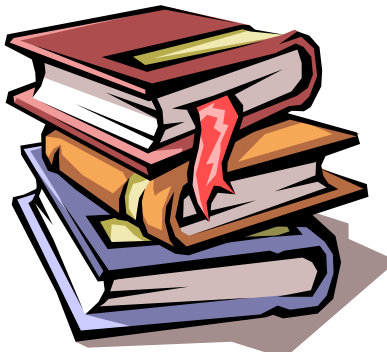


UNIVERSITA' DEGLI STUDI DI NAPOLI FEDERICO II

SCUOLA INTERUNIVERSITARIA CAMPANA
DI SPECIALIZZAZIONE ALL'INSEGNAMENTO
A.N.I.S.N. – ASSOCIAZIONE NAZIONALE INSEGNANTI
SCIENZE NATURALI – SEZIONE CAMPANIA



Schemi per la programmazione
II parte – La valutazione



a cura di

Vincenzo Boccardi

supervisore del tirocinio per la classe

A060 – scienze naturali, chimica e geografia, microbiologia

Il materiale che segue è stato rielaborato in forma personale con riferimento principalmente al testo:
Maurizio Tiriticco, Il concorso a cattedre, Edizioni tecnodid, Napoli, 1998

La valutazione

*La valutazione è un processo che deve esercitarsi costantemente in tutte le azioni educative. La **valutazione degli allievi** costituisce pertanto solo un aspetto della valutazione.*

La valutazione delle prestazioni degli alunni comprende:

- a) La **valutazione iniziale** (*livelli di partenza*): serve all'insegnante per conoscere quali strategie dovrà adottare per la sua azione educativa e didattica.
- b) La **valutazione in itinere** (*valutazione formativa*): serve soprattutto all'alunno ed allo stesso insegnante per correggere il suo percorso di formazione.
- c) La **valutazione finale** (*valutazione sommativa, ma e meglio definirla conclusiva, per evitare ogni equivoco con una semplice somma e media dei voti*): serve all'alunno e alla famiglia e alla società in genere come certificazione delle abilità conseguite.

Misurazione e valutazione di una prova.

Misurazione: è l'**esito oggettivo** della prova stessa (ad esempio sette domande esatte su dodici).

Valutazione: è il **giudizio** che viene formulato per quella prova e si fonda su altre variabili non inerenti alla prova, come le condizioni in cui la prova si è effettuata, il peso che hanno i livelli di partenza, la valenza dell'obiettivo, la difficoltà o la facilità della prova, ecc.

La misurazione può essere bene espressa con un *punteggio*, la valutazione con un *giudizio verbale*.

La **misurazione** costituisce un primo gradino della valutazione. Un elemento importante della misurazione è il **limite di accettabilità**, cioè quel livello individuato a priori al di sotto del quale la prova non si può considerare superata (*soglia di sufficienza*). Esso viene scelto sulla base di una prevalutazione effettuata da chi predispone la prova o di esiti precedenti standardizzati.

La **valutazione** si ha quando a un determinato risultato si attribuisce un giudizio che viene formulato ricorrendo a criteri che vanno al di là sia della prova stessa che della misurazione effettuata. Quando si valuta si adottano pertanto criteri che non sono quelli della misurazione: l'esito della misurazione ci è dato dalla prova stessa, ma la valutazione viene espressa dall'esterno, dal valutatore.

Un esempio: un appartamento di 40 metri quadrati (unica misurazione) può andar bene per un single, ma non per una famiglia numerosa (diversa valutazione). Allo stesso modo una prova con un punteggio basso (misurazione) rispetto al limite di accettabilità può essere giudicata positivamente se denota un miglioramento, negativamente nel caso contrario

Input e output

Ogni verifica implica un **input** (domanda, sollecitazione, richiesta di prestazione) e un **output** (risposta, elaborazione, prestazione).

In base alla loro natura possono aversi i seguenti 4 casi:

STIMOLO APERTO – RISPOSTA APERTA

Es.: Parla dell'evoluzione

STIMOLO CHIUSO – RISPOSTA APERTA

Es.: In quali parti di una cellula sono presenti membrane?

STIMOLO CHIUSO – RISPOSTA CHIUSA

Es.: In quale parte della cellula si svolge il ciclo di Krebs?

STIMOLO APERTO – RISPOSTA CHIUSA

Es.: Come si chiama quell'organulo cellulare circondato da una doppia membrana sede del processo di respirazione cellulare che svolge una funzione essenziale nel rifornire la cellula di ATP e che molto probabilmente costituisce un esempio di endosimbiosi stabilitasi con una grossa cellula a metabolismo fermentativo all'incirca un miliardo di anni fa?

(ridondanza negativa: in pratica la risposta è pre-confezionata dal docente)

Prove non strutturate, strutturate, semistrutturate

a) Le prove non strutturate (tradizionali)

Sono caratterizzate da

- soggettività del docente
- casualità ed episodicità (l' allievo non sa quando la prova sarà somministrata)
- assenza di criteri per la formulazione, la somministrazione, la misurazione e la valutazione della prova

Sono soggette ai seguenti rischi:

- effetto alone
- effetto pigmalione
- stereotipia
- contrasto
- alla stessa prova voti diversificati a secondo del docente

b) Le prove strutturate (oggettive¹)

Non prestano il fianco ad alcuna ambiguità per quanto riguarda la formulazione della risposta e prevedono pertanto una sola risposta esatta (a volte più di una)

Anche se non può esistere una oggettività assoluta, hanno i seguenti vantaggi:

- a tutti gli allievi viene sottoposta la stessa prova
- in sede di correzione tutti gli allievi sono trattati allo stesso modo

¹ In realtà una valutazione completamente oggettiva non esiste né è possibile. Dovrebbero infatti essere oggettivi i fondamenti disciplinari, le scelte didattiche, l'articolazione dei tempi, gli strumenti utilizzati, gli argomenti prescelti, i criteri di valutazione, gli indicatori di livello. Sarebbe pertanto più adeguato parlare di "valutazione scientifica" in quanto possiede i caratteri della comunicabilità, ripetibilità, trasferibilità, confrontabilità.

Tipologie delle prove strutturate

Il test vero/falso (punti 1)

Esempi: i mitocondri contengono clorofilla (Falsa)

Le placche o zolle sono blocchi di litosfera (Vera)

Il ciclo di Krebs si svolge all'interno dello stroma dei mitocondri (Falsa, il compartimento mitocondriale interno costituisce la matrice)

L'energia di attivazione è la differenza tra l'energia del complesso attivato, un intermedio tra reagenti e prodotti molto ricco di energia, e quella dei reagenti (Vera)

Il test a risposta multipla

E' costituito da un'asserzione che viene completata e conclusa da più alternative (generalmente 4 o 5), delle quali una sola è corretta.

Dall'ossidazione di una molecola di $\text{NADH} + \text{H}^+$ si ottiene

- a) una molecola di ATP
- b) due molecole di ATP
- c) tre molecole di ATP
- d) quattro molecole di ATP

Quale di queste molecole non fa parte dei filamenti sottili delle fibre muscolari?

- a) miosina
- b) actina
- c) tropomiosina
- d) troponina

Livello tassonomico di un quesito

I quesiti (item) delle prove strutturate possono essere catalogati in base alla tassonomia di Bloom.

Esempi²

CONOSCENZA (punti 2)

Il giorno solare è più lungo del giorno sidereo di

- a) circa 2 minuti
- b) circa 4 minuti
- c) circa 6 minuti
- d) circa 8 minuti

COMPRESIONE (punti 3)

Per quale motivo esiste questa differenza di durata tra essi

- a) perché la velocità di rivoluzione della Terra è maggiore al perielio e minore all'afelio
- b) perché la distanza della Terra dal Sole cambia durante l'anno
- c) per la combinazione del moto di rotazione con quello di rivoluzione
- d) per l'inclinazione dell'asse terrestre rispetto all'eclittica

APPLICAZIONE (punti 4)

L'asse terrestre è inclinato di circa 66 gradi e mezzo rispetto al piano dell'orbita (eclittica). Cosa avverrebbe se l'asse terrestre fosse perpendicolare all'eclittica?

- a) le stagioni rimarrebbero inalterate
- b) si avrebbero estati più calde ed inverni più freddi
- c) le stagioni non esisterebbero
- d) si avrebbero estati più lunghe ed inverni più corti

² I seguenti esempi sono tratti da: V. Boccardi – Una prova strutturata di verifica per l'insegnamento della geografia astronomica – Didattica delle Scienze, 185, 55-58, ottobre 1996

Il test di completamento (*doze*) (per ogni termine un punto)

E' un brano nel quale sono state cancellate alcune parole chiave che l'allievo deve ricercare in un elenco e collocarle al punto giusto. Nell'elenco deve esserci sempre almeno una parola in più rispetto a quelle da collocare per evitare che l'ultima venga inserita automaticamente.

Esempio

La fotosintesi si svolge nei ed ha come reagenti acqua e I suoi prodotti sono e zuccheri semplici (glucosio).

Anidride carbonica, cloroplasti, mitocondri, ossigeno

Il test di corrispondenza (per ogni abbinamento corretto 1 punto)

Consiste in due liste affiancate di nomi, concetti, frasi da mettere in corrispondenza (per ogni abbinamento 1 punto)

Cloruro di sodio	Legame covalente polare
Ossigeno molecolare	Legame metallico
Acido cloridrico	Legame ionico
Stagno	Legame covalente puro

c) Le prove semistrutturate

Sono prove che pur prevedendo una risposta esatta non propongono alternative da scegliere in quanto la risposta deve essere formulata dall'allievo.

Esempio

Definisci in non più di quattro righe quale significato assume il ciclo di Krebs nell'ambito del metabolismo cellulare.

.....
.....
.....
.....

In questo caso va preparata anche la cosiddetta “risposta campione”

Il significato del ciclo di Krebs e quello di collegare la glicolisi con la catena di trasporto degli elettroni, consentendo così la completa ossidazione del glucosio. Inoltre esso è un'importante fonte di precursori per la biosintesi di numerosi composti.

La terza prova dell'esame di stato³

Riguarda 5 discipline tra cui deve essere previsto l'accertamento della conoscenza della lingua straniera.

Sono previste le seguenti tipologie di prova:

Tipologia A – trattazione sintetica di argomenti: *una serie di domande specifiche relative a un testo o la trattazione di un argomento significativo, anche a carattere pluridisciplinare, con l'indicazione del numero massimo di righe o parole da utilizzare*
Tipicamente è costituita da 5 argomenti, uno per materia.

Esempio: Parla del moto di rotazione terrestre (max 15 righe).

Tipologia B – Quesiti a risposta singola: *riguardano più materie e richiedono che il candidato formuli autonomamente una risposta, rispettando i limiti di estensione indicati dalla commissione. I quesiti possono essere di difficoltà graduata, quindi con valore differente ai fini dell'attribuzione del punteggio.*

Tipicamente riguarda 5 materie con 2-3 quesiti per ciascuna (Totale 10 – 15 quesiti)

Esempio: Spiega la differenza tra giorno solare e giorno sidereo (max 5 righe).

Tipologia C – Quesiti a risposta multipla: *possono concretarsi in vere e proprie prove strutturate su argomenti di tutte le materie dell'ultimo anno di corso. Il candidato sceglie tra più opzioni la risposta giusta.*

Gli item devono essere da 30 a 40 (Tipicamente 8 per materia).

Esempio: Quale di queste condizioni è necessaria affinché si abbia un'eclissi di Sole?

- a) La Luna deve essere al perielio
- b) La Luna deve essere all'afelio
- c) La Luna deve essere in congiunzione
- d) La Luna deve essere in opposizione
- e) La Luna deve essere in un nodo

³ Legge 10 dicembre 1997, n. 425 e DPR 23 luglio 1998, n. 323 e succ. modif.

Tipologia Mista B + C. Per alcune materie (ad esempio italiano, inglese e storia) quesiti a risposta singola; per altre (fisica e geografia) quesiti a scelta multipla per ciascuna. Il numero dei quesiti a risposta singola non può essere inferiore a 8, quello a risposta multipla non può essere inferiore a 16.

Tipologia D – Problemi a soluzione rapida articolati in relazione allo specifico corso di studio (max 2)

Tipologia E - Casi pratici e professionali (istituti Tecnici e professionali) (Max 2)

Tipologia F – Sviluppo di progetti (istituti tecnici e professionali, istituti d'arte e licei artistici) (Un progetto)

REGOLE PER LA COSTRUZIONE DI DOMANDE A SCELTA MULTIPLA⁴

Il quesito deve:

- essere formulato in modo diretto ed esplicito, in modo tale da privilegiare frasi in forma positiva piuttosto che negativa e da evitare l'uso di doppie negazioni;
- essere facilmente leggibile e sintatticamente corretto;
- contenere tutte le informazioni necessarie per la soluzione
- evitare di fornire informazioni non strettamente pertinenti alla soluzione.

La chiave e i distrattori devono

- essere scritti in modo non ambiguo, in modo tale da contenere quanto possibile la loro lunghezza
- essere pressappoco equivalenti come numero di parole
- evitare di fornire appigli linguistici per la soluzione

I distrattori devono presentare alternative

- logicamente plausibili
- omogenee con la chiave dal punto di vista dei contenuti
- diverse tra loro e dalla chiave in modo esplicito e non cavilloso.

⁴ Materiale di lavoro Olimpiadi di Scienze ANISN

Come attribuire un punteggio ad un item

Si possono seguire due criteri: quello della probabilità di dare casualmente una risposta esatta (Tavola A) e quello della difficoltà (livello tassonomico) della domanda (Tavola B).

In genere

- Si premia la risposta corretta
- Si considera positivamente l'astensione o l'omissione
- Si penalizza la risposta sbagliata

Tavola A – Punteggio in base alla probabilità di dare casualmente la risposta esatta

Test	Risposta esatta	Astensione	Risposta sbagliata
Vero/Falso	+ 1	0	- 1
SM a 3 altern.	+ 2	0	- 1
SM a 4 altern.	+ 3	0	- 1
SM a 5 altern.	+ 4	0	- 1

Tavola B – Punteggio in base al livello tassonomico di un item

Test	Risposta esatta	Astensione	Risposta sbagliata
Vero/Falso	+ 1	0	- 1
SM Conoscenza	+ 2	0	- 1
SM Comprensione	+ 3	0	- 1
SM Applicazione	+ 4	0	- 1

Le elaborazioni sui punteggi: alcuni parametri docimologici

La **media**: è la somma di tutti i punteggi considerati divisi per il loro numero.

La **mediana**: è il punteggio che nella distribuzione dal più basso al più alto occupa il posto centrale

La **moda**: è il valore con la massima frequenza nella popolazione di punteggi considerati.

Punto massimo: è il punteggio più alto

Punto minimo: è il punteggio più basso

Range: è la differenza tra punto max e punto min

Se la media supera la mediana la prova ha creato difficoltà negli allievi

Se la mediana supera la media la prova è da considerarsi facile

Esempio

2 – 3 – 4 – 5 – 5 – 6 – 7 – 8 – 10

Media: $50/9 = 5,56$

Mediana 5

Moda 5

Max: 10

Min: 2

Range 8

La deviazione standard (sigma): misura la dispersione dei punteggi rispetto alla media; più la d.s. è alta, più è alta la dispersione. E' data dalla formula:

$$\sigma = \sqrt{\frac{\sum (\text{singolo punteggio} - \text{media})^2}{\text{numero delle prove}}}$$

Si considerino le seguenti due distribuzioni

$$d_{(1)} \quad 5 - 6 - 7$$

$$d_{(2)} \quad 2 - 6 - 10$$

In entrambi i casi la media è 6, ma nel primo caso $\sigma = 0.82$; nel secondo $\sigma = 3.27$

$$\sigma_1 = \text{Radice quadrata } (1^2+0+1^2)/3 = 0.82 \text{ (accettabile)}$$

$$\sigma_2 = \text{Radice quadrata } (4^2+0+4^2)/3 = 3.27 \text{ (non accettabile)}$$

In genere si considerano accettabili valori di σ tali che:

$$3 \sigma \leq \text{media}$$

La deviazione standard non deve pertanto superare la terza parte della media

Il punto Z: questo parametro non si riferisce all'intera popolazione dei valori ma a un singolo punteggio di un alunno. E' dato dal punteggio dell'alunno meno la media diviso σ .

Generalmente è compreso tra -3 e $+3$. Se è inferiore a -3 indica un caso particolarmente preoccupante.

In pratica è un indice per stimare la posizione in cui si colloca il punteggio di un alunno nell'ambito dei risultati dell'intera classe.

Può essere utile anche per attribuire dei voti: poiché Z varia di solito tra -3 e $+3$ aggiungendo 6 a ciascun punto Z è possibile ottenere voti che vanno dal 3 al 9 (valutazione formativa).

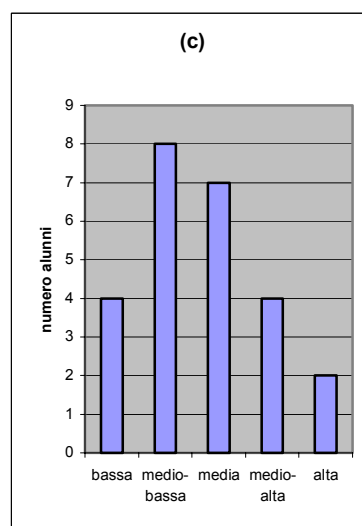
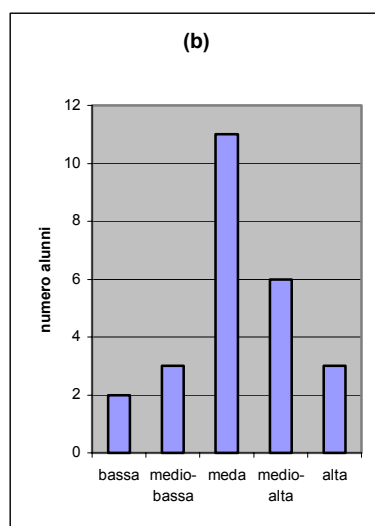
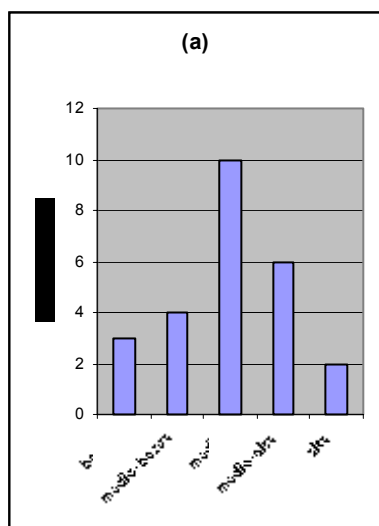
Ma in tal modo ogni studente viene valutato non in base ad una soglia prefissata, ma in relazione all'andamento complessivo della classe, sottostimando i voti se una verifica è andata particolarmente bene, e sovrastimandoli invece per una verifica andata male.

I punti Z costituiscono un modo per standardizzare i punteggi in modo da consentire comparazioni tra prove, alunni, classi diverse.

La distribuzione pentenaria: è molto utile in sede di valutazione formativa. Consiste nell'individuare nella distribuzione dei punteggi 5 fasce così ottenute

Alta:	$\text{punteggi} > \text{media} + 3/2 \sigma/2$
Medio-alta:	$\text{media} + \sigma/2 < \text{punteggi} < \text{media} + 3/2 \sigma/2$
Media:	$\text{media} - \sigma/2 < \text{punteggi} < \text{media} + \sigma/2$
Medio-bassa:	$\text{media} - \sigma/2 < \text{punteggi} < \text{media} - 3/2 \sigma/2$
Bassa:	$\text{punteggi} < \text{media} - 3/2 \sigma/2$

Generalmente si ottiene una curva campana, con la maggioranza dei punteggi nella classe media e due minoranze sui valori alti e bassi (a). Valori spostati sui punteggi alti (b) indicano una ottimizzazione dei risultati; valori spostati sui punteggi bassi un insuccesso della classe (c)



La valutazione dell'efficienza delle domande⁵

Indice di difficoltà (d): stima la difficoltà di un item.

E' dato dal rapporto tra le risposte esatte e il totale delle risposte. Si ottiene così un numero che varia tra 0 (massima difficoltà) e 1 (difficoltà nulla)

Valori minori di 0,2 - 0,3 indicano domande troppo difficili, valori maggiori di 0,7 - 0,8 domande troppo facili.

Indice di discriminazione (Δ): è la differenza tra il numero di risposte esatte ad un determinato quesito nel primo terzile (i migliori) e il numero di risposte esatte nel terzo terzile (i peggiori) diviso il numero di alunni che formano un terzile.

Per poterlo calcolare occorre ordinare i test di una classe dal migliore al peggiore in base ai punteggi complessivi. Si procede poi a dividere i test in tre gruppi e si può quindi comparare il terzo migliore della classe (primo terzile) con il terzo peggiore (terzo terzile).

$\Delta = 1$ significa che solo i migliori hanno risposto correttamente alla domanda

$\Delta = 0$ significa che la domanda in esame non ha fatto alcuna discriminazione tra i peggiori e i migliori

$\Delta = -1$ indica che solo i peggiori hanno risposto alla domanda (discriminazione inversa).

In genere si considerano validi valori di Δ compresi tra 0.3 e 0.7. Valori < 0.3 o addirittura negativi indicano che la domanda non rileva adeguatamente il livello di preparazione dello studente; valori > 0.7 indicano domande inadeguate al livello della classe.

⁵ Per semplificare i seguenti indici sono riferiti ad una prova a scelta multipla.

Esempio (ovviamente il valore di tali indici ha tanto più significato quanto più ampio è il campione).

Alunno	punteggio complessivo	item n. x	
1	45	si	primo terzile
2	43	si	
3	40	no	secondo terzile
4	37	si	
5	33	si	terzo terzile
6	27	no	

$$d = 4/6 = 0.33$$

$$\Delta = (2 - 1)/2 = 1/2 = 0.5$$

(entrambi gli indici sono accettabili)

Bibliografia

- Valutare Come – release di M. Tiriticco, 1996 – Corso per formatori IRRSAE Campania – Fnism A060
- V. Boccardi – Una prova strutturata di verifica per l'insegnamento della geografia astronomica – Didattica delle Scienze, 185, 55-58, ottobre 1996
- M. Tiriticco – Il concorso a cattedre – Tecnodid, 1998
- L. Cozzi – L'utilizzo di strumenti informatici per una valutazione oggettiva nelle scienze naturali – Quaderni Anisn Lombardia, 2002

Esercizi

A - Lavorando con 2-3 colleghi individua un argomento relativo ad un qualsiasi segmento curricolare.

Elabora 5 quesiti relativi ad esso che facciano riferimento a diversi livelli tassonomici.

Attribuisci ad essi punteggi diversi

Fissa un criterio in base al quale si possa considerare raggiunta o meno la competenza minima sull'argomento scelto.

B – Una prova di verifica sommativa costituita da 40 item a scelta multipla è stata somministrata ad un campione di 250 allievi. La seguente tabella mostra l'analisi degli item relativa a 5 quesiti.

Item	Chiave	A	B	C	D	Astensioni	d	Δ
6	C	10	39	120	64	17	0.52	0.08
12	D	20	15	98	80	37	0.68	0.47
18	A	230	8	12	-	-	0.08	0.11
25	B	53	150	13	27	17	0.40	0.37
31	C	31	115	40	26	38	0.84	0.54

Quali considerazioni si possono fare in merito alla qualità dei 5 quesiti analizzati?

3. Una prova di verifica formativa costituita da 10 item a scelta multipla è stata somministrata ad un gruppo di 16 allievi. Le risposte della classe sono riportate nella tabella

allievo		9	14	1	4	11	7	2	16	12	5	13	8	3	10	6	15		
punteggio		9	10	12	13	17	18	18	19	20	21	21	23	24	25	27	28		
Item	Chiave																	d	Δ
1	A	B	B	B					C	D		C			D				
2	C	D	D	D		A		B	A		B								
3	C			A		B	A						D						
4	A	B	B	B	B		B	B	B	B	B	B	B	B	B				
5	C	A		B	D		A	B		D	A	B		B		A			
6	B				C	A		A											
7	B	D	C		A	A													
8	D	C	C	B	A														
9	B															C	A		
10	D	B	C				B			A									

Quali considerazioni è possibile trarre dai dati già disponibili?

Completa la tabella calcolando gli indici d e Δ.

Sulla base dei loro valori, fornisci delle indicazioni sulla qualità dei singoli quesiti.