

ISSN: 2036-5330 (stampa); 2974-9174 (online)

DOI: 10.32076/RA15205

Misurazione standardizzata e valutazione formativa: una sinergia possibile

Standardized measurement and formative assessment: a possible synergy

Roberto Ricci¹

Sintesi

La valutazione formativa necessita di standard e traguardi di riferimento per individuare la meta verso la quale indirizzare le azioni di miglioramento. Attraverso i livelli descrittivi delle prove INVALSI di Italiano e Matematica è possibile fare un passo importante in questa direzione. Il lavoro illustra i principali passaggi tecnico-scientifici necessari per la costruzione dei livelli di risultato e la loro descrizione e le ragioni che hanno portato a definire il livello minimo di accettabilità al termine della scuola secondaria di secondo grado. Sottolineando l'importanza di aprire un dibattito su questa definizione, il contributo illustra alcuni usi possibili dei livelli così definiti a supporto delle politiche scolastiche.

Parole chiave: Livelli descrittivi di risultato; Soglie di accettabilità; Standard e valutazione formativa.

Abstract

Formative assessment needs standards and benchmarks to identify the goal towards which improvement actions should be directed. The descriptive levels of INVALSI tests in Italian and Mathematics are an important step in this direction. The work illustrates the main technical-scientific stages necessary for the construction of the result levels and their description, as well as the reasons for defining the minimum level of acceptability at the end of secondary school. Emphasizing the importance of opening a debate on this definition, the paper illustrates some possible uses of the levels so defined in support of school policies.

Keywords: Descriptive result levels; Acceptability thresholds; Standards and Formative Assessment.

1. INVALSI (Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione) - Le opinioni espresse sono da attribuirsi all'autore e non impegnano la responsabilità dell'Istituto di appartenenza.

1. Prefazione

Le rilevazioni standardizzate sui livelli di apprendimento hanno assunto negli ultimi decenni un ruolo molto importante anche in Italia. Già negli anni Settanta del Novecento la IEA (*International Association for the Evaluation of Educational Achievement*) ha introdotto i primi studi internazionali comparativi (*Six Subjects Survey*) che hanno rappresentato un punto di riferimento per le più recenti e note ricerche della IEA (TIMSS: *Trends in International Mathematics and Science Study*, PIRLS: *Progress in International Reading Literacy Study*, ICILS: *International Computer and Information Literacy Study*, ecc.), ma soprattutto per l'indagine PISA (*Programme for International Student Assessment*) promossa dall'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico). L'Italia ha partecipato a queste indagini sin dalle loro prime edizioni in modo continuativo e sistematico. Tuttavia, gli esiti di queste rilevazioni sono stati oggetto di ricerche e analisi che hanno avuto una limitata diffusione al di fuori della più ristretta cerchia dei ricercatori e dei pochi esperti del settore e sono raramente divenute oggetto del dibattito sulla scuola, rimanendo pressoché sconosciute alla quasi totalità dei docenti e dei dirigenti scolastici.

Tuttavia, la situazione italiana è profondamente mutata nel periodo a cavallo tra la fine del secolo scorso e l'inizio degli anni Duemila, sostanzialmente per due ordini di ragioni. In primo luogo, alla fine degli anni Novanta

del secolo scorso è stata varata la legge sulla autonomia scolastica (D.P.R. 275/1999) e la successiva legge di riforma dei cicli scolastici (Legge n. 53/2003, la cosiddetta riforma Moratti) che ha introdotto le prove nazionali, più note come prove INVALSI, sulla misurazione dei livelli di apprendimento in alcune discipline di base. In secondo luogo, a partire dal 2000, e di seguito ogni tre anni, l'indagine OCSE-PISA ha rilevato in un numero via via crescente di Paesi i livelli di apprendimento nella comprensione della lettura di testi scritti, in matematica e nelle scienze della natura e della terra. Senza alcun dubbio l'indagine OCSE-PISA ha avuto un impatto decisamente rilevante sulle politiche scolastiche degli stati che vi hanno partecipato, determinando un ampliamento del dibattito sui livelli di apprendimento raggiunti dalla popolazione scolastica fino a poco prima nemmeno immaginabile.

In questo lavoro si vuole prendere in esame un aspetto particolare delle rilevazioni standardizzate sugli apprendimenti, ossia la modalità di restituzione degli esiti. Infatti, l'indagine PISA e, a partire dal 2018 anche le rilevazioni INVALSI per la scuola secondaria di primo e di secondo grado, a fianco ai tradizionali punteggi espressi su scala numerica hanno introdotto i cosiddetti livelli di risultato, descritti in termini di prestazioni² corrispondenti a ciascun livello, provando per la prima volta per il panorama italiano a definire una soglia di accettabilità per un determinato grado scolastico. Il passaggio alla rappresentazione dei risultati mediante livelli descrittivi rappresenta un cambiamento di prospettiva di non poco conto poiché si sposta l'atten-

2. In questo paragrafo si usa volutamente il termine generale *prestazione* in una accezione non tecnica, mentre nei paragrafi successivi esso è sostituito da una terminologia più appropriata, cercando di fare chiarezza su cosa si intenda in questa sede quando si parla di misurazione delle *competenze di base*, proponendo anche una soluzione alternativa con l'obiettivo di evitare fraintendimenti che spostano l'oggetto del dibattito ad ambiti scientifici e di ricerca specifici, diversi dal tema che qui si desidera affrontare.

zione verso l'esito di apprendimento oggetto della misurazione, non più intesa come fatto esclusivamente numerico.

Per quanto i riferimenti alle esperienze internazionali siano importanti e di grande aiuto per focalizzare un problema così complesso e rilevante come l'individuazione di livelli di risultato, corredati dalla descrizione delle prestazioni a essi associati, l'attenzione del presente lavoro è rivolta principalmente al caso italiano, ossia alle soluzioni adottate da INVALSI a partire dal 2018.

Le implicazioni di questa scelta tecnico-metodologiche sono diverse e molteplici. In primo luogo, c'è un aspetto tecnico-psicometrico alla base dell'individuazione dei livelli e della loro descrizione. È infatti importante conoscere e valutare tutte le implicazioni delle metodologie adottate, cercando di massimizzare i vantaggi che si possono trarre da esse e minimizzare gli eventuali effetti negativi. Sempre sotto il profilo tecnico, l'operazione effettuata dall'INVALSI per l'individuazione dei livelli su base rigorosamente scientifica rappresenta un elemento di rarità nel panorama nazionale, soprattutto su scala così ampia e generalizzata. Pertanto, per quanto gli aspetti più tecnici esulino dallo scopo di questo lavoro, essi sono trattati nei loro elementi essenziali e nella prospettiva di rendere più esplicite le conseguenze e le potenzialità delle soluzioni proposte. In secondo luogo, l'introduzione dei livelli di risultato e l'individuazione di una (per definizione provvisoria) soglia di corrispondenza con i traguardi delle Indicazioni nazionali per il curricolo e delle Linee guida³ aprono in modo rilevante la di-

scussione sulla opportunità di muoversi nella direzione di una prospettiva criteriiale nella valutazione o, quantomeno, di sollecitare un dibattito sulla necessità di avviare una riflessione aperta, concreta e condivisa sui traguardi che il sistema intende assegnarsi al termine di alcuni momenti cruciali del percorso scolastico. Solo così è possibile costruire un ponte percorribile verso una valutazione realmente formativa (Looney, 2011). In terzo luogo, il caso italiano rappresenta una situazione di particolare interesse sia dal punto di vista strettamente scientifico sia da quello più legato alle implicazioni didattico-pedagogiche. Infatti, l'impianto delle Indicazioni nazionali non contempla per le diverse discipline un sistema di traguardi accompagnati da standard espliciti per la verifica del loro raggiungimento. Tuttavia, l'ordinamento italiano prevede un'importante eccezione poiché per le lingue straniere⁴ i traguardi sono espressi in funzione del QCER (Quadro europeo di riferimento per la conoscenza delle lingue), quindi adottando un sistema propriamente criteriiale, ossia con traguardi e relativi standard per misurarne il raggiungimento. Questa situazione ibrida rappresenta un laboratorio di riflessione straordinario. Se è infatti del tutto condivisibile l'idea che la promozione e la realizzazione di pratiche reali ed effettive di valutazione formativa richiedano un sistema criteriiale di obiettivi e di standard, dall'altro pare eccessivamente ottimistico pensare che la previsione normativa e ordinamentale di tale sistema sia sufficiente per innescare un cambiamento effettivo nelle prassi quotidiane di valutazione. In quarto luogo, ma non da

3. Di seguito si userà l'espressione Indicazioni nazionali per indicare entrambi i documenti, nonostante le loro differenze che però non sono rilevanti nell'economia dei ragionamenti proposti.

4. In questo lavoro si farà riferimento esclusivamente all'insegnamento della lingua Inglese.

ultimo, l'ancoraggio dei livelli di risultato ai traguardi di apprendimento delle Indicazioni nazionali consente di rivedere, ampliandoli, alcuni aspetti fondamentali per le politiche scolastiche, sia considerando l'intero sistema (nazionale, regionale, provinciale, ecc.) sia la singola istituzione scolastica per le sue azioni di miglioramento. Tra questi concetti vi è senza dubbio quello di dispersione scolastica che non può più essere intesa solamente come la quota di giovani che abbandonano la scuola anzitempo, ma anche come quella parte di studentesse e studenti che terminano il loro percorso di studi, ma senza raggiungere, a volte nemmeno lontanamente, le conoscenze e le abilità di base auspicabili al termine del secondo ciclo di istruzione.

2. I livelli descrittivi degli esiti delle prove INVALSI

Per quanto non abbia destato particolare interesse se non in alcuni ristretti ambienti accademico-scientifici, la novità più rilevante introdotta dal D. Lgs. n. 62/2017 relativamente alle prove standardizzate per la misurazione dei livelli di apprendimento di base è la modalità di restituzione degli esiti. L'attenzione generale è stata interamente rivolta ad altri aspetti che, per quanto di grande impatto operativo, non sono così importanti come l'adozione dei cosiddetti livelli di risultato. Il dibattito pubblico, interno ed esterno al mondo della scuola, si è concentrato prevalentemente sulla modalità di somministrazione nella scuola secondaria, elettronica e online, e sull'introduzione di una prova volta

a verificare le competenze ricettive (lettura e ascolto) di Inglese al termine del ciclo primario, secondario di primo e di secondo grado. Tuttavia, superata la fase iniziale in cui la scuola secondaria ha dimostrato di essere perfettamente in grado di svolgere le prove al computer e online, è giunto ormai il momento di concentrare l'attenzione sui livelli descrittivi di risultato e sulle loro diverse implicazioni, sia per la singola scuola sia per l'intero sistema scolastico.

Come già accennato, a partire dal 2018 INVALSI ha modificato la modalità di restituzione degli esiti delle prove della scuola secondaria di primo e di secondo grado, legando la scala sulla quale sono espressi i risultati ai traguardi di apprendimento definiti dalle Indicazioni nazionali. In estrema sintesi, la scala costruita da INVALSI per le prove di Italiano e Matematica si articola su cinque livelli, dal livello 1 al livello 5. I livelli 1 e 2 identificano un risultato non in linea con i traguardi previsti per il grado scolastico oggetto d'interesse, mentre il livello 3 rappresenta un esito della prova adeguato ai traguardi di apprendimento previsti dalle Indicazioni nazionali, infine i livelli 4 e 5 rappresentano i livelli di apprendimento più elevati.

I risultati d'Inglese, anche per la scuola primaria, sono invece espressi mediante la scala a livelli del QCER. Più precisamente: pre-A1 e A1 per la V primaria, pre-A1, A1 e A2 per la III secondaria di primo grado, B1 e B2 per l'ultimo anno della scuola secondaria di secondo grado. Giova ricordare a questo punto una singolarità delle nostre Indicazioni nazionali. Mentre per tutte le discipline, ad eccezione delle lingue straniere, non si è vo-

luto o non si è potuto fissare degli standard in funzione dei quali stabilire se un traguardo di apprendimento sia stato o meno raggiunto, per le lingue straniere si è adottato un approccio rigorosamente criteriale. Per quanto ci sia scarsa o nessuna consapevolezza di questa singolarità, essa dovrebbe, almeno in teoria, produrre effetti considerevoli anche sulla valutazione di scuola. Poiché le Indicazioni nazionali fissano un criterio molto preciso in esito a ciascun ordine scolastico, anche la valutazione di scuola dovrebbe, di conseguenza, attenersi per le lingue straniere agli standard previsti dal QCER. In altre parole, per fare un esempio, se il traguardo previsto al termine del primo ciclo d'istruzione è l'A2, allora la valutazione di scuola dovrebbe riferirsi agli standard descritti dal QCER per l'A2 e specificare quali sono gli standard positivamente raggiunti per ciascuno dei voti assegnati agli studenti (Ricci, 2021). È però del tutto evidente che questa differenza non si riscontra in modo sistematico nelle pratiche valutative, quindi si dimostra che, senza un'adeguata formazione del corpo docente e dei dirigenti scolastici, sarà sempre la prassi consolidata ad avere la meglio. Se si pensa che l'introduzione degli standard determinerà necessariamente e spontaneamente uno sviluppo della valutazione formativa, l'attuale situazione ibrida delle Indicazioni nazionali mostra esattamente il contrario.

In questa sede si vuole però dedicare un'attenzione maggiore alla descrizione, per quanto generale, delle modalità tecniche di costruzione dei livelli di risultato adottate da INVALSI. Infatti, le implicazioni derivanti dal loro uso richiedono una riflessione anche su

come essi sono costruiti e su quali modalità statistico-psicometriche siano state seguite. Tuttavia, gli approfondimenti metodologici esulano ampiamente dallo scopo di questo lavoro e si rimanda a pubblicazioni più specifiche che illustrano approfonditamente tutto il processo di costruzione dei livelli descrittivi di risultato (Desimoni, 2018).

Come già accennato in premessa, anche dal punto di vista tecnico qui considerato, le prove INVALSI rappresentano uno dei pochi esempi in Italia, se non l'unico, in cui si è applicato su larga scala una metodologia rigorosamente psicometrica per la definizione e la costruzione di livelli di risultato (Barbaranelli & Natali, 2005). Tuttavia, prima ancora di descrivere sinteticamente la procedura tecnico-metodologica per la costruzione dei livelli, è necessario effettuare una premessa sui termini che in genere si usano. La terminologia utilizzata nello specifico ambito tecnico-scientifico di riferimento per la costruzione e la descrizione dei livelli di risultato attraverso prove standardizzate ha una derivazione prevalentemente anglosassone e talvolta non coincide con quella più rigorosa delle scienze dell'educazione. Spesso si è usato e si usa l'espressione competenze di base (*basic skills*) per indicare l'oggetto di misurazione delle prove standardizzate, mentre si dovrebbe probabilmente parlare in modo più appropriato di conoscenze e abilità di base. Fare chiarezza su questo punto può contribuire a evitare inutili fraintendimenti e, soprattutto, si evita che il dibattito si incanali lungo direzioni diverse da quelle che invece dovrebbero essere seguite per affrontare i problemi, sovente particolarmente rilevanti, che emergono dagli

esiti delle prove standardizzate. Si aggiunga poi, che esiste un'ampia e da anni consolidata letteratura che sottolinea la forte correlazione associativa tra la dimensione delle conoscenze e abilità di base, da un lato, e le competenze di base, propriamente intese, dall'altro. Pertanto, anche alla luce di questo fatto, risulta evidente il forte valore indicativo che la misurazione delle abilità di base su scala comparativa può avere per stimare la presenza di competenze propriamente intese, tecnicamente molto più complesse da misurare.

La rappresentazione dei risultati delle prove standardizzate INVALSI (si seguito prove INVALSI) mediante livelli descrittivi comporta diversi vantaggi. Tale modalità aiuta a interpretare il significato di un determinato risultato, spostando l'attenzione dall'esito puramente numerico a ciò che sa fare colei o colui che raggiunge quel determinato esito (Van der Linden, 2018). La descrizione del risultato in termini positivi, ossia individuando cosa il rispondente sa fare, presenta potenzialità didattiche enormi (Blömeke & Gustafsson, 2017). Ma c'è un altro aspetto ancora più importante. È auspicabile che la lettura, anche critica, delle scelte adottate da INVALSI rispetto ai traguardi delle Indicazioni nazionali promuova un dibattito a livello di scuola e di sistema su quali debbano essere i traguardi di apprendimento, minimi e più avanzati, da raggiungere e quali siano gli indicatori che permettono di valutare il loro raggiungimento. Questo passaggio potrebbe essere il ponte verso un consolidamento reale e non solo formale delle pratiche di valutazione formativa, fondamentale per passare da una valutazione

che semplicemente certifica a una valutazione che sostiene e promuove il miglioramento e l'apprendimento (Greenstein, 2017).

Nella letteratura di riferimento la soluzione adottata da INVALSI a partire dal 2018 rientra nell'ambito dei *descriptive proficiency levels*, basati su *learning metrics* o *descriptive proficiency scales* (Turner, 2014). In presenza di standard, come nel caso dell'inglese, è possibile un altro approccio, cosiddetto *standard referenced*, in cui si parte da una descrizione di cosa il rispondente deve conoscere e cosa debba essere in grado di fare (Cizek & Bunch, 2007). I due metodi presentano delle forti similitudini da diversi punti di vista, ma mentre il primo ha il suo punto di partenza interno alla banca di domande (*item bank*) dalla quale sono tratte le prove svolte dagli allievi, il secondo, invece, si basa su criteri più o meno cogenti stabiliti esternamente alla prova. In base a quanto detto sino a qui, è evidente che i livelli descrittivi delle prove INVALSI di Italiano e Matematica si basano sulla procedura dei *learning metrics*, mentre quelli di Inglese su quella degli *standard referenced*, ossia gli standard del QCER, richiamati esplicitamente dalle Indicazioni nazionali. Tuttavia, pur nelle loro differenze intrinseche, entrambi i metodi portano alla definizione di livelli che consentono di stabilire una soglia di accettabilità degli esiti. La grande differenza tra le due soluzioni è che mentre per l'Inglese il livello di adeguatezza è stabilito dal legislatore (A1 al termine della V primaria, A2 al termine della secondaria di primo grado e B2 al termine della scuola secondaria di secondo grado⁵), per le prove di Italiano e Matematica tale soglie di accettabilità sono

5. Con effetti a partire dall'a.s. 2022-23, per gli istituti professionali il traguardo è stato fissato al livello B1+ del QCER.

state individuate da INVALSI, nell'auspicio, sinora non realizzato, che si aprisse un dibattito serio e approfondito sull'opportunità di quella scelta, partendo proprio dalle descrizioni sintetiche e analitiche dei livelli di risultato al termine della scuola secondaria di primo e di secondo grado. Tale discussione potrebbe avere effetti estremamente positivi che vanno ampiamente al di là delle funzioni delle prove standardizzate (Cornoy, 2015). Essa consentirebbe di avviare una riflessione approfondita sulla definizione concreta dei traguardi attesi, passaggio fondamentale per garantire la tenuta del sistema, evitando derive verso l'abbassamento progressivo dei livelli di apprendimento richiesti. Inoltre, in assenza di traguardi opportunamente operazionalizzati, è molto difficile pensare che la valutazione di scuola possa veramente essere una valutazione educativa il cui scopo è quello di accompagnare il discente, ma anche il docente, lungo tutto un percorso di miglioramento verso un traguardo condiviso, auspicabilmente definito a un livello più ampio della singola classe o della singola scuola. Tale modalità valutativa realizza, forse si dovrebbe dire realizzerebbe, una sintesi molto efficace tra la necessità di personalizzazione del percorso e quella di armonizzazione dei traguardi che, invece, dovrebbero essere comuni all'intera scuola, ma per quanto riguarda gli apprendimenti di base si dovrebbe forse dire all'intera collettività. Solo così è possibile prevedere seri e appropriati interventi a favore di chi quei traguardi non li raggiunge, adottando le soluzioni più adeguate che, per quanto differenti da caso a caso, sarebbero però dirette verso traguardi espliciti e verificabili,

favorendo i passaggi da un ordine di scuola a un altro e sostenendo una didattica verticale realmente orientativa (INVALSI, 2019).

2.1. Primo passaggio: la costruzione della banca di item

In linea generale, per garantire la validità dell'esito di una rilevazione è necessario, per quanto non sufficiente, che la prova si basi su un ampio numero di domande in grado di coprire adeguatamente l'oggetto dell'indagine stessa. A ben vedere, con le opportune differenze, è la stessa situazione che si verifica quando si prepara una prova di classe in cui l'insegnante predispone un insieme di domande le cui risposte consentono di verificare se il rispondente ha raggiunto gli obiettivi prefissati. Naturalmente, nel caso di una rilevazione standardizzata nazionale tale passaggio ha una strutturazione tecnico-metodologica molto più complessa e articolata. Inoltre, se l'obiettivo è quello di restituire gli esiti attraverso livelli descrittivi articolati e robusti, è necessario disporre di un adeguato numero di quesiti per ciascun livello, in modo che l'oggetto della valutazione possa essere descritto con l'opportuno dettaglio e, soprattutto, con la necessaria gradualità. D'altro canto, però, l'elevato numero di domande da sottoporre a ciascun rispondente per identificare il livello raggiunto richiederebbe prove estremamente lunghe e oggettivamente non sostenibili, anche per un candidato particolarmente competente. La costruzione di una banca di domande (*item bank*) dalla quale trarre molteplici prove secondo criteri di seguito sinteticamente descritti permette di superare il problema,

prevedendo la somministrazione a ciascun rispondente di una prova di ragionevole e sostenibile lunghezza della durata massima di un paio di ore, variabile in base al grado scolastico indagato (Desimoni, 2018; Wright & Stone, 1999).

Tutte le domande che compongono la banca sono calibrate su una stessa scala attraverso la collocazione su un unico *continuum* del loro parametro di difficoltà che, come noto, nel modello di Rasch, adottato anche da INVALSI, corrisponde al livello di abilità necessario per avere una probabilità pari al 50% di rispondere correttamente a una domanda (Rasch, 1980). Questo procedimento garantisce che prove diverse tratte dalla stessa banca forniscano misure equivalenti dell'abilità del rispondente; pertanto, non è più tecnicamente necessario somministrare tutte le domande della banca per ottenere una misura adeguata dell'abilità dello studente, ma è sufficiente che l'allievo risponda a un sottoinsieme di quesiti di congrua e sostenibile lunghezza per avere una misura affidabile del suo livello di preparazione. La costruzione della banca di domande svolge pertanto un ruolo determinante per la descrizione dei livelli di risultato. Essa è formata da quesiti che insistono sui differenti aspetti dei quadri di riferimento (QdR) delle prove INVALSI, opportunamente costruiti in funzione dei traguardi delle Indicazioni nazionali (INVALSI 2018a; INVALSI 2018b).

Dopo un'accurata fase di analisi qualitativa delle domande per verificare su base teorica la rispondenza dei quesiti ai costrutti definiti nei QdR, ma soprattutto la coerenza con i traguardi delle Indicazioni nazionali,

tutte le domande devono passare attraverso un rigoroso vaglio di verifica empirica (*pre test*) delle assunzioni teoriche effettuate. In altre parole, tutte le domande sono verificate mediante la loro somministrazione ad ampi campioni di allievi della medesima età scolare per la quale si intende costruire la banca di domande e con le medesime modalità previste per lo svolgimento delle prove definitive (Falorsi *et al.*, 2019). Mediante gli esiti del pre test vengono poi eliminate dalla banca tutte quelle domande che non garantiscono le proprietà statistico-psicometriche per assicurare un'adeguata capacità misurativa della banca di quesiti: unidimensionalità (la domanda verifica principalmente un solo costrutto), uguale funzionamento⁶ per tutte le topologie di allievi, buon adattamento al modello di Rasch⁷.

Una volta che gli item sono selezionati attraverso i due passaggi della verifica teorica in prima battuta e di quella statistico-sperimentale in seconda, tutti i quesiti sono pre-calibrati su una scala unica attraverso la loro calibrazione concorrente e poi vengono allocati in diverse forme del test, parallele nell'accezione dell'*Item Response Theory* e in modo che esse abbiano stesse caratteristiche di contenuto, stessa lunghezza, stesse tipologie di domande (a risposta chiusa, aperta, semiaperta, ecc.). Inoltre, per garantire stime appropriate e robuste delle abilità dei rispondenti, la definizione delle forme è effettuata seguendo rigorose procedure di ottimizzazione, proprie dell'ambito disciplinare della ricerca operativa (Verschoor, 2007).

6. Assenza di *different item functioning* (DIF).

7. Esclusione di tutti i quesiti che presentano un valore del parametro di *infit* superiore a 1,10.

2.2. Secondo passaggio: l'individuazione e la costruzione dei livelli di Italiano e Matematica

In questa sede, tuttavia, l'aspetto tecnico-metodologico alla base della costruzione dei livelli non è l'unico aspetto da considerare e probabilmente nemmeno il più importante. In altre parole, quanto descritto nel paragrafo 2.1 rappresenta la struttura portante della modalità con la quale sono resi i risultati delle prove INVALSI, ma in termini di uso a livello di scuola e di sistema risulta ancora più rilevante il processo con il quale si giunge alla descrizione dei livelli di risultato. Riprendendo il lavoro di Desimoni del 2018, la descrizione dei livelli avviene attraverso le seguenti fasi, il cui esame ne rende evidente la loro rilevanza per le politiche scolastiche, sia a livello micro sia a livello macro:

- Fase 1: descrizione di ciascun quesito da parte degli esperti disciplinari in relazione ai traguardi delle Indicazioni nazionali e dei QdR INVALSI di Italiano a Matematica.
- Fase 2: calibrazione dei parametri delle domande che devono entrare nella *item bank* e conseguente stima delle abilità dei rispondenti sulla base degli esiti nelle classi (cosiddette classi-campione) in cui la prova si svolge alla presenza di un osservatore esterno, garante del rispetto del protocollo di somministrazione.
- Fase 3: individuazione dei punteggi soglia (*threshold*) che identificano il passaggio da un livello a quello successivo

in base agli esiti nelle classi-campione.

- Fase 4: calcolo per ogni quesito del parametro di difficoltà corrisponde al livello di abilità necessario per rispondere con una certa probabilità⁸ di successo (*response probability*) alla domanda stessa.
- Fase 5: descrizione sintetica e analitica dei livelli di risultato.
- Fase 6: assegnazione a ciascun rispondente del livello raggiunto in funzione dell'esito della prova svolta.

In questa sede le fasi più rilevanti sono certamente la 1 e la 5. Nella fase 1 gli esperti disciplinari descrivono su base teorica le caratteristiche delle domande, giovandosi di tutte le informazioni disponibili, e attribuiscono, sempre su base teorico-qualitativa, ciascun quesito a un determinato livello di risultato. Questa attribuzione è necessariamente provvisoria poiché deve essere confrontata con gli esiti effettivi della prova e del livello di difficoltà di ciascuna domanda definito sulla base delle risposte effettivamente fornite dagli studenti.

Nella fase 5 gli esperti sottopongono le valutazioni effettuate nella fase 1 al vaglio della verifica empirica in base all'esito della somministrazione delle prove e descrivono i livelli di risultato in funzione delle domande il cui parametro di difficoltà si colloca dal punto di vista statistico-psicometrico effettivamente in un determinato livello.

Per l'Inglese la procedura è in parte simile, ma è in un certo qual modo semplificata dall'esistenza di criteri esterni alla prova, ossia il QCER, assunti dal legislatore come standard di riferimento. Per questo motivo,

8. Sulla base delle prassi maggiormente attestate in letteratura la *response probability* è stata fissata al 62%.

per l'Inglese gli esperti sono chiamati a definire con una procedura molto dettagliata le soglie che dividono i diversi livelli del QCER misurati (Consiglio di Europa, 2009). Precisamente, per la prova della quinta primaria la soglia che separa il pre-A1 dall'A1, per la prova della terza secondaria di primo grado le soglie che separano il pre-A1 dall'A1 e l'A1 dall'A2. Infine, per la prova dell'ultimo anno della scuola secondaria di secondo grado sono individuate le seguenti soglie: l'estremo inferiore e quello superiore del B1, dove quest'ultimo identifica il passaggio al B2.

La descrizione delle modalità tecniche di costruzione dei livelli di risultato rende evidenti i vantaggi che essi possono determinare, sia a livello di singola scuola, fornendo uno strumento di riflessione didattica, anche se non l'unico, per verificare l'efficacia delle soluzioni metodologiche e organizzative assunte, sia a livello generale, supportando le decisioni di coloro che hanno la responsabilità di governo e di indirizzo del sistema educativo.

3. L'individuazione del livello di accettabilità

Il problema dell'individuazione di livello di accettabilità non si pone per l'Inglese poiché esso è definito dal legislatore attraverso le Indicazioni nazionali. Per la prova di Italiano e di Matematica la situazione è più complessa e richiede il confronto della descrizione dei livelli delle prove INVALSI con i traguardi delle Indicazioni nazionali. In questa sede si prendono in esame i livelli descrittivi di risultato al termine dell'ultimo anno della scuola secondaria

di secondo grado poiché essi rappresentano un punto di riferimento importante, non solo per le prove INVALSI, ma per tutto il sistema scolastico. Da un punto di vista generale e teorico, si può dire che l'individuazione di una soglia di accettabilità è un'operazione interpretativa delle Indicazioni nazionali operata dall'INVALSI che non ha certamente il valore cogente dei livelli del QCER assunti a riferimento esplicito dal legislatore. Tuttavia, la diversità della situazione non rende impossibile o illegittima l'operazione, per diverse ragioni. In primo luogo, perché essa è resa trasparente attraverso un esplicito raffronto con le Indicazioni nazionali. Inoltre, in secondo luogo, essa è per definizione provvisoria e aperta al dibattito pubblico, nell'auspicio che esso aiuti il sistema, ma anche la singola scuola, a discutere se la scelta operata da INVALSI sia da ritenersi condivisibile o se, invece, vada modificata, anche radicalmente. Già solo un dibattito di questo tipo rappresenterebbe un progresso considerevole verso l'individuazione di standard condivisi che, come più volte indicato in questo lavoro, rappresentano il fondamento della valutazione formativa che ha bisogno di riferirsi a un traguardo ben definito e operazionalizzato al quale tendere e verso il quale si vuole indirizzare l'azione didattica-educativa.

3.1. I livelli di risultato di Italiano

Tenuto conto del rigoroso procedimento tecnico-scientifico alla base della costruzione dei livelli di risultato e di quanto detto nella premessa di questo paragrafo, è necessario analizzare la descrizione

sintetica⁹ dei livelli di Italiano delle prove INVALSI in relazione ai traguardi delle Indicazioni nazionali al termine del secondo ciclo di istruzione.

Descrizione sintetica dei livelli di risultato INVALSI¹⁰

Livello 1. L'allievo/a risponde a domande su testi in prevalenza dal contenuto concreto e familiare, dalla struttura lineare e dal lessico abituale. Individua informazioni *esplicite*, in parti *circoscritte* e *indicate* di un testo, *quando non siano presenti informazioni che potrebbero essere confuse con quelle richieste*. Riconosce un'informazione quando è riformulata con parole familiari e vicine a quelle usate nel testo. Ricava dal contesto il significato di parole o di espressioni di uso *abituale* in parti di testo indicate, e compie *semplici deduzioni* collegando un numero *limitato di informazioni esplicite*. Ricostruisce il significato del testo, o di sue parti, riconoscendone una sintesi che contiene parole chiave. Coglie il punto di vista dell'autore e lo scopo di un testo *se questi sono chiaramente espressi e ripetuti in più punti*. Svolge compiti grammaticali che, sulla base del contesto, richiedono di individuare il significato e la funzione di elementi linguistici fondamentali (es. singole parole, espressioni, frasi).

Livello 2. L'allievo/a risponde a domande su testi in *prevalenza di contenuto concreto e familiare*, caratterizzati da *struttura lineare* e da un lessico non solo di uso abituale, ma anche specialistico. Individua in maniera autonoma le informazioni richieste, date in più punti del testo, anche quando siano presenti altre informazioni che possono essere confuse con quelle da ritrovare. Ricava dal testo il significato di parole o espressioni, anche di uso non frequente, o di carattere tecnico-specialistico. Ricostruisce il significato di parti del testo collegando più informazioni e coglie il tema o l'argomento principale di testi di diverso genere. *Se guidato, coglie il senso del testo*, ad esempio scegliendo tra formulazioni sintetiche alternative. Riconosce il tono (ironico, polemico, ecc.) di frasi o parole e il valore espressivo *di alcune scelte stilistiche* dell'autore in passaggi significativi del testo. *Svolge compiti grammaticali che richiedono il ricorso alla propria esperienza linguistica e alla conoscenza di elementi di base della grammatica* (es. forme di collegamento tra frasi o tra diverse parti di un testo).

Livello 3. L'allievo/a risponde a domande su testi di contenuto anche *astratto* ma vicino alla sua esperienza e alle sue conoscenze, caratterizzati in prevalenza da una struttura lineare, e dall'uso di un lessico anche settoriale o figurato. *Individua più informazioni richieste con parole differenti rispetto a quelle utilizzate nel testo, e distribuite in punti diversi*. Ricava dal testo il significato di espressioni figurate o caratterizzate da un particolare tono (ironico, polemico, ecc.) o che *appartengono a un linguaggio tecnico-specialistico*. Ricostruisce il significato dell'intero testo o di sue parti, anche molto ricche di informazioni. Coglie il modo in cui il testo è organizzato e strutturato. Riconosce il significato o la funzione di alcune scelte stilistiche dell'autore. *Coglie il senso di un testo al di là del suo significato letterale*. *Svolge compiti grammaticali che richiedono il ricorso alla conoscenza spontanea della lingua e alla pratica dei testi unite a una sicura conoscenza di forme e strutture grammaticali fondamentali*.

Tab. 1 - Descrizione sintetica dei primi tre livelli descrittivi di Italiano al termine del secondo ciclo d'istruzione.

(Fonte, INVALSI 2019)

9. Sia per Italiano sia per Matematica esistono due descrizioni dei livelli di risultato, una sintetica e una analitica. La prima fornisce una descrizione generale di ciò che sa fare un allievo che si colloca a un determinato livello, mentre la seconda fornisce informazioni molto più tecniche e specifiche di ogni livello e riporta alcuni esempi particolarmente informativi. Per maggiori dettagli si veda:

https://INVALSI-areaprove.cineca.it/index.php?get=static&pag=g13_descrittori_qualitativi

10. Corsivi dell'autore.

La descrizione dei livelli di risultato avviene in termini positivi, ossia è descritto ciò che l'allievo sa fare se il suo risultato si colloca in un determinato livello. Attraverso i corsivi (dell'autore) si comprende perché la soglia di accettabilità è stata fissata al terzo livello e non al secondo. In via preliminare occorre precisare che la prova di Italiano è così denominata dal legislatore per ovvie ragioni di semplicità, ma essa è in realtà una prova di comprensione del testo scritto e di conoscenza di alcuni elementi di funzionamento della lingua, così come espressamente richiamato dal Quadro di riferimento per la valutazione di INVALSI (INVALSI, 2018a). In questa sede, la predetta precisazione non è un elemento di poco conto poiché sottolinea la forte connessione dell'oggetto di osservazione della prova con tutte le discipline di insegnamento, non solo con l'Italiano. Non a caso si parla di apprendimenti e di abilità di base, spesso di competenze di base nell'accezione non tecnica richiamata in precedenza. La comprensione della lettura è fondamentale per qualsiasi ambito di apprendimento; pertanto, ci si deve attendere che dopo un percorso di tredici anni di scuola l'allievo abbia acquisito un livello di competenza adeguato per inserirsi proficuamente e autonomamente nella società.

Andando nello specifico, è del tutto evidente che il livello 1 non descriva un risultato adeguato di risultato dopo 13 anni di scuola. A ben vedere, esso rappresenta con difficoltà ciò che ci si dovrebbe aspettare al termine del primo ciclo di istruzione. Manca totalmente la capacità di affrontare livelli anche minimi di complessità testuale e grammaticali e non è presente alcuna forma di autonomia interpretativa. Il livello 2 descrive già un grado di complessità maggio-

re, ma mancano ancora alcuni elementi fondamentali richiamati in diversi punti delle Indicazioni nazionali¹¹ per tutti gli indirizzi di studio. In primo luogo, la natura e il contenuto dei testi. Al livello 2 essi devono avere un contenuto prevalentemente concreto e familiare e una struttura lineare. Già questi elementi non rispondono più in modo adeguato alle *linee generali e competenze* e agli *obiettivi specifici di apprendimento* previsti delle Indicazioni nazionali, specie se si considera la natura trasversale della comprensione dei testi. Per tutti gli indirizzi di studio, nessuno escluso, le Indicazioni nazionali sottolineano l'importanza di comprendere la complessità, sia nello specifico dell'apprendimento della lingua, sia trasversalmente a tutte le discipline. Ma c'è un altro elemento che non consente l'identificazione del livello 2 come un livello adeguato dopo 13 anni di scuola: l'autonomia dello studente. Non può certamente considerarsi adeguata al termine della scuola secondaria di secondo grado una comprensione del testo che richiede una guida esterna per garantire che l'allievo lo abbia compreso a pieno, per non parlare del fatto che l'allievo che raggiunge solo questo livello è in grado di cogliere il valore espressivo solo di alcune scelte stilistiche, mentre ci si dovrebbe aspettare abilità decisamente più elevate. Il livello 2 prevede inoltre solo la conoscenza di elementi di base della grammatica, mentre le Indicazioni nazionali richiedono una conoscenza più elevata del funzionamento della lingua poiché esso è alla base delle competenze produttive che al termine del secondo ciclo devono raggiungere elevati livelli di complessità. Tutti gli elementi che mancano nel livello 2 e che sono richiesti dalle Indicazioni nazionali si ritrovano nel livello 3 e i

11. Vedi nota 2.

corsivi ne identificano i principali aspetti distintivi. Naturalmente i livelli 4 e 5 descrivono conoscenze e abilità più elevate ancora, ma l'oggetto della riflessione di questo paragrafo è l'individuazione della soglia di accettabilità e quindi del livello minimo che si può ritenere conforme a quanto richiesto dalle Indicazioni nazionali.

Come chiarito in premessa, dato il quadro normativo di riferimento la soglia di accettabilità può essere identificata solo attraverso una valutazione che si deve aprire al dibattito e al confronto. Ritenere che il livello 2 sia già oltre il livello di accettabilità è ovviamente possibile, ma molto pericoloso poiché significherebbe assumere come accettabili livelli di conoscenze e abilità di base non coerenti con il tenore delle Indicazioni nazionali. Ma avrebbe un effetto ancora più negativo, ossia quello di indulgere ulteriormente verso un facilismo molto pericoloso, specie per i più svantaggiati poiché, esattamente come avviene in econo-

mia, l'inflazione, nel caso in esame dei titoli di studio, determina svantaggi maggiori per le persone meno abbienti e che provengono da contesti meno favorevoli. Tuttavia, ciò non toglie che si possa e forse si debba discutere sulla scelta del livello di accettabilità. Anzi, già solo l'apertura di un dibattito approfondito che entri nel merito di queste questioni sarebbe un risultato di straordinario valore per tutto il sistema educativo, forse per l'intero Paese (Mastrogiovanni *et al.*, 2021).

3.2. I livelli di risultato di Matematica

Un ragionamento in buona parte uguale all'Italiano riguarda anche la Matematica, anche se per quest'ultima il dibattito pubblico è ancora più limitato, se non addirittura quasi assente. La Tab. 2 ripropone i livelli descrittivi sintetici¹² di Matematica al termine del secondo ciclo d'istruzione.

Descrizione sintetica dei livelli di risultato INVALSI
<p>Livello 1. L'allievo/a utilizza conoscenze elementari e procedure di base, prevalentemente acquisite nella scuola secondaria di primo grado e, in parte, alla fine del primo biennio della scuola secondaria di secondo grado. Risponde a domande formulate in maniera semplice usando informazioni direttamente individuabili. Risolve problemi che coinvolgono contesti abituali e che richiedono procedimenti semplici.</p>
<p>Livello 2. L'allievo/a conosce le principali nozioni previste dalle Linee guida e dalle Indicazioni nazionali di matematica per il primo biennio della scuola secondaria di secondo grado. Risponde a domande che richiedono semplici elaborazioni sui dati disponibili (es. confrontare grafici di vario tipo). Risolve problemi in cui occorre scegliere opportunamente i dati dal testo e utilizzare conoscenze matematiche possedute fin dai gradi scolari precedenti.</p>
<p>Livello 3. L'allievo/a usa abilità di base acquisite nel corso della scuola secondaria di secondo grado e collega tra loro conoscenze fondamentali. Riconosce le proprietà dei principali oggetti matematici (es. figure geometriche, grafici e funzioni) e risolve problemi anche utilizzando equazioni e disequazioni elementari o semplici trasformazioni di formule. Riconosce, anche sotto forme diverse, modelli matematici semplici che rappresentano fenomeni o situazioni proposte (es. un modello di crescita lineare).</p>

Tab. 2 - *Descrizione sintetica dei primi tre livelli descrittivi di Matematica al termine del secondo ciclo d'istruzione.*
(Fonte, INVALSI 2019)

12. Si veda nota 8.

Bibliografia

- Barbaranelli, C., & Natali, E.** (2005). *I test psicologici: teorie e modelli psicometrici*, Roma: Carocci Editore.
- Blömeke, S., & Gustafsson J. E.** (2017). *Standard Setting in Education*. The Nordic Countries in an International Perspective. Editors: Blömeke, Sigrid, Gustafsson, Jan-Eric (Eds.).
- Cizek, G. J., & Bunch, M. B.** (2007). *Standard Setting. A guide to Establishing and Evaluating Performance Standards on Tests*. New York (USA): SAGE Publications.
- Carnoy, M.** (2015). *International Test Score Comparisons and Educational Policy*. A Review of the Critiques. Boulder, CO: National Education Policy Center. http://nepc.colorado.edu/files/pb_carnoy_international_test_scores_0.pdf [Accessed 30.09.2023].
- Council of Europe** (2009). Relating Language Examination to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). A manual. Language Policy Division, Strasbourg. www.coe.int/lang [Accessed 30.09.2023].
- Desimoni, M.** (2018). *I livelli per la descrizione degli esiti delle prove INVALSI*. (https://invalsi-areaprove.cineca.it/docs/2018/Livelli_INVALSI_g8.pdf) [Accessed 30.09.2023].
- Falorsi, P. D., Falzetti, P., & Ricci, R.** (2019). *Le metodologie di campo nto e scomposizione della devianza nelle Rilevazioni nazionali dell'INVALSI*, Milano: Franco Angeli, Open Access Catalog.
- Ferretti, F.** (2020). Le prove INVALSI CBT: riflessioni sulle prove della scuola secondaria di primo grado. *L'insegnamento della matematica e delle scienze integrate*, vol. 43, pp. 34-54, ISSN: 1123-7570.
- Giancola, O., & Salmieri, L.** (2023). *La povertà educativa in Italia*. Roma: Carocci Editore.
- Greenstein, L.** (2017). *La valutazione formativa*, Milano: UTET Università.
- INVALSI** (2018a). *Quadro di riferimento delle prove INVALSI di Italiano* (https://invalsi-areaprove.cineca.it/docs/file/QdR_ITALIANO.pdf) [Accessed 30.09.2023].
- INVALSI** (2018b). *Quadro di riferimento delle prove INVALSI di Matematica* (https://invalsi-areaprove.cineca.it/docs/file/QdR_MATEMATICA.pdf) [Accessed 30.09.2023].
- INVALSI** (2019). *I livelli dell'ultimo anno della scuola secondaria di secondo grado* (<https://www.invalsiopen.it/descrittori-livelli-grado-13/>) [Accessed 30.09.2023].
- Looney J. W.** (2011). Integrating Formative and Summative Assessment: Progress Toward a Seamless System?, *OECD Education Working Papers*, 58, OECD Publishing.
- Mastrogiovanni, A., Pieroni, L., & Vendramin, A.** (2021). *Alcuni recenti strumenti INVALSI per l'insegnamento dell'italiano*. Italiano a scuola, 3, 427-432.
- Rasch, G.** (1980). *Probabilistic models for some intelligence and attainment tests* (Copenhagen, Danish Institute for Educational Research), expanded edition with foreword and afterword by B.D. Wright. Chicago: The University of Chicago Press.
- Ricci, R.** (2021). Gli apprendimenti ai tempi della pandemia. *RicercaAzione*, 13 (2), 1-26.
- Ricci, R.** (2023). La dispersione scolastica e la povertà educativa. *Nuova Secondaria Ricerca*, 2 (XLI).
- Turner, R.** (2014). *Described proficiency scales and learning metrics*. Assessment GEMs no. 4. Melbourne, Australia: Australian Council for Educational Research (ACER).
- Van der Linden, W. J.** (2018). *Handbook of Item Response Theory*. Volume Three: Applications. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences Series.
- Verschoor, A.** (2007). *Genetic Algorithms for Automated Test Assembly*. PhD Thesis thesis, University of Twente.
- Wright, B. D., & Stone, M. H.** (1999). *Measurement essentials*. Wilmington, DE: Wide Range, Inc
-