



## Valutazioni e concorsi. Errando discimus

Nicola Paparella

Orcid: 0000-0003-3095-1184

© Nicola Paparella, 2012

Ripubblicato con il titolo: *Valutazioni. Fra competenze e chiacchiericcio docimologico* in D. DE LEO (a cura di), *Pensare il senso. Perché la Filosofia. Scritti in onore di G. Imvitto*, vv.2, Mimesis, Milano, 2014, ISBN 8857518914, 9788857518916. Vol. secondo: pp. 929 – 952.



Come la risacca riporta sulla spiaggia ciò che l'onda aveva sospinto verso il largo, così le mode – anche quelle culturali – riprendono, di tanto in tanto, tecniche e formule che la scienza aveva ormai superato e allontanato, riproponendole come incontrovertibili novità.

E così, nella stagione in cui tutti parlano di rigore e nella quale a molti piace affidarsi ai test, alle prove “oggettive”, alle graduatorie alle stime, alle previsioni, ritrova inattesa credibilità l'idea di poter affidare ad un numero o ad una etichetta il valore di una ricerca o il merito di una persona, spesso a discapito del senso stesso del valutare, oltre che della efficacia e persino della efficienza del processo.

Nell'ambito della scuola e della università, mentre la macchina dei concorsi riprende a muoversi, fra sobbalzi e singhiozzi, condizionata e compromessa dal lungo periodo di inattività, i decisori pubblici e, con loro, le fila mai poco folte, di corifei e turiferari, hanno ripescato come grandi novità ed impareggiabili risorse, procedure di valutazione affidate quasi soltanto, non già alla misurazione, ma all'uso quasi sempre incauto di indicatori numerici. E lo fanno con una tale enfasi da sollecitare consensi ed iniziative anche là dove nessuno è ancora intervenuto ad ordinare e persino in assenza di norme che chiaramente prescrivano, perché, come si sa, le mode hanno una loro autonoma forza cogente alla quale è difficile sottrarsi.

Purtroppo sfugge a molti che la docimologia ha ormai da tempo smorzato il dibattito fra cultori dei metodi quantitativi e cultori dei metodi qualitativi, non soltanto perché ha guadagnato punti di vista più comprensivi, ma anche perché



da tempo ha spiegato e sancito che la base d'appoggio del rigore non sta tanto nelle tabelle e nei numeri - usati purchessia - quanto nell'uso di metodi e tecniche scientificamente accreditati, nella puntualità dell'analisi semantica, nella coerenza dei nessi logici e quindi anche nel filo che lega gli itinerari procedurali alle premesse epistemologiche.

Merita allora riprendere questi temi<sup>1</sup>, non tanto per fare il punto sullo stato della ricerca, quanto per declinarne i risultati sul versante delle operazioni che possono essere affidate, ad esempio, ad una commissione di concorso o, più in generale, a chi abbia l'onere e la responsabilità di valutare e di decidere del destino di altre persone.

### Tre questioni

Ci sono almeno tre questioni essenziali che hanno chiaro carattere fondativo per qualunque discorso riguardante la valutazione e che manifestano tutta quanta la loro efficacia e, insieme, la loro forza epifanica ed ermeneutica, in diversi contesti, anche molto diversi come possono essere, ad esempio, il monitoraggio delle attività didattiche o le operazioni di valutazione di un concorso. Per questo meritano d'essere qui richiamate, anche perché finiscono con il condizionare non soltanto la qualità del processo valutativo e la sua efficacia, ma persino la sua credibilità e la sua legittimità scientifica.

Si tratta di affrontare compiti ben precisi, già nell'avvio di un processo di valutazione, persino quando si operi all'interno di un profilo formale – istituito – e comunque prima di porre mano alle singole prove, ove previste e comunque previste.

Si tratta più esattamente di:

- a) esplicitare i fini (o, anche soltanto gli obiettivi) del lavoro di valutazione e chiarire – in termini formali – il senso da attribuire alle prove, *soprattutto* nella ipotesi che esse siano state già in qualche modo previste ed indicate da decisori diversi da coloro cui spetta il compito di valutare;
- b) determinare le procedure e gli strumenti, determinandone i diversi passaggi in ragione di risultati analiticamente preventivati;
- c) argomentare sui giudizi ovvero sulla valutazione propriamente detta e sul senso e sul peso che essa ha all'interno del processo complessivo.

Si tratta di aspetti decisivi e, come s'è detto, fondanti, che incredibilmente rimangono spesso “sottintesi”, là dove invece l'esplicitazione consentirebbe non soltanto di chiarire e di precisare, assumendone la responsabilità, ma anche di gestire in maniera avveduta, come del resto viene da sempre suggerito in sede scientifica.

---

<sup>1</sup> Ci riferiamo al ns. *Concorsi e valutazioni. Errando discimus*, in G. TREBISACCE, *Valutazione e scuola. Idee per l'innovazione metodologica e didattica*, Quaderni del dipartimento, Università della Calabria, Rende, 2012, pp. 217 e ss.



Già nel 1967, quando M. Gattullo parlava di *controllo* scolastico, insisteva molto perché si procedesse partendo da una attenta distinzione fra la determinazione dell'oggetto, la misurazione e la valutazione propriamente detta<sup>2</sup>.

Ci si muoveva allora in campi appena dischiusi alla ricerca e alla esperienza pedagogica e la docimologia aveva ancora bisogno di distinguere e di precisare i diversi suoi comparti e le diverse sue funzioni, come si andava facendo, anche in Italia, grazie all'assiduo lavoro di una ristretta cerchia di studiosi molto attenti, come B. Vertecchi, che proprio in quegli anni ci dava una definizione essenziale ed efficace, ancora oggi molto accreditata, per la quale la docimologia è da intendersi come "teoria e tecnica del *controllo dei processi formativi*"<sup>3</sup>.

In verità, in quegli anni, si sviluppava, anche da noi, una grande area di interessi, pratici e scientifici, determinata dalla proiezione del tema della valutazione agli ambiti dell'agire sociale, dove per altro è andato imponendosi, sin dall'inizio, un orientamento di tipo operativo, tanto da intendere la valutazione come l'insieme delle attività che possono risultare utili per esprimere un giudizio in forza di affidabili procedure destinate al trattamento delle informazioni utilizzate per formulare il giudizio in questione<sup>4</sup>. In questa prospettiva la valutazione finiva con l'includere, come diceva C. Bucciarelli, "tutte le azioni finalizzate alla raccolta ed alla elaborazione delle informazioni critiche, per le decisioni di proseguimento, arresto o modificazione nelle fasi di processo; nonché i criteri operativi per l'utilizzo di dette informazioni" e questo perché "la finalità del processo valutativo ... è in definitiva l'individuazione dei legami funzionali ... tra i diversi fattori iniziali e le modificazioni avvenute durante l'iter processuale"<sup>5</sup>.

In questo senso si determina una sorta di ampia sovrapposibilità fra due nozioni che di per sé meritano d'esser tenute distinte, quella di valutazione e quella di ricerca valutativa, e principalmente si configura una condizione per la quale l'affidabilità del giudizio valutativo è da ricondurre alle procedure, mentre la credibilità delle procedure è da fondare, per un verso, sugli obiettivi assegnati al processo e, per altro verso, al rigore scientifico della investigazione.

Questa esigenza di rigore e questa attesa di procedure condivise dalla comunità scientifica serpeggia lungo tutta la storia della valutazione, sin dalle prime battute di F.V. Edgeworth<sup>6</sup>, quando ancora non si parlava di docimologia, e sin dai tempi del gruppo che lavorava attorno a H. Piéron<sup>7</sup> e al quale si fa risalire la nascita della docimologia, per giungere ai lavori a noi più vicini, come quelli – in area

---

<sup>2</sup> Cfr. principalmente M. GATTULLO, *Didattica e Docimologia. Misurazione e valutazione nella scuola*, Armando, Roma, 1967.

<sup>3</sup> B. VERTECCHI, in A. VISALBERGHI, R. MARAGLIANO, B. VERTECCHI, *Pedagogia e scienze dell'educazione*, Mondadori, Milano, 1978.

<sup>4</sup> Cfr. C. BEZZI, *Il disegno della ricerca valutativa*, Angeli, Milano, 2001, p. 29.

<sup>5</sup> C. BUCCIARELLI, *Come valutare. La necessità di un approccio sistemico*, in Censis, *Speciale valutazione*, "Note e commenti" n. 1/2, 1991, pp. 9 e 17.

<sup>6</sup> Cfr. F. V. EDGEWORTH, *The element of chance in competitive examinations*, in *Journal of the Royal Statistical Society*, 53, 1890, pp. 460-475, 644-663.

<sup>7</sup> Cfr. H. LAUGIER, H. PIÉRON, É. TOULOUSE e D. WEINBERG, *Études docimologiques sur le perfectionnement des examens et concours*, Conserv. nat. des arts et métiers, Paris, 1934.



pedagogica – di A. Visalberghi, di B. Vertecchi, di R. Maragliano, di G. Domenici<sup>8</sup> e più ancora dei tanti che si sono occupati di processi valutativi applicati al sociale<sup>9</sup>, soprattutto agli investimenti pubblici<sup>10</sup>, alle università<sup>11</sup>, ai progetti europei e alle questioni connesse con l'economia.

### I fini e la coerenza semantica

Il primo aspetto da chiarire, allora, è quello dei *fini* cui non è difficile riconoscere una chiara rilevanza nell'ambito di percorsi valutativi quand'essi siano direttamente e compiutamente progettati, e che però resta ugualmente importante anche quando ci si debba muovere in contesti istituiti nei quali la decisione e la progettazione dell'intervento sono evidentemente disgiunti dalla esecuzione. In un concorso, ad esempio, alcune delle scelte di fondo sono legate a decisioni che l'ordinamento affida a funzionari ben distinti da coloro che sono poi chiamati ad intervenire nel processo valutativo. C'è una evidente distanza tra un bando di concorso (e coloro che lo hanno emanato) e il team dei funzionari che lo debbono gestire. E tuttavia, a questi ultimi non viene chiesto soltanto di eseguire o di adoperare strumenti preconfezionati da gestire a scatola chiusa; tant'è che vengono selezionati con cura, perché si vuole che possano mettere in campo una competenza professionale di apprezzabile livello. Non si chiede loro di tirar fuori una checklist e cavarne qualche numero, ma di garantire rigore e credibilità ad operazioni fondate sul merito comparativo.

Ed è davvero incredibile che apprezzati professionisti possano accingersi a lavori di questo rilievo e a procedure per le quali si richiede l'esercizio di funzioni complesse, senza una qualche riflessione, da parte loro, sul senso da attribuire all'opera che si va a compiere o senza un qualche intervento protesico ad esplicitare finalità ed obiettivi, nessi logici e condizioni di congruità logica fra le diverse operazioni che nel loro insieme chiamiamo valutazione.

Discutere, condividere il senso del lavoro che si compie, confrontarsi su finalità ed obiettivi è la condizione previa di una prestazione almeno dignitosa, altrimenti si preclude, già dalla radice, persino la possibilità di ottenere esiti segnati da efficacia, coerenza e rigore.

Oggi, in sede scientifica, gli addetti ai lavori sanno bene a che cosa si alluda quando si parla di rigore, a proposito dei processi di valutazione, e sanno come

---

<sup>8</sup> Si ricordino in particolare: A. VISALBERGHI, *Misurazione e valutazione nel processo educativo*, Comunità, Milano, 1955; ID, *Didattica e docimologia: misurazione e valutazione nella scuola*, Armando, Roma, 1967; R. MARAGLIANO, *Valutazione e scuola*, Lisciani e Giunti, Teramo, 1989; B. VERTECCHI, *Decisione didattica e valutazione*, La Nuova Italia, Firenze, 1993; G. DOMENICI, *Manuale della valutazione didattica*, Laterza, Bari, 1993.

<sup>9</sup> Cfr. C. CIPOLLA, *Epistemologia della tolleranza*, Angeli, Milano, 1997.

<sup>10</sup> Un esempio per tutti: G. PENNISI, *Tecniche di valutazione degli investimenti pubblici*, con scritti di P.L. SCANDIZZO, Istituto Poligrafico e Zecca dello Stato, seconda edizione aggiornata, Roma, 1991.

<sup>11</sup> Cfr. B. BORDIGNON, *Epistemologia della valutazione universitaria. La valutazione professionale del docente*, Rubbettino, Soveria Mandelli, 2010.



si debba agire e che cosa si debba fare per garantire credibilità ai giudizi di valutazione. Non altrettanto può dirsi per la prassi, per l'esercizio diffuso della valutazione e, in alcuni casi, persino per alcune iniziative ufficiali<sup>12</sup>.

Eppure non sarebbe difficile assicurare qualche buon risultato o, almeno, evitare qualche danno particolarmente grave e comunque del tutto evidente, come può accadere, ad esempio, in sede concorsuale, dove l'uso improprio di strumenti docimologici può determinare un alone negativo nei confronti delle stesse tecniche e delle procedure di misurazione e di controllo.

Ed allora, lo ripetiamo, conviene partire dai fini.

Una volta si sarebbe detto (e in etica lo si dice ancora<sup>13</sup>) che è buona cosa distinguere tra *finis operantis* e *finis operationis*, ossia, per fare un esempio, fra il fine di una commissione di concorso e i fini immanenti alle sue scelte, e fra questi due aspetti e il cosiddetto *finis operis* ossia, nel nostro caso, ciò che l'intera procedura ci consegna. Per dirla in maniera ancora più essenziale, sarebbe bene capire e distinguere *perché* si fanno certe operazioni, *che cosa* si fa ed *a che cosa serve* ciò che si fa.

Ogni opera dell'uomo è mossa da alcune intenzioni e guadagna in trasparenza, efficacia e credibilità se viene accompagnata da una chiara manifestazione dei fini o, almeno, dei punti di particolare rilievo, e, soprattutto in alcune specifiche circostanze, delle ragioni teoriche per le quali ci si orienta in un certo modo e non in un altro. Essere puntuali e precisi a questo riguardo, non è una ridondanza, ma una precisa condizione che conferisce rilievo e coerenza a tutto il lavoro.

Se affermiamo, guidati dal senso comune, che una commissione deve poter essere neutrale, di fatto stiamo dicendo che la *neutralità* dovrebbe poter essere uno dei fini della commissione. Sarebbe allora utile ed opportuno che questo venga chiaramente affermato e poi tradotto in una serie di operazioni capaci di rendere operativamente evidente quanto si è deciso ed affermato.

Se invece, per rimanere nell'esempio, una commissione di concorso si limitasse a dire che intende avvalersi di procedure metriche e che per questo provvederà a costruire schede di rilevazione, se aggiunge che intende controllare le prove con scale opportunamente allestite per l'occasione, di fatto, sta facendo un elenco di strumenti senza fornire alcuna spiegazione che sia in grado di illustrare il perché di quelle tecniche e di quelle procedure. È come scegliere le pagine da leggere senza aver detto perché e che cosa si vuol leggere. È come staccare un biglietto di viaggio prima ancora di progettare il percorso e forse prima d'aver deciso di partire. Che senso ha parlare di procedure metriche se non si dice perché, se non si stabilisce che cosa si vuol misurare, se non si collega tutto questo agli obiettivi dell'intera operazione?

---

<sup>12</sup> Cfr. P. LUCISANO, *Validità e affidabilità delle pratiche valutative: a proposito del progetto pilota 2*, in *Cadmo*, n. 2, 2003, pp. 37-56.

<sup>13</sup> Cfr. S.-L. DUARTE, *A especificação moral dos actos humanos segundo são Tomás de Aquino*, Edusc, Roma, 2008, pp. 440 e ss.



Prima delle tecniche, ci sono i criteri e prima dei criteri trovano spazio gli obiettivi, i quali, a loro volta, richiamano i fini. Perdere questa sequenza o saltarne i passaggi colora di precarietà l'intera procedura.

La definizione dei criteri, l'allestimento di schede, di scale e di tutto quel che segue rientra in operazioni che trovano una loro motivazione, un loro *perché* ufficiale nelle *intenzioni* della commissione ovvero in quello che abbiamo detto il *finis operantis*. Se questo non viene esplicitato (e quindi riportato a verbale, per adoperare il linguaggio della burocrazia), è evidente che ci stiamo muovendo in un contesto di decisioni non ancora pienamente motivate<sup>14</sup>.

Eppure non sarebbe difficile, in un contesto operativo molto burocratizzato, come può essere il sistema scolastico italiano, rifarsi *almeno* alle norme e declinarle nello specifico contesto. Ma non è così. In molte occasioni il sottinteso prevale.

Qualche volta ci è capitato di leggere qualche richiamo alla *trasparenza* che, al pari della *neutralità*, potrebbe essere un buon avvio di discorso sui fini. Anche in questo caso, tuttavia, occorre stare attenti a non trasformare il discorso sulle intenzioni in un ossequio rituale a formule di moda. Si può pensare alla *trasparenza* e limitarsi però alla sola pubblicazione dell'elenco dei criteri da adottare, che è certamente uno degli aspetti richiamati dalla nozione di trasparenza, forse nemmeno il più importante, dovendosi assumere come decisiva e fondamentale, invece, la condizione di *leggibilità di tutte le fasi* del progetto, permettendo così la ricostruzione - anche dall'esterno - del percorso logico che conduce al giudizio finale. È in funzione della trasparenza, inoltre, che vanno esplicitati i *nessi concettuali* che legano le diverse parti del processo e i diversi strumenti ai quali, volta per volta, ci si affida.

Se si dichiara l'opzione di fondo della trasparenza, non si possono poi lasciare come non detti e come sottintesi i nessi che legano un momento all'altro del processo.

Precisiamo ed approfondiamo con un esempio: Se una commissione, nel presentare i criteri con cui intende valutare le prove scritte di un concorso, trascura di considerare che vi sono anche delle prove orali e quando poi, discutendo dei criteri da seguire nella valutazione delle prove orali, non dice in quale rapporto essi stanno con i criteri adoperati per lo scritto, è evidente che non fornisce un segno efficace di trasparenza<sup>15</sup>. Tanto meno manifesta segni comprensibili di coerenza.

---

<sup>14</sup> La giurisprudenza, a questo riguardo, ha ripetutamente censurato qualsiasi procedura valutativa non supportata da esplicite ragioni che consentano di giustificare tanto i punteggi quanto l'aggettivazione utilizzata per esprimere i punteggi che tale aggettivazione ha determinato (cfr., *ex multis*, Cons. Stato, sez. IV, 28 dicembre 1993, n. 1148 e 16 maggio 1994, n. 408). Quel che non sempre appare chiaro è il fatto che le "ragioni" non si possono identificare con i criteri, perché in qualche modo esse vengono prima degli stessi criteri e sono dati dal piano delle *intenzioni* ossia dai *fini*.

<sup>15</sup> Tornerebbe utile, in questo caso, ricordare ed applicare, anche al nostro caso, ciò che affermava Vittorio Masoni nel 1997, quando distingueva la *trasparenza interna* (che comporta l'onere della comunicazione e della documentazione) dalla *trasparenza esterna*



## Nicola Paparella, Pubblicazioni on-line

Esplicitare i fini è sempre un'operazione a doppia valenza. Serve, *ad extra*, a rendere manifesto un certo percorso e serve, *ab intra*, a trovare le linee giustificative, i tratti di collegamento, i nessi di congruità logica...

In questa prospettiva, gioverà capire quali altri possibili fini possono agire, al di là della *neutralità* e della *trasparenza*, in un lavoro di esame dei lavori concorsuali.

Noi pensiamo che non sarebbe inopportuno, in una prova di concorso, invocare l'*equità*, l'intenzione di *discriminare* l'apprezzamento di ciascun elaborato in maniera da poter costruire una graduatoria, l'*imparzialità*, il *buon andamento* della procedura, e così via. Oltre tutto si tratta di "valori" verso i quali orienta anche la norma positiva e per i quali si potrebbe qui ricordare una ricca serie di interventi giurisprudenziali.

Non si tratta di richiami retorici o ridondanti, ma di precondizioni destinate ad assicurare e a fondare operazioni che possono essere complesse e che potrebbero nascondere implicazioni implicite o anche valenze non dichiarate.

L'intento di *discriminare le performance* concorsuali, ad esempio, può giustificare alcune prestazioni e può indurre all'esclusione di altre; può agire come motivo di premura che permette di distribuire i risultati su una graduatoria ben articolata o invece può condurre a risultati addensati attorno a punteggi che per essere espressivi dovranno attendere l'integrazione (esterna) dei criteri di precedenza e di preferenza.

Agire in un modo o nell'altro non può essere casuale, e va deciso con scelte opportunamente esplicitate e motivate.

A nostro parere in un concorso di selezione per dirigenti scolastici, come quello appena concluso nelle regioni italiane<sup>16</sup>, le finalità potevano esser utilmente suggerite dallo stesso Ministero, direttamente o attraverso l'Invalsi, e comunque avrebbero potuto richiamare almeno quattro elementi: *Trasparenza*, *imparzialità*, *coerenza* fra le fasi del processo, *buon andamento* delle procedure. Ma così non è stato. Le commissioni hanno lavorato in maniera del tutto diversa, il che è abbastanza comprensibile, ed hanno quasi sempre trascurato il passaggio sui fini o lo hanno confinato nel limbo delle buone intenzioni, senza ricavarne quella fitta trama di conseguenze operative che sarebbe stato lecito attendersi.

Strettamente connesso al discorso dei fini c'è l'esigenza di restituire alle parole la loro autenticità semantica.

---

(che comporta misure per la facilità di accesso alle fonti e di *intelligibilità* delle informazioni, oltre che di tempestività e di esaustività dei messaggi). Cfr. V. MASONI, *M&V. Monitoraggio e valutazione dei progetti nelle organizzazioni pubbliche e private*, Angeli, Milano, 1997.

<sup>16</sup> Assumiamo alcune vicende, legate al concorso citato, come caso di studio; non per giudicare o ... *valutare*, che non sarebbe questo né il luogo né la forma, ma per cercare di capire e soprattutto, in termini proattivi, per mostrare il tipo di percorso che converrebbe compiere per assicurare una responsabile fruizione di quanto la docimologia illustra e chiarisce.



Su questo sfondo, appena qualche anno fa, considerando anche le attese e, principalmente, le speranze aperte dal web semantico, c'è chi ha avviato la grande impresa di una ontologia, per “*constituire l'oggetto formale* di riferimento, per dirne l'*esistenza*, per predicarne la consistenza, per stimarne le possibili valenze”<sup>17</sup>, per organizzare il lessico e le relazioni semantiche<sup>18</sup> che si stabiliscono all'interno della vasta area coperta dalla valutazione educativa e didattica.

### I criteri, le procedure, gli strumenti

In una procedura valutativa di tipo concorsuale, affidarsi a parole come *Coerenza*, *Efficacia* teorica, *Fondatezza*, *Giustificazione* (delle conclusioni), *Rilevanza*, *Sicurezza*, ... può essere sempre utile.

Ciascuna di queste parole può utilmente contrassegnare un criterio. Si tratta di capire se vengono adoperate nel loro preciso significato e se mantengono ed esprimono una loro chiara correlazione con i fini.

I criteri vanno sempre individuati in stretta correlazione con le intenzioni esplicite e manifeste (il *finis operantis*); è dai fini, infatti, che è possibile rilevare quel fascio di *proprietà* da assumere come costitutive dell'oggetto da esaminare. Di solito, tuttavia, si arriva ai criteri senza passare da una considerazione delle intenzioni che, ad esempio, in una situazione concorsuale consentirebbe di cogliere subito quegli elementi qualitativi (proprietà) che possono essere assunti come espressivi della buona prova o almeno quell'insieme di caratteri da poter associare inequivocabilmente ad altrettante proprietà della prova.

Quelle proprietà o quei caratteri associabili ad alcune proprietà sono i criteri.

E ciascun criterio va proposto e formulato in maniera da...

- a) esprimere *una soltanto* delle proprietà che si vogliono prendere in considerazione,
- b) fare esplicito riferimento al fine (*operantis*),
- c) trovare chiara espressività operativa nel prodotto che si va ad esaminare (ad esempio, gli elaborati d'esame).

Il criterio, in altri termini, deve far da ponte fra il piano dei valori e degli aspetti teorici (e quindi dei fini da perseguire) e il piano fenomenologico, della cosa da valutare.

Operando in questo modo, oltre tutto, diventa abbastanza agevole estendere i criteri anche alla prova orale o almeno trovare un fascio di proprietà (criteri) coerenti fra loro e comunque tali da attraversare il medesimo universo di

---

<sup>17</sup> N. PAPARELLA, *Per costruire una ontologia sul tema della programmazione*, in N. PAPARELLA (a cura di), *Ontologie, simulazioni, competenze*, Amaltea, Melpignano, 2007, p. 102.

<sup>18</sup> Cfr. L. GALLIANI (a cura di), *Web ontology della valutazione educativa*, Pensa Multimedia, Lecce, 2009.





## Nicola Paparella, Pubblicazioni on-line

riferimento (le competenze dei candidati) da tenere sotto controllo sia attraverso le prove scritte che attraverso le prove orali.

A fronte di quanto andiamo dicendo, utilizzeremo come esempi da discutere, alcuni dati che emergono da verbali ufficiali collegati alle prove relative al concorso appena ricordato e destinato alla selezione di un certo numero di dirigenti scolastici. Come abbiamo già detto<sup>19</sup>, non intendiamo, evidentemente, discutere l'operato di una o di più commissioni, vogliamo soltanto trarre spunto da episodi effettivi, per mostrare come sia possibile incappare in alcuni errori (sempre in agguato quando non si manipolano con cura alcuni strumenti) e come sarebbe stato possibile evitarli. In questi casi sarebbe stato forse opportuno garantire alle commissioni un supporto tecnico specialistico da parte dell'Invalsi che probabilmente il Ministero avrebbe potuto sollecitare.

Nel lavoro effettivo delle commissioni, la mancata esplicitazione dei fini non ha permesso di ipotizzare criteri in qualche modo fra loro coerenti.

Come si fa a dire, ad esempio, che si assumono come *criteri* di giudizio, “a) *La padronanza dei temi affrontati*; b) *La qualità dell'articolazione del testo*; c) *La qualità della forma espressiva...* “? Verrebbe da chiedersi: perché proprio 3 criteri e perché proprio quei tre e non altri?

Fermiamoci a considerare almeno uno degli aspetti possibili. Se, ad esempio, si vogliono risultati ben distribuiti in una graduatoria convenientemente articolata (e quindi se si vuole *discriminare* il valore di ciascun elaborato in termini di *imparzialità* e *neutralità*, assicurando, nel contempo, il *buon andamento della procedura*), ci si dovrebbe orientare verso un più largo ventaglio di criteri, da associare a punteggi grezzi da tradurre successivamente in trentesimi. Si è scelto invece un numero ristretto di criteri.

D'altro canto, che potessero non essere soltanto tre, i criteri da adottare, lo crede persino la commissione che li propone, tant'è che la medesima commissione definisce il primo criterio con *due* proposizioni fra loro non sovrapponibili, scrivendo a verbale che il primo criterio è dato da: “Padronanza dei temi affrontati. Ampiezza delle conoscenze possedute”. Dove la seconda proposizione non è esplicativa della prima e neppure integrativa, ma è qualcosa che semplicemente si aggiunge a quanto afferma la prima proposizione. Oltre tutto, se si sta agli aspetti semantici, non è difficile accorgersi che si può essere “padrone dei temi” senza manifestare “ampiezza di conoscenze” e viceversa si può manifestare ampiezza di conoscenze senza segnalare spiccati livelli di padronanza. Quindi si tratta di *due* questioni distinte, di *due proprietà*, di due aspetti e perciò di due criteri, non di un solo criterio. Due proprietà comunque interessanti e importanti, ma due, non una soltanto.

Certamente nel contesto di riferimento – un concorso per la selezione di dirigenti scolastici – non potendo disporre di quelli che in letteratura si dicono *performance*

---

<sup>19</sup> Più innanzi alla n. 16.



*criteria*<sup>20</sup>, perché mancano standard di riferimento, occorre necessariamente rifarsi a quegli aspetti, quei tratti, quelle dimensioni, quelle caratteristiche che possono essere assunte come “elementi importanti”, a condizione, tuttavia, che qualcuno dica quali sono e perché sono da ritenersi importanti. Operazione sicuramente agevole e logicamente credibile se collegata in termini deduttivi alla esplicitazione formale dei fini che sicuramente avrebbe potuto dare consistenza, giustificazione e motivazione ai criteri proposti.

Capita invece che si giunga in maniera inopinata alla definizione dei criteri. Capita anzi, che per conferire al criterio maggiore enfasi (forse soltanto estetica) si perda di vista anche il piano fenomenologico, sganciando, di fatto, il criterio tanto dal quadro dei fini o delle questioni teoriche quanto dal riferimento al contesto fattuale. Se, ad esempio, si parla di *qualità della forma espressiva*, o anche di *eleganza formale* del testo, si allude, evidentemente, non ad un aspetto isolato, ma ad un fascio di proprietà che andrebbero esaminate distintamente per poi riaggregarle, secondo una procedura di ponderazione opportunamente esplicitata. Se così non è, e questo accade tutte le volte in cui il criterio condensa – in maniera più o meno implicita – una pluralità di “elementi importanti”, non soltanto resta nebuloso l’itinerario che ha condotto alla elaborazione dei criteri, ma si hanno pure evidenti conseguenze a livello di ponderazione ossia nell’esercizio effettivo delle “misurazioni” richieste dalle schede di cui di seguito diremo.

Insomma, i criteri debbono potersi riferire ad una sola delle qualità, e debbono poter essere formulati in maniera da non lasciar dubbio alcuno sugli aspetti fenomenologici verso i quali saranno essi orientati.

Ai criteri vanno poi agganciati degli indicatori capaci di dar conto della configurazione effettiva dell’oggetto d’esame (il tema, ad esempio).

## La ponderazione dei criteri

La *ponderazione* è operazione complessa e delicata. Si tratta di associare a ciascun criterio una quota dei punteggi di cui dispone la commissione, in un assetto complessivo coerente, giustificato e immediatamente percepibile dall’esterno (la ... *trasparenza*).

Dal punto di vista procedurale occorre innanzi tutto prendere atto di quante siano le prove previste dalla normativa, dei vincoli eventualmente posti dal bando in ordine alla distribuzione dei punteggi fra le diverse prove, e di *pesare*, conseguentemente, e in maniera coerente i diversi criteri che, in qualche modo, vanno a riguardare il medesimo fascio di competenze e che proprio per questo debbono avere pesi contestualmente giustificati ed equamente distribuiti. Il tutto attraverso un reticolo di argomentazioni che possano essere facilmente condivise e rappresentate in una apposita griglia di ponderazione<sup>21</sup>.

<sup>20</sup> Cfr. J. A. ARTER, J. MCTIGHE, *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*, Corvin Press, Thousand Oaks, (CA), 2001.

<sup>21</sup> Cfr. E. DEL PIANTO, *Assessment center. Tecniche e strumenti per il valutatore*, Angeli, Milano, 2004, pp. 52 e ss.



Se, ad esempio, si dispone di due prove scritte e se si decide (sulla base di intenzioni/fini convenientemente espressi) di assegnare un pari punteggio globale a ciascuna delle due prove, buon senso suggerirebbe di organizzare il set valutativo su un pari numero di criteri, fra le due prove, altrimenti occorrerebbe spiegare e motivare la distribuzione dei pesi. Questo significa che nella ponderazione, occorre tener conto di ciascun criterio in relazione a ciascuno degli altri criteri, della stessa e di altre prove, operando in maniera tale da poter offrire una logica spiegazione alla griglia complessiva della distribuzione dei punteggi.

Nel caso in esame non si ha un quadro complessivo della distribuzione dei pesi, sino al punto di suscitare diversi motivi di perplessità.

La commissione, infatti, dopo aver stabilito e verbalizzato che il criterio 3 della prima prova – indicato come “Qualità della forma espositiva” – corrisponde al criterio 2 della seconda prova (ugualmente indicato come “Qualità della forma espositiva”), inopinatamente attribuisce a questo criterio un peso pari a 8/30 nella prima prova e a 6/30 nella seconda, introducendo una misura di relativizzazione che non si giustifica in procedure di questo tipo. Oltretutto, anche a livello intuitivo, se si tratta di pesare la forma espositiva (criterio 3 della prima prova e criterio 2 della seconda) con indicatori del tutto coincidenti (anzi, dichiaratamente ed intenzionalmente eguali) e per elaborati cui vengono assegnati 30 punti ciascuno, il peso relativo non può che essere eguale. E invece così non è, per la commissione.

Se poi dovessimo pensare ad un peso diverso in ragione di una considerazione distributiva, dovremmo persino ribaltare lo squilibrio, perché si dà il caso che il criterio che pesa meno è collocato nel set valutativo della seconda prova che vanta in totale due criteri soltanto, mentre invece il criterio che pesa di più è collocato nel set che enuclea tre criteri.

Determinazione dei fini, individuazione dei criteri e loro ponderazione sono in ogni caso operazioni da tenere strettamente correlate – distinte, certamente, ma correlate - perché l’una fornisce motivazione all’altra, e da disporre in un assetto di rigorosa coerenza logica per non compromettere la credibilità complessiva del lavoro di valutazione.

## Gli indicatori

Per la natura che gli è propria, il criterio si proietta verso il dato fattuale, ma ne resta al di sopra, per così dire. Ciò che invece seziona e scandisce l’oggetto è l’*indicatore* che in qualche modo ci restituisce una sezione dell’oggetto medesimo, tralasciata attraverso il criterio.

Posto dunque un criterio, l’indicatore è “il livello osservabile di un concetto indagato”<sup>22</sup>.

---

<sup>22</sup> C. BEZZI, *La linea d'ombra. Problemi e soluzioni di ricerca sociale e valutativa*, Angeli – Aiv, Milano, 2011, p 87.



Ne consegue che le principali caratteristiche dell'*indicatore* sono quattro:

- a) La sua formulazione in *termini chiaramente osservabili*;
- b) la sua capacità di segnalare esattamente e rendere chiaramente evidente la proprietà, ovvero il criterio cui esso si riferisce;
- c) la comprensibilità, in maniera che *possa essere univocamente inteso* ed espresso da ciascuno dei componenti della commissione;
- d) *l'assenza di ogni possibile ambiguità*.

In altri termini, ci si attende che attraverso gli indicatori il criterio possa perdere qualsiasi residuo di generalità, per assumere aspetti di specificità.

Se tuttavia i valutatori individuano ed esprimono i criteri in maniera soltanto sommaria, senza quel solido duplice appoggio di cui si diceva - sul quadro teorico di riferimento, da una parte, e sul contesto oggetto di indagine valutativa dall'altra - è evidente che il rischio di produrre indicatori inadeguati è altissimo, o perché ci si affida ad espressioni del tutto incapaci di raggiungere il livello osservabile dei concetti oppure perché ci si aggrappa ad indicatori del tutto scollegati con i criteri.

Nel caso di studio sono stati proposti indicatori semanticamente complessi, capaci più di commentare che di esplicitare i criteri.

È così che sono nati pseudo indicatori, espressi come: “pertinenza del contenuto alla traccia”; “completezza della conoscenza...”, “interpretazione critica (ragionata) dei temi sviluppati” ecc., dove ciascuna di queste voci poteva essere un criterio e viene invece proposta come indicatore.

Ci basti esemplificare sulla nozione di *pertinenza*. Prendiamo in prestito, parafrasandola, una bozza di dialogo che si può facilmente rintracciare in rete, per dimostrare come la pertinenza possa configurarsi in maniera non univoca:

Prima configurazione: Domanda: “Sai l'ora?” Risposta: “Sì”.

Seconda configurazione: Domanda: “Sai l'ora?” Risposta: “Sono le 20.15”.

Terza configurazione: Domanda: “Sai l'ora?” Risposta: “Manca ancora mezzora per la partita”.

Le tre risposte sono tutte e tre giuste e tutte e tre pertinenti. Ma che cosa voleva sapere colui che ha posto la domanda? Nell'analisi valutativa si potrebbe aver interesse a distinguere le tre modalità di elaborazione del messaggio e si potrebbe volerle valutare diversamente. Ovviamente occorrerebbe dirlo.

Il fatto, comunque, che la *pertinenza* possa essere espressa in modi diversi, come del resto ci attesta la bibliografia del settore<sup>23</sup>, significa che *la sola dichiarazione di pertinenza non basta a soddisfare le condizioni poste per gli indicatori*, ossia che essi possano rappresentare il livello osservabile del concetto, che possano segnalare

---

<sup>23</sup> Cfr. D. SPERBER, D. WILSON, *La pertinenza*, tr. it. [sulla I ed. del 1986], Anabasi, Milano, 1993. Si veda pure: F. FERRETTI, D. GAMBARARA (a cura di), *Comunicazione e scienza cognitiva*, Laterza, Bari, 2005.



esattamente il criterio cui si riferiscono, e in maniera univoca, comprensibile e senza ambiguità.

La stessa cosa si può dire per altri presunti indicatori che troviamo nel caso di studio.

Quando, ad esempio, si parla di “originalità dello sviluppo argomentativo”, in riferimento al criterio “qualità dell’articolazione del testo e delle relative argomentazioni” sembra davvero trovarsi in presenza di un commento piuttosto che nella proiezione osservabile della proprietà traguadata dal criterio. E, in ogni caso, resterebbe almeno da chiarire che cosa si voglia intendere per originalità. Si può pensare all’assenza di plagio, si può pensare alla formulazione di problemi nuovi o alla enucleazione di soluzioni convincenti per problemi nuovi o alla proposta di nuove soluzioni per problemi già noti, ... e tutto questo può essere pensato e detto in riferimento ad uno solo dei passaggi nei quali si articola lo svolgimento del tema da esaminare o invece in riferimento a ciascuno degli snodi argomentativi...Si può anche pensare a tutte queste cose, insieme, che però sarebbero prima da esplicitare e poi da disporre in maniera tassonomica, allo scopo di rilevarne punteggi differenziati e comunque tali da giustificare la ponderazione che la Commissione riporta sulla sua *scheda*.

Un ulteriore esempio può essere dato dall’indicatore espresso come “Chiarezza delle argomentazioni. Efficacia ed eleganza formale del testo”, da riferirsi al criterio: “Qualità della forma espositiva”. In questo caso la funzione di commento dell’indicatore è ancor più evidente, perché a fronte di un criterio espresso con una proposizione semplice, l’indicatore adopera due proposizioni lasciando nel dubbio il valutatore che non sa più se debba tener conto di quanto si dice nella prima proposizione (la chiarezza delle argomentazioni) ovvero nella seconda (l’efficacia e l’eleganza formale). Il fatto che si segnalino due proposizioni non riduce, ma aumenta la possibilità di equivocazione. Se poi si considera che la seconda proposizione fa riferimento a due (e non ad una sola) qualità, diventa evidente la complicazione che espone il valutatore all’onere di chi deve considerare, contestualmente, ben tre diverse qualità: la chiarezza, l’efficacia e l’eleganza. Con esiti del tutto problematici, anche perché una espressione verbale può essere chiara, ma non elegante; può essere elegante ed anche chiara, ma non efficace ... E può anche capitare che nello stesso tema ci possano essere (anzi, sicuramente ci sono) non uno, ma diversi passaggi in cui può applicarsi questo tipo di rilevazione. Ed allora che cosa si farà? Si procederà ad una semplice enumerazione delle ricorrenze? E come si passerà dalla conta degli episodi rilevati ai punteggi assegnati dalla scheda? Non si capisce e in verbale nessuno lo spiega.

### La ponderazione e le scale

In un concorso è difficile sfuggire alla misura, considerato che l’intera procedura deve potersi concludere con una graduatoria. Per questa ragione la commissione ha voluto munirsi di apposite *scale ad intervalli*. Poteva optare per altri strumenti, ma quello scelto va sicuramente bene.



La scala ad intervalli riesce a dar conto della distanza che separa i quadri comportamentali osservati; si basa sulla *eguaglianza* delle differenze fra i diversi suoi “intervalli” e non ha uno zero assoluto. La scala ad intervalli nasce da una misura e consente di misurare (ponderare, differenziare) gli oggetti osservati. L’operazione iniziale, fondamentale, da compiere è quella di assumere, nell’ambito della proprietà in esame, e quindi per ciascun indicatore, un quadro comportamentale, un aspetto, una condizione, una ricorrenza, qualcosa che possa connotarsi come unità di misura e quindi come distanza sempre eguale fra ogni coppia di osservazioni (ad esempio: la distanza che separa, in un termometro, ciascuna tacca da quella successiva, è perfettamente coincidente con la distanza che separa qualsiasi altra coppia di marcature consecutive, rilevabili sullo stesso strumento<sup>24</sup>).

Questa condizione permetterebbe di dire, per ciascun indicatore e per tutti i criteri osservati, se il prodotto in esame (il tema, nel nostro caso) meriti un punteggio maggiore di quello assegnato ad un altro analogo prodotto e, soprattutto, di precisare di quante “unità di misura” si collochi più su o più giù, dando comunque per scontato che la proprietà considerata non è mai totalmente assente (manca lo zero assoluto).

Dal momento che le scale adoperate dalla commissione sono da essa medesima costruite, è necessario allora verificare che le condizioni poste siano soddisfatte. E purtroppo nel caso in esame non è così.

D’altro canto, se gli indicatori non rinviano a quadri chiaramente osservabili e se non sono nemmeno accompagnati da alcun *descrittore di livello*, è ovvio che non si riesca a scorgere alcuna “unità di misura”. Che cosa consente, ad esempio, nella fruizione dell’indicatore “compiutezza della conoscenza”, di passare dal primo al secondo gradino della scala? E come possiamo dire che l’intervallo fra il primo e il secondo gradino è perfettamente eguale all’intervallo esistente, poniamo, fra il terzo e il quarto? Dalle scale non si rileva, dai verbali della commissione non risulta, dall’analisi semantica non lo si intuisce.

Nel caso in esame, perciò, non c’è nulla che possa esprimere l’*unità di misura*, conseguentemente non c’è nulla che possa giustificare la graduazione degli indicatori e quindi non c’è nulla che possa giustificare l’associazione degli indicatori ai punteggi che nella scala vengono previsti.

Sui punteggi occorre una ulteriore osservazione. Qualcuno potrebbe dire che sono proprio essi che garantiscono la graduabilità e quindi la condizione della misura. Sarebbe come dire che l’etichetta sostituisce il prodotto etichettato. L’etichetta “dice” il prodotto, ma non è il prodotto. Analogamente: deve essere il descrittore di livello a marcare e a contrassegnare una posizione e con ciò concorrere a costruire la scala. Etichettare con un certo punteggio ciascun

---

<sup>24</sup> Sui problemi della misura si veda M. GATTULLO, *op. cit.*, pp. 34-36. Sulle difficoltà connesse alla misurazione degli aspetti qualitativi della esperienza umana si veda: N. PAPARELLA, *Un reattivo per la misurazione dell’efficacia dell’educazione religiosa nella scuola materna*, in B. VERTECCHI (a cura di), *Per una nuova qualità della scuola*, Tecnodid, Roma 1996, pp. 268 e ss.



gradino della scala è un'operazione successiva al lavoro di costruzione della scala medesima, perché ciò che crea l'intervallo è la distanza fenomenologica fra i fatti espressi dal descrittore di livello; i punteggi servono soltanto a rendere evidente e confrontabile la serie degli intervalli. Oltre tutto, se alla base del discorso ci fosse il punteggio, noi avremmo una *scala di conto* e non una *scala ad intervalli* e l'operazione richiesta al valutatore non sarebbe quella di confrontare il tema con un indicatore, attraverso i suoi descrittori di livello, ma più semplicemente quello di enumerare, di contare determinate occorrenze<sup>25</sup>. Anche questo si poteva fare, evidentemente, ma non è in questa direzione che si è mosso il team dei valutatori.

Nel caso in esame la situazione si fa pesante: man mano che si passa dalle questioni generali, ai criteri, dai criteri agli indicatori, e da questi alla costruzione delle scale e alla stessa attribuzione dei punteggi, si rende sempre più evidente la fragilità delle cosiddette *scale* messe a punto dalla commissione, per ciascuna delle quali le perplessità sono tali da far dubitare che si possano davvero definire *scale ad intervalli*. Anzi, i punti di criticità sono tanti e tali da indurre ad escludere che esse possano in qualche modo giustificare e motivare l'operazione di attribuzione di punteggio, richiesta dalla valutazione delle prove concorsuali.

Potrebbero, in astratto, servire come strumenti che orientano l'attenzione dei commissari, una sorta di inconsueto congegno concettuale capace di funzionare come pro-memoria, come una specie di lista delle cose salienti da osservare, una sorta di *checklist*, realizzata anch'essa al di fuori dei canoni della docimologia, e però funzionale allo scopo. Ma non è così. Anzi, se le schede venissero adoperate soltanto come lista per riscontrare le cosiddette occorrenze, il margine di incertezza del setting valutativo aumenterebbe. Infatti, ciò che si rilevarebbe non sarebbe il "valore" della qualità o della proprietà presente nel tema, ma la proprietà medesima, di cui resterebbe da dare o una enumerazione delle presenze o una relazione con altre proprietà...

Lo si può capire agevolmente se l'esame delle schede (o *scale*, secondo le intenzioni della commissione) viene integrato, come è giusto che sia, da quanto è riportato nel verbale della seduta durante la quale la commissione ha costruito le schede in questione.

Per l'indicatore 1.a (in scheda: "pertinenza del contenuto alla traccia"), si legge: "In merito alle richieste della traccia, le dimensioni pedagogiche devono fare riferimento: 1) alle finalità educative della scuola, quale servizio alla persona di ciascun studente; 2) le dimensioni istituzionali devono fare riferimento alla scuola come istituzione che interagisce con le altre istituzioni della società (...); 3) le dimensioni organizzative devono fare riferimento alla leadership del dirigente". Il tema – è appena il caso di precisarlo – parlando del miglioramento della qualità del servizio pubblico d'istruzione, chiedeva che il candidato evidenziasse le dimensioni pedagogiche, istituzionali, organizzative ecc. – E il verbale, presentando gli indicatori, li esplicitava con frasi che non sono esattamente dei descrittori di livello e neppure un commento, ma una sorta di impropria *checklist*.

---

<sup>25</sup> Cfr. in proposito: A. MARRADI, *Unità di misura ed unità di conto*, Rassegna Italiana di Sociologia, XXVI, 2, 1985



Percezione che viene confermata quando nello stesso verbale si legge: “l’indicatore 1.b richiede i riferimenti normativi” e poi ancora: “l’indicatore 1.c richiede il preciso riferimento all’Invalsi [citato dalla traccia] e alle azioni che il dirigente propone ecc.”

Insomma, se l’indicatore 1.a chiede la *pertinenza del contenuto alla traccia*, quel che si dovrebbe fornire è il sistema di misura di codesta pertinenza, ovvero il dato da riscontrare, in forza del quale si può assegnare l’uno o l’altro dei punteggi previsti (1, 2, 3, 4).

Mancando i descrittori di livello è davvero difficile individuare la misura, in compenso il valutatore trova, in verbale, una serie di questioni ritenute importanti. E, allora, che cosa farà il valutatore? Spunterà su un foglietto personale se ricorrono nel tema le questioni che il verbale riporta? Si potrebbe anche fare, ma come si potrebbe giustificare il punteggio 2, ad esempio, e come lo si discriminerebbe dalla situazione meritevole del punteggio 3? E come si fa a garantire, per questa via, che i punteggi assegnati provengano da una *scala*, dove la distanza fra un gradino e l’altro deve essere garantita come eguale? E poi, se il medesimo indicatore rinvia, a giudizio della Commissione, ad una pluralità di aspetti salienti, come si fa a contenere questa pluralità, composita e disomogenea, entro i quattro gradini della scala?

Insomma, la mancanza di chiari descrittori di livello conferisce precarietà alle *tabelle* prodotte dalla Commissione che, evidentemente, non possono definirsi *scale* e non possono neppure essere adoperate come checklist

Questa circostanza richiede un chiarimento aggiuntivo.

Assumiamo, a mo’ d’esempio, quanto la Commissione riferisce a proposito del primo criterio previsto per la seconda prova, dove si dichiara doversi valutare *due cose (e non una soltanto)*: la “capacità di analisi del caso e di proposta di soluzioni pertinenti” (Criterio n. 1). Si forniscono 3 indicatori:

- 1) Individuazione degli elementi di contesto interni ed esterni in relazione alla trattazione del caso. Punteggi 2 / 4 / 6 / 8
- 2) Completezza dei riferimenti normativi afferenti al Caso. Punteggi 2 / 4 / 6 / 8.
- 3) Capacità di delineare soluzioni coerenti e funzionali alla soluzione del caso. Punteggi 2 / 4 / 6 / 8.

Nella *scala* (o nel verbale di cui la scala è parte integrante) non si dice a quali condizioni si dovrà preferire un punteggio in luogo di un altro, perché, come si è detto, *non si forniscono descrittori* di livello (né in forma esplicita né in maniera implicita), né di essi si trova traccia in alcun verbale.

Questo sottrae efficacia descrittiva e capacità misurativa alle cosiddette scale e non le fa essere funzionali al compito della imparzialità, potendosi sempre ipotizzare che a valutatori diversi o a momenti diversi del medesimo processo di valutazione possano non corrispondere medesime percezioni dei livelli indicati nella scala.





C'è di più. Il fatto che ciascun indicatore aggregi più nuclei semantici (esempio, nel terzo indicatore appena citato, troviamo la capacità di *delineare* soluzioni, la capacità di fornire *soluzioni coerenti*, la capacità di prospettare *soluzioni funzionali* al caso) potrebbe indurre a credere che i descrittori di livello possano essere impliciti alla stessa formulazione dell'indicatore, quasi a voler dire che il livello possa esser dato dalla presenza di uno o di più di uno dei nuclei semantici citati. Ma, nella fattispecie, si tratterebbe di ipotesi poco credibile, stante il fatto che i livelli così implicitamente richiamabili sarebbero soltanto tre e non quattro (come esplicitamente richiesto dalla scala in questione) e, oltre tutto, resterebbe pur sempre l'onere di stabilire una gerarchia (e quindi l'origine del diverso peso valutativo) fra i tre aspetti indicati, rinforzando, in ogni caso, l'idea che ci si trovi dinanzi a qualcosa che non assomiglia a ciò che viene dichiarato *scala ad intervalli*.

### Per una valutazione sensibile

In docimologia si definisce *sensibilità* di uno strumento di valutazione la capacità di rilevare le differenze e di discriminare le diverse configurazioni fenomenologiche.

Questo viene richiesto a tutti gli strumenti di valutazione e principalmente a delle *scale*, che di per sé possono garantire il requisito della sensibilità e quindi la capacità di discriminare le differenze fra gli "oggetti" da valutare. Purtroppo, per le ragioni già riferite, le tabelle proposte dalla commissione non sono delle scale.

Resta però legittimo domandarsi se, non ostante i limiti, gli strumenti adottati possano comunque offrire un supporto valutativo capace di discriminare le differenze, in una prospettiva di trasparenza ed equità.

La risposta è negativa.

C'è nelle tabelle adoperate qualche difficoltà intrinseca di cui forse la medesima commissione si avvede, ponendovi rimedio in maniera empirica, con una procedura immotivata, inefficace e soprattutto inadeguata alla soluzione del problema. Ci riferiamo alla utilizzazione di punteggi non riportati nelle tabelle cosiddette scale.

Attribuendo i punteggi per il primo dei due criteri previsti per la seconda prova, ad esempio, pur utilizzando una "scala a quattro intervalli", la commissione si avvale di sette punteggi, perché adopera anche i punteggi intermedi rispetto a quelli previsti, trasformando di fatto la scala in questione in una scala a 7 intervalli. E se la scala (o tabella) prevedeva soltanto punteggi espressi con numeri pari, la commissione adopera anche i numeri dispari, presumibilmente per dosare meglio il punteggio fra i livelli espressi nella scala, dove però non v'è traccia di graduazione né è possibile rilevare (o anche soltanto intuire) indicatori intermedi fra quelli esplicitamente espressi, né si può affidare ai descrittori del tutto assenti.

Per intendersi, anche nella scala con la quale misuriamo le temperature, è possibile dire che un bambino ha una temperatura di 37° ed un altro ha 37,5°. Ma nel termometro è a tutti chiaro che il mezzo grado di temperatura è dato da



una posizione intermedia fra quelle segnate sul marcatore. Nel caso della “scala” di cui qui ci occupiamo non sono affatto chiari i descrittori di livello, anzi, mancano del tutto, figuriamoci se è possibile individuare le posizioni intermedie<sup>26</sup>.

Le posizioni intermedie, perciò, non soltanto non fanno riferimento ad espliciti descrittori di livello, ma non possono neppure vantare una loro presenza ufficiale nella scala (o tabella che sia), motivo per cui si può persino sostenere che tutte le valutazioni espresse, per questo criterio e per questa prova, con numeri dispari non trovano nella scala della commissione alcuna adeguata motivazione.

### La gestione dei punteggi

Per il buon andamento di una procedura di valutazione giova stabilire un legame logico, chiaro ed evidente, fra tabella di attribuzione dei punteggi, punteggi analitici, indicatori, criteri e ragioni complessive che motivano il processo valutativo. Questo legame nel caso in esame non si evince.

Non si capisce come si arrivi alla misura (meglio, alla stima) finale dei 30 punti, partendo da criteri pesati in maniera non omogenea, con indicatori di complessità semantica non omogenea e con punteggi distribuiti senza il supporto di indicatori di livello.

La controprova di quanto si afferma è data dal fatto che non si riesce a ricostruire, dall'esterno, l'iter del giudizio valutativo, come invece si richiederebbe dalla saggezza docimologica e dalla prassi giurisprudenziale, né si riesce a trovare una spiegazione logica alla imprevedibile distribuzione complessiva dei punteggi, come risulta dall'esame dei risultati complessivi del lavoro di valutazione.

L'anomalo addensarsi di punti – che poi diventano voti – attorno a certi valori non indicano misure di tendenza centrale, ma l'esito, forse persino scontato, di operazioni che di fatto vanno ad azzerare le differenze.

In un concorso è preciso interesse dell'amministrazione rilevare la differenza di performance allo scopo di graduare i candidati e poterli distinguere in ragione dei risultati accertati; se però si adoperano strumenti inadeguati, è sempre possibile il rischio di ottenere il contrario, ossia una graduatoria che assomiglia ad una sorta di massa grigia ed informe, all'interno della quale le differenze sono state di fatto azzerate. Più o meno lo stesso risultato che si otterrebbe quando, dovendo stimare alcune distanze, non si adopera la rotella metrica e ci si affida al computo dei passi che separano alcuni vertici. L'imprevedibilità dei risultati è in questi casi un rischio molto elevato.

Più in generale, ogni volta che ci si trova nella condizione di dover valutare e con questo graduare dei prodotti e quindi delle persone, giova ricordare che la

---

<sup>26</sup> La presenza esplicita di descrittori di livello è comunque condizione necessaria, ma non sufficiente per consentire l'individuazione di posizioni intermedie.



costruzione di una *griglia di ponderazione*<sup>27</sup> è compito da affrontare subito insieme e congiuntamente al processo di elaborazione che conduce alla individuazione dei criteri. Insieme ai criteri vengono selezionati gli indicatori e poi i descrittori e quindi la griglia di ponderazione. Si tratta di passaggi che debbono trovare coerenza e linearità all'interno di un solo processo dal quale deve poter emergere un tessuto di giustificazioni non già per le misure, di cui abbiamo già detto, ma per i pesi medesimi, in una logica di sistema, in forza della quale tutto viene raccolto e finalizzato al risultato complessivo.

Il fatto, ad esempio, che ad un criterio vengano assegnati, poniamo, 8 punti e che poi questi 8 punti debbano andare a concorrere ad una valutazione espressa in trentesimi, è qualcosa che senza opportune motivazioni rischia di apparire arbitraria. La griglia di ponderazione potrebbe essere una buona strategia per sfuggire a questo rischio.

## Contestualizzazione

Il linguaggio della valutazione ha creato una certa confidenza con alcune parole che rinviano a fatti sicuramente apprezzabili, almeno nel loro ambito; se però vengono adoperate al di fuori dei collegamenti e dei rinvii che sono propri della struttura semantica di provenienza, diventano soltanto degli inciampi.

Una di queste parole è *contestualizzazione*.

Nel caso di studio al quale stiamo riferendoci, l'idea di contestualizzazione è stata invocata a giustificazione di una operazione che, almeno nelle intenzioni, apparirebbe condivisibile. I valutatori hanno deliberato di procedere, nei primi giorni, ad una valutazione provvisoria, per ottenere una sorta di sintonia interna al team, riservandosi, ovviamente, di tornare subito dopo ad una nuova lettura degli elaborati d'esame e quindi alla loro valutazione definitiva.

Al di là delle intenzioni, questa procedura, ancorché fragile sotto il profilo giuridico, non giova all'obiettivo o almeno non giova di più di quanto si potrebbe ottenere dalla lettura collegiale e dalla discussione collegiale, mediata dagli strumenti di analisi adottati.

Per contro l'obiettivo della contestualizzazione poteva essere più utilmente perseguito, assumendo la contestualizzazione fra gli indicatori e i descrittori, per capire in quale misura l'elaborato manifestasse un certo aggancio a situazioni percepite nella loro specificità fenomenica; poteva essere richiamata – la contestualizzazione – nel processo di costruzione delle scale, come *premura documentale* a supporto delle scelte che si andavano a compiere, e poteva ancora essere un punto di attenzione nella costruzione di una esplicita griglia di ponderazione.

---

<sup>27</sup> Cfr. P. AUGUGLIARO, *Metodi di progettazione*, in P. AUGUGLIARO, V. MAJER (a cura di), *Assessment center e sviluppo manageriale*, Angeli, Milano, 1993, pp. 53 e ss.



*Conclusivamente* il caso di studio segnala il grande disimpegno dell'amministrazione, che lascia le commissioni del tutto prive di assistenza specialistica, tanto da non ipotizzare qualche supporto tecnico scientifico rispetto agli strumenti valutativi adottati o da adottare.

Lo stesso Invalsi – che pure veniva richiamato dalla traccia di una delle prove – sembra lontanissimo da un *modus operandi* che possa configurarne l'azione come effettiva consulenza offerta al sistema scolastico italiano. Forse non si è accorto che dinanzi a delle prove di selezione così delicate e così importanti, come quelle previste dal concorso per dirigenti scolastici, avrebbe potuto giocare un suo ruolo propositivo, pur nel pieno rispetto delle leggi e dell'autonomia delle commissioni.

Si è preferito lasciare spazio al chiacchiericcio docimologico, creando le condizioni per trasformare il numero in feticcio e per determinare quel vuoto di senso e di significati che ha poi permesso di contrabbandare l'improvvisazione per innovazione metodologica.

A questo punto le strade che si prospettano per il sistema scolastico italiano sono sostanzialmente due e vanno percorse congiuntamente.

Da una parte, occorre trasformare il chiacchiericcio docimologico in competenze effettive e diffuse, per evitare errori vistosi, per creare una generalizzata sensibilità che permetta di capire quanto possano essere complessi i processi di utilizzazione di alcuni strumenti e con quale rigore e dopo quale tirocinio formativo si debba procedere in questo contesto. Dall'altra, occorre che il management nazionale possa configurare la propria azione come funzione ministeriale (da *ministerium*) e quindi di servizio, per dare supporto effettivo e specialistico a chi opera nella scuola e per la scuola.

Negli ultimi tempi i decisori ministeriali hanno avvertito il fascino sottile della grande riforma, dal momento che la scuola e l'università si poggiavano, ed ancora si poggiano, su nessi funzionali sicuramente bisognosi di ripensamento. E tuttavia, se per un momento si lasciassero da parte i grandi disegni palinogenetici e si decidesse di por mano ad una sistematica, puntuale, diffusa, incessante azione di promozione della autenticità, per restituire competenze e responsabilità ad ogni nodo istituzionale, per richiamare tutti e ciascuno all'onere dell'aggiornamento scientifico, al rispetto delle competenze, al gusto della innovazione, nel segno di una progettualità nuova che ritrovi la persona al centro d'ogni premura e il sapere come guida maestra d'ogni opzione, si darebbe vita ad una riforma della scuola (e dell'università) finalmente lontana da proclami e da clamori e però davvero grande e forse persino epocale.

## Indicazioni bibliografiche

J. A. ARTER, J. MCTIGHE, *Scoring Rubrics in the Classroom: Using Performance Criteria for Assessing and Improving Student Performance*, Corwin Press, Thousand Oaks, (CA), 2001.



## Nicola Paparella, Pubblicazioni on-line

- P. AUGUGLIARO, *Metodi di progettazione*, in P. AUGUGLIARO, V. MAJER (a cura di), *Assessment center e sviluppo manageriale*, Angeli, Milano, 1993, pp. 53 e ss.
- C. BEZZI, *Il disegno della ricerca valutativa*, Angeli, Milano, 2001.
- C. BEZZI, *La linea d'ombra. Problemi e soluzioni di ricerca sociale e valutativa*, Angeli – Aiv, Milano, 2011.
- B. BORDIGNON, *Epistemologia della valutazione universitaria. La valutazione professionale del docente*, Rubbettino, Soveria Mandelli, 2010.
- C. BUCCIARELLI, *Come valutare. La necessità di un approccio sistematico*, in Censis, *Speciale valutazione*, “Note e commenti” n. 1/2, 1991.
- C. CIPOLLA, *Epistemologia della tolleranza*, Angeli, Milano, 1997.
- E. DEL PIANTO, *Assessment center. Tecniche e strumenti per il valutatore*, Angeli, Milano, 2004.
- G. DOMENICI, *Manuale della valutazione didattica*, Laterza, Bari, 1993.
- S.-L. DUARTE, *A especificação moral dos actos humanos segundo são Tomás de Aquino*, Edusc, Roma, 2008.
- F. V. EDGEWORTH, *The element of chance in competitive examinations*, in *Journal of the Royal Statistical Society*, 53, 1890, pp. 460-475, 644-663.
- F. FERRETTI, D. GAMBARARA (a cura di), *Comunicazione e scienza cognitiva*, Laterza, Bari, 2005.
- L. GALLIANI (a cura di), *Web ontology della valutazione educativa*, Pensa Multimedia, Lecce, 2009.
- M. GATTULLO, *Didattica e Docimologia. Misurazione e valutazione nella scuola*, Armando, Roma, 1967.
- H. LAUGIER, H. PIÉRON, É. TOULOUSE e D. WEINBERG, *Études docimologiques sur le perfectionnement des examens et concours*, Conserv. nat. des arts et métiers, Paris, 1934.
- P. LUCISANO, *Validità e affidabilità delle pratiche valutative: a proposito del progetto pilota 2*, in *Cadmo*, n. 2, 2003, pp. 37-56.
- R. MARAGLIANO, *Valutazione e scuola*, Lisciani e Giunti, Teramo, 1989.
- A. MARRADI, *Unità di misura ed unità di conto*, *Rassegna Italiana di Sociologia*, XXVI, 2, 1985
- V. MASONI, *M&V. Monitoraggio e valutazione dei progetti nelle organizzazioni pubbliche e private*, Angeli, Milano, 1997.
- N. PAPARELLA, *Un reattivo per la misurazione dell'efficacia dell'educazione religiosa nella scuola materna*, in B. VERTECCHI (a cura di), *Per una nuova qualità della scuola*, Tecnodid, Roma 1996, pp. 268 e ss.



## Nicola Paparella, Pubblicazioni on-line

N. PAPARELLA, *Per costruire una ontologia sul tema della programmazione*, in N. PAPARELLA (a cura di), *Ontologie, simulazioni, competenze*, Amaltea, Melpignano, 2007.

N. PAPARELLA, *Concorsi e valutazioni. Errando discimus*, in G. TREBISACCE, *Valutazione e scuola. Idee per l'innovazione metodologica e didattica*, Quaderni del dipartimento, Università della Calabria, Rende, 2012, pp. 217 e ss.

G. PENNISI, *Tecniche di valutazione degli investimenti pubblici*, con scritti di P.L. SCANDIZZO, Istituto Poligrafico e Zecca dello Stato, seconda edizione aggiornata, Roma, 1991.

D. SPERBER, D. WILSON, *La pertinenza*, tr. it. [sulla I ed. del 1986], Anabasi, Milano, 1993.

A. VISALBERGHI, *Misurazione e valutazione nel processo educativo*, Comunità, Milano, 1955.

A. VISALBERGHI, *Didattica e docimologia: misurazione e valutazione nella scuola*, Armando, Roma, 1967.

A. VISALBERGHI, R. MARAGLIANO, B. VERTECCHI, *Pedagogia e scienze dell'educazione*, Mondadori, Milano, 1978.

B. VERTECCHI, *Decisione didattica e valutazione*, La Nuova Italia, Firenze, 1993.

Sentenza Cons. Stato, sez. IV, 28 dicembre 1993, n. 1148 e 16 maggio 1994, n. 408.