

Il flagello delle griglie analitiche¹

di Giuseppe Rinaldi

Ogni anno, in sede di Esame di Stato, esplose il tormentone delle griglie. Solerti ispettori, compiacenti presidenti e affannati commissari vanno alla ricerca delle griglie che verranno utilizzate per la correzione delle prove scritte e per la valutazione del colloquio. Non è neppure pensabile che la commissione si metta a costruire le sue griglie *ex novo* (tutti hanno fretta di finire e di andare in vacanza), anche se sarebbe logico, poiché ogni commissione dovrebbe avere una sua *politica della valutazione* e niente più dell'insieme delle varie griglie è capace di esprimere effettivamente una politica della valutazione. Le griglie necessarie sono molte, spesso le proposte sono in competizione le une con le altre. Si assiste talvolta a espressioni maniacali di fedeltà per talune griglie contro altre. Non di rado accade che ispettori o presidenti impongano alla commissione le loro griglie preferite. La battaglia delle griglie ricorda un po' quella tra i topi e le rane. Tuttavia spesso gli esiti, soprattutto per gli utenti, non sono soltanto comici.

In effetti sono ormai alcuni anni che nel nostro paese imperversano le griglie. Purtroppo bisogna dire che le varie componenti coinvolte, il corpo docente, il corpo ispettivo e dirigente, possiedono una consapevolezza piuttosto superficiale del meccanismo di funzionamento delle griglie di valutazione. Se le griglie permettono, da un lato, di conferire ai controlli scolastici una specie di legittimazione tecnico scientifica, dall'altro lato, il più delle volte, rischiano di produrre dei veri propri disastri. Una griglia mal fatta o mal utilizzata produce risultati del tutto fuorvianti. I risultati indesiderati suscitano risentimenti, polemiche, perdite di tempo, ricorsi, ecc... E' ovvio che uno scarso controllo sulle griglie di valutazione pone seriissimi problemi di equità, cioè di uguaglianza di trattamento dei candidati davanti all'esame. I risultati erratici e imprevedibili delle griglie sono ben noti al corpo docente. Onde evitare tutte le prevedibili e spiacevoli conseguenze che derivano dalla frettolosa adozione di griglie reperite all'ultimo momento è prassi abbastanza normale nelle commissioni decidere *prima* il punteggio che verrà attribuito al candidato e *poi* utilizzare la griglia per trovare una giustificazione a posteriori. La griglia rimarrà agli atti a testimoniare, se mai qualcuno vorrà aprire il pacco, che tutto il processo di valutazione è avvenuto in forma scientificamente inoppugnabile.

I maggiori danneggiati da questa situazione sono proprio gli utenti, almeno in tutti quei casi in cui i commissari e presidenti prendono *alla lettera* le loro griglie e producono così, senza troppa difficoltà, risultati decisamente assurdi, contro intuitivi e illogici. Se qualche candidato ha avuto all'esame di Stato dei risultati decisamente diversi dal suo normale rendimento scolastico, è probabile che debba questo suo infortunio, oltre a tantissimi altri fattori dovuti alla sua responsabilità, anche ad un uso improvvisato di qualche griglia da parte della commissione.

¹ Poiché mi pare che stiano cominciando a manifestarsi – tra gli insegnanti – reazioni di rigetto verso certi tecnicismi imposti dall'alto in forma becera e pappagallesca, e poiché ritengo che occorra evitare, con ciò, di “buttare il bimbo con l'acqua sporca”, cerco in questo scritto di distinguere accuratamente tra il buon uso e il cattivo uso degli strumenti di valutazione e in particolare delle cosiddette griglie. Chiunque voglia far pervenire all'autore critiche o suggerimenti, può farlo all'indirizzo E-mail: beppexxrinaldiyy@libero.it (antispam: togliere le x e y). Sul mio sito <http://digilander.libero.it/education> si trovano altri materiali concernenti i controlli scolastici, comprese alcune griglie sintetiche per l'Esame di Stato..

=ooOoo=

Eppure per riuscire a dominare le griglie, e a non farsi dominare, è sufficiente conoscere alcuni semplici principi fondamentali che afferiscono a discipline come la statistica, la psicologia e la metodologia della misurazione nell'ambito delle scienze dell'uomo. E più specificatamente, occorre anche conoscere la tecnica per la costruzione delle scale di valutazione (le griglie sono scale di valutazione). Si osserverà che queste conoscenze non spettano ai docenti. Oppure si osserverà che nella cultura pedagogica italiana – spesso ancora vetero gentiliana – tutte queste discipline non sono contemplate, non appartenendo alle scienze dello Spirito, bensì alla vile meccanica. Sarà. Ma è un fatto che le pressioni per l'adozione delle griglie sono assai forti (se non altro per motivi legali) e i commissari si rassegnano. Solo quando i danni saranno così grandi da risultare insopportabili, ci si risolverà a scendere dall'olimpico e a confrontarsi con le questioni di misurazione e valutazione. A beneficio dei volenterosi che vogliono lasciare per un momento lo Spirito nell'angolo, tenterò in breve di sintetizzare quali sono i ragionamenti fondamentali da tenere ben presenti quando ci si accinge a usare le griglie (che in realtà non si dovrebbero neppure chiamare così²).

=ooOoo=

Va subito precisato che le griglie, quando sono utilizzate con cognizione di causa, sono molto utili e sono in grado di migliorare sensibilmente i controlli scolastici. Permettono di esplicitare i criteri di rilevazione e valutazione, permettono un eguale trattamento verso tutti i candidati, permettono a diversi insegnanti di omogeneizzare i loro giudizi. Permettono anche di velocizzare il controllo scolastico. Ma non tutte le griglie si somigliano. Ci sono griglie diverse che ottengono risultati diversi. Anzitutto occorre sapere che ci sono due tipi fondamentali di griglie (questa notizia nel nostro paese suscita spesso stupore, chissà perché): griglie *analitiche* e griglie *globali* (o sintetiche, olistiche). Dirò subito che il tipo meno usato nella scuola italiana (le griglie globali) è anche il più semplice, il più utile³; quello invece più usato (le griglie analitiche) è anche quello più complicato e, guarda un po', anche il più dannoso e incontrollabile. Ci piace davvero complicarci la vita.

=ooOoo=

Le griglie globali collocano la prestazione del candidato (quale essa sia, colloquio, soluzione di problema, tema, saggio breve...) su *una sola* dimensione, composta di solito di non più di *cinque* posizioni. Il fatto che le posizioni siano solitamente cinque non è casuale; è importante sia per motivi fondati sulla psicologia dei valutatori (gli esseri umani non sono in grado di discriminare contemporaneamente più di cinque o sei differenti posizioni⁴), sia per motivi statistici (infatti è conveniente che le distribuzioni delle prestazioni dei candidati vengano rappresentate attraverso una distribuzione che sia semplice, ma che sia anche coerente con la distribuzione normale; una distribuzione che soddisfa queste condizioni è la cosiddetta distribuzione *pentenaria*⁵ – si veda la figura in appendice). Qualunque altro raffinamento che si vorrà fare, qualunque traduzione verso altre scale di punteggi dovrà partire dalla collocazione della prestazione su una di queste cinque posizioni.

² Nel mondo anglosassone, dove le hanno inventate, erano chiamate *rating scales* (noi dovremmo dire “scale di valutazione”). Oggi sono chiamate *rubrics*, *scoring rubrics*, oppure *scoring guidelines*. In italiano ormai imperversa il termine “griglie”. Mi rassegnò all'uso comune, anche se sarebbe da preferire “scale di valutazione”.

³ Questo articolo è dedicato alla critica delle griglie analitiche. Tornerò, in un successivo articolo, sulle griglie globali, che sono molto più utili e hanno molti meno difetti.

⁴ Questa è la famosa legge di Miller.

⁵ Do per scontata la conoscenza di minime nozioni di statistica. Non è possibile in un articolo come questo spiegare nel dettaglio le distribuzioni cui si fa cenno. Tutti i manuali sulla valutazione scolastica spiegano ampiamente e chiaramente questi argomenti.

=ooOoo=

Le griglie analitiche invece pretendono di scomporre la prestazione (quale essa sia, colloquio, soluzione di problema, tema, saggio breve...) in un certo numero di *componenti*, di rilevare il rendimento del candidato su ciascuna componente e poi di sommare i diversi punteggi parziali per ottenere il risultato complessivo. Chi non conosce i meccanismi legati alle scale sarà portato a ritenere che le griglie analitiche, procedendo in questo modo, siano assai più precise di quelle sintetiche o globali (poiché compiono rilevazioni di tante dimensioni diverse rispetto allo stesso oggetto). Ciò forse potrebbe essere vero in teoria, oppure in situazioni sperimentali altamente controllate⁶. Nella realtà della valutazione scolastica, le cose non stanno proprio così. La pretesa di rilevare separatamente diverse caratteristiche della prestazione e di sommarle poi insieme produce un meccanismo inutilmente complicato e, come vedremo, difficilmente controllabile.

=ooOoo=

Per operare consapevolmente con le scale, occorre avere le idee chiare su cosa sia la *soglia di sufficienza*. Ciò che avviene all'esame di Stato non è una semplice rilevazione dei dati, ma una vera propria *valutazione*: ciò significa che è definita una soglia di sufficienza al di sotto della quale la prestazione verrà respinta, verrà dichiarata insufficiente. Dovrebbe essere chiaro come la definizione di una soglia di sufficienza rappresenti uno degli aspetti fondamentali della politica della valutazione. Una soglia troppo larga, finirà per promuovere tutti e non selezionare nessuno; una soglia troppo stretta finirà per bocciare un numero spropositato di candidati. C'è anche la possibilità di una soglia "ballerina" che vada e che venga a seconda dei casi, che finisce per rendere il controllo una vera e propria lotteria. E' anche chiaro che, per ragioni di equità, la soglia di sufficienza dovrebbe essere tendenzialmente uguale ovunque, altrimenti ci sarebbero delle commissioni di manica più larga o di manica più stretta (per questo si determina il fenomeno del "turismo" per andare alla ricerca dei "paradisi" dove le commissioni, per qualche motivo, sono notoriamente di manica larga).

Quale dovrebbe essere la posizione "giusta" della soglia? Evidentemente non c'è una risposta univoca, poiché questa dipende dalla politica della valutazione di ciascuna istituzione. Ma un ancoraggio statistico c'è. E' noto che le prestazioni, in una popolazione qualunque, si distribuiscono secondo una distribuzione nota come *gaussiana*, o *normale* (cfr. figura in appendice). La gran parte della popolazione si concentra intorno ai valori medi di prestazione; abbiamo poi una fascia più ristretta che si collocherà sopra la media; un'altra fascia ancora più ristretta si collocherà nell'area dell'eccellenza. Lo stesso vale per la parte bassa della distribuzione. E' proprio a partire dalla distribuzione gaussiana che è stata ricavata la distribuzione pentenaria, che divide la distribuzione gaussiana in *cinque fasce*. Ragionevolmente, se le cose stanno così si può pensare, in condizioni normali, di bocciare coloro che appartengono alla *fascia bassa* (circa il 7% della popolazione – che poi è quello che avviene più o meno di fatto). Se si vuol coinvolgere anche la fascia sotto la media (24%) si arriverà a un numero di respinti davvero elevato.

Naturalmente la distribuzione pentenaria è solo un riferimento. Molti fattori possono alterare la distribuzione: prove particolarmente difficili possono aumentare la fascia dei respinti. Prove troppo facili aumenteranno la fascia dei promossi. Se i ragazzi copiano faranno lavori tutti uguali e allora saranno più o meno tutti sulla media. Anche un insegnamento di qualità può aumentare a parità di condizioni il numero di coloro che stanno nelle fasce alte.

=ooOoo=

Se, in un modo o nell'altro, la distribuzione gaussiana rappresenta il riferimento per le distribuzioni delle prestazioni, allora dovrebbe anche costituire un riferimento per le scale che vengono usate per le

⁶ In effetti gli psicologi spesso usano misurare indirettamente dimensioni complesse scomponendole in tratti elementari.

valutazioni. Se è comodo identificare cinque fasce sulla normale (di cui una insufficiente) sarà utile che le nostre griglie procedano nello stesso modo. Disgraziatamente questo non succede nell'Esame di Stato (la soglia di sufficienza all'Esame di Stato è definita per legge). Ad esempio, per la prova di italiano dell'Esame di Stato, comune a tutti gli ordini di studio, vengono attribuiti al massimo 15 punti, ma la soglia di sufficienza è 10; ciò significa che con un punteggio di 9/15 il compito viene considerato insufficiente. Anche coloro che non sono del mestiere capiranno che qualcosa non funziona. Approssimativamente due terzi delle posizioni disponibili (comprese tra 1-15) vengono usate per discriminare tra prestazioni insufficienti (che hanno un solo un esito, quello di respingere il lavoro del candidato), mentre solo sei posizioni vengono usate per discriminare tra prestazioni considerate come sufficienti o più che sufficienti. Questa è un'assurdità statistica e metodologica di cui sono colpevoli gli esperti del ministero che si sono inventati, a suo tempo, il meccanismo dell'Esame di Stato⁷. Purtroppo nella PA è difficile risalire ai responsabili dei misfatti.

La prassi corretta dovrebbe essere ben diversa. Se si valuta la prestazione lungo una sola dimensione globale e se si usa come riferimento la distribuzione pentenaria, si conviene normalmente (nel mondo anglosassone si fa così dappertutto) che *uno* solo dei cinque scalini venga usato per l'insufficienza (al di sotto di 1,5 scarti dalla media) e che gli altri quattro vengano utilizzati per discriminare tra le prestazioni sufficienti.

Se invece si valuta la prestazione in forma analitica, prendendo in considerazione diverse dimensioni, c'è il rischio che la delicata soglia di sufficienza non venga rispettata nella rilevazione delle singole dimensioni; in tal modo il significato dell'insufficienza viene perduto nei meandri della somma dei punteggi parziali. Solo alla fine, quando si compongono i punteggi, si scopre se il rendimento è sufficiente o no. Dietro a questa banale questione ci sta il problema della sommabilità delle soglie di sufficienza: in alcuni casi è logico che siano sommabili, in altri no. Facciamo un esempio. Poniamo di valutare un aspirante alla patente automobilistica in base a tre dimensioni: 1) conoscenza del veicolo, 2) conoscenza della segnaletica, 3) destrezza nella guida. Ebbene se questo avesse ottenuto 13/15 nella prima, 6/15 nella seconda (cioè insufficiente) e 15/15 nella terza, avrebbe totalizzato mediamente 11/15 (globalmente sufficiente) e sarebbe promosso. Ma si avrebbe il bel risultato di mettere sulla strada un pirata che non conosce la segnaletica. Da questo esempio si comincia a capire che le griglie analitiche sembrano rigorose, magari lo sono anche, ma possono facilmente sfuggire di mano. Il modo per sfuggire a queste difficoltà ci sarebbe, ma a costo di complicare alquanto tutto il procedimento.

=ooOoo=

Un'altra questione davvero cruciale è quella che riguarda *l'indipendenza* delle dimensioni prese in considerazione. È questo un problema che riguarda solo esclusivamente le scale analitiche e che contribuisce a renderle del tutto inaffidabili, almeno nell'uso scolastico (in altri termini, non si sa davvero che cosa misurino e comunque misurano il loro oggetto in maniera imprecisa⁸). L'argomento sarebbe piuttosto complicato, perché coinvolge questione di metodologia della misurazione e di statistica. Ma cercheremo di rendere l'argomento comprensibile a tutti (anche perché riteniamo che la corretta comprensione di un meccanismo di valutazione che riguarda centinaia di migliaia di persone sia un fatto prima di tutto di democrazia).

⁷ Questa assurdità è dovuta al fatto che i tecnici che si sono inventati l'Esame di Stato hanno voluto scimmiettare la pratica del voto decimale, in uso nella scuola italiana (i voti sono compresi da uno a dieci e la soglia di sufficienza è sei). Ma già quella pratica era statisticamente errata, per cui l'errore si è riprodotto in forma allargata, dai decimi ai centesimi.

⁸ Nel linguaggio tecnico si dice che una misurazione è *valida*, quando misura effettivamente quel che afferma di misurare, e che è *attendibile* se misura con la precisione dovuta. Ci possono essere misurazioni attendibili, ma non valide (si misura con grande precisione una cosa che non c'entra), oppure valide ma non attendibili (si misura davvero la dimensione che si intende misurare, ma lo si fa in modo alquanto impreciso). Poi ci anche sono le misurazioni che non sono né attendibili, né valide, casi non infrequenti.

Se vogliamo misurare una prestazione su diverse dimensioni (di solito si considerano in ballo 3, 4 o 5 dimensioni) e procedere poi a una somma dei punteggi ottenuti, bisogna che questa dimensioni siano in qualche modo *indipendenti* tra loro, altrimenti finiremo per misurare 2, 3 o quattro, volte la stessa cosa⁹. Si dice, in campo statistico, che le dimensioni considerate devono essere *ortogonali* (è un altro modo per dire indipendenti). Ora, siccome le prestazioni degli allievi dipendono da una quantità di fattori imprecisati, ma spesso sono tra di loro fortemente intercorrelate (qualunque insegnante sa che se un allievo conosce bene l'ortografia, di solito si esprime anche bene, produce delle proposizioni coerenti ...).

Insomma, quando tra le varie dimensioni c'è una forte correlazione, misuriamo sempre la stessa cosa (senza sapere davvero cosa). Tanto varrebbe usare una unica dimensione. A questo punto l'ingenuo dice: meglio così, così la misura sarà più precisa. In realtà si possono però fare degli errori gravi perché, qualora l'intercorrelazione non fosse perfetta e sussistesse solo tra alcune dimensioni, si correrebbe il rischio di sovra rappresentare alcune dimensioni rispetto alle altre (senza però sapere esattamente quali siano).

In campo psicologico è difficilissimo che le prestazioni individuali possiedano delle dimensioni fortemente indipendenti tra loro. Già solo questa considerazione scongiurerebbe di usare le scale analitiche. Per sapere se due dimensioni si possono considerare come indipendenti o intercorrelate occorrerebbe fare delle ricerche su popolazioni molto ampie, applicando delle tecniche sofisticate, come ad esempio l'analisi fattoriale.

=ooOoo=

Se le griglie analitiche sono più complicate da utilizzare, se non permettono di avere un chiaro controllo sulla soglia di sufficienza, se sono poco valide e poco attendibili, se si rischia di misurare accuratamente il nulla, allora non resta che utilizzare le scale globali o sintetiche, le quali non hanno nessuno di questi difetti, psicologicamente sono più vicine alle capacità di discriminazione dell'essere umano e sono più facilmente trattabili dal punto di vista statistico. Ma allora perché nella scuola si continuano ad usare le scale analitiche? C'è una sola risposta. Le scale analitiche portano con sé un'*illusione* di precisione, di scientificità e di imparzialità. In nome di questa illusione molti onesti educatori sacrificano sull'altare dei numeri la loro capacità personale di discriminazione delle prestazioni accumulata in anni e anni di esperienza.

=ooOoo=

Allo scopo di provare quanto siamo andati sostenendo, proveremo ad analizzare le caratteristiche (e i difetti) di una scala analitica. Abbiamo scelto come cavia una scala che non è né la migliore né la peggiore, una scala è stata diffusa qualche anno fa da uno dei tanti IRRSAE della nostra penisola (cfr. l'appendice). Di scale di questo genere ne circola una grande quantità. La scala in questione serve per la valutazione della prima prova del Esame di Stato, ovvero la prova di italiano. Abbiamo utilizzato questa disciplina perché è quella più generale, che riguarda tutti gli indirizzi di studio.

=ooOoo=

Anzitutto va detto che questa scala pretenderebbe di connettere insieme l'elemento analitico e l'elemento sintetico. Dal punto di vista della teoria della misurazione ciò è davvero assurdo (vorrebbe dire che si valuta due volte lo stesso oggetto, una volta in modo analitico e una volta in modo sintetico; inoltre si stabilisce anche, piuttosto arbitrariamente, che la valutazione analitica vale un po' di più della valutazione sintetica – per la valutazione sintetica vengono dati al massimo 3 punti su 15, mentre per quella analitica si può arrivare a 12 su 15). In verità, poiché la valutazione globale (“efficacia

⁹ Se misuriamo delle sfere in base a diversi parametri come il raggio, il diametro, la circonferenza, la superficie, ognuno capirà che si misura sempre la stessa cosa, perché ogni parametro è legato all'altro.

complessiva del testo”) non può che essere in gran parte la conseguenza delle prime tre dimensioni analitiche, essa non può che essere una ripetizione (anche perché nelle prime tre dimensioni sono contenuti aspetti che incidono fortemente sulla presentazione globale del testo, ad esempio la sua strutturazione). Se non fosse così, se l’efficacia complessiva avesse invece una sua chiara autonomia (rispetto alle altre tre dimensioni), allora non si capirebbe perché mai chiamarla sintetica (sarebbe una dimensione analitica come le altre; resterebbe semmai da spiegare perché dovrebbe pesare di meno rispetto alle altre tre dimensioni).

=ooOoo=

Prendiamo ora in esame la mini scala dei valori che vengono usati per valutare ciascuna delle dimensioni. Nella legenda vengono chiamati “livelli”. Per l’individuazione dei livelli vengono utilizzati, in forma un po’ sovrabbondante, numeri ed aggettivi. Supponiamo che il valutatore sia invitato a considerare prima di tutto l’aggettivo, più vicino alla semantica comune, e solo dopo il numero, che è assai più astratto. Così vien definito che “basso” vale da zero a un punto; “medio” vale due o tre punti, “alto” vale quattro punti. Tuttavia uno sguardo attento alla tabella ci convincerà che l’indicazione “zero” non ha in realtà nessun significato, poiché nella colonna a destra (dove si sommano i punti) il *range* ammesso va da uno a quattro. Quindi, una realizzazione linguistica “bassa” non può ottenere meno di uno (così sembra almeno). Se questo è vero però, ciò significherebbe che anche un compito consegnato in bianco avrebbe garantiti, secondo la tabella, almeno 4 punti (in questo non vedo nulla di male, ma dovrebbe essere giustificato in qualche modo; in tal caso la presenza dello zero come indicatore numerico del livello basso non si capisce bene cosa ci stia fare).

Nelle prime tre dimensioni (considerate “analitiche”) i livelli (o i punti) sono quattro, anche se due di questi sono contrassegnati dallo stesso aggettivo “medio”. Nell’ultima dimensione, la valutazione globale, i livelli o punti sono invece tre. Questo fatto non è del tutto corretto per le griglie analitiche: infatti il valore tre quando viene assegnato alla valutazione globale significa “alto”; quando viene assegnato alle altre tre dimensioni significa invece “medio”. Fin qui nulla di male. Ma alla fine vengono sommati insieme dei numeri che, lungi dall’essere, come si dovrebbe, privi di connotazioni qualitative, hanno un significato valutativo diverso. Come si vede la girandola di aggettivi e numeri, che si presenta come molto scientifica, in realtà nasconde stravolgimenti indebiti di significato. Una vera e propria manipolazione che, essendo presentata come tecnica, viene *ipso facto* giustificata.

Aggiungo che, dal punto di vista teorico è già difficile individuare delle dimensioni che siano ortogonali. Pretendere poi di sapere quale sia il giusto peso di una dimensione rispetto all’altra (senza una adeguata base empirica) è davvero rischioso.

=ooOoo=

Abbiamo già spiegato che in ambito valutativo è sempre estremamente importante tenere in conto dove si collochi la *soglia di sufficienza*. Negli aggettivi che indicano il significato dei numeri su ciascuna dimensione, non viene assolutamente specificata quale sia la soglia di sufficienza. In effetti, di per sé, il livello “basso” non significa che sia insufficiente. Quel che qui si evita accuratamente di dire, lasciando l’aggiustamento alla cieca logica del calcolo, è quante, delle quattro, dimensioni devono essere insufficienti perché una prestazione sia considerata insufficiente. Si ricordi l’esempio della valutazione dell’automobilista patentando. Se vogliamo essere rigorosi, dovrebbe essere bastare una dimensione insufficiente per conferire l’insufficienza a tutta la prova (insufficiente vuol proprio dire che non basta, non va, che la prova è fallita). Sarebbe questo il caso della correttezza orto – sintattica: si può sostenere che se questa è gravemente insufficiente, qualsiasi altra dimensione non servirà a salvare una prova di italiano. Se invece ragioniamo in maniera più generosa, per una insufficienza generale dovremmo chiedere l’insufficienza per tutte e quattro le dimensioni. E’ chiaro che allora i risultati sarebbero assai diversi. In ogni caso occorrerebbe esplicitare un chiaro criterio. Secondo questa scala invece, solo alla fine, quando si sommano i punteggi (quarti e terzi) e si producono i quindicesimi si

scoprirà se l'elaborato è riuscito a superare la scaglia di sufficienza (che sta implacabilmente tra il 9 e il 10, per decisione ministeriale). Questo confuso meccanismo di somma di aggettivi attraverso numeri produrrà *ciacamente* un livello di sufficiente o insufficiente, magari per un punto in più o in meno (il quale punto però – si ricordi - può valere uno su quattro (prime tre dimensioni), oppure uno su tre (quarta dimensione). Lasciar decidere la soglia di sufficienza al gioco dei decimali e degli arrotondamenti non è molto onesto.

=ooOoo=

Per vedere come funzionano empiricamente le griglie, è sufficiente fare dei casi numerici e osservare le conseguenze. Uno che abbia consegnato il compito in bianco, secondo la griglia esaminata, non può che prendere 4/15. Ma, udite, udite, uno che si sia sempre collocato sul livello medio (nella zona però bassa) totalizzerà 2+2+2+2, cioè 8/15, risultando ancora, secondo la valutazione ministeriale, insufficiente. Ma risulterà insufficiente anche un elaborato che abbia totalizzato 3+2+2+2. Per avere la sufficienza bisogna avere almeno 3+3+2+2, cioè due valori medio alti, un medio basso e un medio medio (quello della quarta dimensione). Poiché, secondo la distribuzione pentenaria i medio-bassi sono il 27% circa della popolazione, con questa griglia si finirebbe per respingere più o meno il 35% dei candidati (poiché vanno aggiunti anche i bassi, che statisticamente ammontano a circa il 7%). D'altro canto, anche i medio alti non vengono trattati tanto bene: il loro voto sarà 3+3+3+2, ovvero un misero 11 (tradotto in quindicesimi, una specie di sei e mezzo!).

Una simile distribuzione dei punteggi risulta assolutamente punitiva per gli allievi. Se si vuol evitare di respingere troppi candidati (e finire sui giornali) si dovranno fare allora i soliti aggiustamenti, oppure ridefinire al ribasso il significato dell'aggettivo "medio"... Tutti espedienti sempre possibili, ma alquanto macchinosi e comunque dagli esiti incontrollabili.

=ooOoo=

Abbiamo già chiarito che la valutazione globale (se è davvero globale) non può sicuramente essere ortogonale, ovvero indipendente, a rispetto alle altre tre dimensioni analitiche. Ma sono ortogonali almeno le tre dimensioni che sono valutate tutte con lo stesso *range* di punteggi e di aggettivi (1-4)? Abbiamo già spiegato che l'ortogonalità effettiva può esser comprovata solo con ricerche *ex post*. Tuttavia almeno nelle descrizioni delle diverse dimensioni considerate si dovrebbe *enfaticamente* l'indipendenza delle dimensioni (evitando di metterci delle cose nei fatti molto simili tra loro). I descrittori delle caratteristiche in esame, in scale di questo genere, dovrebbero poi essere assolutamente chiari, precisi, inequivocabili. Correttori diversi dovrebbero sapere esattamente cosa prendere in considerazione nel momento in cui attribuiscono gli aggettivi e poi i punteggi.

Andiamo allora a leggere i descrittori delle varie caratteristiche. In realtà alcune espressioni definitorie che compaiono sono estremamente generiche e vaghe, di non facile interpretazione (non è molto chiaro esempio che cosa siano gli "elementi di contestualizzazione"). Gli "usi del linguaggio settoriale" che stanno nella prima dimensione sono terribilmente simili alla "presenza di modalità di discorso appropriate alla forma testuale e al contenuto" della seconda dimensione. La valutazione del rapporto fra ampiezza e qualità informativa (quarta dimensione) sembra inoltre già contenuta nella seconda e nella terza dimensione. Possiamo inoltre aggiungere che la voce "La creatività e le capacità critiche personali" è sicuramente frutto di uno svarione, perché la griglia dovrebbe misurare le prestazioni e *non* la persona (purtroppo questo è un errore in cui si incorre molto, molto, ma molto spesso). Dalla misurazione di una prestazione non si dovrebbe mai inferire una caratteristica della persona stessa; le capacità critiche, se ci sono, saranno quelle esibite dal testo esaminato.

=ooOoo=

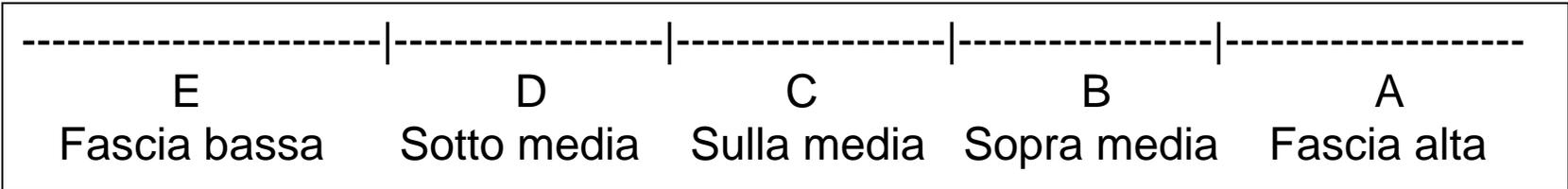
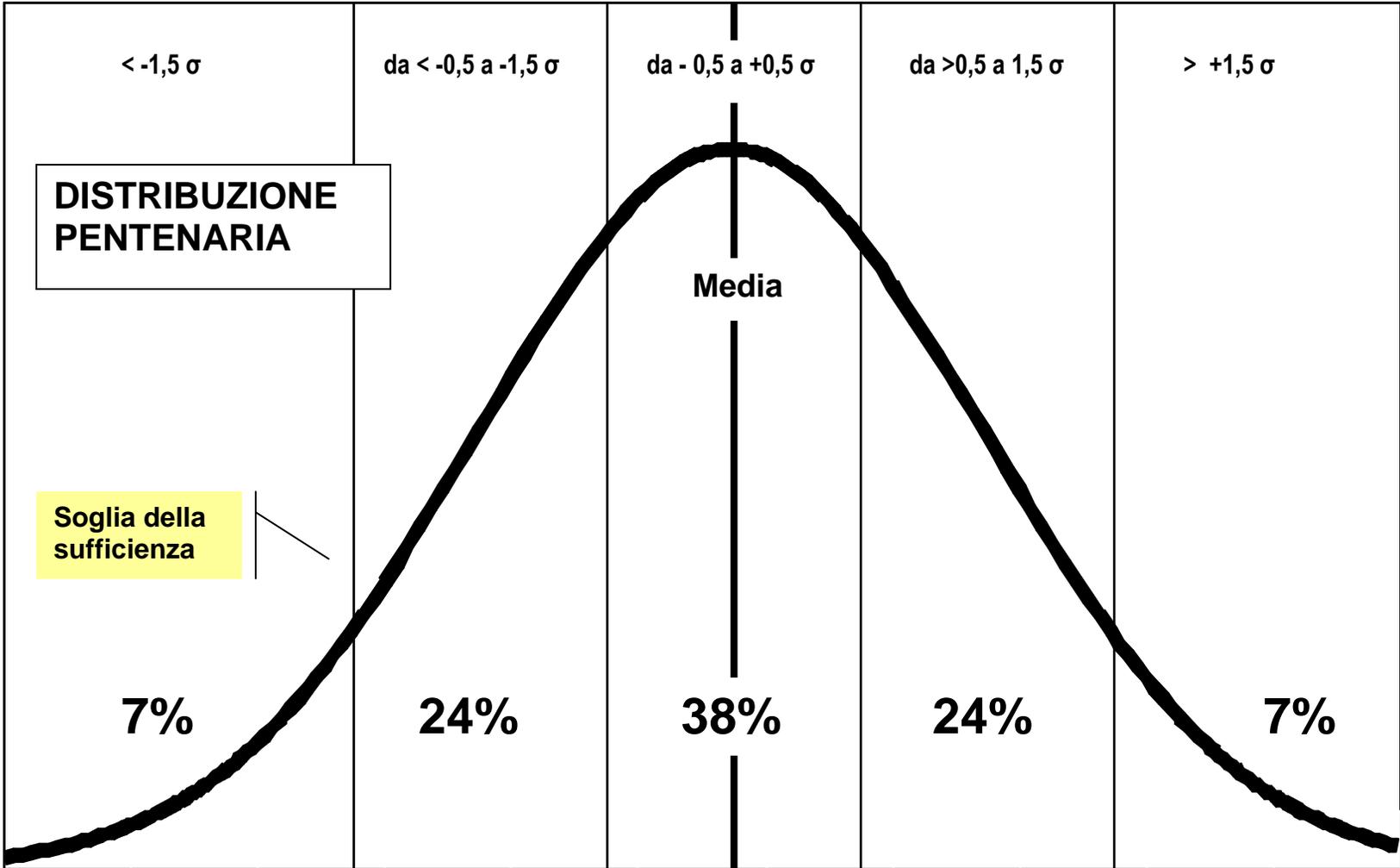
Che valutazione globale possiamo dare, a questo punto, di questa griglia analitica di valutazione? Mette in atto un apparato tecnico molto complicato. Occorre fare molta fatica per familiarizzarsi con le

varie dimensioni da prendere in considerazione e con le diverse scale di rilevazione dei dati parziali (quarti e terzi). Non si ha l'identificazione corretta della soglia di sufficienza effettiva, di ciascuna dimensione, per cui la soglia ministeriale può prendere il sopravvento sugli aggettivi analitici. Le varie dimensioni sono descritte in forma vaga, si sovrappongono in alcuni casi e non sono certamente indipendenti tra loro. Chi ha fabbricato un simile meccanismo mostra senz'altro un ingenuo spirito di servitù nei confronti del tecnicismo e dell'ideologia scienziata, e forse anche un po' di presunzione.

Alla prova dei fatti, griglie di questo genere, spesso imposte dai vertici burocratici, o da fanatici attivisti tecnocrati, vengono aggirate dai correttori dotati di buon senso, i quali prima si mettono d'accordo sulla valutazione finale in quindicesimi (tenendo saggiamente conto della soglia di sufficienza e delle differenze che emergono man mano tra un compito e l'altro¹⁰) e poi provano, più o meno arbitrariamente, a motivare il voto, ovvero a distribuire il punteggio tra le quattro dimensioni (a volte incontrando serie difficoltà nel far quadrare la questione). La griglia così, da strumento utile per migliorare le capacità di discriminazione e per garantire di più l'allievo, finisce per diventare uno strumento ideologico a posteriori che serve solo a garantire i valutatori, ovvero a giustificare una valutazione cui si è giunti attraverso percorsi completamente "extra - grigliari", basati spesso, ciò va riconosciuto, sull'esperienza e sul buon senso degli insegnanti.

¹⁰ Uno dei metodi più sicuri per correggere dei compiti è sempre quello che consiste nel correggere prima alcuni lavori di ragazzi che solitamente fanno bene e alcuni di ragazzi che di solito fanno male (per avere il senso degli estremi). Poi correggere tutti gli altri facendo dei mucchietti di compiti che sembrano simili tra loro (spostando poi da un mucchio all'altro quelli che sono meno simili a ciascun mucchio). Raramente avremo bisogno di fare più di 5 o 6 mucchietti. Ne faremo 7 o 8 se siamo proprio pignoli (ma potremmo fare degli errori di attribuzione).

Appendice



Esame di Stato - Proposta di griglia di valutazione della prima prova

Aspetti del testo	livello rilevato			punti
	basso	medio	alto	
VALUTAZIONE ANALITICA				1-12
REALIZZAZIONE LINGUISTICA	0-1	2-3	4	1-4
Correttezza ortografica, morfosintattica e sintattica; coesione testuale (tra le frasi e le parti più ampie di testo); uso adeguato della punteggiatura; adeguatezza e proprietà lessicale; uso di registro e, dove richiesto, di linguaggi settoriali adeguati alla forma testuale, al destinatario, al contesto e allo scopo.				
COERENZA E ADEGUATEZZA alla forma testuale e alla consegna	0-1	2-3	4	1-4
Capacità di pianificazione: struttura complessiva e articolazione del testo in parti; coerente e lineare organizzazione dell'esposizione e delle argomentazioni; presenza di modalità discorsive appropriate alla forma testuale e al contenuto (ad esempio parafrasi e citazioni nelle prime due tipologie); complessiva aderenza all'insieme delle consegne date.				
CONTROLLO DEI CONTENUTI	0-1	2-3	4	1-4
Ampiezza, padronanza, ricchezza, uso adeguato dei contenuti, in funzione anche delle diverse tipologie di prove e dell'impiego dei materiali forniti: <i>Tipologia A:</i> comprensione e interpretazione del testo proposto; coerenza degli elementi di contestualizzazione; <i>Tipologia B:</i> comprensione dei materiali forniti e loro utilizzo coerente ed efficace <i>Tipologia C e D:</i> coerente esposizione delle conoscenze in proprio possesso in rapporto al tema dato; complessiva capacità di collocare il tema dato nel relativo contesti culturali <i>Per tutte le tipologie:</i> significatività e problematicità degli elementi informativi, delle idee, delle interpretazioni				
VALUTAZIONE GLOBALE	0-1	2	3	1-3
Efficacia complessiva del testo, che tenga conto anche del rapporto fra ampiezza e qualità informativa, della originalità dei contenuti o delle scelte espressive, della globale fruibilità del testo, della creatività e delle capacità critiche personali.				
VALUTAZIONE COMPLESSIVA				max 15