



Università degli Studi di Modena e Reggio Emilia
Centro *E-learning* di Ateneo
Facoltà di Scienze della Formazione

Progettazione, costruzione, somministrazione e valutazione di prove strutturate di conoscenza

Guida alla costruzione e all'uso di strumenti per la valutazione degli apprendimenti

Parte I

Premessa

Queste note nascono dall'esigenza di approfondire i temi valutativi manifestata da alcuni colleghi durante le attività di addestramento all'uso della piattaforma di *e-learning* Dolly (*Didattica on line*), che tra le sue diverse funzionalità ha anche quella che consente di predisporre diverse forme di prove di valutazione degli apprendimenti e di somministrarle automaticamente in qualsiasi momento del corso di studi. A partire da questa esigenza la Facoltà di Scienze della Formazione ha ritenuto utile redigere delle note sulla progettazione, la costruzione, la somministrazione e la valutazione delle prove strutturate di conoscenza e di metterlo a disposizione degli utenti di Dolly che volessero iniziare a progettare e a costruire autonomamente delle prove di valutazione per i propri corsi.

Luciano Cecconi (Facoltà di Scienze della Formazione) e Guido Benvenuto (Università di Roma La Sapienza) hanno predisposto la guida sulla base di un testo precedentemente elaborato per un corso di formazione a distanza sulla valutazione, rivolto ai docenti della scuola impegnati nelle rilevazioni valutative nazionali. Del modulo didattico originario, che comprendeva dieci unità didattiche, è stata conservata per quanto possibile la struttura tipica dell'unità didattica. Sono state invece eliminate le esercitazioni e le prove di verifica per consentire al lettore una consultazione più agile e più veloce dei contenuti. I temi proposti nelle sei unità che compongono questa guida cercano di riprodurre le fasi principali del percorso che intraprende chiunque decida di utilizzare le prove strutturate per la valutazione degli apprendimenti, dalla progettazione della prova (definizione degli obiettivi, selezione del formato della prova) alla sua costruzione (regole per la formulazione dei quesiti), dalla sua somministrazione agli studenti all'elaborazione dei dati, dall'attribuzione di punteggi alla formulazione dei giudizi valutativi.

La seconda parte della guida, in preparazione, affronterà più direttamente l'uso di Dolly per la costruzione e la somministrazione delle prove valutative.

Gli esempi e le indicazioni operative che si è cercato di fornire durante la trattazione si riferiscono prevalentemente al contesto scolastico, tuttavia tale diversità di contesto non è parsa determinante ai fini di una buona comprensione dei concetti e delle procedure. In ogni caso, sono senz'altro gradite segnalazioni di esempi più attinenti alla realtà universitaria o di incongruenze ed errori presenti nel testo (cecconi.luciano@unimore.it).

Chi volesse approfondire gli argomenti trattati nella guida può consultare i seguenti testi:

- Guido Benvenuto, *Mettere i voti a scuola. Introduzione alla docimologia*, Carocci, Roma, 2003.
- Gaetano Domenici, *Manuale della valutazione scolastica*, Laterza, Bari, 2001.
- Benedetto Vertecchi, *Manuale della valutazione*, Franco Angeli, Milano, 2003.

Sommario

UD1. Uno schema per la progettazione di prove di valutazione	Pag. 3
UD2. Determinare gli obiettivi dei quesiti	8
UD3. Diversi formati di quesiti strutturati	17
UD4. Formulare i quesiti (esempi e alcune regole)	25
UD5. Regole per la somministrazione delle prove	30
UD6. Dalla misurazione alla valutazione	35

•

UD1.

Uno schema per la progettazione di prove di valutazione

Obiettivi

- Individuare le diverse fasi per la costruzione di una prova strutturata
- Pre-definire la finalità dello strumento di valutazione

Sommario

Per costruire un efficace strumento di valutazione di tipo strutturato e che possa garantire una rilevazione valida, cioè rispondente allo scopo per il quale è stata costruita, è indispensabile procedere attraverso una serie di fasi tra loro concatenate.

Questa unità delinea lo schema per la pianificazione e la progettazione di una prova strutturata in modo da esplicitare il tipo di coerenza necessario per un'efficace messa a punto dello strumento. Partendo dall'esplicitazione della funzione valutativa via via fino alla verifica sul campo della sua idoneità misurativa, lo schema indicherà i passi che nelle successive unità della guida si svilupperanno in termini di indicazioni operative.

Introduzione al contenuto

La progettazione di una prova strutturata richiede una serie di decisioni teoriche ed operative. Si tratta di entrare nel merito dei singoli contesti di verifica e di esplicitare le finalità generali, gli obiettivi e il formato per i diversi quesiti e di condurre una serie di controlli anche statistici sul prodotto costruito.

La costruzione di una buona prova di valutazione spinge verso la collaborazione e il confronto tra docenti e tra questi e gli esperti nella costruzione delle prove strutturate. In questa unità didattica oltre a fornire lo schema della pianificazione ci soffermeremo sul primo e più decisivo passo da compiere: *esplicitare la finalità complessiva della prova*. Le domande alle quali il docente, o il costruttore di uno strumento strutturato di valutazione, dovrà rispondere preliminarmente sono le seguenti: a che cosa mi serve questa prova? Qual è la sua finalità e quindi la funzione valutativa che deve assolvere? Quanto tempo deve impegnare lo studente? L'esplicitazione di questi punti porta a sviluppare una pianificazione della costruzione dello strumento.

Un esempio per chiarire. Ad inizio di anno accademico un gruppo di docenti decide di pianificare la costruzione di una prova di ingresso (funzione valutativa) per saggiare alcune abilità di base (finalità). La prova sarà somministrata in un certo numero di corsi e impegnerà gli studenti per un'ora (destinatari e tempo di somministrazione).

Le risposte alle domande che definiscono le finalità delle prove (“che cosa vogliamo misurare con questa prova?” oppure “per quale motivo stiamo utilizzando questo tipo di prova?”) ci permettono di catalogare le finalità della valutazione del profitto/prestazioni in quattro grandi categorie:

- di ingresso (*placement*). Per misurare i pre-requisiti e le abilità indispensabili per attivare un'unità formativa;
- diagnostiche (*diagnostic*). Per verificare specifiche difficoltà nell'apprendimento e costruire mirate procedure di recupero;
- regolative del processo (*formative*). Per fornire allo studente un feed-back per autovalutarsi e al docente per, eventualmente, modificare la proposta didattica;
- finali/certificative (*summative*). Per assegnare voti, mettere in graduatoria, certificare competenze.

La finalità generale di una prova si ricava innanzitutto dalla sua collocazione all'interno del tempo didattico e formativo. È indispensabile distinguere il periodo nel quale si colloca la sua utilizzazione (tempo didattico), ma anche quanto tempo dovrebbe o può durare la prova (tempo di svolgimento).

Così, una prova di ingresso sarà diversa da una di tipo diagnostico o sommativo. La finalità della prova, che nel nostro discorso sarà limitata all'accertamento di prestazioni da parte dello studente, direziona quindi le differenti fasi della progettazione generale: l'ampiezza (specifici elementi o un campione rappresentativo dei contenuti) e la struttura dei quesiti che lo compongono.

Cerchiamo ora di esemplificare alcune differenze tra le finalità valutative.

All'inizio di ogni corso, ma possono esserci sensibili differenze per i diversi anni di corso, è indispensabile dal punto di vista della continuità didattica, ricollegarsi a corsi caratterizzati da legami di propedeuticità, individuare i livelli di partenza, cioè i pre-requisiti per affrontare le differenti attività di studio. In questo caso avremo bisogno di pianificare e utilizzare una prova di ingresso per la rilevazione delle conoscenze e delle abilità già padroneggiate dagli studenti. La rilevazione dovrà permettere di raccogliere informazioni sia di tipo quantitativo, per rendersi conto dell'ampiezza di alcune lacune, sia di tipo qualitativo, per individuare le specificità di quelle lacune e pianificare mirati interventi di recupero.

Le *prove di ingresso*, oltre a permettere l'analisi delle conoscenze e delle lacune presenti nel gruppo in cui si effettua la somministrazione, possono aiutare a isolare i punti di forza e quelli di debolezza in modo da impostare le opportune procedure didattiche per valorizzare e recuperare gli apprendimenti tanto dei singoli quanto del gruppo. Le prove di comprensione della lettura nei diversi contesti disciplinari, o specifiche prove di conoscenze di base per le diverse discipline sono un esempio di prove di ingresso. Infatti, all'inizio del corso tutti i docenti sono interessati a individuare le possibili lacune nelle abilità trasversali o in quelle di base specifiche.

Quando le prove sono pensate per individuare difficoltà specifiche che gli studenti possono incontrare nel processo di apprendimento, e quando si concentrano su di uno specifico settore della disciplina o materia, si chiamano prove *diagnostiche*. Esse, infatti, isolano un determinato aspetto dell'apprendimento, si pensi ad esempio a prove specifiche per la comprensione del testo o il calcolo matematico, e aiutano a costruire mirati percorsi di recupero e strategie di riequilibrio.

Le *prove formative*, d'altro canto, sono quelle che si utilizzano durante il percorso delle attività didattiche (*in itinere*), per controllare che gli apprendimenti stiano procedendo gradualmente e per poter operare azioni di rinforzo e di sostegno specifici. In questa famiglia di prove troviamo quelle per l'analisi dell'andamento del profitto dei singoli e per rilevare alcune informazioni relative alle forme di apprendimento dell'intero gruppo. Gli esercizi da svolgere all'interno del corso, individualmente o in gruppo, o come esercitazione extra-lezione, sono un esempio di prova formativa del profitto, in quanto il docente, interagendo a diverso livello (in presenza o a posteriori) nella loro esecuzione, interviene per sostenere lo studente nelle prestazioni e per indicare alcune

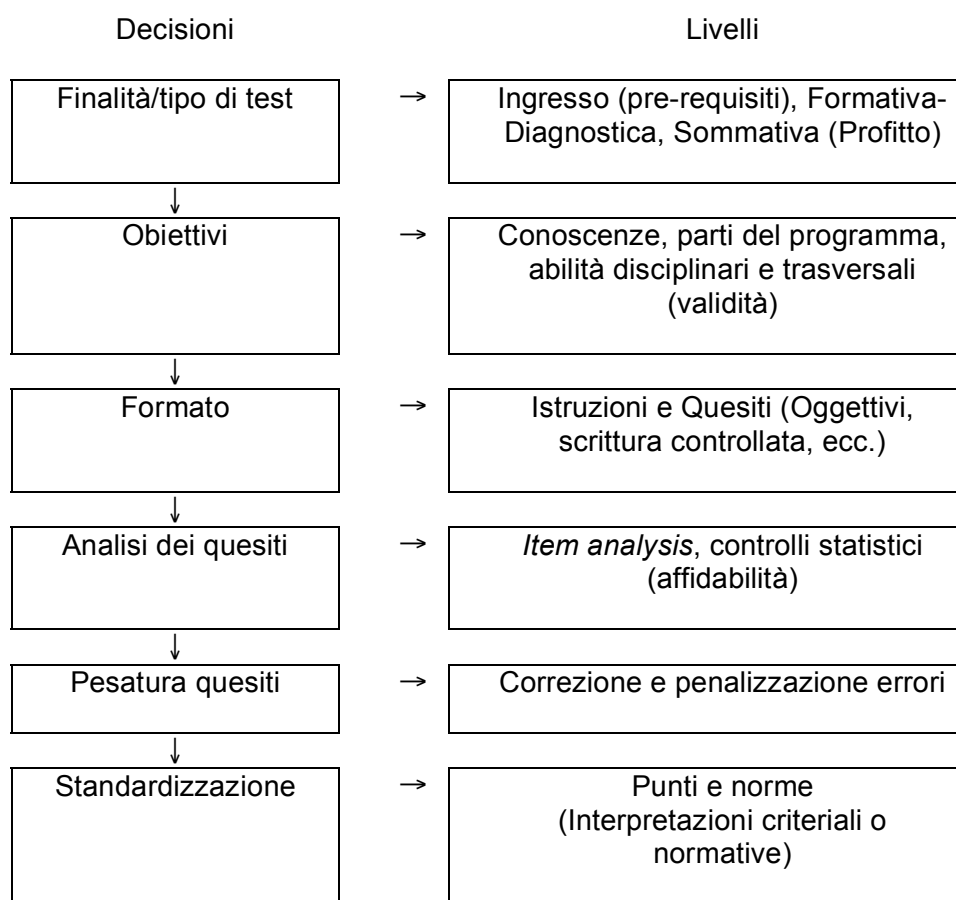
modalità di svolgimento. Accanto alla rilevazione dei livelli raggiunti dai singoli studenti, le prove formative, quando svolte nel gruppo, permettono di analizzare le forme di soluzione dei problemi nell'interazione tra pari, nella costruzione collaborativa degli apprendimenti.

Le *prove finali o sommative* sono invece quelle che consentono di classificare gli studenti rispetto all'apprendimento relativo ad un intero corso. Esse si utilizzano al termine di un modulo o a conclusione di un corso. Stiamo parlando in questo caso delle forme di esame finale e di tutte quelle prove che prevedono lo svolgimento in contemporanea di tutti gli studenti, con indicazioni sul tempo a disposizione per l'esecuzione e che sostanzialmente prevedono l'attribuzione di un giudizio, punteggio o voto finale. Un esempio tra i tanti di prove finali strutturate e/o semi-strutturate sono le prove scritte previste all'interno dell'Esame di Stato a conclusione degli studi secondari superiori, prove finalizzate ad accertare: la padronanza della lingua ufficiale d'insegnamento e le capacità espressive, logico-linguistiche, critiche e creative (prima prova), le conoscenze specifiche relative alla materia caratterizzante il corso di studi seguito (seconda prova), conoscenze, competenze e capacità acquisite relativamente alle materie dell'ultimo anno del corso di studi, nonché le capacità di collegare ed integrare conoscenze e competenze (terza prova).

Le prove di ingresso, formative, diagnostiche e sommative sono prove sostanzialmente mirate alla valutazione delle prestazioni degli studenti sul versante degli apprendimenti cognitivi. Quando si intende invece sondare gli aspetti non cognitivi degli studenti, vale a dire gli interessi, gli atteggiamenti, i valori, per indagare elementi di contesto ambientale e familiare, ricorriamo ad altri tipi di test: test di personalità, di orientamento, di preferenze e atteggiamenti verso lo studio e l'apprendimento in generale. Altre volte, poi, per rilevare le prospettive e/o le potenzialità degli studenti rispetto a determinate abilità o attitudini, ricorriamo a prove *prognostiche e orientative*¹.

¹ Per un approfondimento su queste modalità di rilevazione si veda: Boncori, L. (1993), *Teoria e tecnica dei test*, Boringhieri, Torino; Quagliata, A. (a cura di), (1998), *Per una valutazione formativa e proattiva*, Anicia, Roma; Passolunghi, M.C., De Beni, R. (2001), *I test per la scuola*, Il Mulino, Bologna.

Schema per la pianificazione di prove valutative



UD2.

Determinare gli obiettivi dei quesiti

Obiettivi

- Determinare la tipologia di obiettivo
- Esplicitare gli obiettivi per ogni quesito
- Quantificare il numero di quesiti della prova e per ogni obiettivo

Sommario

Nella fase di costruzione di una prova strutturata, dopo aver definito la finalità generale, si procede con la messa a punto di un quadro di riferimento per la pianificazione dei quesiti che compongono la prova. I singoli quesiti rimandano ai diversi obiettivi di verifica delle prestazioni e quindi si tratta di prevedere la tipologia di obiettivi e il numero di quesiti collegati ai singoli obiettivi.

In questa unità si discute la fase di individuazione e di esplicitazione degli obiettivi. Per facilitare la massima operatività si forniscono alcuni strumenti per la loro messa a punto e alcuni esempi di modelli di riferimento adottati in recenti indagini.

Oltre a questa fase qualitativa di definizione degli obiettivi specifici si richiede di quantificare il numero di quesiti che saranno impiegati per la valutazione, in modo da controllare la coerenza della prova rispetto alle sue finalità valutative.

Introduzione al contenuto

Una buona definizione di obiettivi consiste nell'esatta o almeno più completa definizione dei comportamenti da misurare e valutare. Mager (1972) definisce operativamente un obiettivo ben formulato quello che esplicita i seguenti elementi:

- che cosa deve essere in grado di fare lo studente per dimostrare di aver raggiunto quell'obiettivo (*prestazione, performance*);
- in quali condizioni l'allievo deve dimostrare di aver raggiunto quell'obiettivo (*condizioni*). Si dovrà cioè esplicitare cosa deve fare per risolvere il problema posto dalla domanda;
- quali indicatori/parametri devono essere adottati per correggere/valutare il raggiungimento di quell'obiettivo (*criterio*).

Per costruire una prova di verifica efficace è indispensabile definire, operativamente, il collegamento tra gli obiettivi specifici e i quesiti ad essi riferiti. La tematica relativa agli obiettivi dell'apprendimento è sempre stata cruciale e controversa nel dibattito teorico e nella prassi didattica.

La definizione degli obiettivi, educativi e didattici, investe difatti il profilo culturale delle istituzioni educative e la loro organizzazione curricolare.

I termini utilizzati in questi ultimi decenni sono certamente diversi, in quanto rimandano a diversi sistemi e organizzazioni del sistema dell'istruzione, ma sottolineano una preoccupazione comune: gli obiettivi o traguardi, che definiscono gli "oggetti" sia dei percorsi di studio sia della valutazione, sono da individuare e ridefinire nel rispetto delle condizioni dei singoli studenti e da raggiungere all'interno dei percorsi curricolari. Questo significa che ogni docente indica gli obiettivi disciplinari (conoscenze e abilità) ritenuti fondamentali per il proprio corso.

In quest'ottica, e per chiarire come diversi quesiti di una prova di verifica possano centrare diversi obiettivi di apprendimento, abbiamo ripreso alcuni esempi di formulazione di quesiti seguendo la tassonomia degli obiettivi educativi per l'area cognitiva (Bloom). La riflessione che Bloom ha avviato sin dagli anni '50 del secolo scorso riprende il discorso della programmazione e del profilo curricolare, indicando i diversi livelli nei quali gli obiettivi si possono collocare, dai più semplici ai più complessi: dalle conoscenze (sapere), alle abilità (saper fare).

Ovviamente lo strumento della tassonomia non va inteso in forma rigida, tanto è vero che esistono diverse semplificazioni successive (altre tassonomie), ma in modo flessibile rispetto al contesto nel

quale ci si trova ad operare. Sta il fatto che il costruttore di una prova di verifica, esplicitando gli obiettivi mira al controllo della *validità* (la prova accerta *effettivamente* la conoscenza o l'abilità che si intende valutare²), e definendo il collegamento tra i singoli quesiti e i diversi obiettivi mette a punto uno strumento di misura coerente rispetto ai propri scopi valutativi.

Alcuni esempi di quesiti

Conoscenza

La prima guerra mondiale è iniziata nel ... e si è conclusa nel ...

Comprensione

Dopo aver letto la tabella che segue (Previsioni dello SVIMEZ) indicare se le affermazioni poste sotto la tabella sono ritenute vere (V) o false (F).

Previsioni dello SVIMEZ, per il 1975, dello sviluppo delle forze di lavoro secondo le diverse categorie professionali, sulla base dello sviluppo dal 1951 al 1961 (in migliaia) in Italia:

	1951	1961	1971
Dirigenti e quadri superiori	408,0	631,7	1.444,2
Tecnici	374,0	1.001,7	2.479,9
Addetti al coordinamento	986,0	1.148,5	2.512,0
Capi subalterni	133,0	193,1	855,1
Personale qualificato	3.706,0	5.605,1	11.193,4
Personale generico	11.339,0	10.819,9	3.315,4
	16,946,0	19,400,0	21,800,0

- | | | |
|---|---|---|
| 1) Si prevede che il personale generico debba diminuire, dal 1961 al 1975, di circa un terzo in cifre assolute | V | F |
| 2) Il previsto aumento dei tecnici, dal 1951 al 1975, è in proporzione inferiore a quello degli addetti al coordinamento nello stesso periodo | V | F |
| ... (prosegue) ... | V | F |

Applicazione

In un bicchiere contenente acqua sino all'orlo si trova un cubetto di ghiaccio galleggiante. Al graduale sciogliersi del ghiaccio

- si avrà traboccamento del liquido in quantità proporzionale all'espansione del ghiaccio.
- non si avrà traboccamento del liquido per il principio di Archimede.
- l'acqua traboccherà sino al raggiungimento di una temperatura uniforme nel bicchiere.
- Il liquido non traboccherà a causa del "paradosso idrostatico".
- In base al principio della spinta idrostatica, il liquido traboccherà solo se il peso del ghiaccio è superiore a quello del liquido spostato.

Analisi

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

² Il problema della validità è uno dei più importanti in campo valutativo. Molte prove di valutazione sono scarsamente valide poiché finiscono per sollecitare gli studenti a manifestare abilità diverse da quelle che si intende valutare. Per esempio, se si vuole accertare la capacità di un docente di costruire correttamente una prova di valutazione non è il caso di proporgli di dimostrare questa capacità chiedendogli di scrivere un saggio breve in cui riassumere tutte le regole per costruire le prove. In questo caso si accerterebbero le sue capacità di sintesi e le sue abilità linguistiche e argomentative, non le sue capacità operative di costruire prove. Una prova valida sarebbe la seguente: "Con riferimento ai contenuti presentati nel testo allegato costruisci in trenta minuti cinque quesiti a scelta multipla con quattro alternative di risposta di cui solo una esatta, relativi ai contenuti del testo allegato."

- a) il loro meccanismo di risoluzione facilita la diffusione delle risposte tra gli studenti;
- 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

Sintesi

“Un certo numero di problemi che hanno fissato gli uomini della scienza possono essere classificati sotto il titolo ‘struttura della materia’. In questo ambito le prime ricerche si riferiscono alla struttura discontinua della materia, ipotesi emessa da Dalton nel 1805. Un altro gruppo di ricerche ha per oggetto la teoria cinetica dei gas. Essa era apparsa già nel XVII secolo, poi fu ripresa in maniera molto più completa da Maxwell (1859) e Boltzmann (1906). Dimostrare come la teoria cinetica dei gas costituisca un argomento favorevole all'ipotesi della rappresentazione atomistica della materia. La vostra discussione deve comprendere: a) un'esposizione delle ipotesi fondamentali di questa teoria; b) alcune importanti deduzioni di questa teoria, applicate al caso dei gas perfetti; c) l'interpretazione delle leggi fondamentali di Gay-Lussac e dell'ipotesi di Avogadro”.

Valutazione

La riforma luterana si configura essenzialmente come

- a) una rivoluzione culturale e dei costumi che, in virtù dell'affermazione del principio della libera interpretazione delle Sacre Scritture da parte di ogni credente, ha promosso la lotta all'analfabetismo e l'elevazione del livello di istruzione delle masse.
- b) una sfida contro i canoni interpretativi autentici della dottrina rivelata, stabiliti dai Padri della Chiesa e dalla tradizione scolastica, e contro il valore dei sacramenti, dei precetti e della liturgia.
- c) un attacco, perpetrato spesso in forme violente (come nel caso delle rivolte dei contadini e dei cavalieri), contro la gerarchia ecclesiastica e l'autorità del clero.
- d) una ribellione suscitata, più che da sostanziali divergenze teologiche, dalla reazione morale alla corruzione e mondanizzazione del clero e allo scandalo provocato dalla vendita delle indulgenze.
- e) un movimento ideologico sviluppatosi come diretta conseguenza del disagio sociale avvertito dalle classi subalterne, emarginate dalla partecipazione al potere.

Esempi di quesiti strutturati seguendo la tassonomia di Bloom per l'area cognitiva

La scelta degli obiettivi è ovviamente condizionata dalla fase didattica per cui è pensata la prova e dal tempo a disposizione per il suo svolgimento.

Facciamo un esempio considerando alcuni degli obiettivi di apprendimento per il primo biennio dei licei scientifici (vedi tab. 1)³. Un docente di lettere, per verificare le conoscenze relative alla morfologia e al lessico della lingua italiana, decide di costruire una prova di valutazione degli apprendimenti. In base alla fase didattica nella quale si trova la classe, individua le conoscenze e le abilità da sottoporre a verifica e prevede quanti esercizi/quesiti dovrà formulare in ragione dei diversi obiettivi individuati. In questa operazione diventa indispensabile considerare il tempo a disposizione per la verifica, per calibrare e quantificare il numero di quesiti in relazione agli obiettivi considerati.

³ Cfr OSA (Obiettivi Specifici di Apprendimento) per i licei

PRIMO BIENNIO	
CONOSCENZE	ABILITÀ
<p><i>Il sistema della lingua</i></p> <ul style="list-style-type: none"> - morfosintassi: <ul style="list-style-type: none"> • struttura morfologica della lingua • struttura sintattica della frase semplice (verbo e suoi argomenti; altri elementi connessi; modalità e tipologie di frasi) e della frase complessa (coordinazione, subordinazione; tipi di dipendenti e loro modi verbali) • diatesi del verbo (attiva, passiva, media) • reggenze dei verbi, dei nomi, degli aggettivi • dalle strutture-tipo alle strutture marcate per differenziare le informazioni (sfondo/primo piano; dato/nuovo); ordine degli elementi nella frase; dislocazioni e frasi segmentate, frasi scisse - formazione lessicale delle parole (morfologia lessicale) - principali caratteristiche tipologiche dell'italiano in rapporto ad altre lingue - variabilità linguistica e stratificazione dei registri nell'italiano attuale - lessico: <ul style="list-style-type: none"> • relazioni di significato (sinonimi, antonimi, iperonimi, iponimi, gradazioni); le collocazioni; denotazione e connotazione • etimologia; usi figurati e diversi registri (colloquiale, formale, aulico, etc.) 	<ul style="list-style-type: none"> - analizzare le categorie morfologiche della lingua nei diversi testi - padroneggiare il dinamismo sintattico del verbo - costruire una frase complessa sintatticamente adeguata e comunicativamente efficace - individuare la vitalità di produzione linguistica della lingua (affissazione, suffissazione, etc.) - confrontare le caratteristiche tipologiche dell'italiano con il latino e con le altre lingue studiate - adattare la scelta del registro linguistico al "mezzo" (o al canale) e ai diversi contesti d'uso - usare il lessico della lingua nella sua potenzialità espressiva e in maniera contestualmente adeguata, riconoscendone regolarità, gerarchie interne, diversi registri

Tabella n. 1 - Obiettivi Specifici Apprendimento – Italiano, Lingua e Letteratura (Indicazioni per il Liceo Linguistico)

Per decidere quanti quesiti collegare ad un determinato obiettivo e quanti quesiti costruire per l'intera prova si può considerare il tempo destinato e impegnato nello svolgimento di determinati segmenti di programma. La centralità di una tematica e il tempo dedicato al suo svolgimento dovrebbero essere i parametri per decidere i pesi. Una semplice formula per calcolare velocemente l'importanza da assegnare alle singole parti è la seguente⁴:

$Ip = (Tp/Tc) \times 100$ <p>Ip=Importanza assegnata a ciascuna parte; Tp=Tempo parziale espresso in ore (o in altra unità di misura); Tc=Tempo complessivo (espresso nella stessa unità di misura di Tp).</p>
--

⁴ È Domenici nel suo *Manuale della valutazione scolastica*, (1993, 1^a; 2001, Laterza, Bari) a proporre la formula riportata.

Al termine di questa fase progettuale il docente, o costruttore di una prova strutturata, dovrebbe trovarsi di fronte ad una tabella, come quella riportata di seguito (tab. n.2), che per i singoli obiettivi prescelti e in relazione alle differenti fasi della didattica indichi il totale dei quesiti che la prova prevede di contenere. Ovviamente durante la fase di costruzione della prova, ma anche a seguito della verifica sul campo, è possibile ritornare su questo schema e integrarlo in funzione di successive considerazioni e risultati.

Conoscenze/area di programma	Ob.1	Ob.2	Ob.3	Totale quesiti
Frasi semplici	2		1	3
Frasi complesse	1	1	1	3
Relazioni di significato	1	2	2	5
Etimologia		2	3	5
Totale	4	5	7	16
Ob.1: riconoscere, individuare				
Ob.2: applicare, produrre				
Ob.3: inferire, valutare				

Tabella n.2 – Numero di quesiti per i diversi obiettivi e per l'intera prova

Abbiamo classificato gli obiettivi in tre categorie (ob. 1, ob. 2, ob. 3) che riassumono alcuni dei livelli della tassonomia di Bloom. Ovviamente si può procedere facendo riferimento ad altre tassonomie o schemi di obiettivi specifici di apprendimento messi a punto da studiosi per le diverse aree disciplinari o partendo da quadri di riferimento costruiti a livello internazionale, alla luce di indagini e studi valutativi, come nel caso che di seguito analizziamo. L'importante, dal punto di vista della validità della valutazione, è che si definisca chiaramente il tipo di obiettivo o traguardo nell'apprendimento da sottoporre a verifica, per poi indicare la quantità e qualità dei quesiti da costruire per la valutazione. In questo modo lo strumento di valutazione può prevedere specifici pesi rispetto ai vari obiettivi e rimandare a mirate forme di recupero.

Una recente indagine internazionale (PISA, 2003)⁵ sui livelli di conoscenze e di abilità dei quindicenni in matematica, lettura, scienze e *problem solving*, ha affrontato dal punto di vista

⁵ L'indagine Internazionale Pisa (Programme for International Student Assessment) è stata promossa dall'Organizzazione per la Cooperazione e lo Sviluppo Economico (OCSE). Tale progetto riflette la volontà da parte dei Paesi membri dell'OCSE di monitorare i risultati dei sistemi scolastici in termini di prestazioni degli studenti in un quadro di riferimento a livello internazionale. La valutazione delle prestazioni riguarda gli studenti quindicenni, prossimi alla fine dell'obbligo scolastico, e riguarda le competenze matematiche, di lettura, scientifica e il *problem solving*. Essa si avvale di prove scritte strutturate, costituite da quesiti a scelta multipla o che prevedono che lo studente scriva una risposta. L'indagine è stata realizzata in 43 Paesi nel primo ciclo (32 nel 2000 e 11 nel 2002) e in 42 Paesi nel secondo ciclo (2003). Il Quadro di riferimento di questa indagine è stato recentemente pubblicato: *PISA 2003*, OECD, Roma, Armando Editore, 2004. Per ulteriori informazioni si veda anche Nardi, E. (2002), *Come leggono i quindicenni. Riflessioni sulla ricerca OCSE-PISA*, Franco Angeli, Milano.

teorico e operativo il problema della definizione degli obiettivi da sottoporre a misurazione comparativa.

Per quanto riguarda la lettura sono stati definiti gli obiettivi della valutazione in termini di livelli di competenza per i diversi processi:

- a) individuare informazioni
- b) comprendere il significato generale di un testo
- a) sviluppare un'interpretazione
- b) riflettere sul contenuto di un testo e valutarlo
- c) riflettere sulla forma di un testo e valutarla

Quindi si sono indicate le percentuali dei compiti della prova di valutazione in riferimento ai processi di lettura (v. tabella n. 3)

Processo di lettura	% compiti nella prova
Individuare informazioni	29
Interpretare il testo	50
Riflettere e valutare	21
Totale	100

Tabella n. 3 – Composizione della prova relativa ai processi di lettura nell'indagine OCSE-PISA, 2003.

Allo stesso modo per le prove di area scientifica si è proceduto a selezionare i processi che definiscono tale competenza:

- a) *Descrivere, spiegare e prevedere fenomeni scientifici*
Attraverso tale processo gli studenti dimostrano di comprendere fenomeni scientifici chiamando in causa i concetti appropriati rispetto ad una data situazione.
- b) *Comprendere un'indagine di tipo scientifico*
Significa saper riconoscere e comunicare problemi che possono essere affrontati con l'indagine scientifica e, dunque, sapere cosa comporti tale indagine. Questo processo comprende la capacità di identificare o riconoscere i dati di fatto necessari in una determinata indagine: quali elementi mettere a confronto, quali variabili modificare o controllare, quali ulteriori informazioni sono necessarie, o quali procedure occorre seguire per la raccolta dei dati.
- c) *Interpretare dati e conclusioni scientifiche*
Utilizzare i risultati della ricerca scientifica come prove a supporto di affermazioni e conclusioni.

La prova valutativa è composta tenendo conto dei diversi processi, con percentuali diverse (vedi tabella n. 4).

Processi scientifici	% compiti nella prova
Descrivere, spiegare e prevedere fenomeni scientifici	40-50
Comprendere un'indagine di tipo scientifico	20-25
Interpretare prove di carattere scientifico e trarne conclusioni	20-25
Totale	100

Tabella n. 4 – Composizione della prova relativa ai processi scientifici nell'indagine OCSE-PISA, 2003.

Letture consigliata

Per pianificare la costruzione di prove disciplinari e trasversali, o semplicemente pesare i diversi quesiti di una prova si può utilizzare la Tassonomia di Bloom (Fonte: Block e Anderson, 1975)

OPERAZIONE	DEFINIZIONE ED ESEMPI
Conoscenza	La capacità di ricordare o riconoscere un contenuto in forma praticamente identica a quella nella quale esso è stato presentato originariamente. Può trattarsi di fatti, termini, convenzioni, concetti, regole, generalizzazioni, procedimenti. Per esempio, lo studente deve ricordare da chi e quando fu scoperta l'America oppure che $E=mc^2$
Comprensione	<p>Questa operazione implica tre diverse capacità: la traduzione, l'interpretazione e l'estrapolazione.</p> <p>La traduzione richiede la capacità di trasporre il contenuto presentato da una forma simbolica ad un'altra. Per esempio, lo studente deve essere in grado di spiegare verbalmente le parti di un grafico, di tradurre in cifre un problema espresso in parole, o di fornire l'equivalente inglese di un'espressione tedesca.</p> <p>L'interpretazione implica la capacità di spiegare e riassumere un contenuto presentato. Mentre nella traduzione ciascuna parte del contenuto deve assumere una nuova forma, l'interpretazione richiede che lo studente stabilisca il significato generale del contenuto. Per esempio, lo studente deve essere in grado di riassumere una storia.</p> <p>L'estrapolazione comporta la capacità del discente di andare oltre il contenuto presentato per determinare le probabili applicazioni, conseguenze o effetti date le condizioni descritte. Per esempio, lo studente deve trarre le conclusioni di letture e/o discussioni.</p>
Applicazione	La capacità di utilizzare il contenuto appreso o per risolvere un problema o per apprendere con maggior facilità in una situazione nuova. Per esempio, si chiede allo studente di usare le idee dell'addizione e della sottrazione per trovare un errore nel libretto di assegni del mitico sig. Jones. Ovvero di applicare il principio della "prospettiva" quando impara a disegnare.
Analisi	<p>Questa operazione è sostenuta da due tipi di capacità: l'analisi degli elementi e l'analisi delle relazioni.</p> <p>L'analisi degli elementi richiede che lo studente sia in grado di scomporre nei suoi costituenti un aggregato di contenuto che gli viene presentato (per es. un documento, un oggetto o una comunicazione). Per esempio, lo studente è invitato a trovare tutti i nomi presenti in un paragrafo o a sezionare una rana.</p> <p>L'analisi delle relazioni presuppone nello studente la capacità di individuare la relazione tra una parte di un aggregato di contenuto e le altre parti di esso. Per esempio, lo studente deve collegare un'affermazione incontrata in un romanzo con un avvenimento che si verifica in seguito nello stesso romanzo. Oppure deve trovare la relazione esistente tra l'uso del colore nella pittura e l'evocazione di emozioni.</p>
Sintesi	La capacità di organizzare e combinare il contenuto in modo da produrre una struttura, un modello o un'idea nuovi. A questo livello, lo studente deve comunicare in modo reale. Egli deve altresì esercitare quella che alcuni educatori chiamano capacità creativa. Per esempio, lo studente è tenuto a scrivere un tema su "I tre mortali nemici dell'uomo", o ad allestire una rappresentazione in classe.
Valutazione	La capacità di esprimere giudizi sia qualitativi che quantitativi sul modo in cui particolari elementi o aggregati di contenuto soddisfano criteri interni o esterni. Lo studente deve esprimere un giudizio sul valore e sull'utilità di qualcosa per uno scopo determinato.

Lo studente che compie una valutazione in base a criteri interni deve usare come proprio strumento la coerenza logica. Per esempio, lo studente deve determinare se le conclusioni di uno scrittore discendono dai dati di cui egli era in possesso o se il programma di un computer è logico. Lo studente che compie una valutazione sulla base di criteri esterni deve usare come suo strumento criteri ben specificati, fornitigli da esperti. Per esempio, gli può essere richiesto di criticare un articolo scientifico servendosi della "Checklist for Evaluation Experimental Research in Psychology and Education" di Borg e Gall. Deve anche confrontare il risultato osservato con il risultato a cui si mirava o con il risultato ideale. Per chiarire meglio, lo studente deve confrontare i risultati del suo esperimento con i risultati correnti o il suo lavoro mentale con il modello.

Tassonomia di Bloom (Fonte: Block e Anderson, 1975)

UD3.

Diversi formati di quesiti strutturati

Obiettivi

- Riconoscere le differenze strutturali tra i diversi formati
- Identificare il formato dei quesiti più idoneo agli obiettivi della prova

Sommario

Dopo aver definito la finalità generale di uno strumento di valutazione e aver selezionato gli obiettivi specifici che intende verificare, occorre stabilire le forme dei quesiti che comporranno la prova. Il formato dei quesiti permette di differenziare il grado di analiticità con il quale procedere nella verifica e può rimandare a diverse modalità di correzione e pesatura delle risposte.

Le prove possono avere diversi gradi di strutturazione: dalle scelte con risposte predefinite, con un numero variabile di alternative possibili, a forme che invitano a operazioni di produzione delle risposte.

Introduzione al contenuto

Una volta stabilita la finalità della prova e gli obiettivi generali, la fase successiva consiste nell'individuare il tipo di quesiti da porre e quindi nel pianificare la costruzione della prova. Due sono le forme generali di quesito prevalentemente adoperate per le prove di conoscenza. I quesiti strutturati a scelta predefinita e i quesiti semi-strutturati a risposta scritta. In questa sede saranno presi in esame i primi, per una analisi dettagliata dei secondi si rimanda ad altro contributo.

Quesiti strutturati a scelta pre-definita <i>(selection-type items)</i>	Quesiti a risposta scritta <i>(supply-type items)</i>
1. scelta multipla 2. vero falso 3. corrispondenza 4. completamento	1. risposta sintetica (singola) 2. saggio breve (trattazione sintetica)

In generale possiamo dire che i quesiti strutturati formulati con modalità “oggettiva”, nelle diverse forme di presentazione che vedremo, pongono lo studente di fronte ad un'attività di soluzione di un problema (*problem solving*) e indicano le possibili alternative di risposta o soluzioni.

Se, ad esempio, volessimo verificare l'abilità dello studente nella comprensione di determinate parole o concetti nella lettura di un articolo di giornale, potremmo proporre di risolvere diversi tipi di quesiti, che sostanzialmente mettono lo studente di fronte all'individuazione tra due o più scelte. In alcuni casi potremmo invitare lo studente a individuare la definizione corretta di alcuni termini tra due possibili sinonimi o chiedere se una definizione fornita si possa ritenere corretta o sbagliata. In altri casi inviteremo lo studente a indicare tra le possibili scelte offerte quella che ritiene corretta o migliore. Nel primo caso il *problem solving* pone due modalità di risposta (vero-falso), e chiamiamo tali quesiti dicotomici, nel secondo caso, che offre più possibilità di risposta, i quesiti si definiscono politomici.

Nella scelta della forma di quesito migliore si dovrà quindi considerare: a) il numero di soluzioni che è possibile proporre rispetto al problema; b) la capacità del quesito di discriminare tra chi sa risolvere e chi non sa risolvere il problema posto dal quesito.

Tra le diverse forme di quesiti utilizzabili nelle prove oggettive le principali sono sostanzialmente quattro: vero-falso, scelta multipla, corrispondenza, completamento. Presentiamo di seguito sinteticamente le caratteristiche per ognuno di questi formati.

- Quesiti Vero-Falso

Sono i quesiti più semplici da costruire e fondamentalmente presentano un'affermazione e la richiesta di indicare se tale affermazione va considerata vera o falsa (giusta/errata; sì/no, etc.). La semplicità di costruzione, spesso fraintesa con la rilevazione di conoscenze semplici, ha portato a utilizzare questo formato per saggiare aspetti contenutistici o elementi di ricordo/memorizzazione. E' necessario sottolineare, invece, che il formato vero-falso rileva sostanzialmente la conoscenza verbale, in quanto richiede di segnalare la padronanza di concetti o di determinati elementi di conoscenza attraverso il riconoscimento della verità o falsità di enunciati linguistici. Se ben costruiti, i quesiti del tipo vero/falso permettono anche di rilevare forme complesse di ragionamento, quali la comprensione di elementi in tabelle, come nell'esempio riportato nello schema utilizzato per esemplificare la tassonomia degli obiettivi di Bloom.

Due esempi tipici di quesito del tipo "vero-falso":

<u>Esempi di quesito vero-falso</u>	
La validità di un test indica il grado con cui misura quello che effettivamente si vuole misurare.	V F
Quando <u>verrai</u> a casa mia, ti mostrerò quel libro. venissi	

- Quesiti a scelta multipla

I quesiti a scelta multipla sono costituiti dai seguenti elementi: a) il **corpo della domanda**, che corrisponde alla posizione del problema; b) n **alternative di risposta** (generalmente 4 o 5), che corrispondono alle proposte di soluzione del problema posto.

Il quesito che presentiamo di seguito, per esemplificare un tipo di scelta multipla, è costituito da 4 alternative di risposta di cui solo una è quella corretta. Le tre alternative di risposta non corrette vengono definite **distrattori**, cioè risposte in parte plausibili che hanno la funzione di "distrarre" colui che è chiamato a rispondere durante il processo di individuazione dell'alternativa esatta.

<u>Esempio di quesito a scelta multipla</u>	
Con il termine validità di un test si intende	<div style="border-left: 1px solid black; border-right: 1px solid black; border-bottom: 1px solid black; width: 30px; height: 20px; display: inline-block;"></div> CORPO DELLA DOMANDA
a) il grado di precisione nella misurazione. b) il grado con cui misura ciò che si vuole misurare. c) il grado di accordo tra valutatori. d) il grado di oggettività nella misurazione.	<div style="border-left: 1px solid black; border-right: 1px solid black; border-bottom: 1px solid black; width: 30px; height: 20px; display: inline-block;"></div> RISPOSTE o Alternative di risposta

I quesiti a scelta multipla permettono la misurazione di ragionamenti di un certa complessità, infatti la loro soluzione richiede di scegliere tra differenti risposte che rimandano ciascuna a differenti percorsi di soluzione.

I quesiti a scelta multipla con 4 o più alternative, di cui solo una è corretta, sono quelli più usati e sono ormai considerati “classici”. La letteratura sull’argomento ha studiato analiticamente altre forme di quesiti a scelta multipla che invitino altri tipi di ragionamento, offrendo di graduare le risposte o di combinarle. Queste forme più sofisticate possono essere costruite per specifiche necessità di verifica, e possono rimandare ad una diversa pesatura nell’attribuzione di punteggio. Alcune forme alternative sono⁶:

➤ *scelte multiple con due risposte corrette*

Aristotele usa il termine «sinolo» per indicare

1. l’universo sensibile;
2. l’unione di materia e di forma;
3. l’individuo concreto;
4. la realtà incorruttibile della pura forma;
5. l’unione di anima sensitiva e razionale.

➤ *scelte multiple con riconoscimento dell’esattezza.* Poniamo il seguente quesito:

Il piombo è un metallo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
a causa del fatto che	A	B	C	D	E
il suo peso specifico è alto					

Lo studente potrà indicare A se ritiene che le due affermazioni (“il piombo è un metallo” e “il suo peso specifico è alto”) e il legame (“a causa del fatto che”) siano veri; B se ritiene che solo le due affermazioni siano vere; C se ritiene che la prima affermazione sia vera e la seconda falsa; D se ritiene che la prima sia falsa e la seconda vera; E se ritiene che ambedue le affermazioni siano false.

➤ *scelte multiple con individuazione delle alternative migliore/peggiore.* Di fronte ad un quesito con 4 o 5 alternative di risposta lo studente dovrà indicare quelle che a suo avviso sono la risposta migliore (M) e quella peggiore (P). Così facendo sarà costretto a scegliere tra le diverse alternative, e avrà una combinatoria di soluzioni intermedie (dall’individuazione della M e P, all’inversione di M con P, e tutte le alternative possibili)

⁶ Gattullo propone alcune di queste forme, discutendole analiticamente sia dal punto di vista della costruzione, sia dell’attribuzione dei punteggi (op. cit, 1968, 179-189; 269-273 e tabelle C-D-E).

Le prove orali del controllo scolastico di tipo tradizionale possono essere rese più oggettive se, tra l'altro:

- M a. l'oggetto è determinato in termini di comportamento;
- b. le domande formulate sono molto numerose;
- P c. l'esaminatore sollecita capacità intellettuali diverse da soggetto a soggetto;
- d. la lettura dei risultati è compiuta in modo reciprocamente indipendente da due esaminatori.

- *scelte multiple con classificazione delle alternative.* Con questo quesito si chiede allo studente di mettere in ordine di importanza le diverse alternative. In questo caso, più che nelle forme precedenti, le alternative di risposta sono date da tutte le classificazioni possibili (da quella corretta, come nell'esempio a-b-c-d, a quella rovesciata d-c-b-a, passando per tutte le combinazioni possibili).

Le prove orali del controllo scolastico di tipo tradizionale possono essere rese più oggettive se, tra l'altro:

- 1 a. l'oggetto di esse è determinato in termini di comportamento;
- 2 b. le domande formulate sono molto numerose;
- 4 c. l'esaminatore sollecita capacità intellettuali diverse da soggetto a soggetto;
- 3 d. la lettura dei risultati è compiuta in modo reciprocamente indipendente da due esaminatori.

- Quesiti a completamento

Sono quelle forme di quesiti che si costruiscono eliminando uno o più elementi da una frase o da un testo, presentando agli allievi un testo "bucato" e chiedendo loro di completarlo con gli elementi mancanti (a volte riportati in fondo all'esercizio). Spesso si forniscono anche delle alternative di risposta per i completamenti da effettuare.

Il completamento può riguardare semplici frasi oppure piccoli brani come nell'esempio che riportiamo:

La docimologia è, tra le ormai numerose scienze dell'educazione, quella particolare disciplina che si occupa dei problemi _____. E' sorta attorno agli anni _____, attraverso la sistematizzazione teorica delle prime ricerche sulla _____ dei giudizi di esame. Successivamente si è occupata della _____ delle misurazioni dell'apprendimento e delle _____ didattiche.

- | | | | | |
|---------------|------------------|-----------------|------------------|---------------|
| 1. Trenta | 2. coerenza | 3. precisione | 4. istruzione | 5. Venti |
| 6. valutativi | 7. attendibilità | 8. decisioni | 9. educazione | 10. educativi |
| 11. validità | 12. incoerenza | 13. metodologia | 14. soggettività | |

Esiste una categoria di prove di verifica oggettive all'interno della quale si utilizza la forma del completamento applicandolo però a testi più lunghi, e utilizzando una tecnica di "bucatura" sistematica. Questi strumenti vanno sotto il nome di *cloze test*. Sono prove ampiamente adoperate per la misura della leggibilità e della comprensione dei testi scritti. A differenza dalle prove di completamento, nelle prove di *cloze* i buchi sono sistematici, per esempio, ogni 5 o ogni 10 parole. Il lettore ricostruisce il significato della parola mancante aiutandosi con il testo presente. La costruzione di questo tipo di prove è quindi estremamente semplice. Per la correzione si contano i buchi riempiti, il punteggio complessivo considera il numero di inserimenti corretti sul totale dei buchi previsti (percentuale di completamento).

- Quesiti a corrispondenza

Si tratta di stabilire un collegamento, una corrispondenza tra due o più serie/elenchi di dati. Si pensi ai casi in cui si vuole rilevare l'attribuzione di definizioni, date, o autori. Si costruiscono allora due elenchi (termini/espressioni; autori/date; etc.) e si chiede di far corrispondere, di mettere in relazione, gli elementi dei due elenchi per mezzo di frecce, numerazioni o altre forme di corrispondenza. Nell'esempio che segue viene riportato un quesito a corrispondenza che tende a rilevare le definizioni relativamente ai tipi di quesito che stiamo presentando.

Istruzioni: collega le due liste scrivendo accanto ad ogni lettera della prima lista un solo numero tra quelli elencati nella seconda lista.

Quesiti	Caratteristiche di risposta
a) _____ vero/falso	1 _____ diverse possibilità di risposta di cui solo una è giusta
b) _____ scelta multipla	2 _____ collegamento tra due o più liste di elementi
c) _____ corrispondenza	3 _____ una possibilità di risposta
d) _____ completamento	4 _____ eliminazione di alcuni elementi da una frase o da un testo
	5 _____ diverse possibilità di risposta su più liste di elementi

I quattro formati descritti sono quelli maggiormente usati nella costruzione di prove strutturate. Sempre più spesso nelle indagini comparative internazionali, accanto a forme strutturate di tipo chiuso, si presentano alcuni formati più aperti che richiedono un'attività di giudizio nella correzione. L'importanza di un approccio misto (uso di più forme di quesito in una stessa prova), è largamente presente nelle comparazioni internazionali. Nella tabella n.5 si riporta il quadro dei

quesiti per i processi di lettura adoperati nella recente indagine internazionale sui livelli di competenza in matematica, lettura, scienze e *problem solving* (PISA⁷, 2003). Seguono, nella tabella n. 6 alcuni esempi dei quesiti a risposta aperta utilizzati nell'indagine.

Processo di lettura	% quesiti				Totale
	Scelta multipla semplice (1)	Scelta multipla complessa (2)	Risposta aperta univoca (3)	Risposta aperta articolata (4)	
Individuare informazioni		4	14	11	29
Interpretare il testo	29	3	7	11	50
Riflettere e valutare				21	21
Totale	29	7	21	43	100

Tabella n.5 – Quesiti per i processi di lettura utilizzati nell'indagine OCSE-PISA, 2003

Legenda

- (1) Scelta multipla semplice: scelta tra diverse alternative
 (2) Scelta multipla complessa: scelta di più di un'alternativa
 (3) Aperta univoca: richiedono un'attività minima di giudizio
 (4) Aperta articolata: richiedono il giudizio da parte di un correttore

Domanda aperta a risposta univoca
<p>Susanna vuole costruire un solido che sembri pieno e che abbia 6 cubetti in lunghezza, 5 cubetti in larghezza e 4 cubetti in altezza. Vuole utilizzare il minor numero possibile di cubetti, lasciando più spazio vuoto possibile all'interno del solido. Qual è il numero minimo di cubetti che servono a Susanna per costruire questo solido?</p> <p>Risposta: _____ cubetti.</p> <p><i>Punteggi e commenti</i> Punteggio pieno (Codice 1): risposte che indichino 96 cubetti Nessun punteggio (Codice 0): altre risposte</p>
Domanda aperta a risposta articolata
<p>Già nel XI secolo, i medici cinesi lavoravano sul sistema immunitario. Soffiando nelle narici dei loro pazienti croste seccate e polverizzate prelevate da una vittima del vaiolo, riuscivano spesso a provocare un leggero accesso della malattia, che impediva successivamente un attacco più virulento. Nel XVIII secolo, le persone si sfregavano la pelle con le croste secche delle pustole per proteggersi da questa malattia. Tali pratiche primitive furono introdotte in Inghilterra e nelle colonie dell'America. Tra il 1771 e il 1772, durante un'epidemia di vaiolo, un medico di Boston che si chiamava Zabdiel Boylston graffiò la pelle di suo figlio di 6 anni, come anche quella di altre 285 persone, e sfregò questi graffi con del pus proveniente dalle pustole del vaiolo. Tutti i suoi pazienti sopravvissero, tranne sei.</p> <p>Quale poteva essere l'ipotesi che Zabdiel Boylston stava verificando?</p> <p><i>Punteggi e commenti</i> Punteggio pieno (Codice 2): risposte che fanno riferimento a:</p> <ul style="list-style-type: none"> - l'idea di infettare qualcuno con il vaiolo ne accresce le probabilità di sviluppare un'immunità <p>E</p> <ul style="list-style-type: none"> - l'idea che perforando la pelle il vaiolo venisse introdotto nel flusso sanguigno

⁷ Gli esempi sono tratti dal volume: Ocse, *Pisa 2003, Valutazione dei quindicenni*, Armando Editore, Roma, 2004

Punteggio parziale (Codice 1): nel caso si faccia riferimento a uno solo dei punti illustrati Nessun punteggio (Codice 0): altre risposte
--

Tabella n.6 - Alcuni esempi dei quesiti a risposta aperta utilizzati nell'indagine OCSE-PISA, 2003

Legenda:

- *a risposta univoca* che richiedono la formulazione di una breve risposta che può rientrare in una gamma di possibilità molto ristretta o addirittura essere obbligata;
- *a risposta articolata (o breve)* che richiedono la formulazione di una risposta, breve ma che può essere personalizzata;

UD4.

Formulare i quesiti (esempi e alcune regole)

Obiettivi

- Migliorare la capacità misurativa dei quesiti
- Controllare l'organizzazione dei quesiti nella prova
- Migliorare la formulazione linguistica dei quesiti

Sommario

Le prove strutturate per essere validi strumenti di valutazione degli apprendimenti non devono suscitare ambiguità e non dovrebbero essere delle prove-trabocchetto. Le regole e le indicazioni per la formulazione dei quesiti strutturati proposte in questa unità hanno la funzione di aiutare a costruire prove che abbiano una maggiore validità e una maggiore attendibilità nella misurazione degli apprendimenti.

Le regole sono semplici e permettono di controllare i diversi aspetti distorsivi dei quesiti dal punto di vista dell'organizzazione della prova, della formulazione linguistica e dell'impostazione generale. Per offrire uno strumento operativo, le regole sono presentate in forma di liste di controllo (check-list), che il costruttore di quesiti dovrebbe porsi in termini di autovalutazione per controllare il rispetto della validità e dell'affidabilità della misura.

La costruzione di buoni quesiti, vale a dire coerenti con gli scopi della valutazione e capaci di misurare le diverse prestazioni degli studenti, garantisce l'idoneità complessiva delle prove strutturate. Una scarsa attenzione a questi aspetti banalizza l'uso delle prove strutturate e ne inficia fortemente l'utilizzazione come strumenti utili alla valutazione.

Introduzione al contenuto

Per migliorare la costruzione dei differenti tipi di quesiti è indispensabile rispettare alcune regole generali, valide a prescindere dal tipo di quesiti che si costruiranno. Queste avvertenze, e le relative giustificazioni (vedi tabella n.7: riadattamento della tabella riportata in Gattullo M., Giovannini L. (1989), *Misurare e valutare l'apprendimento nella scuola media*, B. Mondatori, Milano), sostanzialmente sottolineano la necessità di evitare tranelli e di pensare a quesiti che invitino a ragionare, offrendo chiare e plausibili alternative di risposta. Seguendo le diverse indicazioni, e le regole che più avanti si presenteranno, si potranno mettere a punto stimoli intelligenti, utili a sondare le modalità di apprendimento e a comparare le prestazioni dei singoli studenti.

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 sia 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 solo 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 (EVITARE IL NOZIONISMO)	5. Si misurerebbe solo 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.: VFVFVFVFVF; oppure: VVFFVVFFVVFFVVFFVVFF)	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

Tabella n.7 - Indicazioni generali per la costruzione di quesiti oggettivi

Più che di regole⁸ per costruire i quesiti, si offrono di seguito alcune liste di domande che i costruttori di quesiti potranno utilizzare per controllare e revisionare le prove strutturate. Le liste di controllo sono diverse per i differenti formati dei quesiti, e sono organizzate su più piani, per analizzare in dettaglio le loro diverse caratteristiche. Le prime domande riguardano la *posizione e*

⁸ C'è una lunga tradizione di "regole" per la costruzione di quesiti "oggettivi". Per un approfondimento si veda la bibliografia riportata in Benvenuto, *Mettere i voti a scuola*, (Carocci, Roma 2003).

soluzione del problema, per evitare che si sondino conoscenze e abilità secondarie o in modo confuso; seguono le domande riguardanti la *formulazione linguistica*, per migliorare la comprensibilità e ridurre le possibili ambiguità di tipo linguistico; seguono le domande sull'*organizzazione dei quesiti*, per controllare la coerenza e l'eventuale indipendenza dei quesiti nelle prove strutturate; alla fine sono presentate le domande relative ad alcune *convenzioni grafiche*.

Lista di controllo (Check list) per la costruzione di quesiti a scelta multipla

Posizione e soluzione del problema

- 1) A quale obiettivo "mira" la domanda? (cfr. tassonomia Bloom o descrittori di prestazioni)
- 2) La domanda focalizza un solo problema o concetto?
- 3) La domanda è espressa con precisione per evitare incertezze e confusioni nella scelta della risposta?
- 4) I distrattori sono plausibili rispetto al problema considerato? Ogni distrattore deve, per contenuto e natura, risultare in qualche modo collegato alle domande. Un distrattore evidentemente non corretto facilita la sua esclusione (e quindi l'individuazione della risposta esatta).

Formulazione linguistica

- 1) La domanda è formulata in modo semplice ed esplicito? Il corpo della domanda dovrebbe essere composto da una frase in forma interrogativa o affermativa.
- 2) (Per quesiti di comprensione della lettura) Hai usato un linguaggio simile a quello del testo a cui ci si riferisce?
- 3) I distrattori e la risposta corretta sono omogenei per struttura sintattico-grammaticale e per lunghezza? Controlla se la risposta corretta sia significativamente più lunga delle altre alternative di risposta. La disomogeneità della lunghezza facilita la scelta della risposta corretta poiché di solito la risposta corretta è anche quella più lunga.
- 4) Hai evitato che le risposte esatte e i distrattori riprendano termini o formulazioni usate nella domanda? (o nel testo, per i quesiti di comprensione della lettura. A meno che non si richieda di localizzare un punto preciso del testo).
- 5) Hai evitato, nella costruzione delle alternative, l'utilizzazione di espressioni come *sempre, tutti, ogni, dappertutto*, che inducono a credere nell'inesattezza della risposta che le contiene? Oppure, termini come *quasi sempre, spesso, qualche volta*, che inducono a credere nell'esattezza delle alternative. Unica eccezione a quanto detto è il caso in cui tali espressioni siano usate in scala graduata: *sempre, quasi sempre, talvolta, quasi mai, mai*.
- 6) Hai evitato 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.
- 7) Hai disposto in ordine crescente o decrescente le alternative che riportano cifre o quantità?
- 8) Hai evitato alternative del tipo: *nessuna di queste, sia a) che c), tutte le precedenti*? Inserire un'alternativa di questo tipo significherebbe inserirne una evidentemente non omogenea alle altre, facilitando così l'individuazione della risposta corretta.
- 9) Hai evitato che elementi grammaticali o la struttura della frase favoriscano l'individuazione della risposta esatta?
- 10) Nel caso di alternative di risposta che iniziano con lo stesso gruppo di parole hai provato ad includere la parte comune nel corpo della domanda? In questo modo alleggerirai le diverse alternative di risposta.

Organizzazione dei quesiti

- 1) I quesiti sono tra loro indipendenti? Rispondere ad un quesito non deve aiutare a rispondere ad un altro.
- 2) Hai evitato di usare un'alternativa di risposta di un quesito come distrattore o come risposta corretta in altri quesiti?
- 3) Hai previsto la collocazione della risposta corretta in modo che nell'insieme dei quesiti le risposte esatte non corrispondano tutte alla stessa lettera?

Convenzioni grafiche

- 1) Le alternative di risposta dei quesiti che terminano con il punto interrogativo cominciano con la maiuscola?
- 2) Le alternative dei quesiti a completamento iniziano con la minuscola?
- 3) Per convenzione, alla fine di ogni alternativa di risposta hai usato il punto fermo?

- 4) Il corpo della domanda nella forma a completamento non termina mai con i due punti o con i puntini sospensivi. Controlla.
- 5) Le alternative di risposta sono disposte sotto forma di elenco e non di seguito una dietro l'altra?

Tabella n.8: come costruire quesiti oggettivi a scelta multipla

Lista di controllo (Check list) per la costruzione di quesiti vero-falso

Posizione del problema

- 1) A quale obiettivo mira l'insieme dei quesiti V/F?
- 2) A quale obiettivo mirano le singole affermazioni?
- 3) Hai evitato di verificare l'apprendimento mnemonico?
- 4) Le affermazioni sono espresse con precisione per evitare incertezze e confusioni nella scelta della risposta?

Formulazione linguistica

- 1) Hai evitato di utilizzare negazioni nelle affermazioni?
- 2) Le affermazioni sono espresse in forma chiara e concisa, evitando quindi dettagli secondari?
- 3) Hai utilizzato un linguaggio adeguato al livello dei destinatari?
- 4) Hai evitato l'uso di quantificatori quali *sempre*, *spesso*, *qualche volta*, *mai*, etc. in quanto inducono a ritenere più giusta o errata l'affermazione a prescindere dal quesito?

Organizzazione dei quesiti

- 1) Hai controllato l'indipendenza delle affermazioni? Il rispondere ad una non deve facilitare o pregiudicare la risposta ad un'altra.
- 2) Il numero delle affermazioni vere e di quelle false (giuste/esatte) sono bilanciate nel totale dei quesiti?

Tabella n.9: come costruire quesiti oggettivi vero-falso

Lista di controllo (Check list) per la costruzione di quesiti a corrispondenza

Posizione del problema

- 1) Gli elementi (termini, espressioni, enunciati, etc.) dei due elenchi sono omogenei tra loro per contenuto?
- 2) Ad evitare ambiguità nella risposta, almeno uno dei due elenchi contiene semplici elementi (date, singole parole, autori, etc.)? L'altro potrà contenere anche enunciati o elementi più complessi.
- 3) Nel caso di elenchi di diversa complessità, hai adoperato l'elenco con le espressioni più lunghe come premessa del quesito e l'altro per risposte?
- 4) Hai fornito chiare istruzioni per spiegare le modalità di corrispondenza tra gli elenchi (freccette, numeri, ordinamenti, etc.)?
- 5) Ad ogni elemento del primo elenco corrisponde soltanto uno esatto del secondo?

Organizzazione dei quesiti

- 1) Il numero delle corrispondenze da compiere è tra 5 e 10?
- 2) Hai evitato di costruire una corrispondenza "perfetta", cioè le due liste dovrebbero essere disomogenee per numero (una deve contenere più elementi dell'altra), per evitare di rispondere per esclusione?
- 3) Hai controllato che il numero degli elementi di "risposta" è al massimo doppio di quello da completare?

Tabella n.10: come costruire quesiti oggettivi a corrispondenza

Lista di controllo (Check list) per la costruzione di quesiti a completamento e cloze test⁹

Posizione del problema

- 1) Nei completamenti mirati, hai iniziato ad eliminare le parole dal testo dopo il primo periodo (altrimenti tralasciare le prime righe)?
- 2) Nei *cloze*, i periodi iniziali e finali sono rimasti integri? In questo modo si permette di contestualizzare maggiormente e si facilita la comprensibilità complessiva del testo.
- 3) Nei completamenti di frase, hai cercato di posizionare la parola mancante in modo da evitare eccessiva ambiguità?

Formulazione linguistica

- 1) Hai controllato la coerenza sintattica e grammaticale tra testo da completare e termini presentati in calce?
- 2) Hai evitato tra i termini da inserire di utilizzare sinonimi (ugualmente accettabili nel completamento)?
- 3) Tra testo da completare e alternative di completamento c'è omogeneità lessicale?

Organizzazione dei quesiti e convenzioni grafiche

- 1) Hai evitato di costruire prove di cloze con più di 15 completamenti?
- 2) La lunghezza degli spazi vuoti è sempre la stessa e non proporzionale al termine da inserire?
- 3) I termini da inserire sono elencati seguendo un ordine alfabetico? O con quale ordine?
- 4) Per ogni parola da inserire c'è almeno un'alternativa, oltre alla risposta esatta?
- 5) Ognuna delle parole da inserire può essere utilizzata più volte (tecnica del reinserimento tra le alternative)? Se diversamente, prevedere un numero di parole da inserire uguale al numero degli spazi previsti più 4? In questo modo anche l'ultimo inserimento sarà compiuto con almeno 5 alternative.

Tabella n.11: come costruire quesiti oggettivi a completamento e cloze-test

⁹ Per *cloze test* si intendono quelle prove che richiedono all'allievo di inserire le parole mancanti in un testo. A differenza dalle prove di completamento, nelle prove di *cloze* i buchi sono sistematici cioè, ad esempio, ogni 5 o ogni 10 parole. Il lettore ricostruisce il significato della parola mancante aiutandosi con il testo presente. La costruzione di questo tipo di prove è quindi estremamente semplice. Per la correzione si contano i buchi riempiti, e il punteggio complessivo è rappresentato dalla percentuale sul totale dei buchi.

UD5.

Regole per la somministrazione delle prove

Obiettivi

- Pianificare i diversi aspetti della somministrazione delle prove
- Rispettare i singoli elementi della somministrazione

Sommario

Gli strumenti strutturati di valutazione sono prove che sollecitano prestazioni culturali, sono costituiti da un congruo numero di quesiti (item) rispetto ai quali gli studenti si limitano ad indicare le risposte che considerano corrette. Le prove sono effettuate collettivamente, e in un tempo contenuto.

Queste indicazioni che definiscono sinteticamente le caratteristiche delle prove strutturate indicano quanto sia importante che esse vengano utilizzate con la garanzia del rispetto di uguali condizioni ambientali e procedurali per tutti coloro che vi si sottopongono. Le garanzie riguardano una serie di elementi da controllare durante la somministrazione delle prove, pena l'invalidazione dei risultati.

Questa unità si propone di illustrare le principali regole per una buona somministrazione, una somministrazione cioè che sia in grado di garantire uguali condizioni per tutti, quali il rispetto dei tempi e delle consegne, ma anche l'assenza di interferenze di diverso tipo legate sia ai soggetti sia al comportamento del somministratore.

Introduzione al contenuto

Una volta terminata la costruzione di una prova strutturata, dovremmo verificarne l'idoneità complessiva come strumento di misura. La prova migliore consiste nell'analizzare i risultati che un gruppo di rispondenti offrono ad una prima somministrazione, che chiameremo di prova o pilota (*try-out test*). Un proverbio inglese dice che l'unico modo per provare un budino è mangiarlo. Allo stesso modo, per vedere se una prova oggettiva funziona non c'è altra strada che somministrarla ad un gruppo di studenti. Soltanto dopo questa verifica sul campo, e l'eventuale intervento di revisione, si potrà essere sicuri dell'idoneità della prova oggettiva e si potrà quindi procedere ad utilizzarla veramente, cioè per ottenere dati che saranno utilizzati nella valutazione dell'apprendimento e l'assegnazione di un punteggio.

La presenza e l'attenzione alle fasi di analisi dei quesiti (*item analysis*) e di somministrazione segna spesso la differenza tra prove comuni e prove standardizzate. Le prime in molti casi sono costruite e utilizzate in un gruppo senza una sostanziale verifica della loro idoneità come strumento di misura e non sempre evitando distorsioni nella rilevazione (scarsa concentrazione dei singoli e del gruppo, presenza di collaborazione nella risposta, non rispetto dei tempi nell'esecuzione ecc.). Quando invece si procede a rilevazioni e somministrazioni di natura campionaria o per indagini di natura comparativa (nazionali o internazionali) l'attenzione alle due fasi descritte deve essere garantita, altrimenti i risultati non offrono alcuna affidabilità.

Sia la somministrazione pilota, sia quella finale di prove strutturate dovrebbero rispettare una serie di condizioni e tener conto di determinati fattori, al fine di non facilitare alcuni studenti e quindi viziare la rilevazione.

Gli elementi fondamentali da considerare nelle diverse somministrazioni sono:

1. *Gli studenti sono sufficientemente motivati allo svolgimento della prova?* È necessario che tutti coloro che si sottopongono ad una prova valutativa, non solo di natura strutturata, sappiano chiaramente di che cosa si tratta e per quale fine essa è stata pensata. Serve per una verifica formativa o per attribuire voti/giudizi? È una prova di comparazione a livello nazionale o internazionale? In pratica, l'esplicitazione delle finalità serve a far sì che gli studenti diano il giusto peso, sia valutativo sia didattico, alla prova e trovino più motivazioni a svolgere la prova in modo autonomo piuttosto che chiedere la collaborazione di altri e/o prestare aiuto a chi lo chiede.
2. *Ci sono istruzioni predefinite (da leggere prima di sottoporsi alla prova)?* Non sempre sono obbligatorie, soprattutto se gli studenti hanno già familiarizzato con questi strumenti di verifica. In caso contrario è bene prevedere brevi esempi per aiutare a marcare le risposte e a correggere gli eventuali ripensamenti nelle risposte fornite. Le istruzioni possono essere lette collettivamente prima della prova oppure presentate dal somministratore (docente o chi per lui). In ogni caso esse dovranno essere chiare e concise, per evitare che l'allievo sia costretto a usare eccessivo tempo e concentrazione per capire che cosa gli si chiede, e cioè

in attività non direttamente connesse con la conoscenza/abilità che si vuole valutare. Sia i test su carta, sia quelli su computer possono utilizzare fogli di risposta o modalità specifiche di risposta. È meglio abbondare in chiarezza e illustrare analiticamente le modalità e possibilità previste.

3. *In che modo si somministra?* Se la prova richiede una somministrazione collettiva di gruppo, le istruzioni e le condizioni sono uguali per tutti; quando invece la somministrazione è prevista per piccoli gruppi, o addirittura per singoli all'interno del gruppo, oltre a garantire l'uniformità delle istruzioni e condizioni è indispensabile predisporre un piano di gestione del resto del gruppo. A volte, soprattutto nel caso di collaborazione a indagini nazionali e internazionali che prevedono una batteria di prove e diversi strumenti di rilevazione, la somministrazione può impegnare gli studenti per più sessioni, anche in giorni diversi. Per le somministrazioni che si distribuiscono nel tempo (stessa giornata o in diverse giornate) è indispensabile preparare il gruppo interessato per evitare troppe assenze e ridurre l'eventuale affaticamento.
4. *Il somministratore è il docente del corso?* Nella gran parte delle prove la risposta è ovviamente affermativa. Ma qualora volessimo utilizzare prove comuni con altri docenti, o ad esempio, per le prove di ingresso in un corso, il somministratore potrebbe essere anche esterno al gruppo. L'importante è che si rispettino le indicazioni di imparzialità e di chiarezza nell'esecuzione, il rispetto dei tempi e la massima concentrazione da parte di tutti. Quando poi si è interessati a somministrazioni di prove a fini comparativi di tipo nazionale o internazionale, le garanzie di imparzialità dovrebbero aumentare. Eventuali atteggiamenti poco rigorosi, o addirittura "facilitanti" da parte di docenti incaricati della somministrazione invaliderebbero l'intera rilevazione e i progetti che ne utilizzano i dati. In sede di analisi dei dati sono stati rilevati alcuni casi, verificatisi durante alcune esperienze di indagini nazionali, di docenti che hanno assunto, consapevolmente o inconsapevolmente, comportamenti "compiacenti" nei confronti dei loro studenti. Nelle indagini internazionali per evitare possibili distorsioni nella fase di somministrazione, e in base alle finalità generali e alle risorse finanziarie, si ricorre a diverse opzioni: a) utilizzazione di somministratori esterni; b) formazione di somministratori interni; c) controllo dei risultati e dei protocolli di somministrazione; d) affiancamento di somministratori esterni e interni su un campione di enti (o corsi) interessati.
5. *Il test è "carta e matita" o su computer?* Nel primo caso si dovrà evitare l'eventualità che gli studenti copino e quindi prevedere forme parallele (cioè diverse ed equivalenti) o forme mescolate, che presentino cioè gli stessi quesiti ma in ordine diverso, o con una diversa distribuzione delle alternative di risposta. Nel caso dell'uso di forme personalizzate su computer, con item casuali per ogni studente, dovremmo controllare solo che il programma funzioni correttamente. Sia per le prove tradizionali, sia per quelle su computer si dovrà curare l'ambiente in modo da renderlo idoneo allo svolgimento di prove strutturate.
6. *I tempi sono chiari e rispettati da tutti?* I tempi complessivi per lo svolgimento, e quelli per la riconsegna della prova e dell'eventuale modulo per la registrazione delle risposte dovranno essere rigorosamente rispettati. Spesso il tempo aggiuntivo, per far completare i più lenti, o quelli che ne fanno richiesta, risulta dannoso in quanto spesso è usato per consultare i compagni che hanno già finito. D'altra parte si deve prevedere un maggior controllo per quelli che finiscono prima del tempo e si impegnano in attività che possono recare disturbo a chi deve ancora finire. Il rispetto del tempo è un'altra garanzia dell'oggettività della misura. La regola riguarda anche il somministratore, che dovrà garantire che il tempo complessivo per lo svolgimento non sia interrotto o compromesso.

Per quanto riguarda la modalità della somministrazione che può essere collettiva, individuale o automatizzata si faccia riferimento alle indicazioni riportate nello schema seguente.

Somministrazione collettiva	Somministrazione individuale	Somministrazione automatizzata
<ul style="list-style-type: none"> ◆ E' effettuata da uno o più operatori (interni/esterni). ◆ I soggetti rispondono individualmente. ◆ Occorre fare attenzione alle esemplificazioni, alle spiegazioni e ai tempi. ◆ L'interazione individuale con il somministratore è limitata. ◆ I comportamenti non-verbali sono difficilmente osservabili. ◆ La presenza e le funzioni dell'insegnante di classe, nel caso di somministratori esterni, devono essere regolate da precisi accordi. 	<p>L'interazione a due è condizionata da fattori quali:</p> <ul style="list-style-type: none"> ◆ la prossemica, ◆ la personalità del soggetto, ◆ la posizione all'interno della classe. 	<ul style="list-style-type: none"> ◆ L'intera operazione è svolta tramite computer. ◆ Possibilità di personalizzazione delle prove. ◆ Occorre stabilire norme molto precise. ◆ Le istruzioni sono di solito scritte o registrate.

Tabella n.12 – Modalità di somministrazione

La chiarezza e l'esaustività delle istruzioni per lo svolgimento della prova rappresentano nelle somministrazioni aspetti di fondamentale importanza. Ritorniamo brevemente su questo aspetto per offrire qualche esempio specifico.

Nella fase organizzativa della somministrazione occorre decidere se le istruzioni saranno date oralmente, per iscritto o in entrambi i modi, se dovranno essere ripetute e in quale occasione o momento, da chi dovranno essere fornite e se, infine, dovranno contenere esemplificazioni che consentano una migliore comprensione.

Le istruzioni devono contenere tutti quegli elementi di cui lo studente può avere bisogno per completare la prova. Istruzioni chiare su ciò che lo studente deve fare, una volta ricevuta la prova da svolgere, rendono i risultati più precisi ed evitano ambiguità.

Le istruzioni vanno sempre poste all'inizio di ogni prova e lette testualmente o recitate a memoria. Al momento della somministrazione occorre farle leggere anche agli studenti ed eventualmente rileggerle ad alta voce.

Non bisogna dare a voce istruzioni che discordino da quelle scritte sul foglio, ciò potrebbe generare confusione. Dopo avere fornito le istruzioni il docente o il somministratore, prima di dare inizio alla prova, chiederà agli studenti se è tutto chiaro o se sono necessarie ulteriori spiegazioni. In quest'ultimo caso, saranno forniti chiarimenti relativi solo alla struttura della prova e non ai suoi contenuti. Tutte le spiegazioni aggiuntive dovranno essere annotate sul verbale del somministratore. Si riporta di seguito un esempio di istruzioni riguardante un test di esonero previsto in un corso universitario.

Esempio di istruzioni

Test di esonero per il corso di Docimologia (a.a. 2003/2004)

ISTRUZIONI

Questo fascicolo contiene 50 quesiti/item per indagare il livello di studio di tutto il programma previsto.

Ci sono due tipi di quesiti:

- quesiti a scelta multipla (con 4 alternative di risposta)
- quesiti a corrispondenza multipla (per tre quesiti si offrono 5 alternative di risposta)

Leggi con attenzione ogni singolo quesito e quindi rispondi scegliendo una sola alternativa di risposta.

Per rispondere utilizza il **modulo ottico** che ti è stato dato insieme a questo fascicolo nel seguente modo: indica per ogni quesito l'alternativa che ritieni esatta mettendo un pallino nero ben visibile nel quadratino relativo alla lettera che hai scelto in corrispondenza del numero d'ordine del quesito riportato nel modulo ottico.

Esempio:

Quesito 1. (fascicolo) -----> item n.1 (modulo ottico)

Item n.	A	B	C	D	E
1			●		

Se decidi di cambiare la risposta che hai dato fai una croce sul quadratino sbagliato e metti un pallino nero ben visibile nel quadratino corrispondente alla tua nuova scelta.

Esempio:



Item n.	A	B	C	D	E
1	●		●		

Per i quesiti a scelta multipla le alternative di risposta sono solo 4. Fai quindi attenzione a **non utilizzare mai** la colonna dell'alternativa di **risposta E**.

Per i quesiti a corrispondenza multipla le alternative di risposta sono 5. Solo per queste **puoi utilizzare** la colonna dell'alternativa di **risposta E**

Per qualsiasi dubbio sulle modalità di risposta ti puoi rivolgere alle persone presenti in aula che hanno distribuito il fascicolo.

Hai 50 minuti di tempo per completare il lavoro.

Forma 1

UD6.

Dalla misurazione alla valutazione

Obiettivi

- Interpretare i risultati rispetto a criteri o norme
- Attribuire pesi ai diversi quesiti
- Elaborare i punteggi grezzi per trasformarli in valutazioni (voti)

Sommario

Una volta somministrata la prova strutturata si possono raccogliere i risultati e utilizzarli per valutare l'andamento dei singoli e della classe.

Per trasformare i punteggi grezzi in possibili valutazioni è necessario decidere sul tipo di peso che i singoli quesiti possono avere, e quindi procedere a forme di ponderazione dei punteggi, di calcolo per individuare alcuni indici statistici e d'altra parte assumere decisioni su come interpretare i risultati, vale a dire stabilire punteggi soglia o di sufficienza per valutare complessivamente i risultati. L'insieme di queste operazioni implica decisioni di natura quantitativa e qualitativa, che permettono di utilizzare i risultati delle prove strutturate come veri e propri strumenti valutativi.

Introduzione al contenuto

Quando la verifica è compiuta utilizzando prove strutturate possiamo usare i punteggi, cioè i risultati, per attribuire “voti”. Ma per poter leggere, comparare e interpretare i punteggi e ricorrere a diversi calcoli e metodi statistici, dobbiamo avere garanzia del rispetto delle seguenti condizioni:

- i dati siano stati raccolti nel rispetto dell’uniformità della somministrazione e quindi senza distorsioni nella fase di rilevazione;
- lo strumento di misura (prova/test) sia affidabile (*reliable*), cioè permette di rilevare i punteggi senza interferenza della soggettività del correttore o nel più ampio accordo tra possibili valutatori esterni;
- i quesiti mal formulati o poco funzionanti siano stati scartati, revisionati o non considerati nell’attribuzione del punteggio grezzo.

Una volta rispettati i punti precedenti, possiamo assumere che lo strumento/prova permette una lettura affidabile dei risultati e possiamo usare i risultati che emergono dalla sua somministrazione in classe o a gruppi più o meno piccoli di studenti.

Le fasi che permettono di passare dalla misurazione (calcolo di punteggio) alla valutazione (attribuzione di voto o giudizio) sono:

1. correzione e pesatura degli quesiti e delle parti della prova;
2. analisi delle distribuzioni dei risultati ed elaborazione di indici sintetici;
3. scelta di un criterio valutativo.

Di seguito le fasi vengono descritte sequenzialmente per esplicitare le diverse procedure richieste.

1. Correzione e pesatura dei quesiti e delle parti della prova

Il rispetto delle principali avvertenze per la somministrazione elimina i fattori come la collaborazione tra gli studenti per la soluzione dei quesiti. È vero anche che lo studente tende a copiare quanto più è consapevole che i risultati di quella prova possono compromettere il suo percorso scolastico, in caso diverso egli tende a percepire la prova come strumento di valutazione formativa e di conseguenza assume un comportamento più ortodosso. In ogni caso rimane da risolvere uno dei limiti che spesso si attribuiscono alle prove oggettive: la "casualità" delle risposte, tecnica a cui possono ricorrere gli studenti, soprattutto quando non sanno rispondere ai quesiti proposti.

Si sostiene spesso che le risposte ai quesiti a scelta multipla possono essere date a caso, intendendo che il caso può giocare a favore di chi non sa che "pesci prendere". In effetti, seguendo questo

ragionamento, si può sostenere che chi non sa rispondere ad una domanda orale, rimanendo in silenzio dichiara i propri limiti (conoscenze, abilità, etc.), mentre chi non sa rispondere ad un quesito oggettivo ha sempre la possibilità di segnare una delle possibili risposte e quindi di avvalersi del beneficio che può derivare dal caso. Per bilanciare questo limite esistono almeno due possibilità:

- *calcolo delle risposte giuste senza omissioni*. Si fornisce agli studenti l'indicazione di non segnare alcuna risposta nel caso in cui non conoscano la risposta esatta, avvisandoli però che l'omissione di risposta non verrà tenuta in conto. Il punteggio in tal caso si può calcolare percentualizzando il numero di risposte giuste sul totale delle risposte valide (cioè non considerando le omissioni). Questo metodo funziona solo nel caso in cui le omissioni siano ben distribuite tra gli studenti, altrimenti si può verificare il caso in cui lo stesso punteggio è ottenuto dallo studente che ha risposto correttamente ad una sola domanda ($1 \times 100 / 1 = 100$) e da un altro che ha risposto correttamente a tutte le domande ($40 \times 100 / 40 = 100$);

- *pesatura delle risposte considerando la probabilità del caso*. Il ragionamento in questo caso è il seguente: se è vero che esiste la possibilità di affidarsi al caso è indispensabile calcolare la probabilità connessa e tenerne conto nella pesatura delle risposte. In base a queste considerazioni si può ricorrere all'uso di formule.

Una formula per il calcolo della penalizzazione dell'errore è la seguente:

<p>Punteggio con penalizzazione delle risposte sbagliate</p> $P = \left(E - \frac{S}{n - 1} \right) K$ <p>P= punteggio con penalizzazione dell'errore E=numero di risposte esatte dello studente S=numero delle risposte sbagliate N=numero delle possibilità di risposta (a scelta multipla=4/5, vero/falso=2, etc.) K= peso attribuito a ciascun tipo di item</p>
--

La penalizzazione degli errori scoraggia le risposte date a caso, ma talune ricerche sperimentali dimostrano che il comportamento degli studenti, informati della penalizzazione degli errori, non cambia in maniera statisticamente significativa rispetto a quello tenuto nel caso in cui sia invece comunicata loro la decisione opposta (non penalizzazione degli errori).

Un esempio di penalizzazione degli errori in test pubblici è fornita nei criteri adottati dall'Università di Roma, La Sapienza e di Tor Vergata, per l'accesso a diverse lauree di tipo professionale (Medicina e Chirurgia, Odontoiatria e Protesi dentaria, Professioni sanitarie). Nella valutazione delle prove, composte da quesiti a scelta multipla con una sola risposta corretta, ciascuna commissione giudicatrice si attiene ai seguenti criteri: a) 1 punto per ogni risposta esatta; -0.2 punti per ogni risposta sbagliata; 0 punti per ogni risposta non data.

Diversa invece la questione dei pesi da attribuire ai diversi tipi di quesito. Da un lato abbiamo le diverse forme, dall'altro il tipo di obiettivo a cui mirano. Per attribuire pesi relativi dovremmo considerare il grado di difficoltà di soluzione del quesito. E al tal fine potremmo adottare due metodi complementari:

- attribuire pesi in funzione della tassonomia degli obiettivi: al crescere del peso che l'obiettivo ha all'interno di una possibile gerarchia (cfr. Bloom) possiamo attribuire un peso proporzionato;
- attribuire pesi in funzione del numero di alternative possibili che offre il quesito: crescendo il loro numero aumenta il peso che si dovrebbe attribuire alle risposte corrette (maggiore è il numero delle alternative di risposta minore è la probabilità di dare la risposta corretta rispondendo a caso).

Per l'attribuzione dei punteggi in base alla forma del quesito si faccia riferimento alla tabella n.13.

Tipo di quesiti	Punteggio da attribuire
Vero-falso	1
Scelta Multipla	
○ ad una risposta corretta con 4 alternative	3
○ ad una risposta corretta con 5 alternative	4
○ a due risposte corrette con 4 alternative	2 (due corrette) 1 (una corretta)
○ a due risposte corrette con 5 alternative	3 (due corrette) 1 (una corretta)
○ riconoscimento esattezza ^(a)	3 (A), 1 (B), 0 (C e D), -2 (E)
○ alternativa Migliore/Peggiora ^(b)	5 (MP), 2 (solo una), -2 (P per M), -5 (PM)
○ graduatoria alternative ^(c)	da 6 (ABCD) a -6 (DCBA)
Completamento (per ogni inserimento)	
- senza alternative	2
- con alternative	1
Cloze-test (per ogni buco)	1
Corrispondenza (per ogni corrispondenza)	1

Tabella n.13: tipologia di quesiti oggettivi e punteggi

Legenda

- ^(a) v. esempio di quesito riportato nell'Unità 7: il riconoscimento dell'esattezza di ciascuna delle affermazioni d'una coppia di affermazioni date e del legame tra loro (A) porta ad attribuire 3 punti; se lo studente indica solo l'esattezza delle due affermazioni ma non il loro legame (B) si attribuisce 1 punto; se indica l'esattezza di una delle due affermazioni (C o D) si attribuiscono 0 punti perché compie un'individuazione esatta e una errata; se infine indica che entrambe sono errate (E) compie un doppio errore e allora si penalizza con -2 punti.
- ^(b) v. esempio di quesito riportato nell'Unità 7: se lo studente indica correttamente la Migliore e la Peggiora (MP) avrà 5 punti; se ne individua solo una 2 punti; se ne sbaglia una avrà -2 punti; se inverte la Migliore con la Peggiora (PM) avrà -5 punti.
- ^(c) v. esempio di quesito riportato nell'Unità 7: tra la risposta esatta alla quale si attribuiscono 6 punti, cioè l'individuazione della graduatoria corretta (ad esempio ABCD), e quella completamente sbagliata che riceve -6 punti, nel caso di una graduatoria rovesciata (ad esempio DCBA), ci sono diverse soluzioni intermedie con l'attribuzione di penalità per i singoli errori e considerando anche la possibilità di astensioni (cfr. tab. E in Gattullo, 1968).

In alcune prove, ad esempio tra quelle utilizzate nel progetto PISA (*Programme for International Students' Assessment*), sono stati previsti diversi tipi di quesiti, a risposta preformulata o a risposta aperta:

Tra i quesiti a risposta preformulata ce ne sono di diversi tipi:

- *a scelta multipla* con quattro o cinque alternative di risposta di cui una sola corretta;

- o *plurime a scelta multipla* con diverse scelte di tipo binario alle quali si risponde cerchiando una delle due scelte (si/no; una delle due parole/frasi offerte).

Anche tra i quesiti a risposta aperta ce ne sono di diversi tipi:

- o *a risposta univoca* che richiedono la formulazione di una breve risposta che può rientrare in una gamma di possibilità molto ristretta o addirittura essere obbligata;
- o *a risposta articolata (breve o argomentata)* che richiedono la formulazione di una risposta, breve ma che può essere personalizzata; o che permettono di spaziare tra molte possibilità e possono implicare anche l'espressione di un giudizio personale.

Di seguito riportiamo alcuni esempi di quesiti, con le specifiche modalità per l'attribuzione di punteggio, tratti da diverse prove utilizzate nell'ambito dell'Indagine Pisa.

Il lavoro

Il seguente diagramma ad albero mostra la struttura della forza di lavoro di un paese ossia della "popolazione attiva". Nel 1995, la popolazione totale del paese era di circa 3,4 milioni di abitanti.

Struttura della forza di lavoro al 31 marzo 1995 (x1.000s)¹

```

graph TD
    A[Popolazione attiva 2  
2656.5] --> B[Forza di lavoro  
1706.5 64.2%]
    A --> C[Non forza di lavoro 3  
949.9 35.8%]
    B --> D[Occupati  
1578.4 92.5%]
    B --> E[Disoccupati  
128.1 7.5%]
    D --> F[A tempo pieno  
1237.1 78.4%]
    D --> G[A tempo parziale  
341.3 21.6%]
    E --> H[In cerca di lavoro a tempo pieno  
101.6 79.3%]
    E --> I[In cerca di lavoro a tempo parziale  
26.5 20.7%]
    F --> J[In cerca di lavoro a tempo pieno  
23.2 6.8%]
    F --> K[Non in cerca di lavoro a tempo pieno  
318.1 93.2%]
    
```

1. Numero di persone è espresso in migliaia (x 1.000)
 2. La "Popolazione attiva" è composta dalle persone tra i 15 e i 65 anni.
 3. La "Non forza di lavoro" comprende le persone che non cercano un lavoro e/o che non sono in grado di lavorare

Quesito (a risposta univoca: con a punteggio pieno e parziale)

Quante persone della popolazione attiva non facevano parte della forza *force*? (Scrivi il **numero** delle persone, non la percentuale)

Punteggio pieno = Punteggio 2
 Indica che le informazioni relative al numero nel diagramma ad albero e quelle relative alle migliaia nel titolo/nota a piè pagina sono state integrate: 949.000. Accettare le approssimazioni tra 9490.000 e 950.000 espresse in parole o numeri. Accettare anche 900.000 o un milione (in parole o numeri) con qualifica oppure: 949.000; quasi un milione; poco meno di; novecentocin-quantamila; 950.000; 949,9 migliaia; circa 900 mila; 949,9x1000; 949.900; 949(000)

Punteggio parziale = Punteggio 1
 Indica che ha individuato il numero nel diagramma ad albero, ma non ha integrato

l'informazione riguardante le migliaia nel titolo/nota a piè di pagina. Risponde 949.9 0 una cifra approssimata. La risposta può essere espressa in parole o numeri. Accettare le approssimazioni paragonabili a quelle consentite con il punteggio 2: 949,9; 94.900; circa mille; poco meno di 950; circa 900; poco meno di 1000

Nessun punteggio

altre risposte, ad esempio: 35.8% o 7.50%

Quesito (plurimo a scelta multipla)

Supponi che le informazioni sulla forza di lavoro siano presentate ogni anno in un diagramma ad albero come questo.

Di seguito sono elencati quattro elementi del diagramma ad albero. Indica se ci può aspettare che questi elementi cambino da un anno all'altro, tracciando un cerchio intorno a "Cambia" o "Non Cambia". La prima risposta è già fornita come esempio.

Caratteristiche del diagramma ad albero

Risposta

Le etichette in ciascuna casella (es. "Forza di lavoro")

Cambia / Non cambia

Le percentuali (es. "64,2%")

Cambia / Non cambia

Le cifre (es. "2.656,5")

Cambia / Non cambia

Le note in fondo al diagramma ad albero

Cambia / Non cambia

Figura n. 1: Alcuni esempi di quesiti e attribuzione di punteggio (fonte OCDE-PISA, 2001, cfr. Nardi, E. (2002), *Come leggono i quindicenni. Riflessioni sulla ricerca OCSE-PISA*, Franco Angeli, Milano)

Per la correzione dei quesiti a risposta preformulata non si richiede l'intervento di correttori, mentre per quelli a risposta aperta si hanno diversi gradi di complessità. I quesiti a risposta breve prevedono, infatti, sia la correzione dicotomica (corretta/errata) sia una più articolata, con la distinzione tra un punteggio pieno ed uno parziale. I quesiti a risposta argomentata sono invece assimilabili alle forme di produzione scritta vera e propria, e quindi richiedono descrittori o criteri per l'attribuzione dei punteggi e procedure di controllo dell'affidabilità tra correttori. Il diverso punteggio di risposta è dato dal grado di precisione che gli studenti possono presentare. Invece della classica correzione a due valori, corretta/errata (1/0), si prevedeva un massimo di 2 punti per la risposta corretta più esaustiva o completa, 1 punto per quelle risposte corrette in parte, 0 punti per quelle errate. Per individuare e correggere in forma affidabile il grado di esaustività e correttezza

delle risposte è stato necessario, nell'indagine Pisa, addestrare adeguatamente un gruppo di correttori.

Abbiamo detto che i singoli quesiti possono essere pesati in base al loro formato, come abbiamo appena discusso, oppure in base all'importanza che essi hanno nella didattica. In questo secondo caso si può fare riferimento a tassonomie come quella di Bloom, che però rischiano di essere astratte se non ben utilizzate nella fase di costruzione della prova, oppure si possono considerare la complessità e la possibile difficoltà del quesito. Per risolvere tale fase di "pesatura per importanza" possiamo procedere in due modi:

1. *a priori*, assegnando un coefficiente di difficoltà al singolo quesito o a sezioni della prova, in base ai principi della disciplina/didattica specifica;
2. *a posteriori*, utilizzando alcuni indicatori statistici quali la difficoltà media del quesito o indicatori ricavati da analisi degli item non tradizionali.

Si prenda ad esempio lo schema della prova di biologia riportato nella fig. 2. Assegnando un coefficiente (peso) diverso ai diversi obiettivi, il massimo punteggio teorico raggiungibile sarebbe ovviamente differente e di conseguenza il punteggio dei singoli studenti.

Contenuti/area di programma	Ob.1	Ob.2	Ob.3	Punteggio teorico massimo
Biologia 1	2		1	3
Biologia 2	1	1	1	3
Biologia 3	1	2	2	5
Biologia 4		2	3	5
Totale	4	5	7	16
Peso uguale ogni quesito=1	4	5	7	16
Peso diseguale Ob1=2 Ob2=1 Ob3=3	8	5	21	34

Figura n.2 – Schema per una prova di biologia

In conclusione, il punteggio che ricaviamo dalla fase di correzione, con o senza calcolo di penalizzazione, e con l'eventuale pesatura dei quesiti, si chiama *punteggio grezzo*. Esso è la base quantitativa che permette di analizzare l'andamento degli apprendimenti attraverso determinati confronti: dello studente nel tempo (confronto diacronico tra diverse prove), tra gli studenti contestualmente (confronto sincronico di classe/gruppo), tra gruppi di studenti nel tempo (confronto diacronico di classe/gruppo). Per ricavare queste informazioni abbiamo però bisogno di trasformare il punteggio grezzo in punteggi che permettano confronti statistici.

2. Analisi delle distribuzioni dei risultati ed elaborazione di indici sintetici

Una volta calcolato il punteggio grezzo il valutatore, il docente, può calcolare alcuni indici statistici per decidere come interpretare i risultati e assegnare giudizi o voti. A volte, quando la prova non prevede finalità di comparazione o di selezione, ci limitiamo a riportare il punteggio grezzo, e quindi non forniamo un giudizio valutativo. Se invece dobbiamo procedere alla formulazione di un qualche giudizio valutativo possiamo procedere nei seguenti modi:

- a) calcoliamo le fasce valutative. I singoli punteggi sono così riportati in terzili (fascia bassa/media/alta), quartili (bassa/medio-bassa/medio-alta/alta) e così via.
- b) calcoliamo la media e la deviazione standard della distribuzione dei punteggi e in base a questi indicatori costruiamo una distribuzione pentenaria (5 fasce equivalenti);
- c) normalizziamo i punteggi, standardizziamo rispetto a indici statistici o li riportiamo proporzionalmente sulla scala in decimi.

Analizziamo di seguito solo alcune delle procedure elencate.

2.1. Costruzione di una distribuzione pentenaria

La *distribuzione pentenaria* è quella più conosciuta nei contesti di istruzione e formazione¹⁰. In pratica utilizzando intervalli di punteggio pari ad una deviazione standard (1σ) si distribuiscono i diversi punteggi rilevati con una prova strutturata in 5 fasce di livello (E, D, C, B, A). La fascia centrale, contrassegnata con la lettera C, è quella che raccoglie i punteggi intorno alla media. Le due fasce vicino alla media, quella medio-alta, contrassegnata dalla lettera B, e quella medio-bassa, contrassegnata dalla lettera D, raccolgono i punteggi che distano $1,5\sigma$ dalla media, in alto o in basso. Così le due fasce più lontane dalla media, la coda alta con i punteggi migliori, contrassegnata con la lettera A, e la coda bassa con i punteggi peggiori, contrassegnata con la lettera E, raccolgono i punteggi che distano più di $1,5\sigma$ dalla media.

La distribuzione pentenaria permette di ripartire i punteggi ad una prova di classe e di compararli rispetto ad una distribuzione teorica normale quale quella rappresentata dalla curva di Gauss. Nella figura n. 3 si riporta lo schema per raggruppare i punteggi in fasce pentenarie e la curva normale con i punteggi percentuali teorici, come cioè si presenterebbe se sottoponessimo a misura un

¹⁰ Un altro metodo per riportare i risultati di diversi tipi di prove ad un punteggio standard è il sistema *stanine* (*standard nine*), molto popolare nel mondo anglosassone. Esso consiste nel riportare i punteggi grezzi su una scala a 9 valori. La stanine più alta è 9, la più bassa è 1, e la stanine 5 è il punto centrale della distribuzione. Ogni stanine, ad eccezione di 1 e 9 include una fascia di punteggi pari a metà deviazione standard (cfr. Gronlund, 1988; Cunningham, 1998).

fenomeno casuale.

Fasce

Intervalli di punteggio teorici

Intervalli calcolati

$(x_m = 31) (\sigma=7,32)$

Punteggi da esempio

E

inferiori a $x_m - 1,5\sigma$

inferiori a $31-7,3=23,7$

4 studenti: 17,21,22,23

D

compresi tra $x_m - 1,5\sigma$ e $x_m - 0,5\sigma$

compresi tra $31-7,3=23,7$ e $31-3,6=27,4$

Nessuno studente

C

compresi tra $x_m - 0,5\sigma$ e $x_m + 0,5\sigma$

compresi tra $31-3,6=27,4$ e $31+3,6=34,6$

4 studenti:

28,28,28,33

B

compresi tra $x_m + 1,5\sigma$ e $x_m + 0,5\sigma$

compresi tra $31+3,6=34,6$ e $31+7,3=38,3$

5 studenti:

35,35,35,35,36

A

superiori a $x_m + 1,5\sigma$

superiori a $31+7,3=38,3$

3 studenti:

40,40,40

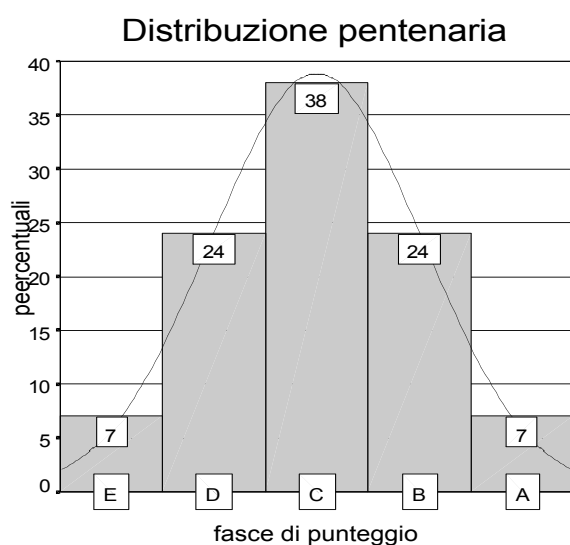


Figura n. 3: Calcolo delle fasce pentenarie e rappresentazione della distribuzione “normale”

Il confronto delle percentuali dei punteggi di una classe con quelle indicate dalla distribuzione normale ci dice se la situazione di apprendimento è casuale o se si discosta da essa e in quale direzione. All’inizio di una procedura formativa è abbastanza scontato che la distribuzione dei punteggi possa assumere una forma “normale”, a indicare la presenza di disomogeneità tra gli studenti. Ben diverso sarebbe se la stessa distribuzione si riscontrasse al termine di una procedura formativa, quando vorremmo che tutti, o almeno la gran parte, avesse raggiunto gli apprendimenti prefissati. Dovremmo trovarci, come esito finale di una efficace didattica, una distribuzione di punteggi che veda la gran parte degli allievi raggiungere punteggi positivi e una quota decisamente piccola con punteggi bassi.

Una distribuzione pentenaria permette in ogni caso di attribuire votazioni su una scala a cinque valori. Se non vogliamo adoperare punti decimali o notazioni troppo dissimili da quelle già adoperate per la comunicazione dell’andamento degli apprendimenti possiamo riportare le 5 voci della distribuzione alla scala di aggettivi o ai voti, come da schema.

Fasce pentenarie	Scala aggettivi	voti
A	Ottimo	8
B	Buono	7
C	Sufficiente	6
D	Quasi sufficiente	5
E	Non sufficiente	4

Abbiamo detto che la possibilità di confrontare i risultati ottenuti dagli studenti di una classe con quelli di una distribuzione ottenuta da un campione rappresentativo non rientra nella quotidiana pratica scolastica. Ma quando si usano strumenti standardizzati a livello nazionale, o internazionale, questo tipo di confronto diventa ricco di significato, in quanto offre la possibilità di parametrarsi ad uno standard. È questo il caso della prove I.E.A. (International Association for the Evaluation of Educational Achievement) costruite in questi ultimi quaranta anni¹¹. Nella figura n. 4 si riportano a titolo di esempio le fasce di livello rilevate su un campione di circa 2000 studenti di IV elementare ad una prova strutturata di alfabetizzazione linguistica, somministrata nei primi anni '90 (cfr. Corda Costa, M, Visalberghi, A., (a cura di) (1995), *Misurare e valutare le competenze linguistiche. Guida scientifico-pratica per gli insegnanti*, La Nuova Italia, Firenze)..

Area geografica	media	dev. std.	n. stud.
Intero campione	27,8	7,83	2119
Nord Est	28,8	7,28	551
Nord Ovest	27,8	6,79	181
Centro	27,3	8,12	429
Sud	26,8	8,17	452
Sud Isole	27,9	8,09	506

Fasce	Scala	Punteggi	%	% cumulata
A	Ottimi	da 40 a 42	3,2	100,0
B	Buoni	Da 33 a 39	30,0	96,8
C	Sufficienti	Da 24 a 32	37,4	66,8
D	Insufficienti	Da 16 a 23	21,3	29,4
E	Gravemente Insufficienti	Da 0 a 15	8,1	8,1

Figura n. 4: Indagine IEA-SAL: Scuola elementare, punteggi medi per area geografica e distribuzione pentenaria dei punteggi (fonte: Corda Costa, Visalberghi, op. cit, 1995).

2.2. Standardizzazione dei punteggi

In alcuni casi abbiamo bisogno di confrontare il punteggio di uno studente in una prova con quelli riportati dallo stesso studente in altre prove. Il confronto non può avvenire utilizzando i punteggi

¹¹ Per una rassegna bibliografica e una dettagliata presentazione delle ricerche condotte fino ai primi anni '90 dall'I.E.A. (*International Association for the Evaluation of Educational Achievement*), si veda Lucisano, Siniscalco (1994); per un aggiornamento si consultino gli Annuari del CEDE (2001a, 2001b).

grezzi, in quanto risentono del tipo di prova o di verifica utilizzata nei diversi casi oggetto di confronto. Una prova potrebbe avere il doppio dei quesiti rispetto ad un'altra, oppure pur in presenza dello stesso numero di quesiti questi potrebbero avere pesi diversi, oppure ancora le prove potrebbero essere diverse sia per quanto riguarda la numerosità dei quesiti sia per quanto riguarda il loro peso. Infine, potremmo avere a che fare con prove miste, composte da quesiti strutturati e forme più aperte con correzione attraverso criteri.

In altri casi, invece, potremmo avere bisogno di rapportare il punteggio degli studenti con una possibile "norma", cioè con risultati di una popolazione di riferimento, ad esempio studenti della stessa classe di corso su scala nazionale o appartenenti ad un campione rappresentativo. Il riferimento a "norme" per un test implica l'assunto che si conosca la distribuzione dei punteggi a quel test su di un campione "normale".

Per risolvere le due richieste, confronto con punteggi precedenti e con punteggi riportati da gruppi rappresentativi, abbiamo bisogno dei punteggi standardizzati, cioè della trasformazione dei punteggi grezzi in punti che esprimano la distanza del singolo punteggio dalla media. Per il calcolo è necessario conoscere la media e la deviazione standard della distribuzione dei punteggi.

I punti standardizzati più conosciuti sono i punti Z e i punti T. Le due trasformazioni differiscono solo nel fatto che la seconda, quella dei punti T, per evitare la presenza di punteggi negativi, aggiunge una costante a tutti i punteggi e quindi esprime tutti valori positivi.

Punti Z standardizzati	Punti T standardizzati
$Z_i = \frac{x_i - x_m}{\sigma}$	$T_i = 50 + 10Z_i$

3. Scelta di un criterio valutativo

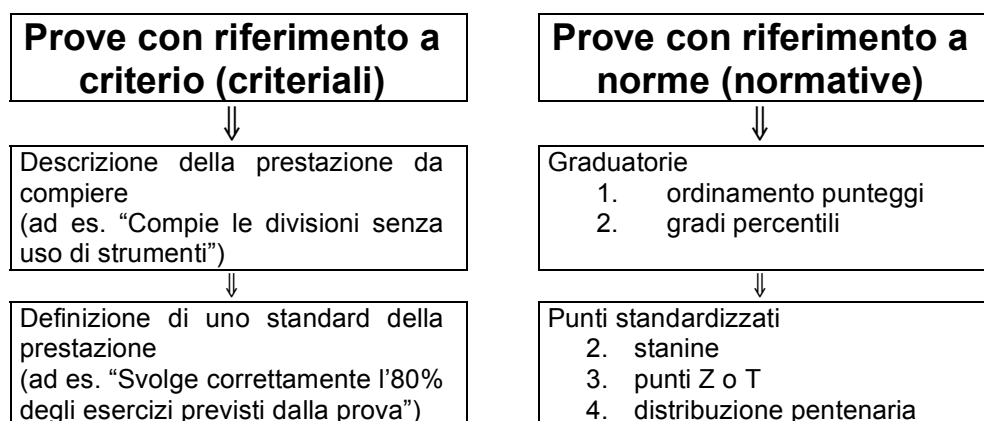
Nella formulazione di un giudizio si possono utilizzare diverse modalità o modelli comparativi. Quelli prevalenti sono i seguenti¹²:

1. confronto tra una prestazione ed un'altra, sempre dello stesso allievo, e successiva nel tempo (giudizio di progresso o profitto).
2. confronto tra le prestazioni di un allievo e gli obiettivi/livelli attesi (giudizio criteriale).

¹² Molti autori hanno affrontato la questione relativa ai diversi livelli o forme del giudizio (cfr. Domenici, 1993 e Gattullo, 1968).

3. confronto tra le prestazioni di un allievo e quelle registrate da altri allievi (campioni/gruppi rappresentativi).

La scelta di uno dei modelli comparativi dipende dai destinatari della nostra prova e della sua funzione valutativa. I diversi approcci per interpretare i risultati sono quindi l'ago della bilancia. In generale i risultati a prove strutturate possono essere interpretati rispetto a criteri o a norme. Nel primo caso parliamo di riferimenti assoluti, in quanto definiamo la soglia che delimita il raggiungimento o meno di una prestazione (*cut-off score*), nel secondo caso il riferimento è relativo alla distribuzione dei risultati. Nello schema seguente si sintetizzano i diversi tipi di metodi che possono essere utilizzati nell'interpretazione dei risultati.



Per interpretare in modo normativo utilizziamo i diversi criteri elencati precedentemente. Quando invece utilizziamo un sistema criterioale vogliamo distinguere chi ha raggiunto un determinato livello, definito esplicitamente e chiaramente. Questo criterio di riferimento si assume come standard o livello soglia di riferimento (*cut-off score*¹³) e potrà fare riferimento alla:

1. *velocità di esecuzione*: nel caso in cui si possa risolvere un numero variabile di compiti (ad es. "rispondere a 5 quesiti in 3 minuti" ecc.)
2. *precisione nella prestazione*: quando le prove prevedono diverse modalità nella soluzione (ad es. "risolvere il problema con un determinato numero di passaggi")
3. *percentuale di quesiti risolti correttamente*: è il criterio più adottato (ad. es "rispondere a non meno dell'80% dei quesiti").

A differenza dell'ampia letteratura sull'uso e l'interpretazione dei risultati a prove normative, per la definizione di standard e prove criteriali la letteratura di natura operativa rimanda all'esperienza sul campo. Gronlund (*How to construct achievement tests*, Prentice-Hall, New Jersey, 1988, 4^a) indica come semplice e pratica procedura quella di "decidere arbitrariamente gli standard e di aggiustarli

¹³ Le dizioni per definire il livello soglia sono molteplici, tra esse: livello criteriale (*criterion level*), punteggio soglia (*cutting score*), passaggio/bocciatura (*pass-fail*), di passaggio (*passing standard*).

in alto o in basso a seconda del contesto”. In ogni caso, offre un esempio utile alla comprensione di questo tipo di prove (vedi figura n. 5), e indica alcuni elementi da tenere in considerazione: 1) definire per le scelte multiple un livello dell’85% di risposte corrette; 2) aumentarlo se il raggiungimento del livello è necessario per proseguire le attività; 3) aumentarlo se il raggiungimento del livello è necessario per la sicurezza (ad es., nei miscugli chimici); 4) aumentarlo se il test o le sue parti sono brevi; 5) abbassarlo se ne è prevista una ripetizione successivamente; 6) abbassarlo se le prove previste non sono particolarmente rilevanti; 7) abbassarlo se le prove previste sono particolarmente difficili; 8) aggiustarlo verso l’alto o il basso secondo l’esperienza di insegnamento.

Prova di	Verifica scolastica	Studente :	Mario Rossi
Padronanza (<i>Mastery</i>)	85% corrette		
Obiettivi	Numero risposte corrette	% corrette	Padronanza (X)
1. Conoscenza dei termini (20)*	18	90	X
2. Conoscenza procedure (20)	18	90	X
3. Comprensione principi (20)	17	85	X
4. Applicazione principi (20)	14	60	
5. Interpretazione dati (20)	17	85	X

Figura n. 5: Profilo di una prova criteriale (fonte: Gronlund, op. cit., 1988)

(* Numero di quesiti per obiettivo)