



Giorgio BOLONDI¹

Valutare una valutazione in Matematica

Assessing an assessment in Mathematics

SUNTO

Questo articolo presenta una riflessione su come valutare la validità di una valutazione.

Parole chiave: Valutazione formativa; valutazione sommativa; valutazioni su larga scala.

¹ Libera Università di Bolzano, ITALIA.

Indirizzi email:

giorgio.bolondi@unibz.it

ABSTRACT

We discuss how to assess the validity of an assessment.

Keywords: Formative assessment; summative assessment; Large-scale assessments.

Ricevuto il 05/09/2022

Accettato il 31/10/2022



INTRODUZIONE

Il problema della *validità* di una valutazione dell'apprendimento in matematica sta ricevendo sempre maggiore attenzione, sia da parte dei ricercatori sia da parte degli insegnanti. Da un lato, molte ricerche comprendono, come parte della propria metodologia, dei momenti di valutazione degli allievi: prima o dopo un *teaching experiment*, per esempio, o durante la fase di osservazione del lavoro di un insegnante. Sempre più frequentemente, inoltre, i ricercatori implementano *mixed methods* nei quali dati e informazioni provenienti da valutazioni su larga scala sono utilizzati assieme a osservazioni e risultati ottenuti con valutazioni locali (Cresweel & Plano-Clark, 2017; Ferretti & Bolondi, 2019; Menck et al., 2017). Il problema di come si modifichi la validità di questi atti di valutazione, nel momento in cui vengono integrati o coordinati, è fondamentale. Da un altro punto di vista, gli insegnanti sono sempre più consapevoli del fatto che la valutazione in matematica è un fatto complesso, apparentemente centrata sul “sapere” da acquisire ma di fatto influenzato dal contesto generale, dalle pratiche diffuse e dalle caratteristiche del sistema, dalle situazioni particolari, dalle convinzioni dell'insegnante, dalle caratteristiche del rapporto instaurato con la classe e con i singoli allievi, in definitiva da elementi della noosfera e da elementi della microsocietà “classe di matematica”. Agenti che “irrompono” in classe dalla noosfera, come le prove standardizzate internazionali (OECD-Pisa, IEA-TIMS) o nazionali (in Italia le *Prove Invalsi*) mettono in luce questa complessità, evidenziando criticità e differenze. La valutazione è un catalizzatore e rivelatore dei diversi elementi che costituiscono l'identità dell'insegnante di matematica (Ferretti et al., 2021; Bolondi et al., 2021). Basti pensare a come l'esperienza della didattica a distanza durante la pandemia abbia portato a roventi discussioni proprio quando si toccava il tema della valutazione: la necessità di valutare gli allievi senza essere fisicamente nello stesso spazio ha portato molti insegnanti a riconsiderare il senso stesso dell'azione valutativa.

La valutazione può quindi essere considerato un elemento centrale in didattica della matematica, sia dal punto di vista teorico sia da quello pratico. Le riflessioni che seguono hanno lo scopo di mettere a

fuoco come guardare alla *validità di una valutazione*, in contesto di ricerca o in contesto di classe.

UN TENTATIVO DI DEFINIZIONE

Al termine “valutazione” vengono abitualmente aggiunti aggettivi o specificazioni (Gipps, 1994) che ne delineano il ruolo, la funzione, la dimensione o altre caratteristiche: si parla quindi di valutazione formativa (Ferretti et al., 2018), valutazione sommativa, valutazione su larga scala (Looney, 2011), valutazione diagnostica, valutazione standardizzata, ...

Tutte queste “valutazioni” sono *processi*. Ogni valutazione dell'apprendimento è un processo articolato in almeno quattro fasi: si tratta sempre di *ricercare, ottenere, organizzare e restituire informazioni*. Informazioni sull'apprendimento, in particolare sull'apprendimento di una particolare disciplina, per esempio la matematica. Ognuna di queste fasi richiede attenzione specifica da parte di chi la mette in atto, sia egli insegnante, ricercatore, o istituzione.

Ricercare. Il primo momento di ogni valutazione è la definizione dell'oggetto della valutazione, l'individuazione di quali informazioni vogliamo ottenere. Può trattarsi di informazioni su conoscenze (definizioni, enunciati, proprietà, ...) o abilità (di calcolo, di rappresentazione, ...), che sono in corso di acquisizione, o che dovrebbero essere acquisite al termine di un determinato percorso. Può trattarsi anche di specifiche competenze matematiche (argomentative, di *problem solving*, di uso del linguaggio tecnico, di applicazione delle conoscenze in contesti specifici) o trasversali e interdisciplinari. In ogni caso, mettere a fuoco che cosa si vuole osservare e a quali informazioni si è interessati è essenziale. Queste informazioni potranno poi essere utilizzate all'interno di una ricerca, per orientare gli studenti, per certificare il raggiungimento di obiettivi formativi, per intervenire proattivamente nei processi di insegnamento o apprendimento, per autovalutare il proprio lavoro di insegnante, per orientare le scelte degli allievi o delle loro famiglie, ...



Ottenere. Una volta definite quali informazioni interessano e per quale scopo, occorre individuare lo strumento migliore per ottenerle. Ogni ricercatore, ogni insegnante, ogni istituzione ha a disposizione una tavolozza di strumenti di valutazione. Ogni sistema scolastico, in particolare, ha una propria tradizione di strumenti e metodi attraverso i quali vengono raccolte le informazioni per la valutazione (solitamente, sommativa); questa tradizione entra in profondità nel vissuto degli allievi prima, e di coloro che diventano insegnanti poi. La scuola italiana, nel suo passato, utilizzava prevalentemente prove scritte a risposta aperta articolata o interrogazioni orali. Ora sempre più ci si rende conto che altri strumenti, come l'osservazione in momenti di "laboratorio matematico", o la discussione matematica in classe, o all'estremo opposto i test a risposta chiusa o le domande come i *cloze*, possono, ciascuno con le proprie caratteristiche peculiari, permettere di raccogliere e ottenere le informazioni che si stanno cercando. È ovvio che un test a risposta chiusa difficilmente può fornirci informazioni attendibili sulla capacità di risolvere un problema complesso, mentre se vogliamo valutare la conoscenza di un enunciato, forse il metodo più rapido è di chiedere direttamente l'enunciato stesso.

Organizzare. Tutte queste informazioni vanno poi organizzate, collegate tra loro, confrontate con i *benchmark* disponibili (la media della classe, i dati di altre classi o le informazioni provenienti dalle prove nazionali, l'esperienza precedente dell'insegnante, i dati delle ricerche precedenti). Vanno analizzate, se possibile, anche in modo longitudinale, per arrivare se necessario e utile anche a una fotografia dell'evoluzione di quello che stiamo osservando.

Restituire. L'informazione va poi restituita. I destinatari della restituzione possono essere i più diversi: l'allievo in primo luogo, la sua famiglia, la scuola, il sistema scolastico nel suo complesso, la società, la comunità dei ricercatori. La scelta della modalità e delle forme della restituzione non è neutra: basti pensare a come, in Italia, il passaggio (nella scuola primaria) dalla valutazione su una scala numerica alla valutazione descrittiva abbia innescato molte riflessioni (e polemiche) sulle prassi valutative messe in campo dagli insegnanti.

La valutazione è quindi un processo che interviene pesantemente in un sistema molto delicato. Questo è vero sia per la valutazione degli insegnanti, sia per le "valutazioni" che compiono i ricercatori (l'intervento dell'osservatore spesso perturba il fenomeno osservato!). È fondamentale quindi disporre di principi per garantire che le informazioni restituite al termine del processo siano valide (Crooks et al., 1996) per gli scopi per i quali sono state ricercate.

VARIABILI DA TENERE IN CONSIDERAZIONE

Le variabili che entrano in gioco per definire la validità di una valutazione sono di diversa natura.

Lo scopo. La validità è strettamente dipendente dallo scopo della valutazione. La validità di una valutazione formativa dipende da elementi strettamente riferiti ai soggetti valutati e valutanti, mentre la validità di una valutazione sommativa è più riferita all'oggetto della valutazione e al contesto (standard locali e nazionali). Nel campo della ricerca didattica, la validità di un *teaching experiment* dipende da elementi diversi da quelli utilizzati per ottenere informazioni sulla diffusione di un costrutto teorico. Le valutazioni su larga scala hanno tra gli altri lo scopo di supportare l'*accountability* delle scelte politiche, tecniche o istituzionali o di individuare i bisogni di una scuola o di un sistema. Anche su questo piano, gli elementi che garantiscono la validità possono differire da un caso all'altro. Un test di ammissione all'università può essere valido per stabilire una soglia minima di conoscenze per i futuri studenti ma non per stabilire una graduatoria per gli aspiranti.

L'ufficialità. Una valutazione può essere un momento formale o un insieme di momenti formali (come spesso avviene per la valutazione sommativa), o un processo più informale (questo è spesso il caso della valutazione formativa). In relazione a questo aspetto una valutazione può essere più o meno *high stake* per il soggetto valutato, vale a dire avere conseguenze più o meno importanti (la bocciatura o la promozione, l'ammissione a una scuola o un corso universitario, l'ottenimento di una borsa di studio, ...).



Ovviamente, la posta in gioco condiziona in maniera decisiva l'azione del soggetto, sia per quanto riguarda il suo impegno sia per le sue emozioni, paure, motivazioni.

La dimensione. La valutazione può avvenire su scale diverse: possiamo osservare un singolo allievo, una coppia che interagisce, un piccolo gruppo; valutare i risultati di una classe o di una scuola; possiamo infine avere valutazioni su larga scala, a livello di regione, nazione o addirittura internazionale. A seconda della scala abbiamo diversi approcci al problema del controllo della validità. Per una valutazione su larga scala, la sua validità dipende pesantemente dalla coerenza e affidabilità dei modelli statistici usati, e si tratta sempre di una validità *statistica*. Per la valutazione in classe gli elementi quantitativi dovrebbero sempre essere considerati con grande cautela.

La relazione tra il soggetto valutante e il soggetto valutato. Il comportamento e le azioni di un allievo le cui conoscenze o capacità matematiche vengono valutate sono influenzati anche da fattori metacognitivi (l'ansia, le convinzioni, la motivazione, la posta in gioco, le aspettative personali e della famiglia, ...). Una variabile importante, che impatta su questi fattori, è la relazione tra chi valuta (l'insegnante della classe, un altro insegnante della scuola, una commissione esterna, una persona o un gruppo di persone dell'istituzione futura dove l'allievo vuole entrare, una commissione nazionale, l'OECD, la IEA, un computer, ...) e chi viene valutato. La ricerca didattica ha mostrato come sia soprattutto in situazioni di valutazione che fenomeni di *contratto didattico* emergono e condizionano i comportamenti di allievi e insegnanti (Brousseau, 1986; si veda anche D'Amore et al, 2020, per un'analisi delle implicazioni di questo costrutto nelle pratiche d'aula).

CATEGORIE DI INDICATORI

Coerenza del processo complessivo. L'intero disegno del processo di valutazione (individuazione delle informazioni, strumento di raccolta, organizzazione dei dati, modalità di restituzione) deve essere coerente internamente, nelle sue diverse fasi. Informazioni raccolte con

uno strumento di valutazione non adatto ai contenuti o alle competenze indagate non forniranno una valutazione valida. Inoltre, il processo deve essere coerente con lo scopo della valutazione. Non si possono infine utilizzare senza cautele e avvertenze i risultati di una valutazione realizzata per un certo scopo in un contesto diverso e con scopi diversi. I risultati di una valutazione statistica su larga scala non sono validi, in linea di massima, per la valutazione sommativa individuale degli allievi, e ovviamente possono solo fornire elementi di contorno utilizzabili per la valutazione formativa. Nella direzione opposta, i voti interni, assegnati dagli insegnanti, non sono solitamente validi per stabilire le graduatorie di accesso al grado successivo, perché troppo dipendenti da fattori di contesto locale. Ovviamente, è possibile utilizzare validamente informazioni di precedenti valutazioni all'interno di nuovi processi valutativi, ma occorre sempre tenere in considerazione gli scopi per i quali queste precedenti informazioni erano state ottenute.

Stabilità. Un indicatore fondamentale per la validità di una valutazione è la sua (potenziale o verificata) stabilità. Questa può essere definita come l'invarianza del risultato, quando il processo viene ripetuto con gli stessi soggetti a distanza di breve tempo. Nel caso di valutazioni su larga scala questa stabilità può essere tenuta sotto controllo con tecniche psicometriche e tecniche di ancoraggio. Ovviamente, nel caso di valutazioni in una classe, questa è un'invarianza "in potenza", in quanto non è possibile somministrare gli stessi task allo stesso soggetto, senza che la prima esperienza influenzi in maniera determinante la seconda.

Legame con il contesto del soggetto. Una valutazione normalmente si avvale di task, proposti in diverse modalità (scritta, orale, interattiva, a casa, in aula, ...). Colui che risponde deve leggere, interpretare e decodificare queste consegne. La relazione di queste consegne con il contesto di provenienza dell'allievo e, se non viene tenuta in considerazione, può portare a risultati fuorvianti. Questo elemento è particolarmente rilevante nel caso di valutazioni standardizzate su larga scala, perché la stessa consegna viene proposta ad allievi provenienti, talvolta, da contesti e situazioni molto diverse. Possono entrare in gioco differenze linguistiche, differenze nella conoscenza del



mondo, diversità di valori e interpretazioni, diversità di esperienze, ... In particolare, vanno considerati:

- il contesto educazionale, e quindi le tradizioni e le abitudini del sistema scolastico in cui avviene la valutazione, o le abitudini e le “norme” della specifica situazione, la classe, la scuola (come il già citato contratto didattico o le norme sociomatematiche) (Yackel and Cobb, 1996);

- il contesto socio-economico;

- il contesto linguistico (con tutte le implicazioni, per esempio, sull’uso di domande tradotte da una lingua all’altra o in più lingue);

- il contesto familiare e individuale dell’allievo, nei suoi aspetti culturali, di tradizione, ...

La considerazione del contesto del soggetto o della popolazione valutata è un elemento essenziale del processo, in particolare per quanto riguarda l’organizzazione e la restituzione delle informazioni.

Indipendenza dal valutatore. A seconda degli scopi, della dimensione o delle altre variabili considerate, una valutazione può essere più o meno dipendente dal soggetto che valuta. Come si è visto, molti fattori metacognitivi dipendono dall’insieme di relazioni che si instaurano nel contesto classe tra gli allievi e l’insegnante. In linea di principio, una valutazione sommativa o di sistema, per essere valida, dovrebbe essere indipendente da chi somministra la prova.

Quadro di riferimento. La valutazione deve essere basata su un ben definito ed esplicito quadro di riferimento. In alcuni casi, può anche essere legata a un syllabus. Il quadro di riferimento permette di definire con chiarezza quali sono le informazioni che interessa ottenere, e per che scopo. Il quadro di riferimento permette di esplicitare anche la cornice epistemologica, didattica, sociale nella quale si effettua la valutazione. Può arrivare a delineare anche l’idea di matematica, di scuola, di allievo. Le valutazioni su larga scala hanno quadri di riferimento espliciti, che arrivano anche ad avere un impatto notevole sui contesti in cui si effettuano. Ogni insegnante ha un suo quadro di riferimento, che spesso è implicito ma sul quale le valutazioni

che realizza nelle proprie classi sono una finestra. Il quadro di riferimento dell’insegnante, parte essenziale della sua identità come insegnante, si costituisce attraverso la sua esperienza di studente, il suo percorso di insegnante in formazione, il suo lavoro in classe, i suoi confronti con altri insegnanti, la sua formazione continua (Bolondi, 2020).

Trasparenza dei risultati. I risultati devono essere organizzati e restituiti, a livello individuale o aggregato, in modo da essere trasparenti e interpretabili da tutti i soggetti interessati. A seconda delle valutazioni, questi soggetti possono essere gli allievi, le loro famiglie, i decisori amministrativi o politici, le istituzioni di riferimento.

Non ambiguità. Le consegne poste agli allievi devono essere non ambigue. Prima ancora che comprendere le domande, l’allievo deve avere chiaro che cosa gli si chiede, che tipo di output e di restituzione. Deve essere messo nelle condizioni che anche le sue risposte possano essere interpretate senza ambiguità.

Chiarezza della formulazione. La formulazione delle singole domande o consegne deve essere:

- adeguata all’età dell’allievo, nella lingua e negli altri mediatori utilizzati (animazioni, figure, tabelle, ...);

- corretta nella lingua utilizzata;

- senza troppo impliciti o non detti.

Disegno delle attività. Il disegno delle singole attività (item, problemi, ...) deve essere adeguato come dimensione e lunghezza all’età degli allievi. Il tempo di lettura delle singole consegne deve essere commisurato alle consegne stesse e al tempo richiesto per la sua risoluzione.

Definizione degli strumenti di rilevazione. Gli strumenti usati per la valutazione devono essere definiti nel disegno originale. Questionari, colloqui orali, osservazioni, interviste, test, attività interattive al computer devono essere definiti fin dall’inizio e implementati in maniera coerente lungo tutto il processo.

Definizione degli strumenti a disposizione degli allievi. L’uso di strumenti di calcolo, formulari,



strumenti di rappresentazione, strumenti di misura, chat con i compagni, ... deve essere permesso/consigliato/richiesto esplicitamente in coerenza con gli scopi della valutazione.

Definizione del formato di restituzione. Soprattutto quando si parla di valutazione sommativa individuale, le categorie di restituzione, i voti, i livelli devono essere ben definiti. Queste categorie possono essere definite a priori, prima del processo, o derivate empiricamente come risultato del processo stesso (come è il caso di alcune valutazioni standardizzate). Nel caso dei “voti” dell’insegnante, o quando il range dei voti è definito a priori, la validità della valutazione può dipendere dalla chiarezza del significato attribuito a ogni voto, o livello. Lo stesso disegno della valutazione dipende da questo significato: che cosa significa “6” in questo contesto? Cosa posso fare per stabilire se il tale mio allievo è al livello del “6”? Quali domande posso porgli? In che modo devo interpretare le sue risposte per poter capire se è al livello del “6”?

Strumenti di registrazione. Le osservazioni, i protocolli, i dati, in generale le risposte o le azioni degli strumenti, vanno registrati per poter essere poi utilizzati. Il rigore e la precisione della registrazione sono fondamentali per la validità del processo.

Autenticità. Questo è un punto che si è rivelato critico durante la recente pandemia. Con questo termine, intendiamo che i dati raccolti siano affidabili e autentici, nel senso che non siano inquinati da *cheating*, suggerimenti, informazioni attinte da fonti non autorizzate durante la valutazione, o altri comportamenti scorretti. Questi comportamenti scorretti possono essere messi in campo sia dagli studenti, sia dai loro insegnanti (per esempio nel caso delle valutazioni istituzionali esterne).

CONCLUSIONI

Ogniquale volta un osservatore interviene in un fenomeno complesso (e i processi di insegnamento-apprendimento lo sono in massimo grado), il suo intervento perturba il fenomeno osservato. Nel momento in cui con la valutazione

si vuole “fotografare” l’evoluzione o il risultato di un processo di questo tipo si rischia di ottenere un’informazione deformata. Non solo, questo intervento influenza l’evoluzione futura del fenomeno, spesso in direzione non prevista o non voluta. Le variabili coinvolte nei processi di valutazione sono molte, e spesso si influenzano reciprocamente: non è in generale possibile individuare chiare relazioni di causa-effetto tra di esse. Per garantire la validità delle informazioni ricavate, chi valuta deve avere un controllo sia degli aspetti globali del processo (la coerenza, il quadro di riferimento complessivo, il contesto) sia con quelli locali e puntuali (la formulazione, gli strumenti, l’autenticità).

RIFERIMENTI BIBLIOGRAFICI

- Bolondi, G. (2020). Quale identità per i docenti di Matematica? *Nuova Secondaria*, XXXVIII/1, 84–87.
- Bolondi, G. (2021). L’insegnante di fronte agli effetti della didattica perturbata. *L’insegnamento della Matematica e delle scienze applicate*, 44 (6), 422-437.
- Bolondi, G., Del Zozzo, A., Ferretti, F., Garzetti, M. & Santi, G.R.P. (2021). Can Formative Assessment Practices Appear Spontaneously during Long Distance Learning? In *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2021) - Volume 1* (pp 625-632). Setubal, Portugal: Scitepress.
- Brousseau, G. (1986), Fondaments et méthodes de la didactique des mathématiques: *Recherches en didactique des mathématiques*, 7(2), 33-115.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). London, UK: Sage.
- Crooks, T.J., Kane, M.T., & Cohen, A.S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education Principles Policy and Practice*, 3(3), 265-286.
- D’Amore, B., Fandiño Pinilla, M.I., Marazzani, I., & Sarrazy, I. (2020). *Gli effetti del contratto didattico in aula. Uno strumento concreto per gli insegnanti di matematica*. Bologna, Italy: Pitagora.
- Fandiño Pinilla M.I. (2020). *Molteplici aspetti dell’apprendimento in matematica. Valutare e*



- intervenire in modo mirato e specifico.*
Bologna, Italy: Pitagora.
- Fandiño Pinilla M.I. (2020). *Diversi aspetti che definiscono l'apprendimento e la valutazione in matematica.* Bologna, Italy: Pitagora.
- Ferretti, F., & Bolondi, G. (2019). This cannot be the result! The didactic phenomenon 'the age of the earth.' *Int. J. Math. Educ.* 52(2), 1–14.
- Ferretti F., Chrysanthou P., & Vannini I. (Eds) (2018). *Formative Assessment for mathematics teaching and learning.* Milano, Italy: Franco Angeli.
- Ferretti, F., Santi, G.R.P., Del Zozzo, A., Garzetti, M., & Bolondi, G. (2021). Assessment Practices and Beliefs: Teachers' Perspectives on Assessment during Long Distance Learning. *Education Science 2021*, 11(6), 264.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment.* London, UK: The Falmer Press.
- Looney J. W. (2011). *Integrating Formative and Summative Assessment: Progress Toward a Seamless System? OECD Education Working Papers*, No. 58. Paris, France: OECD.
- Meinck, S., Neuschmidt, O., & Taneva, M. (2017), Workshop Theme: "Use of Educational Large-Scale Assessment Data for Research on Mathematics Didactics", in Kaiser G., ed., *Proceedings of the 13th International Congress on Mathematical Education, ICME-13 Monographs.* Cham, Switzerland: Springer.
- Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, 27(4), 458–477.