

INDUZIONI è una rivista rivolta agli studenti ed ai docenti di matematica, storia, geografia, osservazioni scientifiche, economia, statistica... delle scuole pre-universitarie, ma anche ai docenti universitari. Il suo scopo è quello di diffondere idee statistiche nella scuola e di illustrare attraverso esempi come la statistica possa essere di ausilio in diverse occasioni della vita pratica.

Il taglio degli articoli dovrebbe essere operativo ed effettivamente utilizzabile nel lavoro scolastico.

INDUZIONI viene pubblicata con il patrocinio della Società Italiana di Statistica (S.I.S.). Inoltre opera in associazione con la rivista «Teaching Statistics».

Comitato scientifico: Giuseppe Burgio, Giuseppe Cicchitelli, Giuseppe De Bartolo, Peter Holmes, Maria Gabriella Ottaviani, Antonella Pinnelli, Silio Rigatti Luchini, Livia Linda Rondini, Carla Rossi, Romano Scozzafava, Alberto Zuliani.

Fondatore: † Enzo Lombardo.

Direttore: Maria Gabriella Ottaviani.

Comitato di redazione: Mario Barra, Anna Maria Birindelli, Giulianella Coletti, Corrado Crocetta, Ornella Giambalvo, Maria Pannone, Isabella Santini, Maria Rita Sebastiani.

Il Direttore e il Comitato di redazione della rivista ringraziano gli esperti esterni che, con la lettura critica e costrutti degli articoli sottoposti per la pubblicazione, hanno contribuito alla qualità di «Induzioni».

INDUZIONI

Demografia, probabilità, statistica
a scuola

32 · 2006



Istituti editoriali
e poligrafici internazionali

Pisa · Roma

Rivista semestrale

Amministrazione e abbonamenti:
Accademia editoriale®
Casella postale n. 1 Succ. n. 8 · I 56123 Pisa

Prezzi di abbonamento (2006):
Italia: € 55,00 (privati), € 110,00 (enti, con edizione *Online*);
Abroad: € 110,00 (*Individuals*), € 140,00 (*Institutions, with Online Edition*);
cumulativo-sostenitore: € 1200,00,
dà diritto a ricevere 35 copie di ogni numero della rivista
Prezzo del fascicolo singolo: € 60,00

Modalità di pagamento: su c/c postale n. 17154550 intestato a
Accademia editoriale · Casella postale n. 1 Succ. n. 8 · I 56123 Pisa;
mediante carta di credito American Express, Eurocard, Mastercard, Visa.
Tel. 050 542332 · Fax 050 574888 · e-mail: iepi@iepi.it

www.libraweb.net

Direttore responsabile: Lucia Corsi.
Autorizzazione del Tribunale di Pisa n. 8 del 2.5.1994.

*

La casa editrice garantisce la massima riservatezza dei dati forniti dagli abbonati e la possibilità di richiederne la rettifica o la cancellazione previa comunicazione alla medesima. Le informazioni custodite dalla casa editrice verranno utilizzate al solo scopo di inviare agli abbonati nuove proposte (Dlgs. 196/2003).

Sono rigorosamente vietati la riproduzione, la traduzione, l'adattamento, anche parziale o per estratti, per qualsiasi uso e con qualsiasi mezzo effettuati, compresi la copia fotostatica, il microfilm, la memorizzazione elettronica, ecc., senza la preventiva autorizzazione scritta degli *Istituti editoriali e poligrafici internazionali*®, Pisa · Roma, un marchio della *Accademia editoriale*®, Pisa · Roma.
Ogni abuso sarà perseguito a norma di legge.

Proprietà riservata · All rights reserved
© Copyright 2006 by *Istituti editoriali e poligrafici internazionali*®, Pisa · Roma,
un marchio della *Accademia editoriale*®, Pisa · Roma

ISSN 1120-690X
ISSN ELETTRONICO 1724-0476

SOMMARIO
N. 32 · GENNAIO-GIUGNO 2006

O. GIAMBALVO, A. M. MILITO, A. M. OLIVERI, <i>La valutazione di una prova di profitto per l'apprendimento della statistica: confronto tra due tecniche</i>	9
M. GNALDI, <i>Conoscenze numeriche di base e conseguenti capacità di ragionamento statistico a livello universitario: i risultati di una ricerca condotta in Inghilterra</i>	29

ESPERIENZE E MATERIALI

A. FREDDI, <i>Alcune considerazioni sull'insegnamento della matematica finanziaria e della matematica attuariale nella scuola secondaria superiore</i>	39
G. P. ZACCOMER, <i>Un'esperienza di integrazione della didattica universitaria della statistica con un'applicazione a dati economici reali: il progetto Movimprese</i>	57
P. MARIANI, <i>Lo statistico in 'miniera'</i>	67
R. RICCI, <i>Elementi per la valutazione di prove strutturate</i>	73
P. QUATTO, <i>Statistica e sondaggi politico-elettorali</i>	85

ATTIVITÀ IN CAMPO DIDATTICO

F. F. MAIORANO, G. BERARDI, <i>Un'indagine di customer satisfaction</i>	93
A. DRIVET, <i>Modelli per analisi ecografiche</i>	107
M. F. LOPORCARO, B. BALDAZZI, S. F. ALLEGRA, <i>Statistica, conoscenza e apprendimento: un'esperienza di incontro tra ricercatori dell'Istat e studenti delle scuole medie superiori</i>	119

QUADRANTE INTERNAZIONALE

P. MONARI, <i>ICOTS7 - 7th International Conference on Teaching Statistics</i>	127
M.G. OTTAVIANI, <i>Joint ICMI/IASE Study: annuncio e richiesta di lavori per la Conferenza</i>	129

LO SCAFFALE DEI LETTORI

M. BERRONDO-AGRELL, <i>Eurêka. 100 Énigmes. Probabilités, décision, jeux</i> (Ludovico Piccinato)	133
---	-----

ELEMENTI PER LA VALUTAZIONE DI PROVE STRUTTURATE

ROBERTO RICCI¹

1. INTRODUZIONE

Negli ultimi anni la valutazione del sistema scolastico e delle diverse azioni formative da esso realizzate ha assunto una rilevanza ed una visibilità molto maggiori rispetto al passato. Sempre più si richiama la necessità di diffondere un'adeguata cultura della valutazione tra tutti i soggetti coinvolti nel processo di insegnamento e apprendimento (Boeti 2005). Inoltre l'esigenza di confrontare risultati ottenuti in situazioni e contesti differenti ha accentuato il bisogno di adeguati metodi qualitativi e quantitativi che siano in grado di garantire la fondatezza di confronti ed analisi comparative.

In questo quadro in rapida evoluzione la Scuola sembra muoversi con qualche difficoltà, probabilmente disorientata dalla diffusione dei risultati di ricerche nazionali ed internazionali le cui finalità e metodologie non sono state presentate in modo adeguatamente approfondito ed esplicito. Una reale cultura della valutazione si basa necessariamente sulla condivisione e sulla comprensione di metodi e finalità che devono essere intesi a fondo da tutti i soggetti che operano a vario titolo nel mondo della Scuola. La ricaduta sulla didattica quotidiana dell'applicazione di metodi statistici per la valutazione dei risultati delle diverse prove può essere veramente utile per i docenti, ma essa non può realizzarsi se non attraverso l'acquisizione della consapevolezza del significato e dei limiti degli strumenti utilizzabili.

Se da un lato le ricerche nazionali ed internazionali che hanno coinvolto la Scuola in questi ultimi anni hanno reso più evidente l'importanza di disporre di adeguati metodi quantitativi di analisi, dall'altro la mancata esplicitazione delle profonde differenze tra i processi di valutazione dell'intero sistema educativo (valutazioni di macrosistema) e quelli realizzabili all'interno di ciascuna classe e delle singole istituzioni scolastiche (valutazioni di microsistema) non ha contribuito certamente alla diffusione ed al consolidamento della cultura della valutazione. Sovente le scuole ed i docenti non riescono a trovare immediata applicazione delle metodologie proposte in alcune ricerche come quelle dell'OCSE-PISA (Mignani e Ricci 2005; Ottaviani *et alii* 2006) e di ALL o quelle dell'Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e Formazione proprio a causa della non sempre chiara distinzione dei due piani di valutazione: quello di macrosistema e quello di microsistema. Se per il primo tipo di valutazione le istituzioni sono i destinatari ed i fruitori dei risultati delle singole ricerche, per la seconda tipologia invece ogni scuola ed ogni docente può esercitare un ruolo attivo in modo da consolidare competenze metodologiche che arricchiscano la sua professionalità specifica.

¹ Istituto Regionale di Ricerca Educativa. Emilia Romagna. E-mail: ricci@irreer.it.

2. PROVE STRUTTURATE

Ogni docente ha tra i suoi compiti quello di valutare il grado di raggiungimento di obiettivi perseguiti dagli studenti tramite il processo di insegnamento/apprendimento. A tale scopo le prove strutturate possono rappresentare un utile supporto per effettuare valutazioni formative e sommative¹.

Delle prove strutturate si mette di solito in evidenza il fatto che sono “oggettive”. Tuttavia la predisposizione di prove “oggettive” richiede un insieme di scelte “soggettive” le cui implicazioni devono essere note al fine di comprendere il valore ed i limiti delle prove “oggettive” stesse (Pannone 2000).

Dal punto di vista tecnico diverse sono le scelte che possono essere operate, in particolare, relativamente alla tipologie delle domande da inserire nella prove. Nel presente lavoro si farà riferimento ad una sola tipologia di domanda (*item*) largamente applicata in diversi contesti educazionali, ovvero agli *item* a risposta multipla. Essi posseggono utili caratteristiche per la valutazione di microsistema, ma richiedono un’attenta analisi per poter cogliere a pieno il valore delle informazioni fornite.

2. 1. Domande a risposta multipla

In diverse fasi della valutazione, sia formativa che sommativa, una delle tipologie di prove che trova ampia applicazione è rappresentata dai questionari con domande a risposta chiusa, note anche come *multiple-choice item*. Essi presentano alcuni indubbi vantaggi che possono costituire un utile punto di partenza, per quanto non esaustivo, per la formulazione di prove per la verifica di abilità e competenze. I *multiple-choice item* sono particolarmente utili per la loro versatilità. Infatti essi possono essere impiegati per formulare domande relative all’analisi di fenomeni di varia natura, alla comprensione di concetti e principi, all’interpretazione di relazioni di causa ed effetto, alla formulazione di inferenze partendo da un insieme di dati, ecc. La maggiore rapidità con la quale è possibile rispondere ai *multiple-choice item* rispetto alle domande a risposta aperta consente di verificare un insieme più ampio di argomenti e quindi il punteggio conseguito dal rispondente può essere maggiormente rappresentativo della sua preparazione in un ambito dell’indagine. I *multiple-choice item* sono strumenti validi e duttili per la valutazione e l’autovalutazione del raggiungimento di determinati obiettivi formativi anche in caso di azioni di insegnamento non in presenza. Tuttavia non è certo possibile immaginare che i *multiple-choice item* siano una panacea per tutte le situazioni di valutazione degli apprendimenti. Essi non rappresentano la scelta più adeguata quando è necessario che il rispondente fornisca argomentazioni ricche ed articolate o proponga elaborazioni personali ed originali, per individuare il processo di apprendimento dello studente su un particolare argomento.

Sovente i *multiple-choice item* non vengono utilizzati poiché si teme che il rispondente scelga le opzioni a caso e non per convincimento, con l’evidente rischio

¹ Come noto in letteratura, per valutazione formativa si intende il processo che accompagna costantemente l’azione didattica in tutte le sue fasi, mentre la valutazione sommativa è l’azione conclusiva del processo di insegnamento/apprendimento.

di conseguire, almeno in linea teorica, una votazione sufficiente anche senza una preparazione adeguata. In primo luogo è opportuno osservare che il problema si pone solo se un tale tipo di questionario è utilizzato per la valutazione sommativa, mentre questa riserva si scioglie qualora ci si trovi ad operare nel contesto della valutazione formativa. In questa condizione lo studente non ha alcun interesse a rispondere a caso poiché il risultato del questionario non gli servirà per conseguire una votazione, ma semplicemente per conoscere il livello di preparazione raggiunta. È tuttavia opportuno osservare che il rischio di ottenere una valutazione almeno sufficiente mediante una scelta a caso è stato spesso sopravvalutato. Infatti, se si immagina di fissare al 60% la percentuale di risposte corrette ad un questionario necessaria per conseguire una votazione sufficiente, la probabilità di raggiungere almeno tale soglia scegliendo le risposte a caso può essere analizzata considerando i dati della TABELLA 1.

TABELLA 1. Probabilità di risposta a caso.

Numero di <i>multiple-choice</i> item in un questionario di item tutti a quattro alternative	Probabilità (in %) di conseguire un punteggio di almeno il 60% fornendo risposte a caso
5	1,57
10	0,037
15	0,00093
20	0,00002

Come si può facilmente notare, anche per un questionario con poche domande (10-15) la probabilità di ottenere una valutazione almeno sufficiente mediante risposte formulate puramente a caso è veramente trascurabile¹.

2. 2. Tipologie di *multiple-choice* item

Anche se solitamente i *multiple-choice* item vengono considerati come una unica tipologia di domande che può essere inserita in un questionario, essi in realtà rappresentano una categoria non omogenea, le cui singole modalità devono essere attentamente valutate.

In primo luogo va detto che un *multiple-choice* item è rappresentato da uno stimolo (*stem*) e da una lista di alternative. Lo stimolo può essere formulato come una domanda o una affermazione incompleta. In termini molto generali la lista delle opzioni contiene la risposta corretta o la migliore alternativa ed un certo numero di risposte scorrette o comunque non preferibili alla migliore. Le opzioni non corrette vengono solitamente definite *distrattori* e la loro formulazione richiede particolare cura ed attenzione.

Le due tipologie di *multiple-choice* item più conosciute ed utilizzate sono quelle che nella lista delle alternative prevedono o una sola risposta esatta oppure una risposta preferibile tra un insieme di alternative. Un altro tipo possibile di *multiple-choice* item è solitamente definito *multiple-choice negativo* in cui il rispondente deve

¹ Si osservi che le probabilità riportate nella seconda colonna della TABELLA 1, a parità di numero di item, decrescono al crescere del numero delle opzioni di risposta per ciascuna domanda.

individuare la risposta errata o, comunque, la peggiore tra quelle proposte. Quest'ultima tipologia mal si adatta alla valutazione degli apprendimenti poiché i risultati raggiunti da uno studente si misurano meglio partendo da ciò che questi conosce positivamente, piuttosto che dalla selezione di un'opzione scorretta che, di per sé, non implica la conoscenza di quella ad essa preferibile (Nitko 2004).

Un altro tipo di *multiple-choice item* è quello con più alternative corrette. I metodi più utilizzati per attribuire il punteggio a questa tipologia di *multiple-choice item* sono fondamentalmente ascrivibili a due categorie. La prima prevede l'attribuzione del punteggio previsto solo nel caso in cui lo studente individui tutte le risposte esatte e non selezioni nemmeno un distrattore. La seconda invece prevede l'assegnazione di un punto per ogni risposta esatta selezionata e uno per ogni distrattore non selezionato. Il secondo metodo è preferibile al primo. Quest'ultimo, infatti, attribuisce lo stesso punteggio ad allievi con performance molto differenti, infatti non riceve alcun punteggio sia lo studente che semplicemente non è stato in grado di individuare una delle opzioni corrette o ne ha indicata una scorretta sia lo studente che non abbia saputo individuare alcuna alternativa corretta.

La maggior parte degli autori che si occupano del problema della valutazione mediante *multiple-choice item* sconsigliano l'uso di item con più alternative corrette di domande poiché determinano un eccessivo innalzamento della difficoltà della domanda ed una scarsa attendibilità dei risultati (Haladyna e Downing 1989).

I *multiple-choice item* a risposta combinata rappresentano un'altra scelta possibile quando si vogliono formulare delle domande a risposta multipla. Questa tipologia prevede la possibilità di inserire fra le opzioni di risposta anche un'alternativa che contenga la combinazione di altre. Tuttavia numerosi studi indicano che *multiple-choice item* così costruiti forniscono risultati scarsamente attendibili e con una limitata capacità di distinguere tra rispondenti con diversa preparazione (Haladyna e Downing 1989).

2. 3. Costruzione dei *multiple-choice item*: alcune indicazioni

La formulazione di un *multiple-choice item*, come di qualsiasi domanda, richiede un'attenta esplicitazione e valutazione degli obiettivi per cui esso viene formulato. Per quanto l'indicazione possa sembrare banale, essa rappresenta un punto fondamentale per la stesura di domande efficaci sia sotto il profilo educativo sia sotto quello statistico-misuratorio (Baker 1971).

È opportuno che la formulazione dello stimolo avvenga in forma concisa, chiara e riservando particolare attenzione alla presenza di negazioni. Quest'ultimo aspetto deve sempre essere considerato congiuntamente alle diverse alternative di risposta per evitare che si vengano a costruire associazioni domanda-risposta con doppie negazioni e quindi difficilmente interpretabili per il rispondente.

Alla formulazione delle alternative deve essere riservata particolare cura poiché essa può esercitare un'influenza notevole sull'attendibilità dei risultati di un questionario costituito da *multiple-choice item*. È opportuno che le diverse opzioni mantengano un certo grado di omogeneità sia per quanto riguarda la forma, la lunghezza e, soprattutto, il contenuto. Il rispetto di questi criteri migliora notevolmente la qualità delle domande, garantendo, di conseguenza, risultati maggior-

mente informativi sul grado di conseguimento degli obiettivi oggetto di valutazione.

Sovente una delle opzioni di risposta di un *multiple-choice item* è “tutte le risposte precedenti” oppure “nessuna delle precedenti”. Nella maggior parte dei casi l’inserimento di una tale alternativa, o persino di entrambe, risponde più ad un’esigenza dell’estensore della domanda che ad una effettiva necessità d’indagine circa l’obiettivo oggetto di studio. Ciò dovrebbe portare ad un ripensamento del numero delle alternative da inserire e ad una sua conseguente riduzione. È infatti noto che opzioni come “tutte le risposte precedenti” oppure “nessuna delle precedenti” determinino una sensibile riduzione della capacità di un item di distinguere tra rispondenti con una preparazione diversa ed un altrettanto forte calo del valore informativo dei risultati ottenuti (Thissen *et alii* 1989).

3. ANALISI STATISTICA DEI RISULTATI

L’analisi statistica dei risultati di un questionario formato da *multiple-choice item* è fondamentale se si vuole utilizzare a pieno l’insieme ricco ed articolato delle informazioni che da esso si possono desumere.

Prima di procedere alla valutazione di alcuni aspetti tecnico-statistici è bene effettuare alcune considerazioni preliminari di carattere generale per chiarire l’ambito metodologico di riferimento e, conseguentemente, la portata effettiva dei risultati delle analisi che si possono condurre.

Sotto il profilo metodologico le analisi statistiche che saranno di seguito proposte si collocano nel contesto della Teoria Classica dei Test psicometrici (Lord e Novick 1968). Essa consente di effettuare elaborazioni ed ottenere risultati estremamente rilevanti dal punto di vista educativo proponendo delle soluzioni operative in buona parte realizzabili con semplici strumenti di calcolo e quindi applicabili in qualsiasi contesto scolastico, sia a livello dell’intera istituzione sia a livello di singola classe o gruppo di classi.

La recente diffusione dei risultati delle ricerche OCSE-PISA, ALL ed INValSI ha notevolmente fatto aumentare l’interesse di tutti gli attori della Scuola verso la valutazione e pone la necessità di comprendere il possibile rapporto di queste indagini con l’azione valutativa quotidiana e periodica attuata istituzionalmente. Le ricerche internazionali si pongono sul piano delle valutazioni di macro-sistema e quindi utilizzano metodologie differenti da quelle attuabili in classe, ovvero in un micro-sistema. Infatti l’analisi dei risultati della rilevazione OCSE-PISA non si colloca nell’ambito della teoria classica dei test psicometrici, ma della *Item Response Theory* (van der Linden e Hambleton 1997). Le assunzioni teoriche di quest’ultima teoria sono differenti rispetto a quelle della teoria classica e, certamente, più coerenti con la valutazione dei processi di insegnamento ed apprendimento, ma richiedono ampie dimensioni campionarie, strumentazioni e competenze statistico-metodologiche che difficilmente possono trovarsi integralmente a livello di singola istituzione scolastica. Va inoltre osservato che per le valutazioni di micro-sistema la teoria classica dei test mantiene tutta la sua validità e può costituire un ottimo punto di partenza per un’efficace ed effettiva diffusione della cultura della valutazione (Muñiz 1998).

TABELLA 2. Valori (unitari) ideali di difficoltà.

Numero alternative	Difficoltà ideale
2	0,75
3	0,66
4	0,62
5	0,60

3. 1. Difficoltà di un item

La misurazione in senso statistico della difficoltà di un item (*ID*) è molto semplice. Infatti la misura di *ID* è la percentuale di studenti che hanno risposto correttamente all'item considerato. Solitamente nella letteratura psicometrica, specie in lingua inglese, la difficoltà è indicata con il termine *p-value*. Tuttavia in questo lavoro tale convenzione terminologica non sarà seguita poiché essa coincide con un altro concetto statistico profondamente differente ed indicato con lo stesso termine anche da strumenti informatici molti diffusi come il foglio elettronico; ciò può ovviamente creare confusione a chi non ha consuetudine con i temi e i concetti della statistica.

Il campo di variazione dell'*ID* è ovviamente da 0% a 100% o, in termini unitari, da 0 ad 1. Tanto più il valore dell'*ID* si avvicina ad 1 tanto più l'item è da considerarsi facile, viceversa se il valore dell'*ID* tende a 0.

La TABELLA 2 riporta i livelli di difficoltà auspicabili in relazione al numero di alternative di una domanda a risposta multipla ormai consolidati e comunemente accettati nella letteratura psicometrica.

In genere item con *ID* maggiori di 0.90 o minori di 0.20 devono essere scartati perché, rispettivamente, troppo facili o troppo difficili e quindi scarsamente informativi sul processo di apprendimento del gruppo di studenti che hanno risposto a quella domanda.

3. 2. Potere discriminante di un item

La valutazione della qualità di una singola domanda deve essere effettuata anche con riferimento all'intera prova di cui essa fa parte. In linea del tutto generale è auspicabile che studenti che conseguono un alto punteggio complessivo dell'intera prova tendano a rispondere positivamente alla domanda oggetto di analisi e viceversa per quelli con un punteggio basso. Tale proprietà è comunemente indicata come *potere discriminante* di un item. Una domanda con un alto potere discriminante sarà quindi in grado di distinguere adeguatamente rispondenti che hanno raggiunto l'obiettivo oggetto di verifica da quelli che invece non l'hanno conseguito. Dal punto di vista statistico il potere discriminante di un item si calcola mediante il *coefficiente di correlazione punto-biserial* (*PBS*), solitamente indicato con r_{pbi} . Un semplice esempio può essere di aiuto per comprendere le modalità di calcolo di r_{pbi} e la sua interpretazione.

Si supponga di prendere in esame un test somministrato a 12 studenti e composto da 10 domande alle quali è attribuito il punteggio 1 nel caso di risposta corretta e 0 altrimenti. La TABELLA 3 riporta i risultati del test.

TABELLA 3. Risultati di un test.

		Item										Punteggio totale
		1	2	3	4	5	6	7	8	9	10	
Studente	1	1	1	1	0	1	1	0	1	0	1	7
	2	1	1	1	0	1	0	0	1	0	1	6
	3	1	1	1	1	1	1	1	1	0	1	9
	4	1	0	0	0	0	0	0	0	0	1	2
	5	0	1	1	0	0	0	1	1	1	1	6
	6	0	0	0	1	1	1	1	1	1	1	7
	7	0	1	1	1	1	1	1	1	1	1	9
	8	0	0	0	0	0	0	1	1	1	1	4
	9	1	1	0	0	0	0	0	0	0	1	3
	10	0	1	0	1	1	1	1	1	1	1	8
	11	0	1	1	1	1	1	1	1	1	0	8
	12	1	1	1	1	1	1	1	0	0	0	7

Il punteggio totale medio (*PTM*) dei 12 studenti è, come noto:

$$PTM = \frac{7 + 6 + 9 + 2 + 6 + \dots + 7}{12} = 6,33$$

La sua deviazione standard (*SD*):

$$SD = \sqrt{\frac{(7 - 6,33)^2 + (6 - 6,33)^2 + \dots + (7 - 6,33)^2}{12}} = 2,17$$

Per ciascuna domanda si possono calcolare le proporzioni (unitarie) di quanti studenti hanno risposto in modo corretto (p) e in modo scorretto (q). Per l'item 1 si ha, ad esempio:

$$p_1 = \frac{1 + 1 + 1 + 1 + 0 + 0 + 0 + 0 + 1 + 0 + 0 + 1}{12} = 0,5$$

$$q_1 = 1 - p_1 = 1 - 0,5 = 0,5$$

il cui significato risulta facilmente interpretabile, ovvero il 50% degli studenti ha risposto correttamente alla prima domanda ed il restante 50% in modo scorretto.

È anche possibile calcolare il punteggio medio complessivo conseguito da coloro che hanno risposto, rispettivamente, in modo corretto (Mp) o scorretto (Mq) all'item considerato. A titolo di esempio si effettua il calcolo di Mp e Mq per il primo item. Alla prima domanda hanno risposto correttamente gli studenti 1, 2, 3, 4, 9 e 12 che hanno conseguito, rispettivamente, il punteggio complessivo nel test in esame di 7, 6, 9, 2, 3 e 7. Si ha quindi che per l'item 1 Mp risulta:

$$Mp_1 = \frac{7 + 6 + 9 + 2 + 3 + 7}{6} = 5,67$$

TABELLA 4. Analisi dei risultati di un test

	Item									
	1	2	3	4	5	6	7	8	9	10
p_j	0,50	0,75	0,58	0,50	0,67	0,58	0,67	0,75	0,50	0,83
q_j	0,50	0,25	0,42	0,50	0,33	0,42	0,33	0,25	0,50	0,17
Mp_j	5,67	7,00	7,43	8,00	7,63	7,86	7,25	7,11	7,00	6,10
Mq_j	7,00	4,33	4,80	4,67	3,75	4,20	4,50	4,00	5,67	7,50

TABELLA 5. Coefficienti di correlazione punto-biseriali

	Item									
	1	2	3	4	5	6	7	8	9	10
${}_j r_{pbi}$	-0,31	0,53	0,60	0,77	0,84	0,83	0,60	0,62	0,31	-0,24

Analogamente, alla prima domanda hanno risposto in modo scorretto gli studenti 5, 6, 7, 8, 10 ed 11 che hanno conseguito, rispettivamente i punteggi complessivi 6, 7, 9, 4, 8 e 8, di conseguenza per l'item 1 Mq risulta:

$$Mq_1 = \frac{6 + 7 + 9 + 4 + 4 + 4}{6} = 7,00$$

. I parametri p , q , Mp e Mq , calcolati per ogni item, sono raccolti nella TABELLA 4.

A questo punto è possibile calcolare per ciascun item il coefficiente di correlazione punto-biseriali la cui formula è:

$$r_{pbi} = \frac{Mp - Mq}{SD} * \sqrt{pq} \quad (18)$$

Per i primi due item si ha:

$${}_1 r_{pbi} = \frac{5,67 - 7}{2,17} \sqrt{0,5 \cdot 0,5} = -0,31$$

$${}_2 r_{pbi} = \frac{7 - 4,33}{2,17} \sqrt{0,75 \cdot 0,25} = 0,53$$

dove il pedice sinistro indica il numero dell'item. La TABELLA 5 riporta i valori degli ${}_j r_{pbi}$ ($j = 1, \dots, 10$) per ciascun item del test, i cui esiti si trovano nella TABELLA 3.

Chiarite le modalità di calcolo di r_{pbi} , è ora necessario interpretarne il significato. In primo luogo risulta evidente come r_{pbi} assuma valori compresi tra -1 e +1. Il coefficiente r_{pbi} esprime la capacità di un item di distinguere tra studenti con preparazioni differenti. Risulta quindi evidente che valori negativi del coefficiente devono portare a scartare l'item esaminato, poiché studenti che conseguono un punteggio complessivo elevato tendono a fornire una risposta scorretta all'item in esame. Viceversa tanto più un item presenta valori prossimi a +1 tanto più si evidenzia un legame positivo tra il punteggio medio complessivo conseguito e la

TABELLA 6. Valori di r_{pbi}

Valori di r_{pbi}	Qualità dell'item
0,40 o superiore	molto buona
da 0,30 a 0,40	buona
da 0,20 a 0,30	modesta
minore di 0,20	inadeguata

risposta fornita. In base all'interpretazione di r_{pbi} per i 10 item proposti si ricava che l'item numero 1 ed il numero 10 sono da scartare.

La TABELLA 6 riporta i livelli di r_{pbi} auspicabili, ormai consolidati e comunemente accettati nella letteratura psicometrica. Essa conferma le conclusioni a cui si era pervenuti utilizzando la conoscenza del significato del coefficiente di correlazione punto-biseriale, ovvero che gli item 1 e 10 vanno scartati.

3. 3. Attendibilità di un questionario

Gli item che formano un questionario od un loro sottogruppo è opportuno siano predisposti per misurare il raggiungimento dello stesso obiettivo. È quindi necessario trovare uno strumento statistico che sia in grado di dare un'adeguata informazione a questo riguardo. Diverse sono le misure di attendibilità proposte in letteratura, ma la più semplice sia dal punto di vista calcolatorio che interpretativo è il coefficiente α di Cronbach, che ha la formula:

$$\alpha = \frac{k}{k-1} \cdot \left(1 - \frac{s_1^2 + s_2^2 + \dots + s_k^2}{SD^2} \right) \quad (19)$$

dove k rappresenta il numero degli item. Il coefficiente α assume, in genere¹, valori compresi tra 0 ed 1. Un valore di α vicino ad 1 indica una forte attendibilità del questionario, ovvero che tutti gli item sono rivolti alla misurazione del raggiungimento dello stesso obiettivo formativo.

In pratica, per calcolare il coefficiente α è necessario determinare la varianza² dell'item j (s_j^2) rispetto al proprio valore medio p_j , per $j = 1, \dots, k$.

Riprendendo i dati delle TABELLE 3 e 4 si possono determinare i valori di s_j^2 , per $j = 1, \dots, k$. A titolo di esempio, si riportano i calcoli per i primi due item.

$$s_1^2 = \frac{(1 - 0,5)^2 + (1 - 0,5)^2 + \dots + (0 - 0,5)^2 + (1 - 0,5)^2}{12} = 0,25$$

$$s_2^2 = \frac{(1 - 0,75)^2 + (1 - 0,75)^2 + \dots + (1 - 0,75)^2 + (1 - 0,75)^2}{12} = 0,19$$

¹ Per un'analisi dettagliata del campo di variazione del coefficiente α di Cronbach, estesa anche ad alcuni casi particolari, si rinvia a L. J. CRONBACH (1951), *Coefficient alpha and the internal structure of tests*, «Psychometrika», 16, pp. 297-333.

² La varianza è il quadrato della deviazione standard.

TABELLA 7. Valore di s^2 per ciascun item

	Item									
	1	2	3	4	5	6	7	8	9	10
s_j^2	0,25	0,19	0,24	0,25	0,22	0,24	0,22	0,19	0,25	0,14

TABELLA 8. Valori di α .

Valori di α	Livello di attendibilità
0,90 o superiore	ottima
da 0,70 a 0,90	buona / molto buona
da 0,50 a 0,70	bassa / medio-bassa
minore di 0,50	inadeguata

La TABELLA 7 riporta il valore di s^2 per ciascun item.

Con riferimento all'intero test il coefficiente α diviene allora:

$$\alpha = \frac{10}{10 - 1} \cdot \left(1 - \frac{0,25 + 0,19 + 0,24 + 0,25 + 0,22 \cdots + 0,14}{2,17^2} \right) = 0,59$$

La TABELLA 8 riporta i livelli di α auspicabili comunemente proposti nella letteratura psicometrica.

In genere l'attendibilità di un questionario può essere migliorata mediante l'inserimento nella prova di ulteriori domande con un elevato potere discriminante. Nell'esempio analizzato in precedenza sarebbe opportuno eliminare le domande con un coefficiente di correlazione punto-biseriale negativo ed introdurne di nuove con un adeguato potere discriminante.

3. 4. Valutazione dei distrattori

Dal punto di vista educativo anche i distrattori possono apportare informazioni molto rilevanti che possono essere valutate con diversi strumenti statistici. Tuttavia essi richiedono competenze statistico-metodologiche piuttosto elevate e non sempre disponibili in una istituzione scolastica. Si osserva però che anche la semplice analisi delle frequenze di selezione di ciascun distrattore per i rispondenti suddivisi per livelli di punteggio complessivo può fornire utilissime indicazioni, sia per decidere se sostituire o modificare un distrattore sia per comprendere il grado di avanzamento dei rispondenti nel processo di insegnamento-apprendimento.

3. 5. Limiti ed applicazioni

Come indicato all'inizio del paragrafo, la teoria classica dei test psicometrici costituisce un utile supporto alla valutazione, pur non rappresentando ovviamente la soluzione a tutti i problemi. In primo luogo i risultati da essa desumibili sono legati al particolare insieme di studenti analizzati e, quindi, è opportuno prestare a questo aspetto particolare attenzione, specie se si intende effettuare delle comparazioni.

Tuttavia, se il contesto applicativo di riferimento è la valutazione di micro-sistema, tale problema perde gran parte della sua rilevanza, specie se le analisi effettuate non hanno un fine inferenziale.

Gli strumenti di analisi presentati, oltre che essere relativamente semplici nel loro contenuto e nella loro interpretazione e dunque facilmente implementabili per farne applicazioni in classe, sono anche agilmente automatizzabili con la predisposizione di semplici fogli elettronici che possono essere utilizzati dai docenti di tutte le discipline¹.

4. CONSIDERAZIONI FINALI

Le riflessioni proposte nei paragrafi precedenti vogliono essere uno spunto per attuare concretamente nella vita quotidiana pratiche che possano favorire una diffusione positiva della cultura della valutazione, pur nella consapevolezza che gli aspetti quantitativo-misuratori ne rappresentano solo un aspetto, importante ma non esaustivo.

Il grosso dibattito sviluppatosi negli ultimi anni nella Scuola italiana circa la valutazione e le sue metodologie non ha sempre tenuto in adeguata considerazione gli aspetti quantitativo-misuratori, trascurando sovente che semplici ed interessanti risultati possono essere ottenuti anche nella prassi quotidiana, nella didassi, e non solo mediante ricerche nazionali ed internazionali. Queste hanno infatti lo scopo di valutare e confrontare i sistemi scolastici e non di valutare e seguire i processi di insegnamento/apprendimento in una classe.

Una maggiore attenzione anche agli aspetti statistici della misurazione può favorire una concreta e positiva diffusione della cultura della valutazione e può fare anche da volano ad una rivisitazione della didattica delle materie scientifiche in generale e della matematica in particolare (Ottaviani 2003).

BIBLIOGRAFIA

- E. L. BAKER (1971), *The effects of manipulated item writing constraints on the homogeneity of test items*, «Journal of Educational Measurement», 8, pp. 305-309.
- L. BOETI (2005), *La statistica dai banchi di scuola alla cattedra. Uno strumento per gestire il processo formativo*, «Induzioni», 30, pp. 75-91.
- T. M. HALADYNA, S. M. DOWNING (1989), *A taxonomy of multiple-choice item-writing rules*, «Applied Psychological Measurement», 2 (1), pp. 51-78.
- F. LORD, M. NOVICK, (1968), *Statistical Theories of Mental Test Scores*, Reading, MA, Addison-Wesley Publishing.
- S. MIGNANI, R. RICCI (2005), *Il ruolo del metodo statistico nel progetto PISA*, «Induzioni», 30, pp. 59-73.
- J. F. MUÑIZ (1998), *Teoría clásica de los tests*, Madrid, Pirámide.
- A. J. NITKO (2004), *Educational assessment of students*, Columbus Ohio, Perason.
- M. G. OTTAVIANI (2003), *Didattica della statistica: un campo di ricerca in evoluzione*, «Induzioni», 26, pp. 65-71.
- M. G. OTTAVIANI, S. MIGNANI, R. RICCI (2006), *Metodi statistici per la valutazione di abilità*

¹ Un esempio di foglio elettronico per l'esecuzione automatica del calcolo degli indicatori proposti in questo lavoro può essere scaricato all'indirizzo: <http://www.irreer.it/~ricci/>.

- e competenze: uno studio di caso che riguarda la matematica*, in *Atti del XXV Convegno UMI-CIIM "Valutare in Matematica"*, (in stampa).
- A. M. PANNONE (2000), *Validazioni statistiche di prove oggettive di profitto*, «Induzioni», 20, pp. 145-150.
- D. THISSEN, L. STEINBERG, J. MOONEY, (1989), *Trace lines for testlets: A use of multiple-categorical-response models*, «Journal of Educational Measurement», 26, pp. 247-260.
- W. J. VAN DER LINDEN, R. K. HAMBLETON, (Eds.) (1997), *Handbook of modern item response theory*, New York, Springer.