

Quale quantità e quale qualità per la comprensione delle esperienze educative

Pietro Lucisano - Andrea Marco De Luca

«Sapienza» Università di Roma, Dipartimento di Ricerche Storico-filosofiche e Pedagogiche

doi: 10.7358/ecps-2014-009-luci

pietro.lucisano@uniroma1.it

WHICH QUANTITY AND QUALITY FOR UNDERSTANDING EDUCATIONAL EXPERIENCES

ABSTRACT

This paper illustrates the methodology adopted to design and calibrate a test for measuring the student's linguistic abilities at the end of the third year of vocational training courses (10th grade). The work is part of a major ISFOL study also including tests on math and strategic abilities. The test was calibrated on a judgement sample of 1046 students. It measures listening, reading comprehension, mechanics and written composition through several tasks. This paper presents methodologies of two item analysis procedures the CTT («Classical Test Theory») and IRT («Item Response Theory»), and the procedures for designing derivate variables from different tasks. In particular, it presents the advantages of using the Rasch model for a better measure of the latent trait of ability and for calibrating the test items. The paper rounds off by illustrating the main results of the multivariate analysis of the study. These results should obviously be interpreted in the light of the characteristics of the target population, its social and cultural background, which is weak in this case, and the collection of its school failures.

Keywords: Evaluation, Exams, Measurement, Quality of measures, Text.

Se davvero non te ne importa niente non ti accorgi nemmeno che c'è qualcosa di sbagliato. Non ti viene neanche in mente. Il solo fatto di dire che c'è qualcosa di sbagliato significa che ci tieni.

Pirsig, 1974, p. 164

Quando Pirsig percorreva con la sua moto le Montagne Rocciose alla ricerca della qualità, ragionando assieme al fantasma di Fedro, forse ancora non era così diffusa l'idea che la qualità potesse essere ridotta alla rispondenza di un prodotto agli standard attesi. Ma la tecnologia avanza e oggi ingegneri ed economisti hanno conquistato una egemonia così estesa da espropriare la forza dirompente di questo termine e da oscurare anni di riflessioni filosofiche e i travagli che ha provocato la critica del giudizio, sia che la si affrontasse in termini analitici sia che la si affrontasse in termini dialettici.

La riduzione semplifica. La riduzione della qualità a misura delle caratteristiche o delle proprietà di una entità (una persona, un prodotto, un processo, un progetto) in confronto a quanto ci si attende da tale entità, per un determinato impiego, consente di riportare la qualità a quantità, di misurarla, etichettarla, sottoporla a verifiche e certificazioni. Questa riduzione è però così radicale da lasciare fuori proprio la qualità stessa. Quella stessa qualità che Lalande (1971) definisce in modo sintetico «Qualità: specie, stato naturale, bontà» specificando che «All'opposto della quantità questa non è misurabile».

La difficoltà di definire la qualità, prima che tutto fosse risolto dagli standard «ISO qualche cosa», aveva attraversato infiniti dibattiti: «Come si fa a dire che una cosa è bella?» si chiedevano i filosofi; «Come si fa a dire che una legge è giusta?» si chiedevano i giuristi; «Come si fa a dire che una mozzarella è buona?» si chiedevano i buongustai. Certo è che se è difficile definire se una mozzarella è buona, rimane assai meno difficile riconoscerlo. Gli studenti di Fedro erano in grado di riconoscere un buon tema da un tema mal concepito e anche uno sprovveduto riesce a comprendere la differenza tra un'opera d'arte e una crosta. Quando però qualcuno si affanna a definire la differenza, a spiegarne le ragioni, subito qualcosa si incrina.

«Ma come diavolo si fa a giustificare secondo ragione il rifiuto di definire qualcosa? Le definizioni sono il fondamento della ragione». La spiegazione si muove alla ricerca delle ragioni della qualità sulla base della ragione e la ragione si pone al di fuori e al di sopra della qualità per conoscerla e non si accontenta di contemplarla, di percepirla o di comprenderla. Non si accontenta di essere in interazione con essa o, come direbbe Dewey, di farne esperienza.

In questa operazione così ardita la ragione – pur di mantenersi fedele all'assunto metafisico della separazione tra soggetto e oggetto – rischia di perdere anche uno dei suoi strumenti più efficaci, la capacità di misurare. Così le misure della qualità intaccano la qualità della misura, alla quale, personalmente, rimango molto affezionato. Abbiamo impiegato anni a costruire strumenti di misura che ci consentono di avere elementi informativi rilevanti e la comunità scientifica ha definito regole per l'uso di questi strumenti e

per l'uso dei risultati che ci rendono disponibili. Anche nel nostro piccolo, come studiosi di educazione, abbiamo vissuto dibattiti molto vivaci su queste tematiche. Come Fedro, nella mia ricerca ho cercato di utilizzare al meglio i metodi quantitativi, di costruire strumenti di misura efficaci. Ho imparato a comprendere che i dati, prima di essere dati debbono essere presi, che il modo in cui vengono presi e il perché abbiamo deciso di raccoglierci, sono inseparabili dalla loro natura, e che solo la loro contestualizzazione ci aiuta a essere aiutati dai dati. Che, una volta raccolti, i dati vanno osservati, e ruminati, come San Girolamo diceva della parola, prima di trarne conclusioni. E ancora che queste conclusioni, ancorché utili, non hanno mai un carattere definitivo.

Nel dibattito attuale l'egemonia della tecnocrazia rischia di privarci del significato stesso delle parole che sono lo strumento principale di comprensione della nostra esperienza: valutazione, misura, apprendimento, non significano più nei media, e di conseguenza nella coscienza delle persone, quello che hanno significato per Dewey, Visalberghi, Calonghi, Gattullo. Così quando oggi parliamo dell'importanza della valutazione e delle misure, rischiamo di essere confusi con i promotori di graduatorie, di classifiche, di concorsi di bellezza.

La facilità che i computer hanno consentito nel computo, ha prodotto una sorta di passione per le misure anche quando queste sono del tutto prive di senso. I tifosi di calcio hanno siti in cui trovano dati su tutto ciò che avviene in ogni partita, secondo per secondo, ma sapere quante volte un giocatore tocca la palla non coincide con il suo avere giocato bene o meno. Noi abbiamo studiato che i dati acquistano significato in relazione al modello interpretativo nel quale sono presi in considerazione e Occam aveva capito ben prima della teoria della comunicazione che le informazioni, quando sono troppe e mal organizzate, costituiscono un rumore dal quale ci si deve difendere con una lama affilata.

In questa confusione di significati si rischia di perdere di vista il senso stesso dei processi conoscitivi che attengono a fenomeni complessi, come nel nostro caso l'educazione, il cui compito è comprendere e valorizzare la Φύσις ovvero la natura o in altri termini le qualità dei ragazzi. L'alternativa è ridurre l'azione educativa al mero addestramento del capitale umano ai fini di disporne in un mercato del lavoro, che bontà sua, non è in grado di definire con chiarezza neanche che cosa gli occorre.

L'approccio tecnocratico all'educazione muove dunque da una definizione soggettiva della qualità in esame alla quale vengono applicati indicatori quantitativi, che poi vengono composti in modo abbastanza arbitrario per rappresentare entità qualitative. La giustificazione di questo approccio è riconducibile al fastidio della complessità e all'esigenza del fare e del fare

presto, prescindendo dalla fatica di darsi la pena di raccogliere tutti gli elementi necessari per agire. L'emergenza che rende indispensabile agire subito si trasforma di conseguenza in una serie di tentativi e di errori dai quali non si riesce ad apprendere perché non ci si è dati un metodo. Spesso questi errori si compiono solo perché non vengono poste in atto quelle procedure che il pensiero riflessivo e il metodo scientifico richiedono e che possono essere riassunte nel «pensare prima di agire». Gli esempi sono infiniti: dai decreti *omnibus* pieni di confusioni e che costantemente debbono essere riveduti, al *bonus* per la maturità previsto, abrogato, corretto, riabrogato, utilizzato in forma di sanatoria per poi essere di nuovo riabrogato (stavo scrivendo definitivamente riabrogato e poi mi sono ricordato che è meglio «mai dire mai»). I processi che ne derivano, non consentendo di trarre insegnamento dall'esperienza, sono proprio il contrario dell'educazione e dell'apprendimento.

L'osservazione di quanto succede ci consente di verificare che le catastrofi annunciate sono indifferenti ai provvedimenti presi dal fare emergenziale; allo stesso tempo è abbastanza evidente che questo modo di procedere produce tutta una serie di danni collaterali, dei quali gli autori cercano di non assumersi le responsabilità. Ad esempio l'impianto valutativo dell'INVALSI ha generato il *cheating* e l'insegnamento sui test, ma nessuno dei responsabili si ritiene responsabile: la colpa è della gente che non capisce. Il modello di valutazione dell'ANVUR ha prodotto case editrici e riviste che pubblicano a pagamento secondo i criteri previsti e convegni in cui pagando si realizzano pubblicazioni internazionali e colleghi che scrivono libri e libri che, pur essendo sul nulla, portano le mediane dei loro settori disciplinari sempre più in alto. Qualità? Se tutti si fossero comportati secondo gli auspici di chi ha pensato questo sistema, i danni collaterali non si sarebbero verificati: dunque la colpa è della gente che non capisce. O c'è una colpa anche nel non capire la gente?

Noi professori universitari queste cose le conosciamo bene. Abbiamo colleghi che fanno lezioni meravigliose, solo che gli studenti non capiscono e allora si affannano come vestali a difendere il fuoco sacro del sapere da questi barbari, ma se il fuoco del sapere non scalda nessuno, allora non è fuoco.

Essendo oggi i dibattiti culturali condotti in modo molto rumoroso, ogni osservazione sul merito di quanto si va portando avanti viene considerata in termini conflittuali e percepita come il suo opposto, per cui se qualcuno osserva che i test dovrebbero essere meglio costruiti, finisce per essere accusato di essere contrario ai test e alla valutazione, mentre ciò che si stava valutando, proprio perché alla valutazione ci teniamo, è l'uso improprio di strumenti di misura. Insomma credo sia legittimo affermare che non è con la bilancia che si misura l'altezza di una persona, senza rischiare di essere accusato di rifiutare le misure *tout court*.

Abbiamo detto che la qualità definita rischia una riduzione della complessità e della stessa qualità percepita. Il concetto di qualità si riferisce alla dimensione esperienziale, alla percezione di un oggetto e può essere rappresentata da un aggettivo (uno studente competente, una scuola efficiente) o da un avverbio se riferita a un processo (studiare approfonditamente, misurare seriamente).

La qualità dunque è costantemente in bilico tra una dimensione soggettiva «romantica», che ne coglie la dimensione profonda, e una dimensione razionale «classica», che cerca di realizzare la quadratura del cerchio per approssimazioni successive. Non credo che si possa rinunciare a queste due prospettive, né sono in grado di proporre una sintesi.

Il problema che vorrei porre è tuttavia la denuncia di una via approssimativa alla definizione e alla misura che procede senza cautele scientifiche. Che dà i numeri alle cose e alle persone e applica alle cose e alle persone le proprietà dei numeri senza rispetto né per i numeri, né per le cose, e tantomeno per le persone e che definisce i suoi punti di arrivo «misura della qualità».

Al nostro tempo la necessità di sottrarre all'arbitrio il problema della valutazione educativa è oggetto di due discipline: Docimologia e Docimanzia. La Docimologia è una disciplina rigorosa, cosciente dei suoi limiti che tenta di trovare un percorso condiviso al problema della valutazione e in particolare della valutazione educativa, interrogandosi sugli obiettivi, sui processi e sui risultati.

La Docimanzia è una disciplina al servizio della politica, che attribuisce valori numerici a realtà empiriche in modo piuttosto disordinato e poi si sbizzarrisce nell'effettuare su questi numeri inconsistenti le più ardite analisi di tipo statistico al fine di stabilire graduatorie, formulare leggi, standardizzare la realtà, fino a ricavarne indicazioni strategiche per la gestione di interi comparti educativi. Il vantaggio della Docimanzia è la semplificazione. Essa offre certezze sia ai suoi proparatori, sia al popolo che, come bene spiega il grande Inquisitore, è ben lontano dal volere la libertà e preferisce attenersi ai voti piuttosto che affrontare la croce di doversi assumere responsabilità personali o collegiali di decisioni educative difficili e di esito incerto.

Alla base di queste modellizzazioni c'è inoltre l'assunto che il valutatore umano, essendo per natura corruttibile, debba esercitare la sua funzione all'interno di una procedura che riduca al massimo i suoi margini di libertà e di intelligenza. Così lo sforzo di limitare l'arbitrio, ovviamente condotto a fin di bene per evitare cadute ed errori, si trasforma nella negazione dogmatica dell'esercizio del libero arbitrio. L'effetto di questo modello è da un lato la rigidità deterministica che trasforma istruzione e educazione in processi di addestramento, dall'altro la spinta verso comportamenti di reazione alla

omogeneizzazione normativa. Ogni proibizione che non proviene da una intima convinzione contiene in sé i germi di reazioni di segno opposto che, nel nostro caso, si sono manifestati con il *cheating* di massa e con la riconsiderazione da parte degli insegnanti dei voti in relazione ai parametri.

Questo modello contrasta con la libertà di insegnamento e con la libertà dell'intelligenza e produce danni sia in termini di qualità dell'insegnamento (*teaching to the test*) sia in termini morali (*cheating*). Inoltre, come il peccato originale, traccia un solco tra chi monopolizza la conoscenza del bene e del male per il bene di tutti e chi vorrebbe accedervi e per questo stesso motivo ne viene escluso. La povertà delle soluzioni che derivano da questo approccio alla conoscenza è visibile, l'idea che la società e la scuola possano essere riformate sulla base di una meritocrazia «bastone e la carota», con pochi soldi di incentivo e minacce di punizioni è sciocca e volgare. Ma soprattutto questa visione contiene un uso irragionevole della razionalità che per comprendere la $\Phi\upsilon\sigma\iota\varsigma$ la priva della $\Phi\upsilon\sigma\iota\varsigma$ stessa. E qui ancora può soccorrerci una immagine di Pirsig «Squareness. Proprio così, quando si sottrae la Qualità si ottiene la squareness»¹. Ma la *squareness* non è solo il noioso regno della conformità che Pirsig associa alla società sovietica e al Mondo nuovo di Huxley, è anche nell'entusiasmo rivoluzionario che vorrebbe un mondo più conforme alla ragione.

Don Milani in una intervista filmata sosteneva con convinzione che i suoi ragazzi non fumavano perché non era ragionevole: «Qui i piccoli non fumano perché glielo proibisco. I più grandi non fumano perché hanno capito che era stupido fumare. A questo punto non si può più parlare di obbedienza ma di libera scelta». Come non è ragionevole fumare, non è ragionevole mangiare dolci, innamorarsi come Paolo e Francesca, bere un bicchiere di più. A Fedro «vennero in mente anche persone di sua conoscenza alle quali questo mondo privo di qualità sarebbe andato benissimo. Erano gli stessi che cercavano di farlo smettere di fumare. Gli chiedevano delle giustificazioni razionali per il suo vizio». «Dovevano avere delle motivazioni, dei piani, delle soluzioni per tutto» (Pirsig, 1974, p. 216).

Queste soluzioni nel nostro tempo sono incarnate dagli standard che nascono a fin di bene, nascono come ben sottolinea Kozol dall'irresistibile impulso a dire «I know what's good, because it's good for me, and I will damn well make you good in my own image» (Kozol, 2000, p. viii).

C'è un pericolo nella democrazia – un lato oscuro potremmo dire. Un'idea positiva di libertà comporta la rappresentazione di ciò che è buono per gli altri, che in qualche misura giustifica di imporla agli altri. Quindi quello che all'ori-

¹ «La squareness può essere descritta, in modo succinto ma esauriente, come l'incapacità a vedere la Qualità prima che essa venga definita intellettualmente» (Pirsig, 1974, p. 217).

gine comincia come una celebrazione dell'identità umana e creatività, può portare alla dominazione dell'individuo attraverso una dettagliata prescrittiva e imposta concezione di quello che è la verità e la retta via. (Woods, 2005, p. 4)

E se è vero, come affermava Pirsig, che «Le leggi della natura sono *invenzioni* umane, come i fantasmi. E così le leggi della logica e della matematica. Tutte queste benedette cose sono un'invenzione dell'uomo, compresa l'idea che *non lo sono*», allora anche Docimologia e Docimanzia sono invenzioni umane, e tuttavia, producono leggi ed effetti sia in chi ci crede sia in chi non ci crede e queste leggi e questi effetti sono terreno possibile di investigazione valutativa. Investigazione sorretta dalla convinzione che la valutazione rimane fondamentale. Tra i fantasmi che producono effetti merita ricordare le analisi di «valore aggiunto e gli effetti che producono attraverso una classificazione di 'qualità' delle scuole» (Corsini, 2012).

Pertanto la seconda parte di questa riflessione propone un'analisi «classica», docimologica, di una soluzione docimastica, quella adottata per determinare deterministicamente il voto dell'esame finale della scuola secondaria di secondo grado, dell'esame, cioè di terza media.

La riforma della scuola media (Legge 31/12/1962, n. 1859) prevedeva che l'Esame di Stato riguardasse tutte le materie previste nel piano di studi: «italiano, storia ed educazione civica, geografia, matematica, osservazioni ed elementi di scienze naturali, lingua straniera, educazione artistica, educazione fisica» (art. 6). All'esame si veniva ammessi sulla base di un giudizio collegiale formulato dal Consiglio di classe in sede di scrutinio finale. «Il Consiglio di classe, [...] formula un giudizio complessivo e attribuisce i voti in decimi relativi a ciascuna delle materie anzidette. Dichiara, quindi, l'alunno ammesso agli esami di licenza a meno che non riscontri una insufficienza notevolmente grave di formazione e di sviluppo della personalità» (D.P.R. 14/05/1966, n. 362, art. 2). Il giudizio finale prevedeva la sintesi da parte della commissione «sulla base delle risultanze dell'esame degli atti dello scrutinio finale e di ogni altro elemento a sua disposizione» e riguardava «il grado di formazione e di sviluppo della personalità raggiunto da ogni candidato». La formulazione finale del giudizio avveniva attraverso l'attribuzione del giudizio sintetico di «ottimo», «distinto», «buono», «sufficiente»; se negativo con la dichiarazione di «non licenziato» (art. 11). La formulazione in fasce espresse con aggettivi piuttosto che in numeri era stata determinata da lunghissime discussioni e studi che avevano messo in luce i limiti dei voti scolastici in termini di misurazioni².

² Ornella Andreani Dentici (1968, p. 49), spiegava: «Le abilità e le prestazioni dei ragazzi ricevono una valutazione numerica graduata su un continuo: è questo il caso dei voti scolastici.

L'attuale impianto dell'Esame di scuola media deriva da un lento e progressivo sgretolamento della riforma e dei suoi portati innovativi, compresi quelli relativi all'uso di voti e giudizi (Benvenuto, 2003).

Il D.P.R. 22/06/2009, n. 122, riduce l'ammissione ad un computo di voti e il giudizio di ammissione, che la riforma del 1962 prevedeva avesse anche un valore orientativo, a un unico voto (anche se specificamente non si parla di un calcolo di media): «[...] una votazione non inferiore a sei decimi in ciascuna disciplina o gruppo di discipline valutate con l'attribuzione di un unico voto secondo l'ordinamento vigente», «Il giudizio di idoneità di cui all'articolo 11, comma 4-bis, del decreto legislativo n. 59 del 2004, e successive modificazioni, è espresso dal Consiglio di classe in decimi, considerando il percorso scolastico compiuto dall'allievo nella scuola secondaria di primo grado» (art. 3, comma 2). Seguendo alla lettera la norma, uno studente con un voto 4 in tecnologia e tutti 10 nelle altre materie non può essere ammesso all'esame (salvo la modifica del voto da parte dell'insegnante spontanea o forzata).

Il testo introduce poi nella valutazione conclusiva dell'Esame di Stato «l'esito della prova scritta nazionale di cui all'articolo 11, comma 4-ter, del decreto legislativo n. 59 del 2004, e successive modificazioni. I testi della prova sono scelti dal Ministro tra quelli predisposti annualmente dall'Istituto nazionale per la valutazione del sistema di istruzione (INVALSI), ai sensi del predetto comma 4-ter».

Se ne ricava un meccanismo rigido per cui la valutazione finisce per configurarsi nel modo seguente: «All'esito dell'Esame di Stato concorrono gli esiti delle prove scritte e orali, ivi compresa la prova di cui al comma 4, e il giudizio di idoneità di cui al comma 2. Il voto finale è costituito dalla media dei voti in decimi ottenuti nelle singole prove e nel giudizio di idoneità arrotondata all'unità superiore per frazione pari o superiore a 0,5». Dunque il voto finale non è più affidato all'arbitrio e alle valutazioni educative che tengono conto del grado di formazione e di sviluppo della personalità di ogni singolo allievo, ma è finalmente semplice e preciso, espresso in numeri che tutti capiscono e alla commissione resta solo la libertà di attribuire una eventuale lode ove il computo restituisca il voto di 10.

Questa semplicità e precisione sono ancora più evidenti se si esaminano le istruzioni per l'attribuzione del punteggio. La Circolare Ministeriale n. 48 del 31 maggio 2012 precisa che «All'esito dell'Esame di Stato concorrono gli esiti delle prove scritte e orali, ivi compresa la prova nazionale INVALSI, e il giudizio di idoneità all'ammissione. Il voto finale è costituito dalla media

Essi sono estremamente ingannevoli, perché in realtà sono valutazioni qualitative [...], mentre d'altra parte la loro veste esteriore spinge a manipolarli con operazioni come la media, eccetera».

dei voti in decimi ottenuti nelle singole prove e nel giudizio di idoneità, arrotondata all'unità superiore per frazione pari o superiore a 0,5. Per media dei voti deve intendersi la media aritmetica, dovendosi attribuire a tutte le prove d'esame il medesimo rilievo. Si esclude pertanto ogni possibilità di ricorrere alla media ponderata».

Tuttavia in modo curioso, dopo aver escluso la media ponderata, la circolare stessa, fa un appello ai docenti a fare sì «che il voto conclusivo sia il frutto meditato di una valutazione collegiale delle diverse prove e del complessivo percorso scolastico dei giovani candidati». Si chiede di «evitare possibili appiattimenti, che rischierebbero di penalizzare potenziali 'eccellenze' e di evidenziare i punti di forza nella preparazione dei candidati, anche in funzione orientativa rispetto al proseguimento degli studi». Si esclude tuttavia che per farlo i commissari possano introdurre un *bonus* «In merito alla ipotesi di applicare un *bonus* in analogia all'Esame di Stato conclusivo del secondo ciclo d'istruzione, si fa presente che tale istituto non è contemplato da alcuna norma per l'esame finale del primo ciclo. Pertanto, è da escludere che le commissioni d'esame possano operare in tale senso». Senza ponderazioni e senza *bonus* c'è da chiedersi che cosa possa fare la commissione se non alterare i voti o il risultato del test INVALSI.

Le eccellenze sono nel cuore della tecnocrazia meritocratica, tuttavia l'eccellenza a cui si riferiscono è l'eccellenza standardizzata. Infatti l'algoritmo consente di raggiungere il punteggio 10 solo se si è raggiunto il 10 in quasi tutte le materie considerate.

In sintesi gli elementi che concorrono alla valutazione dell'esame sono 7:

1. voto d'ammissione;
2. prova scritta di italiano;
3. prova scritta di matematica;
4. prova scritta di inglese;
5. prova scritta di seconda lingua comunitaria;
6. colloquio interdisciplinare;
7. prova nazionale INVALSI.

Ciascuno di questi accertamenti dà luogo a un punteggio in decimi che poi deve essere moltiplicato per 0,143 per dare luogo al punteggio finale.

I docimologi osserverebbero che questo approccio presenta problemi di validità del contenuto. Infatti se confrontiamo il valore attribuito a ciascuno degli aspetti sottoposti a verifica con i contenuti del *curriculum* scolastico sono evidenti sopravvalutazioni di alcuni aspetti e sottovalutazioni di altri.

Nella Tabella 1 sono messi a confronto l'organizzazione oraria della Scuola secondaria di primo grado per le classi a tempo normale con i punti su cui verte la valutazione dell'Esame di Stato. Come si può vedere, l'esame

considera due volte l'insieme delle discipline nel voto d'ammissione e nel colloquio finale (anche se realisticamente il ruolo di alcune discipline nel colloquio finale finisce per essere ancillare).

Altre discipline invece contribuiscono specificamente al voto finale. Ad esempio la seconda lingua, pur avendo un orario analogo a Tecnologia, Arte e immagine, Musica, Scienze e Storia, contribuisce autonomamente al voto finale. Altre, infine, come l'italiano e la matematica contribuiscono quattro volte alla determinazione del punteggio finale.

Questo realizza una gerarchia delle discipline evidente a insegnanti, studenti e famiglie che non può non avere ricadute sulla didattica.

È istruttivo mostrare alcune simulazioni dello sviluppo della tecnica di attribuzione dei punteggi e di come l'impianto rigido rischi di far sovrastimare o sottostimare ciò che ci si propone di valutare in modo rigoroso.

Tabella 1. – Confronto Organizzazione oraria - Punti di valutazione.

DISCIPLINE	ORE SETTIMANALI	AMMISSIONE	ITALIANO	MATEMATICA	INGLESE	2 ^a LINGUA	COLLOQUIO	INVALSI
Italiano	6	√	√				√	√
Approfondimento in materie letterarie	1	√	√			-	√	
Storia	2	√					√	
Geografia	1	√					√	
Matematica	4	√		√			√	√
Scienze	2	√					√	
Tecnologia	2	√					√	
Lingua inglese	3	√			√		√	
2 ^a lingua comunitaria	2	√				√	√	
Arte e immagine	2	√					√	
Scienze motorie e sportive	2	√					√	
Musica	2	√					√	
Religione cattolica / Materia alternativa	1							
Totale delle ore settimanali	30							

Prendiamo ad esempio due ipotetici percorsi di studenti con i relativi voti di esame: il primo descrive uno studente che ha un percorso scolastico discreto, e che ha spiccate capacità nelle lingue, il secondo descrive uno studente che ha seguito con ottimo profitto la maggior parte delle discipline, che ha risultati abbastanza buoni nelle lingue straniere (Tabella 2).

In questo caso è evidente la sottovalutazione del giudizio complessivo sull'esperienza scolastica dello studente e la sopravvalutazione del risultato nelle due lingue straniere. L'effetto è illustrato nella Figura 1.

Probabilmente la commissione avrebbe ragionato diversamente, ma sappiamo che la commissione non ha margini per ragionare svolgendo quello che Mars (1982, pp. 66-88) definisce un «lavoro da Asino». Se tuttavia, anche escludendo la possibilità di fidarsi del giudizio degli insegnanti, si fosse utilizzata una ponderazione dei parametri diversa, ad esempio che tenesse conto del peso orario di ciascuna disciplina sul *curriculum* (Tabella 3) attribuendo alla prova INVALSI il peso orario di italiano e matematica, il risultato sarebbe stato diverso e il voto finale dei due studenti si sarebbe invertito.

L'approccio descritto è altrettanto rigido di quello imposto dal Ministero, ma permette di dare più evidenza agli aspetti complessivi della preparazione dello studente, al suo percorso scolastico ed al colloquio finale su tutte le materie. A stime dunque più collegiali e meno soggette a errore della correzione di compiti scritti disciplinari e, come vedremo più avanti, anche della prova nazionale INVALSI.

La rigidità della procedura crea problemi anche per l'effetto delle approssimazioni; infatti il gioco dei decimali interviene per un valore significativo sul voto complessivo.

Tabella 2. – Confronto medie aritmetiche.

	STUDENTE A	STUDENTE B
1 Ammissione	7	10
2 Italiano	7	9
3 Matematica	7	8
4 Inglese	10	7
5 2ª lingua comunitaria	10	7
6 Colloquio	9	10
7 INVALSI	10	8
Media aritmetica	8,57	8,43
Voto finale (arrotondato)	9	8

**Esame finale Secondaria 1° Grado
media aritmetica
Studenti con voto finale 9/10 e 8/10**

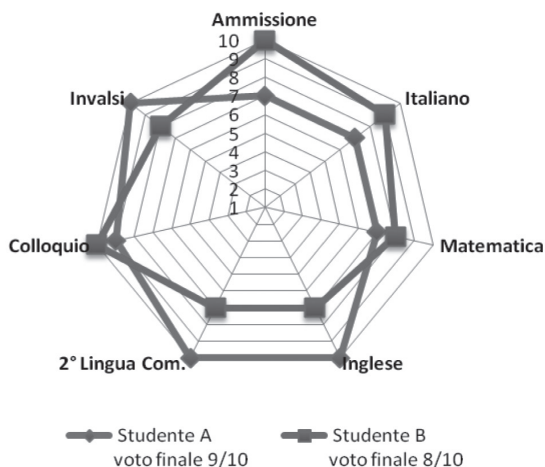


Figura 1. – Rappresentazione dei voti.

Tabella 3. – Valori di ponderazione e confronto medie ponderate.

VALORI PONDERALI	PESO	PESO %	CRITERIO	STUDENTE A (PESO * VOTO)	STUDENTE B (PESO * VOTO)
Ammissione	29	0,34	ore settimanali totali	203	290
Italiano	7	0,08	6 ore italiano + 1 ora approf.	49	63
Matematica	4	0,05	4 ore settimanali	28	32
Lingua inglese	3	0,04	3 ore settimanali	30	21
2ª lingua com.	2	0,02	2 ore settimanali	20	14
Colloquio	29	0,34	ore settimanali totali	261	290
INVALSI	11	0,13	ore Italiano + ore Matematica	110	88
Media ponderata				8,25	9,39
Voto finale (arrotondato)				8	9

Osservando attentamente la Tabella 2 si può notare che la differenza finale di un voto tra i due soggetti è giustificata da una differenza nei punteggi medi di 0,14. Il peso dell'approssimazione incide sul voto finale fino al valore di 3,5 punti rispetto ai parametri di origine.

Nella Tabella 4 vediamo che due studenti con tre punti di differenza nei voti di origine ottengono lo stesso voto finale. Dunque il sistema «preciso» non tiene conto dei due punti in più nel giudizio di ammissione e del punto in più nel voto di italiano.

Tabella 4. – Esempio degli effetti dell'arrotondamento dei decimali.

	STUDENTE A	STUDENTE B
1 Ammissione	8	10
2 Italiano	8	9
3 Matematica	8	8
4 Inglese	7	7
5 2ª lingua comunitaria	7	7
6 Colloquio	8	8
7 INVALSI	10	10
Media aritmetica	8,00	8,43
Voto finale (arrotondato)	8	8

Purtroppo occorrerebbe più spazio per altre simulazioni, ma è evidente che, prima di affidarsi alla rigidità dei calcoli, meglio sarebbe fare un po' di calcoli.

Poiché, chi di calcoli ferisce merita di essere affrontato con i calcoli, vale la pena valutare il ruolo che gioca in questo meccanismo perverso la prova INVALSI. L'INVALSI sin dalle prove pilota ha reso evidenti i limiti di un approccio alla misurazione in cui la necessità di fare presto e tanto si contrapponeva alla esigenza di fare bene (Lucisano, 2003 e 2012b).

Questa prova nazionale interviene misurando due dimensioni la prima chiamata impropriamente «Italiano» (dato che si tratta di comprensione della lettura e di grammatica), la seconda «Matematica». Ora questa prova ha limiti evidenti di validità rispetto al suo stesso impianto teorico (Corsini, 2013) e limiti nella capacità di misurare riferiti alla popolazione in esame.

Nella Tabella 5 vengono riportati i risultati della prova di italiano del 2012, così come restituiti dall'INVALSI (2012/13, fig. 5.7: *Placement della prova di Italiano*, III Scuola sec. I gr.). La figura propone in parallelo la distribuzione degli studenti secondo la misura di abilità (parametro theta) del

modello di Rasch e la distribuzione degli item secondo la difficoltà. Difficoltà e abilità vengono misurate in questa analisi sulla stessa scala di misura. La scala di misura nella figura va da 3 a -3. In parole semplici il valore 0, se lo consideriamo come indicatore di difficoltà, corrisponde a un item di media difficoltà e, se lo consideriamo come misura dell'abilità, a uno studente di medie capacità.

Gli studenti sono rappresentati graficamente dalle X che esprimono il numero riproposto nella colonnina *STUD*, mentre gli item vengono riportati nella colonna *item* con il numero che corrisponde alla loro difficoltà nella prova. Nella colonna *STUDCUM* vengono sommati progressivamente gli studenti e nella colonna *%CUM* viene riportata la percentuale cumulata relativa allo stesso dato.

Ora ci si attenderebbe che la distribuzione dei soggetti e degli item sia tale da proporre due curve normali opposte, invece il nostro grafico mostra come il 41% degli studenti presenti un livello di abilità coperto solo da 3 domande della prova, mentre in fondo al grafico troviamo ben 11 domande a misurare un livello di abilità molto basso attribuibile solo al 3% degli studenti. Si tratta dunque di uno strumento di misura fortemente carente in particolare per la misura delle capacità medio alte, che si sarebbe dovuto rivedere sostanzialmente e ritarare.

L'INVALSI o non ha avuto tempo per effettuare *try out* o ha sbagliato il campione per la taratura della prova, o ancora aveva queste informazioni e ha deciso o dovuto usare comunque la prova perché ormai è un adempimento di legge. Questa prova misura male. Misura come una bilancia per pesare persone che avesse come valore massimo 50 chili e che dunque restituisse il valore 50 a tutti quelli che lo superano.

Ancor più interessante è vedere come vengono utilizzati i punteggi di questa prova. Infatti, sebbene si disponga dei punteggi ricavati dall'analisi di Rasch che pesa il valore delle singole domande per l'uso nelle prove finali, si procede alla semplice somma delle risposte corrette, che poi viene convertita in un voto in decimi sulla base della Tabella 6.

Come si può osservare, la conversione procede dalla somma dei punteggi ottenuti nelle due prove alla traduzione in decimi per fasce di punteggio. Qui il calcolo confonde i dati rilevati: uno studente eccellente nella comprensione della lettura e assolutamente non in grado di collaborare con i numeri finisce per avere lo stesso punteggio di uno studente di sufficienti capacità nell'una e nell'altra area di abilità.

Questa prova, tuttavia, nel giudizio finale equivale come valore al voto di ammissione e al risultato complessivo della prova di esame. Un'eventuale difficoltà momentanea di uno studente in questa prova non potrà essere considerato dalla commissione che è chiamata solo a fare calcoli.

Tabella 5. – INVALSI, 2012/13, p. 69: rielaborazione della figura 5.7, «Placement della prova di Italiano», classe III Scuola secondaria di I grado.

MISURA		STUD	ITEM	STUDCUM	%CUM
3	X	16		16	0,2
	XXX	47		63	0,6
			37		
2	XXXXXXXXXX	141		204	2,1
	XXXXXXXXXXXXXXXXXX	220		424	4,4
	XXXXXXXXXXXXXXXXXXXXXX	298		722	7,4
	XXXXXXXXXXXXXXXXXXXXXXXXXX	361		1083	11,1
1	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	424		1507	15,5
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	487		1994	20,5
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	502		2496	25,6
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	502	22 27	2999	30,8
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	518		3517	36,1
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	487		4004	41,1
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	502	2	4506	46,3
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	471	26 47	4977	51,1
0	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	471	14 38	5448	56
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	455	4 18 44	5903	60,6
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	408	30 48	6311	64,8
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	393	29	6704	68,9
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	408	8 17	7112	73,1
	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	628	24 41 49	7740	79,5
	XXXXXXXXXXXXXXXXXXXXXX	251	10	7991	82,1
	XXXXXXXXXXXXXXXXXXXXXX	236	1	8227	84,5
-1	XXXXXXXXXXXXXXXXXXXXXX	220	13 15 20 25 33 34	8447	86,8
	XXXXXXXXXXXXXXXXXXXXXX	204	23 43	8651	88,9
	XXXXXXXXXXXXXXXXXXXXXXXXXX	314	16 42 45	8965	92,1
	XXXXXXXXXX	126	12 35 39	9090	93,4
	XXXXXXX	110	19 21	9200	94,5
	XXXXXXX	94		9294	95,5
	XXXXXX	79		9373	96,3
	XXXXX	63	6	9436	96,9
-2	XXX	47	40	9483	97,4
	XXX	47	28	9530	97,9
	XXX	47	3 9	9577	98,4
	XX	31	32	9608	98,7
	XX	31	5 11 46	9640	99
	XX	31	36	9671	99,4
	X	16		9687	99,5
	X	16	31	9703	99,7
	X	16		9718	99,8
				9718	99,8
-3				9718	99,8
	X	16		9734	100

Tabella 6. – Conversione Punteggio-Voto.

SOMMA DEI PUNTEGGI DI ITALIANO E MATEMATICA	VOTO UNICO IN DECIMI
Minore o uguale a 40	4
41-54	5
55-64	6
65-73	7
74-82	8
83-91	9
92-100	10

Si capiscono le lamentele e gli appelli che vengono dal mondo degli insegnanti che avvertono il sistema come ingiusto, dei genitori e dell'associazionismo scolastico. Si capisce come la richiesta della stessa circolare incoraggi il *cheating*³ finalizzato a valorizzare le eccellenze.

Una prova strutturata che restituisca ai docenti misure sui risultati ottenuti dagli studenti per singola area di competenza, può essere un riferimento utile per le commissioni chiamate a valutare l'esito di del percorso scolastico di ciascun singolo studente, utile per l'autovalutazione delle scuole, utile per chi è chiamato a studiare e a valutare i fenomeni scolastici. Ridurre l'esito di prove su aspetti diversi della preparazione dei ragazzi a un parametro unico da utilizzare in un algoritmo contrasta, invece, in primo luogo con la regola che misure di dati su aspetti diversi (matematica e italiano o peso e altezza), non consentono l'uso della media, e poi con il buon senso che dice che nel valutare un percorso triennale di lavoro l'insuccesso in una delle prove non può condizionare il giudizio sul percorso stesso.

Ora si tratta di tornare al tema di questo lavoro: la qualità e in particolare la qualità della quantità. Abbiamo dimostrato che la quantità non qualificata produce maggiori danni della stima operata sulla base del buon senso. Non per questo si deve abbandonare la ricerca di elementi quantitativi sui quali formulare dei giudizi. Compito della ricerca è conseguire dove possibile quantificazioni qualificate, anche se il giudizio, soprattutto il giudizio educa-

³ Mars mette in luce come un lavoro in cui si è chiamati ad una attività esecutiva in cui il controllo è elevato e la coesione di gruppo basse conduce a risposte di *cheating*. «Se una persona non può adattare i suoi valori, credenze e atteggiamenti a ciò che gli viene richiesto di fare sul lavoro, può in primo luogo rassegnarsi e rimuovere il conflitto di coscienza, in secondo luogo può avere un crollo emotivo o può cercare di sopravvivere alle contraddizioni in questo stato di conflitto che può essere chiamato alienazione [...]. In questa situazione egli può rispondere attraverso [...], sabotaggio, assenteismo o l'imbroglione» (Mars, 1982, p. 35).

tivo non discende unicamente dalla considerazione di parametri quantitativi. La ricerca non solo deve misurare con misura, intesa nel senso suggerito da Visalberghi (1955, p. 11) «come abito di equilibrio e di discrezione»; deve anche misurare con misure qualificate e non usare come misure cose che con le misure hanno poco o niente a che vedere.

Gli elementi quantitativi debbono essere messi a disposizione della commissione, che però dovrebbe avere il compito di formulare un giudizio con fini principalmente educativi, capace di premiare il merito, ma anche l'impegno, capace di valorizzare eccellenze che, proprio per la loro natura di eccellenze, possono non raggiungere valori altrettanto soddisfacenti in tutte le discipline. Uno studente eccellente non è uno studente normale (standard) anche in un livello scolastico che si è chiamato per anni scuola media. Quale è il fine dell'esame al termine del ciclo secondario inferiore? Serve come recita lo Statuto dell'INVALSI a valorizzare il capitale umano o a orientare i ragazzi in una fase difficile dello sviluppo della loro personalità? Serve a premiarli e punirli o a incoraggiarli ad apprendere e valorizzare le loro attitudini e le loro inclinazioni⁴?

Il ruolo del fine nella formulazione di valutazioni è sottolineato in tutta la letteratura su questo argomento dalla *Critica del giudizio* di Kant alla *Teoria della valutazione* di John Dewey e forse sarebbe ragionevole chiedersi seriamente qual è il fine del giudizio che si esprime su uno studente al termine di un ciclo di studi e riesaminare serenamente alla luce del fine tutta la procedura da utilizzare per raggiungerlo al meglio. E se i mezzi non sono indipendenti dai fini merita anche riconsiderare il ruolo dei ricercatori e della loro responsabilità deontologica nel predisporre mezzi (misurare) per fini definiti da altri e che dichiarano di non condividere⁵ (Lucisano, 2012a).

⁴ «La mercificazione dell'apprendimento genera una contraddizione fondamentale fra i valori di uso e di scambio del prodotto dell'apprendimento, che si manifesta sotto forma di conflitti fra l'apprendere per sapere e l'apprendere per esibire conoscenze nel contesto di una valutazione [...]. L'esame diventa quindi una nuova pratica parassitaria, il cui scopo è aumentare il valore di scambio dell'apprendimento, indipendentemente dal suo valore d'uso» (Lave & Wenger, 2006, p. 16).

⁵ «Far dipendere esclusivamente dai risultati di poche prove standardizzate relative ad alcuni ambiti tematici e disciplinari il successo scolastico del singolo studente (e/o la valutazione del singolo insegnante) rischierebbe di indurre una eccessiva focalizzazione sulle stesse dell'attenzione degli operatori della scuola» (Cipollone, Montanaro, & Sestito, 2012, p. 22).

RIFERIMENTI BIBLIOGRAFICI

- Andreani Dentici, O. (1968). *Abilità mentale e rendimento scolastico*. Firenze: La Nuova Italia.
- Benvenuto, G. (2003). *Mettere i «voti» a scuola. Introduzione alla docimologia*. Roma: Carocci.
- Cipollone, P., Montanaro, P., & Sestito, P. (2012). Il capitale umano per la crescita economica: possibili percorsi di miglioramento del sistema d'istruzione, in Italia. *Questioni di Economia e Finanza (Occasional Papers)*, 122, Aprile.
- Corsini, C. (2012). La lezione americana: l'impiego del Valore Aggiunto nella valutazione di scuole e insegnanti. *Scuola Democratica*, n.s., 6, 60-75.
- Corsini, C. (2013). Validità di contenuto delle prove INVALSI di comprensione della lettura. *Giornale Italiano della Ricerca Educativa*, 6(10), Giugno, 46-60.
- Dewey, J. (1939a). *Theory of valuation*. Chicago, IL: University of Chicago Press (trad. it., *Teoria della valutazione*. Firenze: La Nuova Italia, 1967).
- Dewey, J. (1939b). Unity of science as social problem. In O. Neurath, R. Carnap, & C. Morris (Eds.), *Foundations of the unity of science. Toward an international encyclopedia of unified science* (pp. 29-38). Chicago, IL - London: University of Chicago Press (trad. it. in P. Lucisano, L'unità della scienza come problema sociale di John Dewey. *Cadmo*, 22, 2000, 33-37).
- Kozol, J. (2000). Foreword. In D. Meier, *Will standards save public education?* Boston: Beacon Press.
- INVALSI (2012/13). *Rilevazioni nazionali sugli apprendimenti 2012/13*. http://www.invalsi.it/snvpn2013/rapporti/Rapporto_SNV_PN_2013_DEF_11_07_2013.pdf.
- Lalande, A. (1971). *Dizionario critico di filosofia*. Milano: ISEDI.
- Lave, J., & Wenger, E. (1991). *Situated learning. Legittimate peripheral participation*. Cambridge: Cambridge University Press (trad. it., *L'apprendimento situato. Dall'osservazione alla partecipazione attiva nei contesti sociali*. Trento: Erikson, 2006).
- Lucisano, P. (2003). Validità e affidabilità delle pratiche valutative: a proposito del Progetto Pilota 2. *Cadmo*, 37-56.
- Lucisano, P. (2012a). Responsabilità sociale, valutazione e ricerca educativa. *Giornale Italiano della Ricerca Educativa*, Numero speciale, 5, Ottobre, 13-20.
- Lucisano, P. (2012b). Prove di comprensione. *Scuola Democratica*, 6, Ottobre.
- Mars, G. (1982). *Cheats at work: An anthropology of workplace crime*. London: UNWIN.
- Pirsig, R. M. (1974). Zen and the art of motorcycle maintenance (trad. it., *Lo Zen e l'arte della manutenzione della motocicletta*. Milano: Adelphi, 1990).
- Visalberghi, A. (1955). *Misurazione e valutazione nel processo educativo*. Milano: Edizioni di Comunità.
- Woods, P. A. (2005). *Democratic leadership in education*. London: Paul Chapman Publishing.

RIASSUNTO

L'articolo presenta la metodologia seguita per la costruzione e la taratura di una prova di abilità linguistiche per studenti in uscita dal triennio della formazione professionale iniziale. Il lavoro si è sviluppato all'interno di una indagine dell'ISFOL che considerava tre aree di abilità, linguistiche, matematiche e abilità strategiche per lo studio. Lo strumento tarato su un campione di giudizio di 1046 studenti misura capacità di ascolto, comprensione della lettura, uso corretto della lingua e produzione scritta. Vengono illustrate le metodologie di item analisi CTT («Classical Test Theory») and IRT («Item Response Theory») e di costruzione di variabili derivate dalle diverse prove. In particolare vengono presentati i vantaggi dell'uso del modello di Rasch per una più efficace misura del tratto latente abilità e per la taratura della prova. Nella parte conclusiva del lavoro vengono illustrati i principali risultati dell'analisi multivariata della ricerca, risultati vanno letti ed interpretati, ovviamente, alla luce delle caratteristiche della popolazione esaminata, della estrazione sociale e culturale complessivamente piuttosto debole, e della storia di fallimenti scolastici che caratterizza l'esperienza di questi studenti.

Parole chiave: Esami, Misurazione, Qualità delle misure, Test, Valutazione.