹¹⁷ Cfr. Lastrucci, E. (2001), *Autovalutazione di istituto*, Roma, Anicia, pp. 155-184.

il vasto insieme che viene rappresentato nella tabella riportata nella pagina successiva (che pone in relazione le classi di dati da rilevare con le rispettive fonti). Lo scopo principale di tale rilevazione è di entrare in possesso di tutti i dati che consentono una programmazione dell'intervento educativo "perfettamente" calibrata sulle caratteristiche in ingresso degli utenti del servizio scolastico, aumentando conseguentemente le probabilità di successo e degli studenti e dello stesso intervento¹¹⁸.

PROFILO COMPLESSIVO	
Componente	Fonte dei dati
1) Notizie intorno all'ambiente socio- culturale e familiare di provenienza, nonché all'esperienza extrascolastica	- Analisi del contesto (generale) - Colloqui con la famiglia (individuali) - Colloqui con l'alunno (individuali)
2) Notizie intorno alla precedente carriera scolastica	- Analisi documenti scolastici
3) Pre-requisiti disciplinari (conoscenze e abilità relative alla materia specifica)	- Prove di ingresso per l'analisi dei pre-requisiti
4) Pre-requisiti generali (conoscenze e abilità derivate dall'apprendimento di altre materie)	- Prove di ingresso per l'analisi dei pre-requisiti
5) Profilo dello sviluppo cognitivo	- Test di sviluppo cognitivo
6) Profilo linguistico	- Test di comprensione della lettura
	- Elaborati scritti
	- Colloqui con l'alunno
	- Glotto-kit
7) Profilo motivazionale	- Test di padronanza lessicale
8) Interessi	- Scale di atteggiamento
	- Questionari di interessi

Tab. 1 – Profilo complessivo dello studente¹¹⁹.

¹¹⁸ Come abbiamo visto, nella proposta di Vertecchi, viene chiamata *valutazione diagnostica* l'attività di accertamento che ha lo scopo di rilevare, sia nelle fasi iniziali sia in quelle intermedie di un intervento di istruzione o formazione, il possesso da parte degli studenti delle conoscenze sufficienti per intraprendere o proseguire certi studi, al fine di decidere se è necessario svolgere attività di recupero o sostegno. Ci sembra evidente la differenza tra gli scopi dei due insiemi di operazioni. Nonostante tutto riteniamo che la relazione tra le due definizioni sia di inclusione, per cui la prima rappresenta un sottoinsieme della seconda. Per quanto riguarda la *valutazione diagnostica* e il *profilo complessivo* degli studenti cfr. anche Lastrucci, E. (2007²), *Progettare, programmare e valutare l'attività formativa*, cit., pp. 163-209 e Idem (2004), *Valutazione diagnostica. Metodi e strumenti per costruire il profilo in ingresso dell'alunno*, Roma, Anicia. Per l'applicazione del concetto di *profilo* all'autovalutazione d'istituto cfr. anche Melchiori, R. (2001), *Valutazione interna d'istituto scolastico o di microsistema. Sintesi del progetto. Modello concettuale, modello operativo, modello degli strumenti, modello sistema informativo*, SNQI, progetti 7/8, INValSI, Frascati (RM), ottobre, reperibile sul sito Internet <http://validi.invalsi.it>

¹¹⁹ Lastrucci, E. (2001), *Autovalutazione di istituto*, Roma, Anicia, p. 168. Per motivi di spazio, non riportiamo in questa sede la versione aggiornata del profilo complessivo dell'allievo. Essa, infatti prevede altre quattro componenti (riguardanti gli atteggiamenti, la personalità, la socialità e lo sviluppo affettivo e la socializzazione civico-politica e la coscienza storica degli studenti). Inoltre,

Il profilo proposto non ha ovviamente un carattere prescrittivo. A nostro avviso ogni équipe docente dovrebbe elaborare il proprio. Quello presentato costituisce un'ottima base, integrabile in varie direzioni. In esso per esempio non si considerano alcune abilità trasversali, quali quelle prosociali¹²⁰, e non si prevede un articolato schema di registrazione delle informazioni rilevate. Ogni componente del profilo, infatti, come suggerisce Giuseppe Boncori nel brano riportato (lettura 13), potrebbe essere strutturata in *voci* o *criteri*, che in ambito scolastico coincidono o possono coincidere con gli obiettivi dell'attività di insegnamento. Per ogni criterio, inoltre, bisognerebbe stabilire quali sono i *segni indicatori* ed i *livelli*, ovvero i comportamenti che indicano in modo intersoggettivamente condivisibile in che grado gli stessi obiettivi sono stati conseguiti¹²¹. Un apposito sistema notazionale consentirebbe, infine, di riassumere in un quadro sinottico l'insieme dei dati posseduti.

In questa direzione si muove, ad esempio, Domenici, che, rilevate le insufficienze nell'applicazione della norma istitutiva del *fascicolo personale dell'allievo*¹²², propone di utilizzare un *dossier personale dell'allievo*, la cui struttura nelle sue linee essenziali è molto simile a quella che abbiamo appena descritto¹²³. Ogni insegnante, infatti, produce per la propria disciplina due griglie (la prima riguardante gli aspetti cognitivi dell'apprendimento e la seconda quelli affettivo-motivazionali e relazionali) che riportano in ascissa la descrizione comportamentale degli obiettivi previsti, in ordinata il numero delle misurazioni effettuate e nelle celle la presenza o l'assenza dei comportamenti che segnalano il conseguimento dei traguardi educativi. Il consiglio di classe sintetizza le griglie disciplinari in un unico documento, che ha la stessa struttura di quelli elaborati dai docenti, ma sull'asse delle ascisse riporta, al posto dei descrittori dell'apprendimento, i nomi delle discipline insegnate. Un ulteriore schema riassuntivo registra l'evoluzione diacronica delle competenze degli studenti e prevede le ulteriori tendenze di sviluppo¹²⁴.

la tabella che sintetizza il profilo è stata arricchita sia con l'introduzione di una colonna in cui vengono esplicitate le articolazioni di ogni componente sia con un ampliamento delle fonti da cui possono essere raccolti i dati. Cfr. Lastrucci, E. (2004), *Valutazione diagnostica. Metodi e strumenti per costruire il profilo in ingresso dell'alunno*, cit., pp. 44-49.

¹²⁰ Bellacicco, D., Cellamare, S. (1999), *Introduzione all'osservazione del bambino. Come rilevare le abilità prosociali nella scuola dell'infanzia*, Roma, Monolite.

¹²¹ Cfr. Calonghi, L. (1992), *Profili e livelli nella valutazione del profitto*, in «Orientamenti pedagogici», a. XXXIX, n. 3, Maggio-Giugno, pp. 605-616; Boncori, G. (1992), *Profilo*, in *Enciclopedia pedagogica*, diretta da M. Laeng, Brescia, La Scuola, col. 9446-9450.

¹²² Cfr. Domenici, G. (1998), *Manuale dell'orientamento e della didattica modulare*, Roma-Bari, Laterza, pp. 67-70. Il *fascicolo personale dell'allievo* è stato istituito con il D. M. del 16/11/1992.

¹²³ Cfr. Domenici, G. (1996), *Continuità e orientamento. Il «dossier» personale dell'allievo*, in «Cadmo», a. IV, n. 10-11, Aprile, pp. 91-99.

¹²⁴ Gandolfo, M. T. (1998), *Il dossier dell'allievo*, in Domenici, G. (1998), *Manuale dell'orientamento e della didattica modulare*, cit., 159-177.

I profili presentati hanno il merito indiscutibile di garantire la continuità orizzontale e verticale attraverso un insieme di rappresentazioni di stato che, aggiornate in tempo reale, descrivono l'evoluzione delle caratteristiche dei singoli studenti, permettendo di calibrare e ricalibrare gli interventi didattici sui reali bisogni degli utenti del servizio scolastico. Entrambi, però, sono composti da voci che non trovano riferimento in alcuna moderna teoria psicologica. Il profilo proposto da Domenico, infatti, consiglia di elaborare i descrittori dell'apprendimento disciplinare utilizzando la tassonomia di Bloom sia per gli aspetti cognitivi sia per quelli affettivo-motivazionali e relazionali¹²⁵. La fecondità di tale tassonomia risiede sicuramente nella sua semplicità e utilizzabilità; ma le teorie psicologiche ad essa sottostanti «[...] di marca behaviourista e neobehaviourista (sia nelle primitive versioni di Watson e Skinner, sia in quelle evolute di Bloom, Gagnè, Bandura ecc.) hanno lasciato il campo – anche nelle applicazioni tecnologiche – alle teorie «human information processing» e, soprattutto, alle teorie cognitive che – a partire da Vygotsky, Piaget, Bruner fino a Olson, Nelson, Gardner – hanno rimesso in evidenza il ruolo dei sistemi simbolici nella strutturazione delle prestazioni e quindi inducono a spostare l'attenzione dagli esiti ai processi, agli oggetti/contenuti dell'apprendimento oltre che ai soggetti»¹²⁶.

Il limite di questo approccio risiede nella mancanza di un'analisi critica dei differenti strumenti concettuali e operativi (prodotti dalle diverse teorie psicologiche e sociologiche) che vengono adottati in sede didattica. Se consideriamo, infatti, che ogni teoria possiede un apparato concettuale che, da un lato, segmenta diversamente la realtà empirica e, dall'altro, si riflette inevitabilmente sulle ipotesi elaborate per risolvere i quotidiani problemi d'insegnamento, comprendiamo come sia necessaria un'analisi storico-critica delle categorie adottate per almeno due diversi motivi: 1) la mancanza di un paradigma dominante nelle scienze umane non assicura che gli strumenti adottati siano tra loro coerenti; sul piano pratico ne consegue il pericolo di fornire agli studenti stimoli tra loro contraddittori o la cui interazione potrebbe produrre conseguenze non prevedibili e non necessariamente positive; 2) la mancanza di un esplicito riferimento a determinate teorie comporta l'impossibilità di contribuire significativamente e da un punto di vista pedagogico al loro sviluppo.

¹²⁵ Cfr. Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R. (1956), *Taxonomy of educational objectives. The classification of educational goals, Handbook I: cognitive domains*, New York, David McKay Company; trad. it. *Tassonomia degli obiettivi educativi. La classificazione delle mete dell'educazione*, vol. I, *Area cognitiva*, Teramo, Giunti & Lisciani, 1983; Bloom, B. S., Krathwohl, D. R., Masia, B. B. (1964), *Taxonomy of educational objectives. The classification of educational goals, Handbook II: affective domains*, New York, David McKay Company; trad. it. *Tassonomia degli obiettivi educativi. La classificazione delle mete dell'educazione*, vol. II, *Area affettiva*, Teramo, Giunti & Lisciani, 1986.

¹²⁶ Calidoni, P. (1993), "Genesi, significato, limiti e prospettive della cosiddetta «pedagogia per obiettivi»", cit., p. 385.

Ne consegue un ulteriore rinforzo alla separazione esistente tra i diversi campi di ricerca scientifica e tra questi ultimi e la reale attività educativa¹²⁷.

L'uso della tecnica del *testing*, si è detto, richiede l'elaborazione di strumenti di misura validi e attendibili. L'importanza e la delicatezza di questa operazione non deve essere sottovalutata, pena la rilevazione di dati inutili o, peggio ancora, il compimento di errori di valutazione; per questo motivo due lunghi brani, uno di Domenici (lettura 16) e l'altro di Benvenuto¹²⁸ (lettura 17), illustrano come costruire le prove oggettive e i relativi quesiti. Anche attraverso una lettura superficiale si comprende immediatamente che le attività richieste non sono banali o poco impegnative. Al termine della misurazione, inoltre, è necessario trattare i dati statisticamente sia per formulare ipotesi circa la bontà della prova sia per valutare la prestazione dei singoli allievi e della classe. Il testo di Vertecchi (lettura 18), con un linguaggio facilmente comprensibile seppur rigorosissimo, suggerisce alcune semplici ma efficaci soluzioni tecniche. Ovviamente queste operazioni richiedono del tempo e una conoscenza elementare della statistica.

L'uso del *testing* potrebbe conseguentemente sembrare irrealizzabile: i momenti e le funzioni della valutazione sono numerosi; la costruzione degli strumenti di misurazione e l'elaborazione dei dati sono operazioni che richiedono tempo, concentrazione e una certa familiarità con la statistica; la definizione delle mete educative e delle finalità in obiettivi comportamentali implica uno sforzo di analisi cui non siamo abituati. Le prassi invalse negli istituti scolastici sono molto diverse: da un lato, non assicurano le risorse necessarie per l'implementazione di processi così complessi, che spesso vengono semplificati rifugiandosi ancora nella tradizionale struttura bifasica dell'attività didattica (trasmissione del sapere e accertamento delle conoscenze acquisite); dall'altro, oppongono all'innovazione una resistenza culturale e organizzativa, che si trasforma nell'esercizio di una pressione sociale finalizzata a scoraggiare i promotori di nuove idee.

Considerare la forma di valutazione presentata in queste pagine come poco concreta o irrealizzabile denota un atteggiamento di rassegnata accettazione dello *status quo*, sul cui significato politico e filosofico non è il caso di indagare in questa sede. Noi riteniamo che la relazione tra le pratiche valutative invalse e quella proposta sia simile a quella che intercorre tra realtà empirica e mondo dei valori, tra essere e dover essere. Già nel 1958 Visalberghi affermava che un giudizio di valore è tale se le sue conseguenze «[...] 1) pur potendo non venire generalmente apprezzate dalla gente che resta attaccata ai suoi attuali punti di vista e moduli valutativi, 2) siano tuttavia destinate alla lunga a determinare uno stato di maggiore soddisfazione ge-

¹²⁷ Cfr. Tornatore, L. (1984), *Sperimentalismo educativo e conoscenza scientifica*, in Egle Becchi, B. Vertecchi (a cura di) (1984), *Manuale critico della sperimentazione e della ricerca educativa*, Milano, Franco Angeli, pp. 37-48; Lumbelli, L. (1984), *Qualità e quantità nella ricerca empirica in pedagogia*, in *ibidem*, pp. 101-133.

¹²⁸ Cui si riferisce in questo caso al secondo contributo di Benvenuto, intitolato *Formulare i quesiti (esempi e alcune regole per la costruzione)*.

nerale, tramite anche la modifica di precedenti opinioni e assunzioni»¹²⁹. I valori stessi, quali mezzi procedurali, suggeriscono come promuovere tali modifiche: essi, infatti, rappresentano criteri d'ordine delle attività da svolgere per conseguire i traguardi che gli stessi valori definiscono; indicano, in altri termini, come organizzare le operazioni in organici piani di azione finalizzati alla realizzazione di un nuovo stato di fatto.

Concretamente i docenti che adottano la forma di valutazione qui presentata devono creare le condizioni affinché essa possa essere implementata nella loro scuola. Questo comporta: 1) che essi debbano possedere una *visione strategica*, capace di identificare gli strumenti concettuali e operativi che nei momenti opportuni possono essere adottati dal Collegio dei docenti o dai consigli di classe; 2) tale visione diviene *pedagogica* se si tiene a mente che le innovazioni devono sempre porsi nell' "area di sviluppo prossimale" dell'organizzazione scolastica; 3) diviene *organizzativa* se le proposte sono sempre sostenibili per ciò che riguarda, da un lato, le questioni culturali e, dall'altro, l'impiego delle risorse disponibili; si pensi, in proposito, a quant'è importante ottimizzare queste ultime in vista della stesura e dell'aggiornamento di un profilo complessivo per ogni studente.

Tutto ciò richiede tra le altre cose il superamento di una visione oramai vetusta del principio della libertà d'insegnamento. La forma di valutazione presentata richiede che ogni Consiglio di classe faccia lo sforzo di definire quali informazioni rilevare, chi lo deve fare, perché, quando, con quali strumenti e, soprattutto, come utilizzare i dati per migliorare l'efficacia degli interventi. Questi ultimi, ricordiamo, sono gli strumenti con cui assicurare agli utenti il successo formativo. Le stesse materie di insegnamento devono essere concepite in modo dinamico e funzionale come mezzo per il raggiungimento di questo scopo. La progettazione e la programmazione sono, quindi, da un certo punto di vista, attività tecniche di elaborazione delle iniziative didattiche, da un altro, il momento della concertazione, della negoziazione di traguardi, obiettivi, strumenti ecc., cui la società e gli stessi studenti hanno il diritto di partecipare nelle forme dovute. Nella scuola dell'autonomia la libertà d'insegnamento, pertanto, non consiste più nella scelta di metodi o di "stili d'insegnamento" per realizzare programmi stabiliti centralmente, ma nella cooperazione all'elaborazione prima di tutto del Piano dell'Offerta Formativa (Pof) e, in seguito, alla progettazione delle attività da svolgere in classe con i singoli studenti. Non si configura, quindi, nei termini individuali di spazi d'insegnamento impermeabili a ingerenze esterne, ma, da un lato, come diritto di definizione collettiva del senso del proprio operato, dei traguardi che desidera conseguire l'organizzazione scolastica di appartenenza, dall'altro, come possibilità e capacità di lavorare in gruppo, di contribuire con il proprio apporto alla definizione di mete e strategie¹³⁰.

¹²⁹ Visalberghi, A. (1958), *Esperienza e valutazione*, cit., pp. 91-92.

¹³⁰ Cfr. Lastrucci, E. (2001), *Autovalutazione di istituto*, cit., pp. 26-27; Visalberghi, A. (1986²), *Pedagogia e scienze dell'educazione*, cit., pp. 55-59.

Questa interpretazione della libertà d'insegnamento comporta un cambiamento culturale di vasta portata, che in questa sede è emerso in modo episodico in rapporto al tema della valutazione. Ma porre il successo formativo della maggior parte degli allievi come obiettivo esplicito che deve conseguire l'organizzazione scolastica, significa permettere ad ogni studente di conseguire le mete educative dichiarate nel Pof, ovvero di acquisire le competenze necessarie per esercitare i propri diritti di cittadinanza tanto nella locale comunità di appartenenza, quanto a livello ecumenico o cosmopolita. Si richiede, in altri termini, che i risultati della maggioranza degli studenti siano equivalenti, ovvero che siano qualitativamente omogenei sebbene riguardanti aree diverse di interesse. Ne consegue la diversa *nozione di sufficienza* presentata da Vertecchi nell'ultimo brano di questa raccolta (lettura 19). Per garantire l'equivalenza dei risultati gli obiettivi dell'attività di insegnamento devono essere conseguiti, i comportamenti o le prestazioni degli studenti devono raggiungere predeterminati standard di accettabilità. Se sono al di sotto, dichiarare adeguate o sufficienti le loro conoscenze in ragione di altri criteri (il progresso, l'impegno) significa accettare il rischio che alcuni studenti non arrivino a padroneggiare gli strumenti concettuali e operativi per interpretare e agire nella nostra società.

12. Una tipologia dell'osservazione (Postic, M., De Ketele, J. M., Osservare le situazioni educative)¹³¹

Ogni strumento (metodo, tecnica, attrezzi...) *deve essere sempre* subordinato all'obiettivo ricercato. Il problema della scelta di uno strumento è dunque un problema di *adeguatezza* "mezzo-obiettivo".

Una volta stabilito bene l'obiettivo, la scelta di un mezzo è in funzione - tra le altre cose - della conoscenza del ventaglio di strumenti disponibili e delle loro caratteristiche (possibilità e limiti, vantaggi e svantaggi) [...].

Per stabilire la nostra tipologia, noi utilizzeremo dei parametri che obbediscono alle seguenti esigenze:

1. rispondere ai seguenti interrogativi basilari: perché osservare? chi osserva? osservare chi? con quali strumenti osservare? quale unità di osservazione scegliere? Osservare con quale grado di inferenza? come annotare l'osservazione? in quale tipo di situazione è meglio osservare? quale grado di libertà bisogna lasciare all'osservatore? come trattare le informazioni raccolte? [...].

LE FUNZIONI DELL'OSSERVAZIONE

Perché osservare? [...]

Certamente, uno stesso strumento può rivestire simultaneamente numerose funzioni, ma la scelta e l'utilizzazione di uno strumento di osservazione devono essere pensate prima di tutto in relazione ad una funzione prioritaria.

La funzione descrittiva. Con la funzione descrittiva si osserva per descrivere i fenomeni o una situazione [...].

La funzione formativa. In questa prospettiva, si osserva per *agire sulla base della situazione osservata* e si agisce per *formare* [...].

La funzione valutativa. In questa prospettiva, siamo nella dinamica seguente: si osserva per *valutare*, si valuta per *decidere*, si decide per *agire*. L'azione sarà a sua volta sottoposta alla valutazione (e dunque all'osservazione) per una nuova decisione...

La funzione euristica. Si parla di funzione euristica o di *osservazione invocata* quando l'attività è orientata verso l'emergere di ipotesi pertinenti che saranno ulteriormente sottoposte ad attività di controllo [...].

¹³¹ Tratto da: Postic, M., De Ketele, J. M. (1988), *Observer les situations éducatives*, Presser Universitaires de Frances; trad. it. *Osservare le situazioni educative. I metodi osservativi nella ricerca e nella valutazione*, Torino, SEI, 1993, pp. 25-46.

La funzione di verifica. Si parla qui di *osservazione provocata*, nel senso che una situazione è provocata, ricercata o manipolata, al fine di verificare un'ipotesi. L'osservazione rientra nelle attività di misura delle variabili riguardanti l'ipotesi [...].

L'AUTORE DELL'OSSERVAZIONE

[...] abbiamo definito l'osservazione come un'operazione sia di rilevamento che di strutturazione dei dati in modo da far emergere una rete di significati. Rispondere alla domanda "chi osserva?" è già un primo tentativo per sapere quali dati hanno la possibilità di essere prelevati, in quale modo saranno strutturati e in quale direzione sarà ricercato il significato. Allo stesso modo, per permettere all'osservazione di svolgere una certa funzione, è bene porsi la domanda "chi deve osservare?" [...]

L'osservatore è *indipendente* quando osserva un gruppo senza integrarsi: un futuro insegnante osserva la classe in funzione della sua formazione [...].

L'osservatore è *partecipante* quando si integra al gruppo e alla vita di questo. Gli etologi che hanno spesso utilizzato e studiato l'osservazione partecipante, ritengono sia importante distinguere numerose forme.

Così Margaret Mead descrive *l'osservazione partecipante passiva* come segue:

«L'osservatore partecipante entra in gioco, osserva ma non intacca niente [...] (Mead, 1966, p. 32)¹³².

[...] Essa è particolarmente indicata - e a volte la sola possibile - a scoprire e studiare un "mondo" estraneo e non familiare [...]. Nell'ambiente scolastico, si possono immaginare alcune situazioni: l'osservatore si fa studiare, poco a poco si integra in un'équipe pedagogica, ecc.

Ma si parlerà anche di *osservazione partecipante attiva* quando l'osservatore gioca dei ruoli effettivi capaci di modificare radicalmente alcuni aspetti della vita di gruppo. È il caso in cui l'osservatore è il professore stesso, che si può porre come osservatore per se stesso o semplicemente come "l'informatore" per un osservatore esterno; è dunque il caso in cui l'investigatore o alcuni alunni sono assunti come "informatori" dall'investigatore esterno [...]. Ci sembra importante che all'inizio l'investigatore consideri queste informazioni come provenienti dal campo rappresentativo e faccia i controlli necessari se vuole stabilire la corrispondenza tra il campo rappresentativo dell'informatore ed il campo percettivo dei comportamenti manifesti [...].

Questa distinzione e la nostra non ci sono nemmeno senza far pensare ad un'altra distinzione divenuta ormai classica tra ricerca-azione e azione-ricerca.

¹³² Mead, M. (1966), *L'un et l'autre sexe*, Paris, Denoël [n.d.c.].

Notiamo, infine, che estraneo o partecipante, l'osservatore sarà in alcuni casi *perceptivo* esplicitamente come osservatore; in altri casi, no. Ovviamente questo comporta delle conseguenze sul problema della reattività (effetti dovuti alla consapevolezza di essere osservati).

L'OGGETTO DELL'OSSERVAZIONE

Tre distinzioni permettono di delineare meglio l'oggetto dell'osservazione e, di conseguenza, di caratterizzare i diversi metodi d'osservazione:

1. l'osservazione può fondarsi su dei *fatti* o, al contrario, su delle *rappresentazioni*.
2. l'osservazione può essere *attributiva* o, al contrario, *narrativa*.
3. l'osservazione può essere *allospettiva* o, al contrario, *introspettiva*.

L'OSSERVAZIONE CHE SI FONDA SU FATTI O SU RAPPRESENTAZIONI

Si dirà che l'osservazione poggia su *fatti* (in senso largo) quando è orientata sulle caratteristiche della situazione, sui comportamenti o sulle interazioni tra le persone. Si dirà che l'osservazione si fonda sulle rappresentazioni quando è mirata a raccogliere opinioni, modi di percepire le cose o i comportamenti, o a dare loro un significato o ancora ad attribuire loro una causa.

[...] ci sembra importante distinguere diverse forme di rappresentazioni che molto spesso possono non coincidere e, perciò, meritano un'attenzione particolare.

Un primo livello di distinzione poggia sulle risposte a tre interrogativi: *chi emette? a proposito di chi o di che cosa? per quale destinatario?* Prendendo come esempio la situazione scolastica, posso interessarmi a ciò che il soggetto-alunno pensa di se stesso, degli altri alunni, del suo gruppo-classe, di un altro gruppo-classe, del suo professore, [...] di un certo comportamento vissuto, di un comportamento in generale, di un valore, ecc., e ciò in funzione del destinatario della comunicazione [...].

Un altro livello di distinzione consiste nel differenziare le *rappresentazioni dirette* (come nel caso precedente) dalle *rappresentazioni indirette* (che noi chiamiamo anche rappresentazioni sociali). Attraverso queste ultime bisogna capire le immagini che un emittente si crea della rappresentazione degli altri; come ad esempio: l'immagine che tale alunno ha della rappresentazione di un altro alunno nei suoi riguardi, quella che egli ha della rappresentazione del suo professore nei confronti della disciplina, ecc. A nostro avviso le rappresentazioni indirette giocano un ruolo spesso più decisivo delle rappresentazioni dirette: in quanto alunno io agisco spesso più in funzione di ciò che il professore pensa di me, che in funzione della mia percezione verso di me.

Un terzo livello di distinzione si impone soprattutto quando l'oggetto di studio ha portato ad una problematica di cambiamento. L'oggetto della rappresentazione

può essere collegato al piano *dell'ideale* (ad esempio: è idealmente importante per me terminare un consiglio di classe con una decisione concreta?), sul piano del *reale* o del vissuto (ad esempio: lo facciamo abitualmente?), sul piano del *possibile* (ad esempio: è possibile nel mio contesto?), sul piano del *volere* (ad esempio: è ciò che voglio veramente?) [...].

L'OSSERVAZIONE ATTRIBUTIVA O NARRATIVA

A seconda del tipo di oggetto, l'osservazione sarà **attributiva** quando si concentra l'attenzione:

1. sulla presenza o assenza di uno o più "oggetti";
2. o/e sulla presenza o assenza (è il grado eventuale) di caratteristiche in un "oggetto" (soggetto, cosa, situazione, interazione, ecc.) [...];
3. o/e sulla presenza o assenza di un'azione su un "oggetto" [...].

Nell'osservazione attributiva, lo sforzo riconduce a ciò che può essere affermato o negato circa l'oggetto da osservare.

L'osservazione sarà invece **narrativa** quando l'attenzione è concentrata:

1. sullo svolgimento delle azioni;
2. e/o sulla successione degli stati (fisici, affettivi o altri) che accompagnano lo svolgimento delle azioni;
3. e/o sugli effetti dell'azione;
4. e/o sulle conseguenze ulteriori (ad esempio, le decisioni, i comportamenti, le modifiche della situazione ... che hanno seguito l'azione ed i suoi effetti immediati raggiunti).

L'OSSERVAZIONE ALLOSPETTIVA O INTROSPETTIVA

Per osservazione introspettiva (o autospettiva) bisogna qui intendere l'osservazione del soggetto (o della situazione nella quale egli è coinvolto) da parte di se stesso.

Al contrario, noi parleremo di osservazione allospettiva quando l'osservazione di un soggetto (o della situazione) è condotta da un altro [...].

Così tutti i questionari di interesse o di attitudine sono di natura introspettiva. Lo stesso si può dire per la maggior parte dei colloqui, più o meno sistematici. E la volta anche di alcuni inventari (*check list*) sistematici. Allo stesso modo molte scale di valutazione (*rating scales*) e numerose tecniche non sistematiche devono essere classificate come introspettive [...].

Quando il problema e gli obiettivi sono posti bene e una tecnica di osservazione introspettiva sembra *a prima vista* più adeguata, alcune ricerche raccomandano tuttavia un ricorso alle tecniche allospettive piuttosto che introspettive nei seguenti casi:

- è difficile per un bambino "introspezzarsi" e restare attento al compito presentato (Gellert, 1955)¹³³;
- il coinvolgimento e l'interesse per un'attività rende difficile ad un soggetto raccontare ciò che egli fa (Weick, 1968)¹³⁴;
- il coinvolgimento in una situazione non permette al soggetto di essere consapevole delle dimensioni dell'avvenimento, così come appaiono a un osservatore distaccato (Richardson, Dohrenwend e Klein, 1965)¹³⁵;
- la persona impegnata in una situazione perde di vista le azioni o gli attributi troppo familiari (id.);
- alcune tecniche introspettive evitano difficilmente le risposte stereotipate o culturalizzate (id., ciò si collega anche al problema del campo normativo);
- alcune persone sono incapaci di comunicare le loro esperienze (Riesman e Watson, 1964)¹³⁶;
- o, molto semplicemente, le persone sono poco motivate, soprattutto in alcuni campi (id.).

GLI STRUMENTI D'OSSERVAZIONE

[...] ci sembra interessante distinguere [cinque] tecniche principali. Le prime tre sono spesso raggruppate sotto il termine di "griglia d'osservazione": (1) i sistemi di categorie; (2) gli inventari o sistemi di segni; (3) le scale di valutazione. Si tratta di tecniche *d'osservazione sistematica* dove gli attributi da osservare sono prestabiliti. Le ultime [due] sono tecniche narrative, chiaramente più aperte e meno predeterminate. Noi le raggrupperemo sotto il termine di *osservazione esperienziale*, perché esse si applicano all'indagine condotta a partire dall'esperienza di una situazione vissuta in un arco di tempo da se stessi o da altre persone, questo spiega come il modello di indagine sia di predominanza clinica [...].

¹³³ Gellert, E. (1955), *Systematic observation: a method in child study*, Harvard Educational Review, 25 [n.d.c.].

¹³⁴ Weick, E. (1968), *Systematic observational methods*, in Lindsey, G., Gardner, Arowson, E. (1968), *The handbook of social psychology*, New York, Addison-Wesley, 54 (vol. 2) [n.d.c.].

¹³⁵ Richardson, S. A., Dohrenwend, B. S., Klein, D. (1965), *Interviewing: its forms and function*, New York, Basic Books.

¹³⁶ Riesman, D., Watson, J. (1964), *The sociability project: a chronicle of frustration and achievement*, in Hammond P. E. (a cura di) (1964), *Sociologist at work*, New York, Harcourt, Brace & World [n.d.c.].

I SISTEMI DI CATEGORIE

Frequentemente utilizzata, la distinzione tra sistemi di categorie e sistemi di segni ha dato adito a numerose discussioni [...]. Così, l'item "il professore pone una domanda" sarà a volte una categoria (come nella griglia di Flanders), altre volte un segno (come negli altri numerosi strumenti d'osservazione).

Dal nostro punto di vista, il criterio di maggiore distinzione risiede nel fatto che, in un sistema di categorie, ogni avvenimento osservato deve essere codificato e classificato, mentre non avviene la stessa cosa in un sistema di segni.

Questo criterio di distinzione permette di spiegare che i sistemi di categorie comprendono in generale molti meno item di osservazione (per lo più una dozzina) dei sistemi di segni: il compito dell'osservatore sarebbe veramente troppo difficile se egli dovesse classificare tutti gli avvenimenti osservati in una lista troppo vasta di categorie.

Questo spiega dunque che un sistema di categorie è un insieme finito di categorie mutualmente *escludenti* (Medley e Mitzel, 1963, p. 298; Postic, 1977, pp. 49-50): ogni avvenimento può e deve essere classificato in una categoria e uno solo alla volta [...].

Ciò spiega dunque che l'uso dei sistemi di categorie comprende delle unità d'osservazione "*molecolari*", cioè relativamente strette e corte [...].

Facciamo notare che allo stesso modo un sistema di categorie non implica necessariamente una codifica *unidimensionale*. Si è perfino raccomandato di utilizzare una codifica *multidimensionale*. Così, ad esempio, un sistema di categorie potrebbe comprendere più dimensioni: le categorie dei locutori, le categorie dei destinatari, le categorie contenuto della comunicazione [...]

Facciamo notare infine che, in un sistema di categorie, i risultati codificati sono abitualmente trattati in termini di numero di unità che sono state classificate in una determinata categoria [...].

GLI INVENTARI O SISTEMI DI SEGNI

Come abbiamo anticipato precedentemente, gli inventari o sistemi di segni sono liste di azioni o di avvenimenti specifici che possono apparire o meno durante un determinato periodo d'osservazione [...].

Poiché ogni avvenimento osservato non deve necessariamente essere codificato, un osservatore potrà prendere in considerazione un numero più grande di item d'osservazione [...]. Se gli inventari o sistemi di segni richiedono uno scarso allenamento da parte degli osservatori e se la loro affidabilità è spesso migliore, al contrario il problema maggiore risiede nella *polisemia comportamentale* degli avvenimenti osservati, cosa che pone a volte dei problemi di validità (in termini di legami con le variabili che si desidera osservare).

Con un sistema di segni, possono essere utilizzate numerose modalità di codifica dei risultati, a seconda dei casi: (a) annotare un'azione una sola volta se compare durante il periodo di tempo; (b) annotare la frequenza di apparizione di questa azione durante il periodo; (c) registrare per mezzo di un cronografo la somma delle durate di apparizione di questo atto durante il periodo. Si comprende facilmente che la scelta di una procedura non è senza ripercussioni sul significato della misura.

Mentre nei sistemi di categorie si parla di codifica uni- o multidimensionale, si parlerà qui piuttosto di *organizzazione uni- o multidimensionale* delle azioni o dei segni da osservare. Così, ad esempio, un inventario di segni potrebbe essere organizzato secondo le dimensioni cognitiva, motivazionale e relazionale [...]. Un'organizzazione multidimensionale dei segni è auspicabile non solo per facilitare la memorizzazione di una lista più o meno lunga da parte degli osservatori, ma anche per mettere meglio in evidenza il quadro concettuale che sottende lo strumento (cosa che offre una certa garanzia di validità).

LE SCALE DI VALUTAZIONE

Le scale di valutazione ([...] in inglese *rating scales*) designano una tecnica che comprende un insieme prestabilito di categorie o di segni per ognuna o ognuno dei quali è richiesto un giudizio equilibrato. Questi giudizi si esprimono sia con enunciati descrittivi, sia con numeri, sia con una forma grafica, sia con una combinazione di queste modalità. Un buon esempio di combinazione delle tre forme è fornito da Léon, Cambon, Lumbroso e Winnykamen¹³⁷ a proposito dell'osservazione delle attività dei bambini durante la ricreazione:

¹³⁷ Si troverà questo esempio nel loro *Manuel de psychopédagogie expérimentale*, Paris, PUF, 1977, 168-169.

-3	-2	-1	+1	+2	+3
Inoperoso	Solitario	Osservatore	Gioco parallelo	Gioco associativo	Gioco cooperativo

Servendosi di una scala appropriata (di frequenza, d'intensità ...), ciascun sistema di categorie o inventario può essere utilizzato secondo la tecnica delle scale di valutazione, a condizione di ricorrere a delle unità di osservazione "molari" (dopo una sequenza che forma un tutto, al termine di una classe...).

L'utilizzazione delle scale di valutazione riveste un certo numero di vantaggi non trascurabili: carico mentale meno elevato per l'osservatore, rapidità di annotazione, possibilità di ricorrere ad un numero molto elevato di categorie, trattamenti quantitativi più numerosi o più sofisticati, correlazioni statistiche con le variabili dipendenti generalmente significative (vedi Medley e Mitzel, 1963)...

Come hanno sottolineato molti autori, la loro apparente facilità non deve farci dimenticare che esse sono soggette a numerosi errori sistematici¹³⁸ [...]:

- *l'errore da abbreviazione*: l'osservatore ricostruisce il "messaggio osservato" e-
liminando i dettagli;
- *l'errore per perdita della parte "centrale" del messaggio*: soprattutto se l'unità
d'osservazione è lunga, l'osservatore esprime una valutazione a partire dai
comportamenti osservati all'inizio e alla fine del periodo di tempo;
- *l'errore di chiusura*: quando il messaggio è poco chiaro, l'osservatore ha la
tendenza a "chiuderlo" per renderlo significativo in una determinata direzio-
ne;
- *l'errore di simmetria*: quando il messaggio è poco chiaro e la valutazione deve
essere fornita più volte successivamente, alcuni osservatori hanno la tenden-
za ad annotare le loro valutazioni una volta in un senso, una volta in un al-
tro;
- *l'errore di tendenza centrale*: quando i gradi estremi della scala di valutazione
sono stati mal operazionalizzati o quando il messaggio è poco chiaro, un'al-
tra tendenza frequente degli osservatori consiste nel ricorrere ai gradi centra-
li della scala;
- *l'errore per incremento di contrasto*: esso consiste nel trascurare alcuni compor-
tamenti e nell'accentuarne altri;
- *l'errore per assimilazione alle prime osservazioni*: se l'osservatore deve valutare a
più riprese una stessa categoria, può avere la tendenza a lasciarsi influenzare
dalla sua prima valutazione;
- *l'effetto alone* [...];
- *l'errore per assimilazione alla misurazione attesa* [...];

¹³⁸ Si trova un elenco dettagliato di questi errori sistematici in: De Ketele, J. M. (1977), *L'observation au service de la gestion du processus éducatif*, Louvain, Thèse inédite.

- *l'errore legato alla posizione dell'osservatore*: a seconda se la valutazione è emessa in modo introspettivo o allospettivo, si può assistere alle distorsioni messe in evidenza dalle teorie dell'attribuzione [...].

LA TECNICA DEGLI EPISODI CRITICI

La tecnica degli episodi critici [...] consiste nell'annotare, il più rapidamente possibile, dopo che hanno avuto luogo degli episodi critici, cioè le parole significative che sono state pronunciate o le azioni rilevanti che si sono svolte nelle situazioni concrete e ciò facendo molta attenzione a riportare soltanto fatti esatti e non opinioni o giudizi. Solo la scelta dell'episodio quanto al suo carattere significativo può essere connotato dalla "soggettività".

[...] risulta che:

1. il valore della tecnica aumenta con l'aumento di osservazioni oggettive successive, la cui varietà ed evidente continuità abbozzeranno un quadro delle strutture di comportamento dell'alunno (o del soggetto osservato), della sua evoluzione, dei suoi interessi, delle sue attitudini, delle sue forze e debolezze.
2. all'inizio si è raccomandato di impiegare la tecnica solo su alcuni soggetti e di estenderne poi l'uso man mano che si è acquisita maggiore esperienza.
3. la tecnica deve essere utilizzata soprattutto nella situazione in cui il bambino (il soggetto) ha un'ampia possibilità di scelta nell'espressione delle sue relazioni personali e sociali [...].
4. l'annotazione deve essere molto semplice, ma deve includere: data dell'avvenimento; enunciazione della situazione nella quale l'avvenimento ha avuto luogo; descrizione dei fatti che costituiscono l'avvenimento; resoconto di ogni informazione relativa all'episodio.
5. l'impiego della tecnica è da evitare nelle situazioni troppo complesse (interazioni multiple poco distinguibili).

Gli errori più comuni da evitare sono i seguenti:

1. valutazione del comportamento piuttosto che descrizione dell'episodio specifico.
2. registrazione solo dei dati negativi.
3. generalizzazione, prima che un numero sufficiente di dati siano raccolti.
4. mancanza di controllo dei dati eccedenti alla luce eventuale di dati ulteriormente significativi.

Rileviamo che l'utilizzazione della tecnica degli episodi critici può essere una tappa preliminare interessante per l'elaborazione di una griglia d'osservazione sistematica [...].

LA TECNICA DI REGISTRAZIONE DI UNA PARTE RAPPRESENTATIVA DI VITA

La tecnica di registrazione di una parte rappresentativa di vita presenta numerosi vantaggi. Heyns e Lippit (1954)¹³⁹ sottolineano che per la sua applicazione non occorre che gli osservatori siano dotati di attitudini speciali. Wright (1960)¹⁴⁰ riconosce che essa fornisce dati teoricamente neutri, molto dettagliati, isomorfi con il comportamento reale che permettono una quantificazione ulteriore [...]. A questi vantaggi noi aggiungeremo che la tecnica è interessante negli studi etologici o ecologici, nella fase induttiva della ricerca, per rilevare le variabili pertinenti e nella fase di controllo, mediante un sistema di categorie prestabilite, ma non conosciute dagli osservatori.

Ma, accanto a questi vantaggi, Weick (1968)¹⁴¹ rileva alcuni inconvenienti non trascurabili: una certa difficoltà di utilizzazione nell'ambito di interazioni complesse e veloci; in alcuni casi, la difficoltà di trovare il linguaggio adatto per fornire una descrizione sufficientemente ricca; l'eterogeneità del linguaggio (funzione dell'esperienza anteriore) utilizzato da osservatori diversi per descrivere uno stesso fenomeno; la validità a livello dell'esattezza secondo gli osservatori. Se tali inconvenienti possono essere limitati mediante il ricorso ad un co-osservatore o ad un allenamento adeguato dobbiamo invece sottolineare il "costo" di una tale tecnica, tanto più che il lavoro fondamentale di ricerca del significato rimane da elaborare ... su un materiale imponente [...].

IL GRADO D'INFERENZA DELL'OSSERVAZIONE

[...] A seconda degli strumenti utilizzati e delle condizioni di utilizzazione, l'osservazione comporterà una parte più o meno ampia di inferenza.

In un'osservazione con *inferenza debole*, l'osservatore enuncia scrupolosamente ciò che vede o ciò che sente senza preoccuparsi del significato che questi fatti rivestono.

¹³⁹ Heynes, R., Lippit, R., *Systematic observational techniques*, in Lindsey, G., Gardner, Arowson, E. (1968), *The handbook of social psychology*, New York, Addison-Wesley, 1954, 1968 (vol. 2) [n.d.c.].

¹⁴⁰ Wright, H. F. (1960), *Observational child study*, in Mussen, P. H. (a cura di) (1960), *Handbook of research methods in child development*, New York, Wiley [n. d. c.].

¹⁴¹ Weick, E. (1968), *Systematic observational methods*, cit. [n.d.c.].

In un'osservazione con *inferenza forte*, l'osservatore enuncia sia intenzioni, sia motivi, sia sentimenti ..., in una parola attribuisce un significato a ciò che osserva. Questo processo di inferenza si fonda sugli effetti del comportamento, le relazioni tra gli antecedenti ed i conseguenti ... e sulle esperienze personali dell'osservatore.

I TIPI DI ANNOTAZIONE [...]

L'ANNOTAZIONE IMMEDIATA E L'ANNOTAZIONE DIFFERITA

L'annotazione è *immediata* quando è fatta subito dopo l'osservazione del comportamento o dell'oggetto. Al contrario, l'annotazione è *differita* quando un tempo più o meno lungo separa l'osservazione dall'annotazione [...] *più l'annotazione è differita, più può entrare in azione il processo d'inferenza* [...].

Molti dei sistemi di osservazione basati sulle annotazioni differite [...] sono [...] sottomessi ai seguenti rischi:

1. dettagli importanti possono essere dimenticati (Wrightstone, 1960)¹⁴², soprattutto i fenomeni transitori (Weick, 1968)¹⁴³;
2. processi inferenziali possono stabilirsi e arrivare a confondere osservazione e impressione.

L'OSSERVAZIONE DIRETTA E L'OSSERVAZIONE MEDIATA

[...] I vantaggi dell'osservazione mediata, normalmente riconosciuti, sono i seguenti:

1. è particolarmente indicata quando l'osservazione diretta si rivela impossibile o difficile (ad esempio: carattere troppo transitorio degli oggetti d'osservazione).
2. riduce l'esigenza d'attenzione dell'osservatore.
3. permette l'immagazzinamento dell'informazione grezza e dunque:
 - la codifica ulteriore;
 - il ritorno indietro;
 - il controllo della codifica;
 - nuovi trattamenti dell'osservazione;
 - la codifica multipla, ecc.

I modi più comuni per rendere immediata l'osservazione sono la registrazione sonora, la trascrizione scritta, la foto e il film (o video) [...].

¹⁴² Wrightstone, J. M. (1960), *Observational techniques*, in Harris, C. W. (a cura di) (1960³), *Encyclopedia of educational research*, New York, Mac Millian [n.d.c.].

¹⁴³ Weick, E. (1968), *Systematic observational methods*, cit. [n.d.c.].

LA SITUAZIONE DI OSSERVAZIONE

[...] la situazione studiata può essere naturale o fittizia, manipolata o libera da ogni tipo di manipolazione.

La situazione è *naturale* quando i soggetti (osservati nell'ambito di un'indagine con o senza il controllo di un'ipotesi; con o senza piano sperimentale stretto ...) si trovano in una situazione di vita "abituale" o "familiare".

La situazione è *fittizia* quando i soggetti studiati si trovano in una situazione che esce dal quadro "abituale" o "familiare".

Questo si verifica se [il ricercatore] sottopone un soggetto ad un'esperienza in laboratorio o se gli somministra un test [...].

In effetti, contrariamente ad un pregiudizio corrente, la situazione naturale può essere manipolata. Definire cos'è una situazione *manipolata* ci sembra tuttavia difficile se non si distinguono due tipi di alterazioni:

1. la manipolazione attraverso la *scelta intenzionale delle situazioni* da studiare [...];
2. la manipolazione nel *corso del processo di ricerca*, dove il ricercatore interviene sulla situazione o sul soggetto per modificarne il significato [...];

Queste distinzioni ci permettono di tracciare un ventaglio di situazioni diverse delle quali *a priori* non possiamo dire che l'una è migliore dell'altra: il valore di ogni situazione dipende dagli obiettivi perseguiti e da tutti i tipi di contingenze [...].

IL GRADO DI LIBERTÀ LASCIATO ALL'OSSERVATORE

A seconda del grado di libertà lasciato all'osservatore, l'osservazione sarà più o meno *sistematica*. L'osservazione tende ad essere sistematica quando soddisfa le seguenti condizioni: (1) impiego di procedure coerenti e ripetibili; (2) determinazione delle condizioni dell'osservazione (standardizzazione); (3) impiego di tecniche rigorose di osservazione, di annotazione e di codifica [...].

IL TRATTAMENTO DELLE OSSERVAZIONI

[...] Il trattamento si basa su delle unità d'osservazione più o meno *molecolari* o *molari*. Esse saranno tanto più molecolari quanto più condurranno su questioni incidentali o categorie molto specifiche, decontestualizzate, non collegate agli antecedenti, concomitanti e conseguenti... Esse saranno tanto più molari quanto più formeranno direttamente un'unità di significato e quanto più saranno collocate nello spazio e nel tempo.

Il trattamento può poggiare su meccanismi di *quantificazione* o al contrario su un processo di ricerca *qualitativa* del significato [...].

Il trattamento può essere parzialmente o interamente - *predeterminato* (cioè stabilito passo dopo passo durante il processo di raccolta dei dati) o *postdeterminato* (cioè determinato dopo la raccolta dei dati).

Noi vorremmo insistere di nuovo sul fatto che non esiste *un* modo scientifico di elaborare i risultati, ma modalità più o meno adeguate a seconda dei momenti, degli oggetti e delle funzioni alle quali si mira [...].

Tavola 3.2 - Tipologia delle modalità d'osservazione

PARAMETRI	DISTINZIONI
Funzioni dell'osservazione	1. L'osservazione può essere: <i>descrittiva, formativa. valutativa, euristica, di verifica</i>
Autore dell'osservazione	L'osservatore può essere: 2. <i>indipendente vs partecipante</i> (passivo o attivo) 3. <i>percepito vs non percepito</i>
L'oggetto dell'osservazione	L'osservazione può: 4. <i>vertere sui fatti vs le rappresentazioni</i> 5. <i>essere attributiva vs narrativa</i> 6. <i>essere allospettiva vs introspettiva</i>
Gli strumenti d'osservazione	7. L'osservazione può basarsi: a) <i>sulle tecniche d'osservazione sistematica:</i> - <i>sistemi di categorie</i> - <i>gli inventari o sistemi di segni</i> - <i>le scale di valutazione</i> b) <i>sulle tecniche d'osservazione esperienziale:</i> - <i>i diari</i> - <i>la tecnica degli episodi critici</i> - <i>le tecniche di registrazione di una parte rappresentativa di vita</i> - <i>il diario di bordo</i>
Il grado d'inferenza della osservazione	8. Il livello d'inferenza può essere: <i>debole o forte</i>
L'annotazione	L'annotazione dell'osservazione può essere: 9. <i>immediata vs differita</i> 10. <i>diretta vs mediata</i>
La situazione d'osservazione	La situazione d'osservazione può essere: 11. <i>creata vs naturale</i> 12. <i>manipolata vs non manipolata</i>
Il grado di libertà lasciato all'osservatore	13. L'osservazione può essere: <i>sistematica vs non sistematica</i>
Il trattamento dell'osservazione	Il trattamento dell'osservazione può poggiare su: 14. <i>unità molecolari vs molari</i> 15. <i>procedure quantitative vs processi qualitativi</i> 16. <i>analisi sequenziali vs non sequenziali</i> 17. <i>processi predeterminati o determinati nel corso dell'osservazione o post-determinati</i>

13. La costruzione di un "profilo" (Boncori, G., *Profilo*, in *Enciclopedia pedagogica*)¹⁴⁴

Prospetto di un soggetto che ne delinea le caratteristiche essenziali in base a rilevazioni oggettive, osservazioni sistematiche, valutazioni, in vista di interventi educativi.

ETIM.: da «filo», linea sottile o estrema, la parte più piccola del vestito (F); anche da *findo* (taglio); e *pro* (gr.), "davanti". La radice semantica mette in luce l'idea di tracciare e mettere in evidenza, in primo piano, le linee essenziali.

Trad.: ingl. *Profile*; fr. *profil*; ted. *profil*; sp. *perfil*

L'uso dei p., o giudizi, in ambito educativo prevede molti campi di utilizzazione e può presentarsi in forme diverse. Spesso vengono usati a conclusione di un'osservazione o di una valutazione, per sintetizzare l'essenziale di una serie di rilevazioni, ad esempio al termine di un'interrogazione, di un esame, di un test d'ingresso, di un colloquio, di un corso scolastico (ad es. nella scuola media).

Possono essere espressi in forma verbale, con la possibilità di usare simboli e rappresentazioni grafiche di vario tipo. La loro importanza educativa è collegata alla funzione di sintesi, al termine di un processo di osservazione [...], indagine, valutazione che ne costituisce il presupposto metodologico: i p. possono quindi far riferimento a dati valutativi in funzione di interventi formativi, oppure possono assumere una prospettiva più "sommativa", come nel caso di esami d'idoneità a classi superiori. In ogni caso, le procedure specifiche con cui vengono elaborati sono condizionate dalla validità e dall'attendibilità dei processi che essi esprimono: ad esempio, un p. orientativo, basato sulla somministrazione di test, fonda la sua *validità* non solo sulla metodologia docimologica per la costruzione dei p., ma anche su una metodologia psicologica e psicometrica che assicuri la scelta dei test adatti alla situazione, la corretta somministrazione, correzione, ecc.

L'elaborazione di un p. prevede alcuni momenti essenziali che riguardano la scelta dei criteri e dei segni indicatori, la predisposizione dei livelli, la scelta del tipo di espressione in funzione della comunicazione.

Il *criterio*, come caratteristica che determina il confronto con altre, è ciò che guida la sintesi dei dati e l'elaborazione del p.: nelle situazioni educative è spesso rappresentato dagli obiettivi che ci si era prefissi. In una situazione scolastica gli obiettivi-criterio possono riguardare apprendimenti prerequisiti comuni a diverse discipline, come ad esempio la capacità di lettura, la comprensione elementare della lingua

¹⁴⁴ Tratto da: Boncori, G. (1992), *Profilo*, in *Enciclopedia pedagogica*, diretta da M. Laeng, Brescia, La Scuola, coll. 9446-9450.

[...], oppure possono essere diversi a seconda delle materie; in questo caso i criteri si definiscono, in modo più analitico, con riferimento agli obiettivi ed ai contenuti specifici: ad esempio, "avere un lessico appropriato" è un obiettivo che include la capacità di usare parole in modo congruo alla realtà di cui si parla. Il contenuto può essere riferito a numeri e simboli matematici, o a parole, segni grafici e punteggiatura di tipo letterario. Nel primo caso il criterio della valutazione e del p. è costituito dall'obiettivo "proprietà lessicale" + contenuto matematico, e potrà essere usato in profili relativi a questa materia; nel secondo caso, il p. ha un criterio e un uso connesso con discipline di tipo umanistico (italiano, lingue straniere, ecc.). In ogni caso, è importante definire, ed eventualmente concordare (nel caso di più operatori, docenti) i criteri su cui fondarsi.

Il p., analogamente al processo valutativo e osservativo, si esprime riguardo al raggiungimento di criteri (obiettivi e contenuti) o al possesso di capacità (intellettuali, personali): tale pronunciamento si basa su comportamenti che sono ritenuti segni validi e attendibili che danno quindi un'indicazione qualitativa. Perché il p. sia corretto è necessario scegliere ed eventualmente concordare (nel caso di p. costruiti collegialmente, ad es. da un consiglio di classe) quali segni vanno presi come indicatori, in una situazione concreta, tenendo conto di variabili evolutive, differenziali, di una determinata programmazione educativa o scolastica. Ad esempio, un p. relativo alla capacità critica si fonderà su segni diversi nel caso di un bambino di scuola elementare o di un adolescente: nel primo caso gli indicatori potranno privilegiare segni più "concreti", nel secondo più "logico-formali" (G. Boncori, 1989).

I *livelli* danno un'indicazione quantitativa sull'intensità delle caratteristiche presenti e sono espressi generalmente attraverso una scala. Il punto di riferimento, necessario per dare significato alle affermazioni del p., può essere ancorato alla media degli individui, a valori massimi o minimi, a seconda delle scelte che si ritiene opportuno fare, decidendo il criterio di normalità o di accettabilità sociale al quale far riferimento. Un uso scolastico è talora riferito a *valori medi* (tendenza centrale): la "normalità" e rappresentata, in questo caso, da circa i 2/3 degli alunni di quella classe, con 1/3 dei soggetti "fuori della media" o per eccesso (i più bravi) o per difetto (i meno bravi). L'applicazione corretta ("giusta") di questo criterio di "normalità", che accentua la dimensione di competizione nell'ambito del gruppo (questa distribuzione evidenzia infatti le misure che si discostano dalla media, in alto e in basso, e permette meno distinzioni nei punteggi intermedi), presuppone una sostanziale omogeneità nelle situazioni iniziali e negli interventi educativi.

Il punto di riferimento può anche essere costituito da un *livello minimo*: in questo caso, la "normalità" è in relazione a condizioni preliminari minime per svolgere un certo compito (almeno in un'ottica educativa). Ad esempio, nel quadro della teoria piagetiana, un bambino di 5 anni non ha, generalmente, un livello minimo di maturità mentale per ragionare in modo "logico-formale", capire un sillogismo, ecc.; per capire un testo è necessario un livello minimo di conoscenza lessicale. Le prove di

accertamento possono quindi essere impiegate per impartire determinati insegnamenti solo quando ci siano le condizioni minime.

Nel caso invece di un riferimento ad un "*livello massimo*", la "normalità" è costituita dal completo raggiungimento dell'obiettivo da parte della stragrande maggioranza (80-90%; modello del mastery learning - distribuzione a J). Ad esempio, certi aspetti fondamentali di tipo linguistico (correttezza ortografica), matematico (calcolo elementare) e altri (cultura generale, ecc.), possono giustificare l'uso di questo modello. I livelli possono anche avere come punto di riferimento un'accettabilità sociale, per cui il possesso o la maturazione di caratteristiche personali e intellettuali (cognitive e non cognitive), atteggiamenti e comportamenti è tale da non ostacolare la socializzazione, il profitto scolastico, le possibilità lavorative. Molti obiettivi sulla "partecipazione alla vita della scuola" e al "livello di maturazione", tipici della scuola elementare e media, sono del tipo suddetto. I livelli possono riferirsi, specialmente nelle prospettive di valutazione formativa, ai diversi *tipi di intervento* necessari per raggiungere gli obiettivi: ordinari o individualizzati e, in quest'ultimo caso, differenziati per la fascia superiore e inferiore.

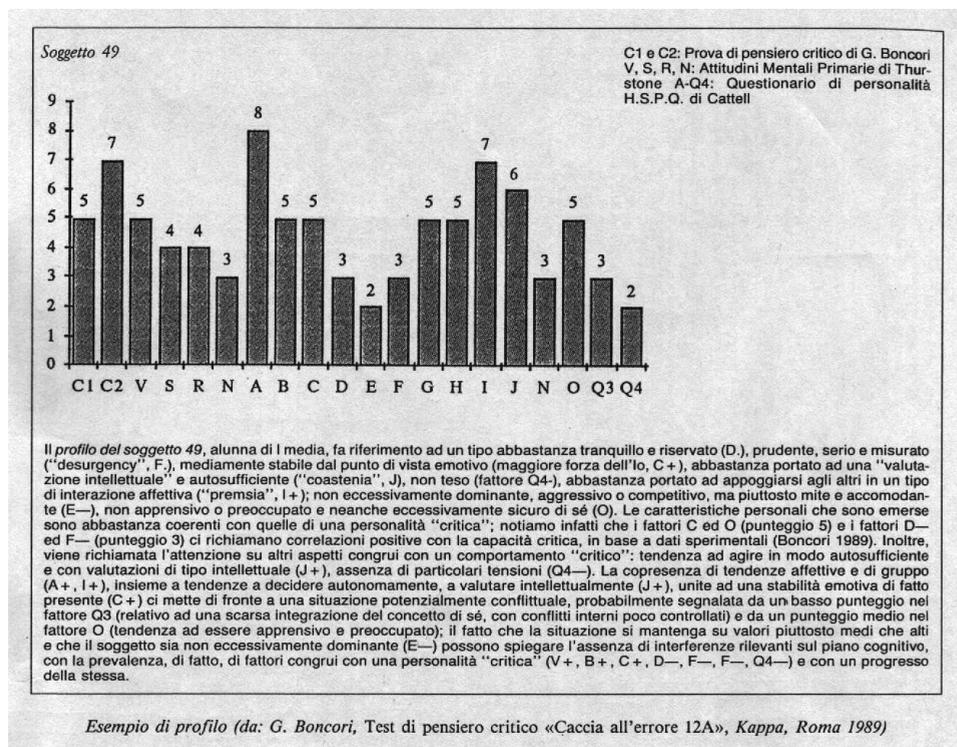
La precisione metrologica dei livelli dipende anche dal numero e dall'ampiezza dei gradini: questi sintetizzano e accorpano con regolarità diversi punti di un continuum (più preciso ma di uso più frequente nella ricerca). Il numero ottimale, con cui si ottiene il massimo dell'informazione, è situato generalmente tra i 5 e i 9 gradini, con una distribuzione simmetrica (uguale numero di livelli al di sopra e al di sotto della media) e unimodale (che denota un significato costante della variabile per una data popolazione in esame). Anche l'ampiezza dei livelli va definita con esattezza in funzione della distribuzione (numerosità, forma, errore standard di misura) e usata in modo costante.

Infine, i livelli (e i punteggi) presenti in un profilo devono essere standardizzati nello stesso tipo di scala, rispettando le norme statistiche relative: ad esempio, è possibile esprimere un profilo in termini di "gradi centili" (che esprimono un'indicazione sulla percentuale della popolazione al di sopra e al di sotto) nell'ambito di una scala ordinale, in cui è possibile fare una graduatoria dei soggetti; questo tipo di livelli non si possono alternare, nello stesso p. con altri tipi di scale, per es. punti standard [...].

Il p. dà le informazioni essenziali sia sui singoli criteri, sia sulla loro intercorrelazione, con una sintesi che dice qualcosa di più delle singole componenti prese separatamente. Ad esempio, un p. che si esprime sulla presenza di capacità matematiche, grafiche e spaziali al di sopra della media, al termine della scuola secondaria, può fornire elementi di orientamento verso scelte universitarie di tipo scientifico piuttosto che umanistico.

L'espressione del p. deve essere chiara e in funzione della situazione concreta: ad esempio, i p. alla fine della scuola media sono "giudizi di media", sintesi della maturazione e dell'apprendimento valutata da più docenti (rappresentativo, quindi, e non troppo generico da annullare l'informazione), con prospettive orientative e co-

municata in modo significativo (ad es. con termini non eccessivamente "tecnici") ad alunni e famiglie.



I p. possono essere espressi in vari modi, utilizzando forme più o meno schematiche, discorsive, simboliche, grafiche. Gli *aggettivi*, rispondenti ai criteri e ai livelli predisposti, devono essere scelti in base ad una corretta metodologia docimologica, con significato costante e validazioni sperimentali; ad esempio, se si usa l'aggettivo "sufficiente", come gradino medio in una scala di 3 livelli, con riferimento alla distribuzione normale, questa parola si riferirà ai due terzi circa dei soggetti; gli altri (1/3) saranno equidistribuiti tra i migliori e i peggiori. Gli aggettivi, e le scale che ne derivano, possono esprimere solo il livello (ad es. sufficiente, buono, insufficiente, ecc.) o sia il criterio che il livello (ad es. molto originale, poco originale, ecc.). Anche in questi casi le parole vanno usate in modo univoco e sistematico, con validazioni sperimentali che ne assicurino la validità, l'attendibilità, la possibilità di comunicazione in modo univoco e significativo.

I *simboli possono essere di vario tipo* (lettere, numeri cardinali, ordinali, ecc.) come pure le *rappresentazioni grafiche* (istogrammi, poligoni di frequenza, ecc.). Ogni simbolo utilizzato va spiegato nel suo significato concreto. È anche possibile elaborare p. con diverse modalità espressive (discorsiva, grafica, ecc.). Ad esempio, dispen-

do di valutazioni oggettive riguardanti la capacità critica, l'intelligenza e la personalità, è possibile elaborare un p. in cui siano espressi i punteggi standardizzati ottenuti nelle diverse prove, una rappresentazione grafica con istogrammi e una verbalizzazione discorsiva che sintetizzi i diversi modi e stili cognitivi in cui si può manifestare la "capacità critica".

BIBL.: AA.VV., *Pupils in Profile*, SCRE Publication 67, London 1977; G. DOMENICI, *Descrittori dell'apprendimento*, Giunti & Lisciani, Teramo 1981-82; L. CALONGHI, *Valutare*, De Agostini, Novara 1983; H.D. BLACK - W.B. DOCKRELL, *Criterion Referenced Assessment in the Classroom*, The Scottish Council for Research in Education, Edinburgh 1984; B. VERTECCHI, *Manuale della valutazione*, Editori Riuniti, Roma 1985; B.R. WORTHEN - J.R. SANDERS, *Educational Evaluation*, Longman, New York and London 1987; G. BONCORI, *Test di pensiero critico «Caccia all'errore 12»*, Kappa, Roma 1989; M. GATTULLO - M.L. GIOVANNINI, *Misurare e valutare l'apprendimento nella scuola media*, B. Mondadori, Milano 1989; L. BONCORI, *Teoria e tecniche dei test*, Boringhieri, Torino 1990.

14. Strumenti di misurazione e prove oggettive: definizioni e tipologie (Melchiori, R., Adas. Il laboratorio della valutazione)¹⁴⁵

Stimoli, osservazioni e prestazioni

L'effettuazione del controllo del processo formativo, o di una sua parte, comporta l'utilizzo di una grande varietà di strumenti per l'accertamento del raggiungimento degli obiettivi scolastici o generali oppure specifici - operativi relativi a una disciplina. Uno stesso strumento difficilmente è utilizzato per controllare o l'acquisizione di conoscenze o la capacità di risolvere problemi o la dimostrazione di abilità all'inizio o durante oppure al termine di un processo di istruzione, o di una sua parte significativa.

Un *controllo* corrisponde quindi alla determinazione di una situazione artificiale nella quale sulla base di una *prova* o *stimolo*¹⁴⁶ si manifesta e può essere osservata la *prestazione* o il *comportamento* di uno studente¹⁴⁷.

La prova è utilizzata per sollecitare lo studente affinché riveli e manifesti, tramite o il comportamento o la prestazione, ciò che ha appreso o sa fare; il comportamento o la prestazione, quindi, sono apprezzati, tramite l'uso di strumenti o prove, sulla base di criteri di valutazione derivati e congruenti con gli obiettivi operativi stabiliti per controllare il livello di padronanza raggiunto in una particolare disciplina o area disciplinare o tematica oppure argomento. Nella tavola 2 è riportata una classificazione degli stimoli e delle risposte da utilizzare per le prove di controllo della situazione scolastica degli studenti [...].

Le prove oggettive

Nell'ambito della valutazione realizzata, o all'interno di una classe da parte di un insegnante o attuata all'interno di una istituzione scolastica, relativamente al controllo e alla valutazione delle prestazioni degli studenti¹⁴⁸, le *prove oggettive* sono degli strumenti di misura in grado di restringere la soggettività della correzione e quindi migliorare la valutazione dei risultati. Le prove oggettive scolastiche, che si

¹⁴⁵ Tratto da: Melchiori, R. (2002), *Adas. Il laboratorio della valutazione. 1. Aspetti concettuali*, Milano, Franco Angeli, pp. 25-27 e 30-32.

¹⁴⁶ Gli stimoli sono le sollecitazioni che si forniscono allo studente sotto forma di prova: si possono classificare gli stimoli, con le relative risposte e quindi le prove di controllo.

¹⁴⁷ Il comportamento è identificabile con le azioni compiute da uno studente; la prestazione è il risultato o prodotto di un comportamento. Quest'ultima può essere misurata e valutata, dopo che è stata rilevata, anche in assenza dello studente.

¹⁴⁸ In letteratura si utilizzano anche i seguenti termini: prove di profitto, *achievement test* o *objective test*, per controllare l'apprendimento degli studenti; *assessment* per le attività di valutazione dei risultati dell'attività d'istruzione o di formazione.

riferiscono essenzialmente al profitto, possono essere validamente utilizzate: per *stimolare* l'interesse dello studente; per *controllare* il grado di preparazione dello studente prima di iniziare un corso scolastico; per *misurare* il progresso fatto durante il processo d'istruzione; per *determinare* o la necessità o l'efficacia del lavoro di recupero; per *consentire* qualunque altra forma d'accertamento che l'insegnante o il consiglio di classe o il comitato di valutazione della scuola vuole attuare. In tutti questi casi lo scopo e il taglio della prova oggettiva sono diversi e debbono riflettere le specifiche necessità ed esigenze.

Una *prova oggettiva* di profitto, quindi, è costituita da un insieme, non prestabilito, di *elementi di prova*¹⁴⁹ o *quesiti* espressi in modo da provocare una risposta rigidamente predeterminata – esatta, errata, astensione -. Essa presenta, come strumento, vantaggi metrologici con alto livello di *rappresentatività* del curriculum svolto, cioè in un arco di tempo relativamente breve si possono presentare più quesiti di una qualunque altra prova, e di omogeneità, cioè le stesse domande sono poste nello stesso momento a tutti gli studenti¹⁵⁰.

Per ottenere i benefici desiderati, la prova oggettiva deve essere ben costruita; utilizzare una prova oggettiva che fornisca indicazioni errate sullo studente, o sulla classe, è peggio che non usarla affatto. Nessun insegnante utilizzerebbe nella propria classe, o nella scuola, una prova oggettiva che contenga istruzioni errate o elementi di prova o quesiti mal posti; tuttavia, molte prove oggettive, proposte da insegnanti, presentano punti deboli che pregiudicano seriamente la validità e l'utilità dei punteggi rilevati.

La lunghezza della prova oggettiva e il momento scelto per somministrarla possono essere facilmente adattate alle varie situazioni; la prova oggettiva, indipendentemente dall'ambito di uso – o classe o sezioni/indirizzi, oppure istituto scolastico - può essere di rilevante aiuto per accompagnare lo sviluppo del processo d'istruzione degli studenti e per misurare l'*efficacia* e gli *effetti*, o impatto interno complessivo, dell'istruzione impartita.

Le prove oggettive sono state classificate nel tempo a seconda del *fuoco* su cui si punta l'attenzione e dello *scopo* che ci si propone¹⁵¹.

¹⁴⁹ Per *elemento di prova* si intende un costrutto semplice o composito a cui deve essere associata una o più risposte. Il costrutto di un elemento di prova semplice comprende una consegna - cioè il testo della domanda e le indicazioni riguardo la risposta - e alcune risposte (una o più a secondo del tipo di quesito). Il costrutto composito, invece, comprende una consegna e in successione alcuni quesiti, che riguardano il testo della consegna, la cui struttura dipende dal tipo di elemento composito. Sono da considerare elementi di prova semplici il tipo a scelta multipla (semplice) e il tipo vero/falso; i tipi compositi sono, invece, a corrispondenza, a completamento, a comprensione della lettura [...].

¹⁵⁰ Uno strumento di misura è buono se è *oggettivo*, cioè se esiste concordanza di giudizio tra valutatori competenti, *valido*, cioè se misura effettivamente ciò per cui è stato costruito, *attendibile*, cioè se fornisce lo stesso risultato in condizioni o periodi diversi [...].

¹⁵¹ Per una maggiore esemplificazione vedi: Mauro Laeng (1995), *Enciclopedia dell'educazione*, Torino, UTET.

RISPOSTA	APERTA	CHIUSA
STIMOLO		
APERTO	<p>Lo stimolo consiste nel fornire indicazioni di orientamento su aree problematiche.</p> <p>La risposta prevede capacità di raccolta di informazioni, di ri-analisi espresse in forme argomentative, di proiezioni in aree limitrofe.</p> <p>Esempi: interrogazioni ampie, temi, relazioni su esperienze, redazioni di articoli.</p>	<p>Lo stimolo è ampio, ma non prevede una prestazione dello studente.</p> <p>La risposta non rileva un comportamento o una prestazione di abilità o conoscenze.</p> <p>Esempi: formulazioni di argomenti dove lo studente deve esprimere solo consenso, interrogazioni dove lo studente ascolta ed esprime una risposta non verbale ma gestuale o mimica.</p>
CHIUSO	<p>Lo stimolo è strutturato e predispone la prestazione richiesta.</p> <p>La risposta non è predeterminata ma uno studente è in grado costruire una risposta organizzata e articolata, e più omogenea rispetto a quella di altri.</p> <p>Esempi: composizioni e saggi brevi, attività di ricerca, relazioni su esperienze di laboratorio.</p>	<p>Lo stimolo contiene completamente il modello della risposta.</p> <p>La risposta corrisponde a una prestazione già organizzata.</p> <p>Esempi: esercizi di grammatica, risoluzione di problemi, prove oggettive o prove oggettive di profitto</p>

Tavola 2. Classificazione degli stimoli e delle risposte e relative prove di controllo¹⁵².

In questo volume si propongono tre classificazioni delle prove oggettive che riguardano il *procedimento* utilizzato per la loro costruzione, la *posizione* temporale utilizzata per il loro uso e, infine, la *lettura* dei risultati. Nel primo caso le prove oggettive possono essere considerate *informali*, come quelle di classe, oppure *standardizzate*. Nel secondo caso le prove oggettive possono essere *iniziali* – o di prerequisiti –, *formative* o *sommative*; infine, in rapporto alla lettura dei risultati, possono essere di *rango* o *competenza*.

Le tre classificazioni presentano caratterizzazioni diverse, ma una singola prova oggettiva può anche essere considerata contemporaneamente informale, iniziale e di rango; infatti si può realizzare una prova oggettiva di matematica costruita per una classe di terza media su obiettivi generali della seconda media in modo da accertare il livello di padronanza o di soglia iniziale delle conoscenze in funzione della predisposizione di un lavoro didattico di compensazione – diagnosi -.

¹⁵² Adattamento di schema tratto da B. Vertecchi (1984), *Manuale della valutazione. Analisi degli apprendimenti*, cit., t. 6, p. 107.

In generale le particolarità delle tre classificazioni sono:

- a) le prove oggettive realizzate da un insegnante per una classe di studenti, anche in collaborazione con altri docenti o dello stesso consiglio di classe o dello stessa materia o dello stesso indirizzo oppure della stessa scuola, sono da considerare *prove oggettive informali* in quanto non sono previste, né richieste effettivamente, attività di standardizzazione e i risultati sono utilizzati soltanto all'interno della classe o delle classi oppure della scuola. Dove la costruzione delle prove oggettive prevede azioni formalizzate di scelta degli obiettivi specifici, la formulazione dei quesiti, la sperimentazione della prova oggettiva su un campione ristretto di studenti, l'analisi dei quesiti con revisione della prova oggettiva, la somministrazione su di un campione più ampio, l'analisi dei risultati con definizione dei punteggi standardizzati e delle norme di applicazione e l'uso dei punteggi calcolati come riferimento per altre somministrazioni, allora la prova oggettiva si può considerare *standardizzata* [...]. Le prove informali sono molto utili in quanto si adattano meglio alle singole classi per le quali vengono appositamente preparate, possono essere adoperate spesso e sono molto utili per accertare abilità e lacune dei singoli studenti. Le prove standardizzate, invece, sono utili quando occorre confrontare su un piano territoriale ampio - distrettuale, provinciale, regionale e nazionale - campioni significativi di studenti oppure quando si vogliono confrontare i risultati degli studenti di una classe con valori standard per verificare la posizione della stessa classe. Queste prove oggettive possono essere usate poche volte durante un anno scolastico;
- b) le prove oggettive, informali o standardizzate, classificate rispetto alla loro utilizzazione temporale possono essere di tipo: *iniziale* - o di prerequisiti -, *formative* e *sommative*. Le prime sono preparate e utilizzate per controllare se gli studenti possiedono le conoscenze necessarie, ovvero hanno raggiunto gli obiettivi specifici, per affrontare un nuovo corso d'istruzione. Nel caso in cui l'obiettivo della prova oggettiva sia quello di constatare le lacune degli studenti circa gli argomenti o la parte svolta di una disciplina allora la prova si può considerare *diagnostica*. Generalmente queste prove sono composte da quesiti che mirano ad accertare il possesso delle abilità e delle conoscenze indispensabili ai fini degli apprendimenti futuri. Nello svolgimento di una procedura d'istruzione, relativa ad argomenti di una disciplina, si possono utilizzare prove oggettive di tipo *formativo* che permettono di rilevare le eventuali difficoltà che gli studenti possono avere incontrato. Per questo motivo sono simili a quelle diagnostiche e generalmente intercambiabili con esse. Le prove diagnostiche e quelle formative comprendono quesiti posti in ordine di difficoltà crescente in modo da permettere di

rilevare esattamente abilità e lacune. Infine sono *sommative* le prove oggettive che permettono di rilevare la preparazione complessiva raggiunta al termine di un segmento significativo del percorso d'istruzione. Le prove oggettive sommative sono composte da quesiti che riguardano solo gli aspetti più generali e importanti delle discipline studiate e che quindi consentono di ottenere una visione d'insieme della preparazione degli studenti.

- c) Le prove oggettive classificate rispetto ai risultati possono essere o di *rango* o di *competenza*. Sono prove oggettive di *rango* o *posizione* quelle che permettono di realizzare una gerarchia della classe sulla base dei punteggi ottenuti nella prova stessa in modo da evidenziare il rapporto esistente tra la prestazione di ogni studente e quella del resto della classe. Rispetto alla *competenza* le prove oggettive permettono di stabilire se sono stati conseguiti gli obiettivi scolastici programmati dal docente. In generale, comunque, la differenza tra le due forme non è così netta: in certi casi una prova di competenza può essere utilizzata anche per accertare la posizione di ogni singolo studente e viceversa. Le prove di competenza sono composte da quesiti che saggiavano gli obiettivi scolastici operativi, mentre le prove di rango contengono quesiti che saggiavano gli obiettivi specifici più importanti di una disciplina.

Le prove oggettive sono, in generale, da considerare insieme alle funzioni della valutazione e da definire nell'ambito della pianificazione e programmazione del curriculum da svolgere.

15. Un quadro di sintesi degli strumenti di verifica scolastica (Benvenuto, G., *Mettere i voti a scuola. Introduzione alla docimologia*)¹⁵³

Gli strumenti che possono essere costruiti ed utilizzati per le diverse esigenze valutative sono molteplici. Per approfondire le principali pratiche e riflessioni che i docenti hanno sviluppato in questi ultimi anni sulla valutazione degli apprendimenti e dell'insegnamento utilizziamo alcuni dati emersi dall'indagine Iard sulle condizioni di vita e di lavoro nella scuola italiana (Cavalli, 2000)¹⁵⁴. Da essa traiamo alcune annotazioni relative alle forme e criteri più usati per la valutazione in Italia¹⁵⁵.

Gli strumenti di verifica più utilizzati continuano ad essere quelli più tradizionali (v. tabella 3.1): le prove scritte sono presenti nell'82% delle elementari nel 74% delle superiori, le interrogazioni non programmate nel 51% delle elementari ad nel 62% delle superiori.

I test di profitto costruiti dagli insegnanti (39-48%) sembrano prevalere su quelli suggeriti dai libri di testo (29-39%) e sicuramente su quelli forniti da istituti specializzati (3-7%). Probabilmente questi ultimi sono ancora poco conosciuti, diffusi o reperibili, ma anche la buona percentuale relativa ai test costruiti dagli insegnanti lascia molti dubbi interpretativi. Come avanza Cesare Scurati nel commento a questa tabella, "occorrerebbe stabilire se non siano forse stati confusi con gli usuali compiti in classe". Ad una domanda successiva, sui problemi incontrati nella valutazione, gli insegnanti segnalano tra i principali proprio "la formulazione di criteri oggettivi (69,2% alle elementari, 55,3% alle medie, 55,0 alle superiori), il trovare tempo per fare verifiche congrue per ogni alunno (47,4% alle elementari, 51,0% alle medie, 64,9 alle superiori), ed il condensare impressioni e convincimenti in un giudizio o in un voto (56,6% alle elementari, 37,2% alle medie, 31,5 alle superiori)".

Un altro elemento generale che si rileva dai dati è la diminuzione graduale tra la varietà e quantità di forme regolarmente usate nel ciclo elementare e quelle alle superiori. Scompaiono le *relazioni orali concordate* (dal 35% al 20%), le *relazioni scritte di gruppo* (dal 27% all'11%), i *saggi brevi* (dal 43% al 30%) e soprattutto le *annotazioni sul comportamento sociale* (dal 59% al 22%).

¹⁵³ Tratto da: Benvenuto, G. (2003), *Mettere i voti a scuola. Introduzione alla docimologia*, Carocci, Roma, pp.101-105.

¹⁵⁴ Cavalli, A. (a cura di) (2000), *Gli insegnanti nella scuola che cambia. Seconda indagine IARD sulle condizioni di vita e di lavoro nella scuola italiana*, Bologna, il Mulino [n.d.c.].

¹⁵⁵ Le risposte sono state fornite dai docenti dei diversi ordini scolastici tra il mese di dicembre 1998 e il mese di marzo 1999.

	Elementare		Media		Superiore	
	Uso regolare	Non uso	Uso regolare	Non uso	Uso regolare	Non uso
Interrogazioni programmate	34,6	28,0	22,9	22,2	25,9	32,7
Interrogazioni non programmate	51,1	9,5	57,7	8,2	62,3	11,4
Relazioni orali concordate	35,6	22,6	34,6	18,6	19,9	31,3
Relazioni scritte individuali	46,0	16,4	41,4	12,5	32,9	22,0
Relazioni scritte di gruppo	27,5	19,4	16,7	27,3	10,6	42,0
Prove scritte tradizionali	81,7	2,8	75,0	10,1	74,0	13,7
Saggi brevi	42,9	12,9	39,6	16,8	29,8	27,2
Test di profitto costruiti dagli insegnanti	41,5	19,0	39,3	20,0	48,1	15,8
Test di profitto suggeriti dai libri di testo	33,9	15,9	39,4	13,8	29,1	25,2
Test di profitto forniti da istituti specializzati	7,1	56,2	3,4	68,0	5,2	66,2
Annotazioni sul comportamento sociale	59,0	7,4	59,9	11,1	22,5	47,7
<i>Base (n. docenti)</i>	965		1448		2393	

Tabella 3.1: Uso di strumenti e procedure di valutazioni (%). (Fonte: Cavalli, A. (a cura di) (2000), *Gli insegnanti nella scuola che cambia. Seconda indagine IARD sulle condizioni di vita e di lavoro nella scuola italiana*, Bologna, il Mulino, p. 188)

Per riepilogare la varietà di strumenti di verifica, o prove, utili nei diversi momenti e funzioni della valutazione, possiamo far riferimento al livello di strutturazione che essi possono assumere (cfr. Vertecchi, 1984; Domenici, 1993; Melchiori, 2002)¹⁵⁶. Per livello di strutturazione si può intendere da un lato il livello di stimolo che le prove presentano, aperto o chiuso, e dall'altro il tipo di risposta che essi richiedono, che può essere anch'esso di tipo chiuso o aperto.

Nella figura 3.12 si esemplificano alcune delle principali forme in uso nei contesti scolastici rispetto al grado di apertura e chiusura degli stimoli e delle risposte. Mentre nella figura 3.13 le diverse forme di verifica sono presentate incrociando i livelli di strutturazione delle prove con le diverse finalità e i tempi della valutazione. In questa seconda schematizzazione le stesse tipologie di prove sono ripetute in quanto possono essere utilizzate per diverse funzioni: i quesiti oggettivi, ad esempio, possono servire a conoscere la situazione di partenza, ma anche certificare gli apprendimenti al termine di una procedura didattica; così come i quesiti a risposta

¹⁵⁶ Vertecchi, B. (1984), *Manuale della valutazione. Analisi degli apprendimenti*, cit.; Domenici, G. (1993), *Manuale della valutazione scolastica*, cit.; Melchiori, R. (2002), *ADAS. Il laboratorio della valutazione*, cit. [n.d.c.].

aperta possono essere utilizzati per tutte le funzioni della valutazione in quanto permettono un ampio spettro di rilevazione.

Nei due schemi si propongono, ovviamente, solo alcune tra le forme di verifica scolastica, lasciando al lettore il compito di integrare con altre forme. Il tentativo di costruire un quadro di sintesi va comunque nella direzione di definire maggiormente le caratteristiche peculiari dei diversi strumenti di verifica e di indicare l'uso preferenziale che ognuno di essi può avere.

	Stimolo aperto	Stimolo chiuso
Risposta aperta	interrogazioni dialogiche scritture libere temi tradizionali riflessioni/commenti protocolli descrittivi	interrogazioni strutturate (colloquio strutturato) prove di scrittura con istruzioni prove pratiche ricerche su temi relazioni (di laboratorio) quesiti a risposta aperta
Risposta chiusa	false domande aperte	quesiti oggettivi (Scelta multipla, Vero/falso, completamento, corrispondenza), soluzione problemi cloze test

Figura 3.12: tipologia di prove rispetto al grado di apertura/chiusura degli stimoli e delle risposte

[...]

Due sono i principi che orientano nella scelta degli strumenti di valutazione e relativamente alla loro funzionalità docimologia:

- non esiste lo strumento di verifica migliore, ma solo quello più adatto allo scopo per il quale è pensato. E il compito del docente è allora quello di prendere decisioni sullo strumento che meglio si adatta, coerentemente agli obiettivi di verifica individuati e ai limiti imposti dal contesto in cui opera;
- l'affidabilità dei diversi strumenti di rilevazione si ottiene controllando le diverse fasi di costruzione, utilizzazione, ma anche disponendo di un accordo nelle scale di misura a cui essi rimandano.

[Riflettere su questi due principi consente] di affrontare più onestamente un dilemma che spesso assale il docente o chi si trova a svolgere verifiche dell'apprendimento: adottare forme strutturate, che aiutano nella rilevazione tendenzialmente oggettiva o forme semi-strutturate, o aperte, che offrono la possibilità di articolare e approfondire le risposte?

Le prove strutturate, offrendo sia uno stimolo chiuso sia modalità di risposta predefinite, offrono maggiori garanzie di affidabilità nell'attribuzione di punteggi e

voti, ma spesso sono di difficile costruzione o non permettono di indagare specifiche abilità. Stiamo parlando fondamentalmente dei quesiti oggettivi, che possono presentarsi sotto forme di quesiti a scelta multipla, vero/falso, corrispondenze e completamenti.

Le prove semi-strutturate sono quelle che offrono uno stimolo o delle istruzioni di tipo chiuso e autonoma elaborazione delle risposte. La chiusura dello stimolo presenta quei vincoli che permettono di presentare a tutti la stessa domanda o situazione problematica, l'apertura della risposta permette invece di esibire i diversi gradi di strutturazione (comprensione e studio) e di originalità che gli studenti possono avere relativamente ai diversi livelli di apprendimento.

Le prove libere, o aperte, ancor più di quelle semi-strutturate sono particolarmente adatte per alcune fasi formative della didattica, e per un approccio dialogico e maggiormente interattivo nella valutazione, ma di difficile utilizzazione in sede di misurazione collegiale.

Forma degli strumenti di verifica dell'apprendimento scolastico				
Funzioni della valutazione	Strutturata, oggettiva <i>Stimolo (domande) chiuso e risposte predeterminate (chiuso)</i>	Semi-strutturata <i>Stimolo (istruzioni) chiuso e autonoma elaborazione delle risposte</i>	Aperta o non strutturata <i>Stimolo (domande/istruzioni) aperto e autonoma elaborazione delle risposte</i>	Tempi della valutazione
CONOSCITIVA - esplorativa → - diagnostica → - ri-orientativa →	- osservazione sistematica - quesiti oggettivi (V/F, corrispondenze, scelta multipla, complementi) - cloze-test - test socio-psicologico	- quesiti a risposta aperta - elaborati scritti (riflessioni, descrizioni, narrazioni)	- protocolli descrittivi - colloquio/discussione	Iniziale (pre-requisiti, ingresso)
REGOLATIVA formativa →	- quesiti oggettivi - soluzione di problemi	- interrogazione strutturata - prova di scrittura (saggi, riassunti, relazioni, ecc.) - quesiti a risposta aperta - esercitazioni pratiche	- interrogazione dialogica - temi tradizionali - riflessioni/commenti	Intermedia
SOMMATIVA - orientativa / prognostica → - certificativa / selettiva →	- quesiti oggettivi - soluzione di problemi - test socio-psicologico	- colloqui strutturati - prove pratiche - quesiti a risposta aperta	- interrogazione - colloquio pluridisciplinare	Finale
Raccolta delle diverse rilevazioni	<i>Fascicolo studente / port-folio</i>			Continua/diacronica

Figura 3.13: strumenti, funzioni e tempi nella valutazione degli apprendimenti scolastici

16. La costruzione delle prove oggettive (Domenici, G., *Manuale della valutazione scolastica*)¹⁵⁷

[...]

La sequenza delle operazioni [...] da compiersi [per l'elaborazione e la messa a punto delle prove oggettive] è la seguente.

1. Determinazione degli ambiti disciplinari oggetto di verifica, delle caratteristiche degli allievi cui si somministrerà lo strumento - ovvero livello di difficoltà complessiva della prova -, e delle specifiche funzioni valutative che si vogliono attuare.
2. Determinazione dell'estensione del segmento curricolare o dell'area delle competenze da esaminarsi.
3. Definizione, analisi ed eventuale campionamento degli obiettivi specifici della verifica.
4. Determinazione preventiva della lunghezza e/o durata del test.
5. Scelta della tipologia più opportuna degli item da utilizzare.
6. Elaborazione dei quesiti e definizione della durata della prova.
7. Determinazione dei punteggi grezzi da assegnarsi a ogni tipo di quesiti a seconda che si registrino risposte esatte, sbagliate o omesse.

[...] Volendo perciò chiudere il circolo delle operazioni da svolgere dal momento della decisione di costruire una prova fino a quello della lettura dei dati ricavati dalla sua somministrazione, ai precedenti sette punti occorrerebbe aggiungere anche i seguenti.

8. Somministrazione del test.
9. Correzione e attribuzione dei punteggi.
10. Analisi e interpretazione dei risultati e analisi della qualità dello strumento impiegato (*item analysis*).

[...] Quei dieci punti, come si sarà notato, rappresentano peraltro, e con buona approssimazione, le fasi che comunque si dovrebbero seguire per la messa a punto di una qualsiasi prova di verifica dell'apprendimento, persino di quelle tradizionali – beninteso quando si senta la necessità di costruirne di accettabili nonostante i loro limiti intrinseci - oltre che di quelle semistrutturate. Certo, cambiano gli oggetti specifici su cui si opera e le modalità d'intervento. Tuttavia, dato che, come abbiamo più volte ribadito e torneremo a insistere, si devono poter utilizzare tutti gli strumenti di verifica reputati volta a volta accettabili per assolvere al meglio le plurime

¹⁵⁷ Tratto da: Domenici, G. (2001²), *Manuale della valutazione scolastica*, Roma-Bari, Laterza, pp. 88-98.

e articolate funzioni valutative, analizzare quelle che dovrebbero essere comuni è cosa assai utile.

In questo capitolo prenderemo in esame solo quelle sottese dai primi sette punti, cioè quelle che si riferiscono strettamente alla costruzione delle prove oggettive [...].

3.1. *Determinazione degli ambiti disciplinari, delle caratteristiche degli allievi - o del livello complessivo di difficoltà della prova - e delle funzioni valutative (punti 1.-2.)*

Quando si decida di elaborare una prova oggettiva, ovviamente i primi e più immediati punti di riferimento che vengono in mente sono l'ambito o gli ambiti disciplinari per i quali si vuol compiere la verifica (prove disciplinari, multi o interdisciplinari, prove di comprensione della lettura, logico-linguistiche, ecc.); le caratteristiche degli allievi (classe frequentata, livelli generali o specifici di competenze possedute o supposte in quello o in quegli ambiti disciplinari, ecc.) che contribuiscono a definire il livello complessivo di difficoltà che la prova deve presentare o in relazione al quale si struttureranno i quesiti, e le funzioni specifiche o specialistiche della valutazione (prove diagnostiche per la verifica dei prerequisiti, prove di verifica formativa, di verifica sommativa, ecc.) [...].

Una volta stabiliti quei tre punti di riferimento (risulta utile fissarli preliminarmente su un foglio di carta secondo il seguente esempio: *Scuola media... Prova d'ingresso di matematica. Classi prime*), sarà allora necessario determinare l'estensione del segmento curricolare o dell'area delle competenze in rapporto ai quali si compirà la verifica. Se si vorrà mettere a punto uno strumento di verifica formativa l'area di riferimento dovrà necessariamente essere breve e circoscritta, più o meno corrispondente al lavoro di dieci-quindici giorni scolastici. Se la prova dovrà viceversa verificare al termine di un anno scolastico le più importanti competenze, il possesso di reticoli di conoscenze, ecc., relativi a una disciplina di studio, allora sarà l'intero curriculum annuale ad assumere il valore di riferimento, e così via [...].

3.2 *Definizione, analisi e campionamento degli obiettivi. Determinazione della lunghezza e/o durata della prova (punti 3.-4.)*

Una volta definita l'area disciplinare specifica in relazione alla quale si dovranno accertare le conoscenze possedute dagli allievi, si potrà finalmente procedere alla determinazione degli obiettivi analitici della verifica. Si potrà, in altri termini, pervenire a un'accorta identificazione di tutte quelle abilità reputate dal o dai docenti esperti disciplinari che costruiscono la prova, come davvero rappresentative della padronanza delle questioni, dei saperi relativi all'argomento o all'area tematica connessi al segmento curricolare in rapporto al quale si dovrà compiere la verifica [...].

Per determinare gli obiettivi da sottoporre a verifica, occorrerà quindi ripercorrere a ritroso o smontare, magari secondo criteri logico-sequenziali, l'ambito disciplinare - o pluri, multi, inter transdisciplinare - oggetto del controllo, avendo però sempre come punto di riferimento le caratteristiche psicologico-strutturali, in particolare cognitive, degli allievi cui la prova verrà somministrata e il corrispondente livello di difficoltà.

Per la determinazione delle categorie o delle classi di diversa complessità degli obiettivi cognitivi della verifica, risultano di grande aiuto le ormai numerose tassonomie disponibili, anche se è pericoloso considerarle, come purtroppo assai spesso accade, totalmente vincolanti e perciò esaustive di tutta la possibile gamma delle tipologie di abilità da sottoporre a controllo prima, durante e dopo un certo periodo di formazione e per qualsiasi ambito o sub-ambito disciplinare.

Più correttamente vanno invece interpretate come utilissimi schemi di riferimento capaci di agevolare, senza banalizzare, l'operazione di identificazione delle abilità da sottoporre a controllo e persino, come vedremo, la stessa stesura dei quesiti. Come si è già detto, da questi punti di vista la già citata tassonomia semplificata di Bloom si presta in modo del tutto particolare per la messa a punto delle prove di accertamento delle conoscenze.

Così, in una griglia analoga a quella qui di seguito riportata, strutturata da tante colonne quante sono le classi di obiettivi della tassonomia semplificata, si possono trascrivere sinteticamente le corrispondenti abilità-obiettivo via via che vengano identificate attraverso l'analisi del particolare segmento curricolare preso in esame.

Il numero tendenziale degli obiettivi indicati in ciascuna colonna decresce progressivamente passando dalla prima all'ultima, perché generalmente, ma non necessariamente, i «termini» che hanno significato specifico e che gli allievi devono saper usare con pertinenza, riferiti alle ipotetiche unità di apprendimento o all'intero curricolo annuale, risultano numericamente superiori ai «fatti» che bisogna conoscere, e questi alle «regole», ai «principi», alle «leggi», che occorre pariteticamente padroneggiare per quella stessa sezione curricolare.

Bisogna inoltre ribadire che i primi tre livelli tassonomici indicati definiscono abilità che assai spesso sono solo, o prevalentemente, di tipo riproduttivo, analogico, cioè conoscenze che manifestano soprattutto l'operare del e col «pensiero convergente».

Se quindi nel processo di verifica ci si limitasse, come spesso accade, a rilevare esclusivamente abilità sottese a quei primi tre livelli, si coglierebbero solo conoscenze di tipo meccanico, riproduttivo. Non anche quelle conoscenze in grado di attivare un più articolato e significativo processo di apprendimento, di produrre modificazioni significative della mappa cognitiva dell'allievo, del suo modo di operare [...].

A questo punto bisogna quindi procedere all'analisi degli obiettivi così classificati, per escludere quelli ridondanti e per integrare l'elenco con quelli casualmente omessi, in modo che la prova verifichi efficacemente tutte le abilità, le conoscenze, le competenze reputate importanti.

Tab. 5.1. *Uno schema di riferimento per la determinazione degli obiettivi della verifica*

Conoscenza dei termini	Conoscenza dei fatti	Conoscenza di regole e principi	Capacità di effettuare trasformazioni e adattamenti	Capacità di effettuare applicazioni
A1.....	B1	C1	D1	E1
A2	B2	C2	D2	E2
A3.....	B3.....	C3.....	D3.....	E3.....
.....
.....	En
.....	Cn	Dn
.....	Bn
An

La procedura di analisi è utile anche per definire a quale livello di complessità si vuole verificare la padronanza di certe abilità. Peraltro, obiettivi classificati nella categoria «conoscenza di termini» specifici possono rimandare a concetti, regole, fatti (e viceversa), e perciò la stessa dizione di obiettivo può risultare presente in due o più colonne o livelli tassonomici. Si dovrà quindi stabilire se si vogliono verificare solo conoscenze relative alla competenza lessicale oppure dell'apparato concettuale cui l'eventuale termine rimanda (magari perché il livello tassonomico più elevato include assai spesso la padronanza delle abilità corrispondenti ai livelli che lo precedono), ovvero se si vogliono rilevare le conoscenze distinte cui entrambi fanno riferimento, e così via [...].

Compiuta l'operazione, potrebbe accadere di trovarsi di fronte un numero così elevato di obiettivi da verificare, che la prova che ne deriverebbe sarebbe troppo ampia. In tali casi le vie che si possono seguire sono essenzialmente due. La prima consiste nel mettere a punto due o più prove da somministrarsi in tempi diversi; l'altra suggerisce di operare una vera e propria campionatura rappresentativa di tutte le abilità-obiettivo determinate per quella specifica sezione curricolare.

In questa, come in molte altre decisioni da assumersi, è utile distinguere [...] tra le differenti funzioni della verifica [...].

[N]el caso di prove a verifica sommativa di classe risultando infatti piuttosto esteso il segmento curricolare corrispondente, bisognerà fare in modo che si possano cogliere le abilità più rappresentative di esso, che per ovvie ragioni saranno abbastanza aggregate.

Per verifiche diagnostiche e formative si dovranno invece cogliere analiticamente tutte quelle abilità capaci di dar conto delle caratteristiche conoscitive dell'allievo o del processo di formazione: il campionamento non è necessario proprio perché bisogna poter conoscere persino la più minuscola unità di apprendimento non strutturata al fine di attivare opportuni interventi di recupero. Le prove devono pertanto essere analitiche ed estese, e richiederanno l'utilizzo di tutti gli obiettivi e persino dei sotto-obiettivi identificati.

Per la determinazione dell'ampiezza o estensione dei test di verifica sommativa o finale si può procedere in modi diversi. Quello semplice fa riferimento alla quantità di tempo utilizzato per la trattazione delle questioni tematiche e dei contenuti relativi a ciascun obiettivo o a gruppi omogenei di essi. L'ipotesi di fondo consiste nel dare maggior peso alla verifica di quelle conoscenze che per essere acquisite hanno comportato il maggior impegno relativo sia nei docenti che negli allievi. A un più forte impegno, misurato soprattutto attraverso la variabile tempo, si presume che corrispondano maggiori complessità delle questioni.

Anche se, come si può intuire, l'asserzione non è affatto sempre fondata, rappresenta tuttavia un criterio utile per la determinazione del peso relativo da assegnarsi alla verifica di questo o quell'obiettivo. Criterio che andrebbe però sempre e opportunamente corretto non solo in forza del peso specifico che determinati gruppi di obiettivi assumono nell'economia formativa di quel particolare ambito conoscitivo oggetto di verifica, ma anche, e soprattutto, nella verifica diagnostico-formativa, sulla base degli esiti di uno *screening*, di una sorta di setacciamento più o meno empirico delle difficoltà modali, degli errori più diffusi se non sistematici che gli allievi di quella particolare scuola, di quel particolare contesto socio-culturale solitamente incontrano e/o compiono.

Se allora, con le correzioni appena suggerite, si considerano i tempi impiegati nella trattazione delle questioni, degli argomenti, delle parti di curriculum corrispondenti ai diversi obiettivi, espressi in ore (o in altre unità di misura del tempo), si può determinare il grado percentuale di importanza da assegnarsi a quelle parti e ovviamente alle omologhe abilità-obiettivo della verifica, attraverso il calcolo suggerito dalla seguente formula:

$$Ip = \frac{Tp}{Tc} \times 100$$

dove:

Ip = Importanza assegnata a ciascuna delle parti curriculari

Tp = Tempo dedicato a ciascuna parte

Tc = Tempo complessivo dedicato a tutte le parti curriculari in esame.

Bisogna sempre tener presente che per garantire misure *valide* cioè capaci di rispecchiare le reali competenze degli allievi, occorre, tra l'altro, che la prova non duri troppo. Infatti, se l'impegno chiesto con la somministrazione va troppo oltre una certa soglia, finisce col rilevare anche, in taluni casi soprattutto, il livello di affaticabilità dei singoli, la capacità e la durata di concentrazione, che rappresenteranno pure elementi importanti della formazione, ma da rilevarsi semmai in apposite circostanze, non certo quando si desidera avere informazioni affidabili su quanto e come hanno appreso gli allievi.

Per tali ragioni e visto che la curva dell'attenzione cade bruscamente dopo una ventina di minuti (per gli adulti), che dopo due o tre riprese il livello della stessa attenzione intensiva diventa piuttosto basso, è buona norma non far durare oltre i cinquanta-sessanta minuti l'impegno richiesto in una specifica prestazione [...]. Dato che da molte indagini empiriche (svolte con allievi che vanno dalla quinta elementare alla maturità) è emerso che anche con item a scelta multipla abbastanza impegnativi il tempo per rispondere a ciascuno di essi risulta essere mediamente di uno o al massimo di due minuti, un test sommativo non dovrebbe contenere più di 40-50 item.

Tornando al problema della determinazione del peso da assegnarsi a ciascuna questione, o parte del curriculum, la sua soluzione presuppone che queste vengano circoscritte e definite prima o dopo l'individuazione degli obiettivi, secondo l'esempio indicato nella tab. 5.2, nel quale, pur non essendo state descritte le parti, si suppone che siano sette e che a ciascuna di esse sia stato dedicato il tempo parziale indicato accanto per facilitare il calcolo dell'importanza percentuale.

Se si suppone, ancora, che il test da elaborarsi sia costituito da item a scelta multipla, si potrà finalmente calcolare il numero di quegli item che andranno a verificare le abilità corrispondenti a ognuna delle parti considerate.

La scelta, invece, dei livelli tassonomici in rapporto ai quali si struttureranno i quesiti potrà compiersi con una opzione empirica, buon senso, oppure applicando la stessa formula alla griglia degli obiettivi che è stata elaborata. Si saprà così quante domande dovranno formularsi per rilevare, in rapporto a ciascuna parte del curriculum, la conoscenza dei termini, dei fatti, delle regole e principi e le capacità sia di effettuare trasformazioni e adattamenti, sia di compiere applicazioni.

Tab. 5.2. Schema per la determinazione del peso da assegnarsi alle questioni oggetto di verifica

Parti e/o questioni e/o argomenti	Tempo parziale (in ore)	Importanza percentuale
1).....	2	10
2).....	1	5
3).....	3	15
4).....	4	20
5).....	3	15
6).....	2	10
7).....	5	25
<i>T_c</i>	20	100

17. **Formulare i quesiti (esempi e alcune regole per la costruzione)**(Benvenuto, G., Lopriore, L., *Costruzione e somministrazione degli strumenti*)¹⁵⁸

Prima di affrontare la costruzione dei differenti tipi di quesiti è indispensabile sottolineare alcune regole generali, valide a prescindere dal tipo di quesiti che si costruiranno. Queste avvertenze, e relative giustificazioni (v. tabella 4.3) servono a ricordare che l'uso delle prove oggettive non deve suscitare ambiguità, altrimenti vengono a mancare i presupposti di validità, attendibilità e oggettività degli strumenti di misurazione costruiti [...].

Nel presentare adesso le differenti regole per la formulazione dei quesiti le suddivideremo rispetto a 4 grandi categorie:

- a. **posizione e soluzione del problema** (domanda e specificità della richiesta);
- b. **formulazione linguistica** (uso della lingua);
- c. **organizzazione dei quesiti** (ordine e presentazione dei quesiti);
- d. **convenzione grafica** (aspetti tipografici e di struttura della prova).

Tabella 4.3

Indicazioni generali per la costruzione di quesiti oggettivi nelle prove di apprendimento	
Avvertenze	Giustificazioni
1. Il linguaggio (termini e strutture delle proposizioni) non sia inutilmente complicato, ma adeguato ai destinatari.	1. La misurazione degli obiettivi sarebbe accompagnata o filtrata dalla misurazione delle abilità linguistiche riguardanti tali complicazioni.
2. Gli stimoli siano brevi ed essenziali, nella misura del possibile.	2. Si perde meno tempo, si provoca minore stanchezza.
3. Non fare tranelli.	3. Si misurerebbe l'abilità di sfuggire ai tranelli.
4. Non chiedere cose banali o sciocche.	4. Si misurerebbero conoscenze inutili.
5. Non fare domande alle quali si possa rispondere solo in base al buon senso o alla cultura generale.	5. Si misurerebbe il buon senso o la cultura generale.
6. Le risposte sbagliate non siano ingenuie né raffinate, ma adeguate alla preparazione degli scolari.	6. Si misurerebbero acquisizioni estranee agli obiettivi perseguiti. Taluni studenti potrebbero trovarsi in difficoltà.
7. Citare un autore, se si fanno riferimenti a quanto egli abbia detto o scritto.	7. Non si saprebbe altrimenti riconoscere come vere o false le asserzioni riportate.
8. La collocazione delle risposte esatte non deve essere preordinata (per es.: VFVVFVVFV; oppure: VVFFVVFFVVFFVVFFVFF).	8. Qualcuno potrebbe scoprirla.
9. Non utilizzare mai "pezzi" presi tali e quali dai libri usati per lo studio.	9. Qualcuno li potrebbe aver memorizzati, o ritrovarvi più facilmente la risposta esatta.
10. Ciascuna domanda sia indipendente dalle altre.	10. Chi scoprisse i collegamenti potrebbe essere facilitato.

*Fonte: riadattamento della tabella riportata in Gattullo-Giovannini, 1989, p.103-217¹⁵⁹

¹⁵⁸ Tratto da: Benvenuto, G., Lopriore, L. (1999), *Costruzione e somministrazione degli strumenti*, in Pagnoncelli L. (a cura di) (1999), *Formazione e valutazione dell'apprendimento. Metodologie e strumenti*, Roma, Anicia, pp. 216-233.

[...]

4.1.5.1 QUESITI VERO-FALSO

[...]

Posizione del problema		Quesiti mal formulati	
1) Evitare di verificare l'apprendimento mnemonico.		[...]	V F
2) Le affermazioni devono far riferimento ad una sola idea (elemento/concetto).		Il contratto si conclude quando il proponente dà comunicazione dell'accettazione o del deposito.	V F
Formulazione linguistica			
3) Evitare di utilizzare negazioni nelle affermazioni.		Il diritto facoltativo di assentarsi dal lavoro nei primi 3 anni di vita del bambino non spetta solo alla lavoratrice madre.	V F
4) Le affermazioni devono essere espresse in forma chiara e concisa, evitando quindi dettagli secondari.		Nell'attuale ordinamento internazionale, vigente nei singoli Stati aderenti d'ONU, la norma giuridica è obbligatoria.	V F
5) Utilizzare un linguaggio adeguato al livello dei destinatari e delle tematiche da valutare.		(in relazione ai destinatari)	V F
6) Evitare l'uso di quantificatori quali <i>sempre</i> , <i>spesso</i> , <i>qualche volta</i> , <i>mai</i> , etc. in quanto inducono a ritenere più giusta o errata l'affermazione a prescindere dal quesito.		I regolamenti esecutivi e autonomi sono spesso sottoposti a limiti sostanziali e formali.	V F
Organizzazione dei quesiti			
7) Controllare l'indipendenza delle affermazioni: il rispondere ad una non deve facilitare o pregiudicare la risposta ad un'altra.			
8) Il numero delle affermazioni vere e di quelle false (giuste/esatte) dovrebbero bilanciarsi nel totale dei quesiti.			

Tabella 4.4 Regole per la costruzione di quesiti vero-falso

Negli esempi che seguono si affronta la tipologia delle rilevazioni, adoperando quesiti vero-falso.

[...] **Esempio 3: per la rilevazione di utilizzazione di forme di ragionamento complesse (analogie, sillogismi, *problem solving* ecc.)**

¹⁵⁹ Gattullo, M., Giovannini M. L. (1989), *Misurare e valutare l'apprendimento nella scuola media*, Milano, B. Mondatori [n.d.c.].

Secondo la teoria di Darwin l'ambiente seleziona le variazioni genetiche favorevoli.
(Introduzione al vero-falso)

Se questa teoria è vera, allora le giraffe moderne hanno tutte colli lunghi perché solo queste ultime sono sopravvissute. (V) (F)

La lepre delle nevi ha acquisito un mantello bianco in risposta all'avvicinarsi dell'inverno.
(V) (F)

Data la serie 1, 5, 12 il valore successivo a 2,6 è 24. (V) (F)

Esempio 4: vero-falso multiple

È possibile utilizzare altre forme di quesiti vero-falso, che rimandano più da vicino ai quesiti a scelta multipla. Stiamo parlando delle vero-falso multiple, che fanno riferimento ad un unico corpo di domanda o relative ad un unico problema.

Durante il suo viaggio sulla Beagle, Darwin fa colpito da diverse osservazioni (per ognuna segna se vera o falsa)

	Vera	Falsa
a) procedendo verso sud certe specie erano gradualmente sostituite da altre molto affini		
b) non c'era alcuna relazione tra i fossili trovati e le forme viventi [...]		

4.1.5.2 I QUESITI A SCELTA MULTIPLA

Per molti anni i quesiti a scelta multipla (d'ora in poi: S/M) sono stati utilizzati preferenzialmente nella costruzione delle prove oggettive [...].

Quanto più i quesiti a S/M riescono a simulare le possibili risposte ad una domanda ritenuta centrale nella valutazione, tanto più essi rilevano forme di apprendimento complesse e non di carattere mnemonico. Nella comprensione della lettura, di differenti tipi di testi, i quesiti a S/M permettono di sondare diverse forme di ragionamento, a patto che nella loro costruzione si considerino proprio queste differenti forme di ragionamento e non la scelta per esclusione. Ad esemplificare invece la duttilità dei quesiti S/M nella misurazione di differenti casi si vedano gli esempi seguenti (cfr. Gattullo, 1968)¹⁶⁰:

1. L'obiezione che le prove oggettive portano a una facilitazione eccessiva delle risposte e dell'apprendimento scolastico in generale, si basa sulla premessa

¹⁶⁰ Gattullo, M. (1968), *Didattica e docimologia*, cit. [n.d.c.].

che

- a) il loro meccanismo di risoluzione facilita la diffusione delle risposte tra gli allievi.
- b) c'è identità strutturale tra risposta e meccanismo di risposta.
- c) l'analisi statistica delle risposte alle prove oggettive ne conferma realmente la facilità.
- d) i meccanismi di risposta sono realmente di facile apprendimento.

[...] [Il quesito è un esempio] di misurazione della capacità di analisi: [...] rimanda all'analisi di elementi costitutivi e vuole accertare la premessa implicita di un'obiezione o di un pregiudizio nei confronti delle prove oggettive [...].

Esaminiamo di seguito le regole che riguardano i 4 elementi caratterizzanti la costruzione di quesiti a scelta multipla, presentando per ognuna un esempio di formulazione errata di un quesito.

Regole per la costruzione di quesiti a scelta multipla

Posizione e soluzione del problema	Quesiti mal formulati
1) La domanda deve focalizzare un solo problema o concetto, esprimendolo con precisione per evitare incertezze e confusioni nella scelta della risposta (nell'esempio il quesito è mal formulato in quanto il corpo della domanda non focalizza alcun preciso problema, di conseguenza tutti i distrattori potrebbero essere giusti).	Nell'Africa meridionale a) si pratica l'estrazione dei diamanti. b) abbonda la foresta tropicale. c) carovane di cammelli trasportano merci. d) i trasporti moderni sono molto carenti.
2) Usare distrattori plausibili rispetto al problema considerato. Ogni distrattore deve, per contenuto e natura, risultare in qualche modo collegato alle domande. (nell'esempio il quesito è mal formulato in quanto il quarto distrattore introduce un "genere" differente, e la domanda non tende a verificare il significato del termine "estrazione").	L'Africa meridionale ha il primo posto nell'estrazione di a) bauxite. b) ferro. c) carbone. d) alberi.
Formulazione linguistica	Quesiti mal formulati
3) Formulare la domanda in forma semplice ed esplicita. Il corpo della domanda dovrebbe essere composto da una frase in forma interrogativa o affermativa. (nell'esempio il quesito è mal formulato in quanto il corpo della domanda presenta un doppia subordinazione che introduce elementi di complessificazione. In questo caso, per semplificare, basterebbe eliminare gli incisi e alleggerire la domanda delle informazioni aggiuntive non necessarie).	La Rivoluzione industriale, avviatasi nell'Inghilterra alla fine del 1700 e diffusa nei Paesi dell'Europa occidentale, negli Stati Uniti e nel Giappone, comportò l'accelerazione del progresso economico. Con il termine accelerazione si intende a) l'aumento. b) la diminuzione. c) l'integrazione. d) la distribuzione.

[Continua]

[Segue]

4) Nei quesiti relativi alla comprensione della lettura, usare un linguaggio simile a quello del testo a cui ci si riferisce.	(In relazione ai testi e ai destinatari specifici)
5) Formulare i distrattori e la risposta corretta in modo omogeneo. Questi devono essere relativamente simili sia per struttura sintattica sia per lunghezza. <i>(nell'esempio il quesito è mal formulato in quanto le alternative non sono omogenee per lunghezza e livello di informazione. La "a", "c" e "d" sono omogenee tra loro, e riprendono il termine "Paesi" contenuto nel corpo della domanda, mentre la risposta giusta, la "b", oltre ad essere più lunga, presenta un contenuto decisamente più informativo).</i>	In quali Paesi opera prevalentemente il fondo delle Nazioni Unite l'infanzia? a) Nei Paesi dell'Africa. b) Ovunque ci sono bambini bisognosi di cure, cibo e assistenza. c) Nei Paesi europei. d) Nei Paesi dell'Asia.
6) Evitare che le risposte esatte e i distrattori riprendano termini o formulazioni usate nella domanda (o nel testo se quesiti per la comprensione della lettura e quindi si richieda di localizzare un punto preciso del testo). <i>(nell'esempio il quesito è mal formulato in quanto l'alternativa "b" è l'unica a riprendere il termine "istituto" presentato nel corpo della domanda e a candidarsi quale risposta giusta per esclusione).</i>	(v. anche esempio precedente) Quali istituti compiono operazioni a medio e a lungo termine per favorire le imprese industriali e permettere la trasformazione in capitali fissi? a) Le banche di credito commerciale. b) Gli istituti di credito mobiliare. c) Le Casse comunali di credito agrario. d) Le società bancarie di credito fondiario.
7) Evitare, nella costruzione delle alternative, l'utilizzazione di espressioni come <i>sempre, tutti, ogni, dappertutto</i> , che inducono a credere nell'inesattezza della risposta che le contiene. Evitare ugualmente termini come <i>quasi sempre, spesso, qualche volta</i> , che inducono a credere nell'esattezza delle alternative. Unica eccezione a quanto detto è il caso in cui tali espressioni sono usate in scala graduata: <i>_sempre, quasi sempre, talvolta, quasi mai, mai</i> .	Fumare è dannoso al DNA poiché le sostanze catramose a) ossidano a tal punto questa molecola da renderla immobile. b) qualche volta si depositano nei polmoni. c) aumentano le possibilità di errore nella sua ricostruzione. d) trasmettono quasi sempre errate informazioni genetiche alla molecola.
8) Evitare di inserire negazioni semplici o doppie nel corpo della domanda. Se è proprio indispensabile, evidenziare la negazione in neretto, con sottolineatura o usando la lettera maiuscola.	La scuola dell'obbligo non dovrebbe preparare gli allievi all'attività professionale in quanto a) la sua finalità è semplicemente di tipo educativo. b) tra i suoi obiettivi non compare la professionalizzazione c) si deve occupare di processi di socializzazione. d) il suo scopo è di istruire

[Continua]

[Segue]

<p>9) Quando si tratta di proporre risposte che indicano cifre o quantità disporle in ordine crescente o decrescente. <i>(nell'esempio il quesito è mal formulato in quanto la disposizione non ordinata - crescente o decrescente- delle cifre complessifica ingiustamente il quesito) [...].</i></p>	<p>L'età media per i maschi in Italia per il primo matrimonio è a) 27,5 anni. b) 24,5 anni c) 29,0 anni d) 23,8 anni.</p>
<p>10) Evitare alternative del tipo: <i>nessuna di queste, sia a) che c), tutte le precedenti</i>, perché non omogenee alle altre. <i>(nell'esempio il quesito è mal formulato in quanto l'ultima alternativa o viene esclusa in quanto disomogenea con le altre, oppure si presenta come forte distrattore, in quanto offre un livello più articolato di risposta e tendenzialmente più esaustivo).</i></p>	<p>Con il termine validità di un test si intende a) il grado di precisione della misurazione. b) il grado con cui si misura ciò che si vuol misurare. c) il grado di accordo tra valutatori esperti nella sua utilizzazione. d) Sia a) sia c).</p>
<p>11) Evitare che elementi grammaticali o la struttura della frase favoriscano l'individuazione della risposta esatta. <i>(nell'esempio il quesito è mal formulato in quanto solo le alternative "a" e "c" concordano grammaticalmente con il corpo della domanda. In questo modo, ragionando per esclusione il quesito diventa tra 2 alternative).</i></p>	<p>Con il termine "contemporaneamente" si intende nel a) futuro. b) stesso tempo. c) tempo. d) insieme.</p>
<p>12) Quando le alternative di risposta iniziano con lo stesso gruppo di parole è preferibile includerlo nel corpo della domanda. <i>(nell'esempio il quesito è mal formulato in quanto tutte le alternative iniziano con l'espressione "il grado di" che potrebbe essere inclusa nel corpo della domanda e alleggerire il quesito).</i></p>	<p>Con il termine validità di un test si intende a) il grado di precisione della misurazione. b) il grado con cui si misura ciò che si vuol misurare. c) il grado di accordo tra valutatori esperti nella sua utilizzazione. d) il grado di coerenza interna tra i singoli quesiti.</p>
Organizzazione dei quesiti	
<p>13) Fare in modo che i quesiti siano tra loro indipendenti.</p>	
<p>14) Non usare alcuna risposta precedente come distrattore o come risposta corretta in altri quesiti.</p>	
<p>15) Prevedere la collocazione della risposta corretta in modo che nell'insieme dei quesiti le risposte esatte non corrispondano tutte alla stessa lettera.</p>	
Convenzioni grafiche.	
<p>16) Le alternative di risposta dei quesiti che terminano con il punto interrogativo devono cominciare con la maiuscola [...].</p>	[...]
<p>17) I distrattori dei quesiti a completamento devono iniziare con la minuscola [...].</p>	

[Continua]

[Segue]

18) Per convenzione, alla fine di ogni alternativa di risposta si usa il punto [...].	[...]
19) Il corpo della domanda nella forma a completamento non termina mai con i due punti o con i puntini sospensivi. <i>(nell'esempio il quesito è mal formulato in quanto il corpo della domanda termina con i due punti; l'esempio corretto è quello riportato per esemplificare la regola 11).</i>	Con il termine "contemporaneamente" si intende nel: (...) a) futuro. b) stesso tempo. c) tempo. d) insieme.
20) Le alternative di risposta vanno elencate e non disposte di seguito una dietro l'altra. <i>(nell'esempio il quesito è mal formulato in quanto le alternative sono una di seguito all'altra e rendono palesemente più complessa la soluzione; l'esempio corretto è quello riportato per esemplificare la regola 7).</i>	[...]

Tabella 4.5

4.1.5.3 I QUESITI A CORRISPONDENZA

Questa forma di quesito permette di misurare differenti obiettivi educativi. L'uso più semplice è naturalmente quello della misurazione delle conoscenze.

[...] Nelle prove di comprensione della lettura questo tipo di quesito è particolarmente indicato per la misurazione dell'abilità lessicale (due liste, una dei termini da spiegare, l'altra con le definizioni). Spesso l'abilità linguistica di definizione è richiesta relativamente a illustrazioni di oggetti (apparecchiature, macchine ecc.) o procedimenti (sequenze di azioni, istruzioni ecc.) [...].

Nell'esempio che segue si propone l'associazione, l'accoppiamento (*matching*) di elementi posti in due elenchi, per rilevare, successivamente alla lettura di un breve brano, la comprensione delle unità narrative fondamentali.

Istruzioni: Completa le frasi della prima colonna collegandole con una della seconda

(Usa delle frecce per collegare le frasi, oppure segna accanto ad ogni lettera il numero corrispondente al completamento)

___ a) Un ladro entrò nella casa di un mercante	1) rimise al suo posto la mercanzia rubata.
___ b) Il ladro cadde a terra	2) ma non notò nulla.
___ c) Il ladro senti che veniva gente	3) e accese una candela.
___ d) Il mercante passò davanti al mucchio delle foglie	4) e si introdusse nel granaio.
___ e) Il ladro alla fine non potè più trattenersi	5) e vi si nascose sotto.
	6) e andò al piano superiore.
	7) e starnutì.
	8) facendo un gran rumore.

[...]

Regole per la costruzione di quesiti a corrispondenza

Posizione del problema
1) Gli elementi (termini, espressioni, enunciati ecc.) dei due elenchi devono essere omogenei fra loro per contenuto.
2) Ad evitare ambiguità nella risposta, almeno uno dei due elenchi dovrà contenere semplici elementi (date, singole parole, autori ecc.) mentre l'altro potrà contenere anche enunciati o elementi più complessi.
3) Nel caso di elenchi di diversa complessità, adopera l'elenco con le espressioni più lunghe come premessa del quesito e l'altro per risposte.
4) Fornire chiare istruzioni per spiegare le modalità di corrispondenza tra gli elenchi (frecce, numeri, ordinamenti ecc.).
5) Ad ogni elemento del primo elenco ne deve corrispondere soltanto uno esatto del secondo.
Organizzazione dei quesiti
6) Il numero delle corrispondenze da compiere dovrebbe essere tra 5 e 15.
7) Evitare di costruire una corrispondenza "perfetta", cioè le due liste dovrebbero essere disomogenee per numero, per evitare la dipendenza nelle risposte.
8) Il numero degli elementi di "risposta" sarà al massimo doppio di quello da completare.

Tab. 4.6

4.1.5.4 I QUESITI A COMPLETAMENTO

I quesiti a completamento sono molto utilizzati per competenze lessicali e la capacità di comprensione linguistica generale. Un particolare tipo di prova di completamento è quello definito di *cloze* [...]. Esso si basa sulla cancellazione di una selezione di parole da un testo originale, cancellazione di parole che avviene in forma sistematica (ad esempio una ogni 5 o più). In questa forma le parole vengono eliminate senza tener conto del loro significato e della loro particolare funzione e quindi per individuare le parole cancellate si farà riferimento alla ridondanza linguistica del testo oltre che alla conoscenza di quelle parole stesse.

Alcuni esempi di quesiti a completamento sono riportati di seguito [...]:

3. Completamento di un testo senza alternative di risposta

I beni e i servizi vengono contrattati nei, ossia in quei luoghi dove si radunano appositamente e compratori. Nel significato comune il termine individua un luogo nel quale vengono scambiate In senso strettamente economico la nozione non richiede necessariamente un luogo determinato. Infatti esso è il complesso delle e delle di un dato bene.

In questo caso si offre la lettura di un breve testo nel quale sono state eliminate alcune parole. Non si offrono alternative di risposta e quindi la difficoltà di riempire

i "buchi" è direttamente proporzionale alla comprensibilità del testo leggibile. Una volta individuata la tematica (il mercato, la domanda e l'offerta) risulta più semplice risolvere questo tipo di quesiti.

4. Completamento di un testo con alternative di risposta

I beni e i servizi vengono contrattati nei, ossia in quei luoghi dove si radunano appositamente e compratori. Nel significato comune il termine individua un luogo nel quale vengono scambiate In senso strettamente economico la nozione non richiede necessariamente un luogo determinato. Infatti esso è il complesso delle e delle di un dato bene.

a)venditori b)beni c)mercati d)domande e)merci f)offerte g)compratori

Le diverse alternative di risposta, in questo quesito, sono formulate al termine del testo. Esse sono in numero superiore a quello dei completamenti da svolgere, per evitare le risposte per esclusione e quelle condizionate.

Regole per la costruzione di quesiti a completamento

Posizione del problema
1) Nei completamenti mirati, iniziare ad eliminare le parole dal testo dopo il primo periodo (altrimenti tralasciare le prime righe).
2) Nei cloze, i periodi iniziali e finali dovrebbero rimanere integri.
3) Eliminare al massimo 2 o 3 parole per periodo (far dipendere il numero dalla lunghezza del periodo).
4) Se è un completamento di frase, cercare di collocare verso la fine la parola mancante, così da evitare eccessiva ambiguità.
Formulazione linguistica
5) Controllare la coerenza sintattica e grammaticale tra testo da completare e termini presentati in calce.
6) Evitare tra i termini da inserire di utilizzare sinonimi (ugualmente accettabili nel completamento).
7) Garantire l'omogeneità lessicale tra testo da completare e alternative di completamento.
Organizzazione dei quesiti
8) Evitare di costruire quesiti con più di 15 completamenti.
9) La lunghezza degli spazi vuoti sia sempre la stessa e non proporzionale al termine da inserire.
10) I termini da inserire siano elencati seguendo un ordine alfabetico.
11) Per ogni parola da completare ci sia, oltre alla risposta esatta, almeno un'alternativa.
12) Presentare un numero di parole da inserire al massimo doppio degli spazi previsti, e nel caso di un alto numero disporle su più colonne.

Tabella 4.7

18. Correzione delle prove oggettive di profitto e analisi delle prestazioni (Vertecchi, B., *Decisione didattica e valutazione*)¹⁶¹

La correzione di una prova oggettiva di profitto si effettua in modo estremamente rapido, attraverso una serie di *griglie*. Ciascuna griglia è uguale alla corrispondente pagina della prova, ma reca in corrispondenza della risposta esatta una perforazione. Sovrapponendo la griglia alla prova, l'insegnante accerta se l'allievo ha indicato la risposta esatta; annota quindi ogni risposta esatta (per esempio con un tratto di matita colorata), ripetendo la stessa operazione per tutte le pagine del test. Si contano quindi quante sono le risposte esatte per ciascuna parte della prova (vero/falso, scelte multiple ecc.). Si procede infine all'attribuzione di punteggi, che può essere effettuata scegliendo uno dei due seguenti criteri:

1) non si penalizza l'errore: in questo caso il punteggio viene determinato computando le risposte esatte, ed eventualmente modificando il computo per mezzo di moltiplicatori fissi, che consentano di attribuire pesi diversi alle risposte, a seconda che si tratti di vero/falso, scelte multiple ecc.;

2) si penalizza l'errore: ciò viene fatto modificando il computo del punto attraverso la detrazione per ogni risposta errata di una frazione di punteggio, corrispondente alla probabilità di rispondere bene a caso. Il punteggio penalizzato viene quindi espresso dalla formula

$$P = E - \frac{S}{N - 1}$$

in cui P = punteggio, E = numero delle risposte esatte, S = numero delle risposte errate, n = numero delle possibilità di risposta.

Supponiamo che un allievo risponda correttamente a 7 scelte multiple su 10 (scelte multiple a una soluzione corretta su cinque proposte) e che a ciascuna scelta multipla siano attribuiti due punti. Se non si penalizza l'errore il punteggio corrispondente è 14. Se invece si penalizza l'errore bisogna considerare la probabilità di rispondere bene a caso, che è di 1 su 5; perciò

$$P = 2 (7 - 3/4) = 12,5$$

¹⁶¹ Tratto da: Vertecchi, B. (1993), *Decisione didattica e valutazione*, Scandicci (FI), La Nuova Italia, pp. 145-148.

I due modi di attribuire i punteggi sono sostanzialmente equivalenti da un punto di vista pratico, anche se la penalizzazione dell'errore consente di depurare il punteggio di quella frazione che può essere attribuita alle risposte casuali. Ciò che conta infatti è che l'allievo sia coinvolto responsabilmente nell'attività di verifica senza che essa assuma quel carattere inquisitorio che può indurre ad atteggiamenti di rifiuto, più o meno espliciti (copiature, risposte a caso ecc.).

Il punteggio di una prova acquista un preciso significato solo se confrontato con quelli ottenuti da tutti gli allievi. Si devono cioè considerare le caratteristiche della distribuzione dei punteggi. Supponiamo che una prova sia costituita da 10 quesiti vero/falso, 12 scelte multiple ad una soluzione, 5 a due soluzioni e da un completamento di 15 quesiti, e che si attribuisca un punto a ciascun vero/falso, 2 punti a ciascuna scelta multipla a una soluzione, 3 punti alla coppia di soluzioni delle scelte multiple doppie, 1 punto a ogni completamento corretto: si dispone quindi di un totale di 64 punti. Una volta corretta la prova, l'insegnante potrà formare una graduatoria, assegnando un punteggio a ciascun allievo. Questo punteggio viene detto grezzo: esso costituisce [...] il punto di partenza per analisi più accurate [...]. Informazioni molto dettagliate relative alle specifiche difficoltà incontrate dagli allievi nel rispondere ai quesiti di una prova oggettiva possono ricavarsi dall'analisi dei singoli quesiti (*item analysis*). A questo fine è necessario eseguire la *tabulazione* completa delle risposte. Essa si effettua nel modo seguente (v. figura 11):

1. si individuano su un foglio quadrettato le due semirette di un quadrante cartesiano;
2. si stabilisce la graduatoria dei punteggi grezzi, dal più basso al più alto;
3. si trascrivono i punteggi così ordinati sull'asse delle ascisse;
4. sull'asse delle ordinate si indicano i singoli quesiti della prova;
5. all'intersezione di ciascun punteggio con ogni quesito si indica qual è stata la risposta: per semplicità è sufficiente annotare soltanto gli errori.

La tabulazione fornisce una immagine anche visiva dell'andamento di una prova. Si osserva, com'è ovvio, che gli errori sono più numerosi in corrispondenza dei punteggi più bassi; ma ci si rende anche conto di quali quesiti abbiano dato luogo a errori o fraintendimenti. Se si osserva la figura 11, si vede che al quesito n. 2 hanno risposto in modo errato tutti gli allievi, indicando la stessa risposta b): evidentemente il quesito era mal formulato. Il quesito n. 5, pur avendo fatto registrare un buon numero di risposte corrette, presenta una concentrazione di errori sul distrattore c): se ne può ricavare che si tratta un distrattore particolarmente fuorviante, o che gli altri distrattori sono poco credibili.

La tabulazione delle risposte rappresenta un passaggio necessario anche in funzione di più accurate considerazioni sulle caratteristiche di un test: si può stabilire quanto ciascun quesito sia *discriminativo* (confrontando le prestazioni degli allievi che hanno ottenuto punteggi più elevati con quelle corrispondenti ai punteggi più bassi),

quale sia la *difficoltà* di ciascun quesito (e cioè quale frazione degli allievi ha fornito la risposta corretta).

ITEM scelte multiple	PUNTEGGIO																						
	20	21	22	22	24	25	25	27	27	28	28	28	29	29	30	30	30	31	31	32	34	36	
1	a	b		d	b	e			e		a	b			d				e				
2	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
3	c	d		e			c		a			c				e							
4	d		b		d			d			a			c							c		
5		c	c			c				c			c				c						
--																							

Figura 11. *Tabulazione delle risposte*. In ascissa è indicato il punteggio conseguito da ciascun allievo, in ordine crescente o decrescente. In ordinata si legge il numero che contraddistingue i singoli quesiti della prova. All'intersezione tra il punteggio e un determinato quesito si riscontra quale è stata la risposta: se la casella è bianca significa che l'allievo ha risposto correttamente; in caso contrario è indicato quale risposta è stata fornita.

Dall'esempio di tabulazione riportato nella figura 11 si può ricavare che il quesito n. 2 è molto difficile, per il fatto che nessun allievo ha fornito la risposta esatta, ma per niente discriminativo, dal momento che sia ai punteggi migliori, sia a quelli più bassi corrisponde una risposta errata. Il quesito n. 1 appare invece di media difficoltà, ma piuttosto discriminativo: c'è infatti una concentrazione di errori in corrispondenza dei punteggi più bassi, mentre gli allievi che hanno ottenuto i punteggi più alti hanno fornito quasi tutti una risposta esatta.

19. Una diversa nozione di "sufficienza" (Vertecchi, B., *Decisione didattica e valutazione*)¹⁶²

Un concetto tanto generalmente accettato, quanto privo di reali giustificazioni, è quello di *sufficienza*. Esso dovrebbe designare le condizioni minime che consentono di considerare positive le prestazioni di un allievo. La sufficienza costituirebbe dunque un termine di separazione all'interno della scala di misurazione del profitto, al di qua del quale si estenderebbe la semidistribuzione positiva, variabile tra un minimo e un massimo, e al di là quella negativa, anch'essa variabile tra un minimo e un massimo. La critica di questo concetto della sufficienza può essere condotta su due piani, il primo tecnico-metodologico e l'altro pedagogico.

Dal punto di vista tecnico-metodologico la capacità di un termine di discriminare nettamente tra due serie di misure è affidata soprattutto alla qualità delle misure stesse. Ciò significa che la discriminazione è tanto più esatta quanto più valide e attendibili sono le misure. [...] Tali non sono le misure che si ottengono nelle situazioni didattiche tradizionali, e che anzi riesce persino difficile stabilire a che tipo di scala esse diano luogo¹⁶³. Questa indeterminatezza delle proprietà metriche della scala comporta una probabilità di errore piuttosto estesa; anche se non si vuole respingere del tutto il giudizio che viene espresso sulle prestazioni degli allievi, esso deve essere riferito piuttosto a una fascia di rendimento che a una precisa posizione su una scala.

Ne viene che se si individua sulla scala una posizione che funge da termine di separazione, ci sono molte probabilità che i punteggi prossimi a tale termine corrispondano a prestazioni che altrimenti misurate si collocherebbero in posizione diversa rispetto al termine stesso: in altre parole, i giudizi che si fondano su misurazioni prossime alla sufficienza sono sostanzialmente arbitrari, per il fatto che esiste un margine di oscillazione delle misure tale da renderle sostanzialmente casuali all'interno di una certa fascia. Si tratta di un limite della valutazione tradizionale che è specialmente grave se si considera che attorno alla sufficienza tende a concentrarsi la maggior parte dei punteggi; i cinque e i sei nella scala in decimi sono i voti che ricorrono più spesso in quanto, dovendosi discriminare tra prestazioni più o meno

¹⁶² Tratto da: Vertecchi, B. (1993), *Decisione didattica e valutazione*, Scandicci (FI), La Nuova Italia, pp. 125-127.

¹⁶³ Usualmente si ritiene che la scala sia ad intervalli, ma più probabilmente è ordinale. Cfr., Domenici, G. (2001²), *Manuale della valutazione scolastica*, cit., p. 129. Per una definizione delle diverse scale di misura cfr. Lucisano, P., Salerni, A. (2002), *Metodologia della ricerca in educazione e formazione*, Roma, Anicia, pp. 62-65 [n.d.c.].

positive, è inevitabile che vi sia un accentramento su posizioni intermedie (vedi figura 8).

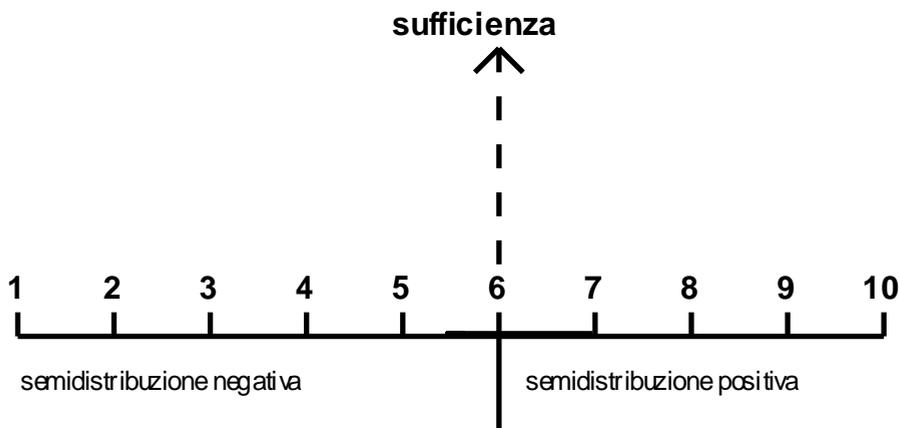


Figura 8. *La sufficienza*. Nella scala in decimi il voto che divide la semidistribuzione negativa della scala da quella positiva è il 6. Tuttavia, come è facile constatare effettuando una rilevazione empirica nelle scuole secondarie superiori dove è in uso tale scala, il massimo dei voti si concentra proprio a ridosso della linea di sufficienza. Ne deriva che, anche supponendo un errore modesto negli apprezzamenti delle prestazioni degli allievi, in un numero elevato di casi è abbastanza casuale che il voto si collochi nella semidistribuzione negativa o in quella positiva.

La critica tecnico-metodologica del concetto di sufficienza potrebbe essere in parte considerevole superata attraverso l'adozione di modalità di misurazione più precise. Restano intatte però le riserve che si possono muovere da un punto di vista pedagogico. C'è da chiedersi infatti quale sia il senso delle differenziazioni che si stabiliscono all'interno delle due semidistribuzioni, quella positiva e quella negativa.

Supponiamo che si valuti l'apprendimento di un allievo relativo a un determinato contenuto, come potrebbe essere la dimostrazione del teorema di Pitagora. Un'analisi del contenuto rivela che ci sono alcune conoscenze che possono considerarsi preliminari rispetto al contenuto vero e proprio (*prerequisiti dell'apprendimento*): così l'allievo dovrebbe già essere in grado di distinguere un triangolo rettangolo; di indicare l'altezza, i cateti, l'ipotenusa; di calcolare il perimetro e la superficie.

Se l'insegnante effettua una valutazione preliminare dei prerequisiti, il suo giudizio dovrà essere univoco, e cioè ridursi all'alternativa "l'allievo possiede/non possiede i prerequisiti". Non avrebbe molto senso affermare che li possiede di più o di meno. Sarebbe infatti come dire "l'allievo A sa cos'è l'ipotenusa, ma l'allievo B lo sa meglio dell'allievo A, mentre l'allievo C lo sa più o meno". Del tutto privo di logica apparirebbe il giudizio sul versante negativo: "l'allievo D non sa cos'è l'ipotenusa, ma l'allievo E lo sa ancora di meno". È chiaro che simili differenziazioni non hanno senso: se nella scuola esse vengono effettuate, ciò accade perché il giudizio non si

riferisce all'effettiva conoscenza dell'allievo, ma alla facilità con cui apprende, alla proprietà e al garbo con cui espone ecc.

Se si passa alla valutazione terminale dell'apprendimento, una volta che la procedura didattica relativa al teorema di Pitagora si sia conclusa, le alternative rimangono le stesse: un allievo sa o non sa dimostrare il teorema di Pitagora, ma non è possibile che lo sappia o non lo sappia dimostrare di più o di meno.

La soluzione del problema così impostato richiede una radicale revisione del concetto di sufficienza: essa non dovrebbe più essere considerata un limite minimo, collocato nella parte centrale della distribuzione dei voti, ma semplicemente una condizione di adeguatezza, che consenta di distinguere quando la prestazione dell'allievo corrisponde al compito che gli è richiesto, e quando non lo soddisfa. Se volessimo identificare la posizione che nella scala corrisponde all'adeguatezza, non potremmo che collocarla nella parte alta, decisamente sopra al 6. Potremmo anche dire che una differenziazione più ampia è possibile nell'area dell'insufficienza che in quella della sufficienza: infatti, mentre quest'ultima è credibile solo se è definita in modo univoco, possiamo attenderci che le prestazioni fornite dagli allievi presentino varie misure di inadeguatezza: anzi, è proprio l'intervallo che separa ciascuna dalla condizione di adeguatezza che segnala il problema didattico che occorre affrontare. È evidente che una simile netta distinzione è possibile solo se gli obiettivi sono stati indicati con precisione e se sono stati accuratamente descritti i comportamenti che si attendono: le prestazioni degli allievi devono infatti essere confrontate con essi.

Finito di stampare nel mese di ottobre 2008
dal Centro Stampa Nuova Cultura, Roma