UNIVERSITÀ DEGLI STUDI DI VERONA

DIPARTIMENTO DI SCIENZE UMANE

PROGRAMMA DOTTORALE IN EDUCATION

CONTRIBUTO FINANZIARIO INVALSI

Ciclo 38°

TITOLO TESI DI DOTTORATO

La descrizione dei livelli di competenza di comprensione del testo delle prove INVALSI nella scuola secondaria di primo e secondo grado: ricadute educative per docenti e studenti

(S.S.D. PAED-02/B Pedagogia sperimentale)

Coordinatore: Prof. Lorenzo Bernini

Tutor: Prof.ssa Roberta Silva

Co-tutor: Prof.ssa Alessia Bevilacqua

Dottoranda: Dott.ssa Antonella Mastrogiovanni

INDICE

Abstract 7

SUL PERCHÉ DI QUESTA RICERCA 11

Introduzione 11

PARTE I: QUADRO TEORICO DI RIFERIMENTO DELLA RICERCA 15

CAPITOLO 1: FUNZIONI E FORME DI VALUTAZIONE *SULLA* SCUOLA 15

1.1 La valutazione tra *assessment*, *evaluation* e *accountability* 15

1.1.1 Le indagini internazionali IEA e OCSE 17

1.1.2 Le indagini IEA 19

1.1.3 Le indagini dell'OCSE 22

1.1.4 Le indagini nazionali condotte dall'INVALSI 24

1.1.5 Il Sistema Nazionale di Valutazione 30

1.2 Le indagini nazionali e internazionali a confronto 32

1.2.1 Valutazione esterna e valutazione interna nel contesto scolastico 35

1.3 Gli aspetti etici della valutazione standardizzata nei contesti educativi 38

CAPITOLO 2: FUNZIONI E FORME DI VALUTAZIONE *NELLA* SCUOLA 43

2.1 Dalla pedagogia per obiettivi alla centralità dello studente 43

2.2 La valutazione nella legislazione scolastica italiana tra primo e secondo ciclo d'istruzione 45

2.3 Elementi di docimologia: le riflessioni sulle modalità di valutazione 48

2.4 Il rapporto tra misurazione e valutazione 51

2.5 Tra forme e funzioni valutative 55

2.6 Il feedback educativo: definizioni e indicazioni operative 56

2.7 Le prove INVALSI: dalla valutazione di sistema alla restituzione degli esiti al singolo studente 58

CAPITOLO 3: IL CONTESTO DELLA RICERCA 63

3.1 Il costrutto della prova di Italiano INVALSI 63

3.2 Attendibilità e validità della prova: dalla teoria psicometrica alla discussione critica 67

3.3 Dal punteggio numerico ai livelli descrittivi 71

3.4 I metodi di lavoro del gruppo di italiano per la costruzione delle prove nazionali 74

CAPITOLO 4: LA CORNICE METODOLOGICA DELLA RICERCA 77

4.1 Domande e obiettivi di ricerca 77

4.2 Il paradigma della ricerca: la visione costruttivista e la visione sistemica dell'azione didattica e valutativa 79

4.3 La metodologia della ricerca: lo studio di caso 86

4.3.1 Posizionamento della ricercatrice 89

4.4 Il disegno della ricerca 93

4.5 La selezione dei partecipanti alla ricerca 95

4.6 La metodologia della raccolta dati 98
4.6.1 La revisione sistematica della letteratura 98
4.6.2 L'analisi documentale 99
4.6.3 L'intervista semi strutturata 100
4.6.4 Le tracce delle interviste 103
4.6.5 La trascrizione delle interviste 106
4.7 La metodologia per l'analisi delle interviste 107
4.7.1 L'Analisi Tematica Riflessiva 107
4.8 La dimensione etica della ricerca 111
CAPITOLO 5: ANALISI E INTERPRETAZIONE DEI DATI 113
5.1 Revisione sistematica della letteratura: risultati 113
5.2 L'analisi documentale: risultati 122
5.3 Dalle Categorie Tematiche ai Temi Generali emergenti nelle interviste di docenti e studenti ... 130
5.3.1 Le Categorie Tematiche individuate nelle interviste di docenti e studenti 131
5.3.1.1 Analisi delle interviste dei docenti 131
5.3.1.2 Analisi delle interviste degli studenti 140
5.4 Definizione dei Temi Generali 144
CAPITOLO 6: DISCUSSIONE DEI RISULTATI 151
6.1 Lettura integrata dei risultati 153
6.2 Le percezioni di docenti e studenti rispetto agli esiti descrittivi 154
6.3 L'incidenza degli esiti descrittivi sui processi di insegnamento e apprendimento 156
6.4 Conclusioni 159
CAPITOLO 7: CONCLUSIONI 163
7.1 Limiti della ricerca e proposte future 165
BIBLIOGRAFIA 169
APPENDICE 1 Tabella di sintesi degli studi inclusi nella revisione sistematica 189
APPENDICE 2 – Moduli consenso informato 192

Dedico questo lavoro all'intero sistema di affetti di cui sono parte, Alessandro, Alessandra, Alba le famiglie a cui appartengo, a tutti gli amici di vita, ma in particolare alle mie figlie,
Elena e Caterina
perché vedano con chiarezza che una famiglia non serve a trattenere, ma a dare radici per poter andare. Vi auguro di non restare ferme, mai, di attraversare la vita con coraggio, curiosità e libertà, sapendo che appartenere non significa rinunciare a sé stesse.

Vorrei ringraziare tutte le persone che mi hanno accompagnata, supportata, sopportata, consolata, stimolata, formata. Ciascuno ha avuto un'importante funzione in questo lungo, intenso, articolato percorso. Ringrazio le mie amiche e amici colleghe/colleghi INVALSI tutte/i, in particolare Rita, Cristiana, Alessia, Antonella, Francesca, Laura, Marta, Monica, Stefania. Ringrazio anche Roberto e Cinzia per avermi dato questa opportunità. Ringrazio le mie amiche di quartiere Barbara, Ambra, Claudia. Ringrazio le mie compagne di viaggio del 38° ciclo e non solo, Elena, Giulia, Federica, Giada, Manuela, Letizia, Martina. Ringrazio le professoresse e i professori del collegio dottorale, in particolare le tutor che mi hanno seguita in questo percorso, prof.ssa Roberta Silva e prof.ssa Alessia Bevilacqua. Ringrazio tutti i/le docenti e gli/le studenti/studentesse che hanno contribuito alla realizzazione di questa ricerca.

Abstract

In Italia, a partire dal 2018, la valutazione nazionale degli apprendimenti relativa alla comprensione del testo prevede la restituzione di un feedback descrittivo individuale del livello raggiunto nelle prove standardizzate (Brown, 2022; Desimoni, 2018; DPR 28 marzo 2013, n. 80 – Articolo 6, comma 3) per ciascuno studente al termine dell'anno scolastico della scuola secondaria di primo e secondo grado. Questo passaggio, da una restituzione del dato esclusivamente a livello di sistema (nazionale, scuola, classe) a una restituzione anche al singolo studente, rappresenta un cambiamento di prospettiva per cui lo studente assume una posizione nuova, ancor più centrale perché la valutazione del sistema di istruzione non utilizza solo i suoi esiti come informazione del "benessere" del sistema, ma rimanda direttamente allo studente il livello raggiunto nella comprensione della lettura, rispetto al suo grado scolastico e rispetto a specifiche aree di conoscenze e abilità, in relazione alla normativa nazionale.

Con questo studio si è cercato di comprendere quali fossero le ricadute di tale descrizione individuale sull'azione didattica e scolastica. A partire da due domande di ricerca, 1. *Quali sono le percezioni degli studenti e dei docenti riguardo agli esiti descrittivi individuali delle prove nazionali standardizzate sulla comprensione del testo*? 2. *In che modo, a partire dall'esperienza dei soggetti coinvolti, studenti e docenti, gli esiti descrittivi delle prove standardizzate nazionali relativi ai livelli raggiunti nella prova di comprensione del testo possono incidere sui processi di insegnamento e apprendimento?,* è stata sviluppata una ricerca esplorativa di tipo qualitativo (Creswell & Poth, 2024; Denzin & Lincoln, 1994) contestualizzata all'interno di un paradigma costruttivista (Lincoln & Guba, 1985; Mortari, 2007) e una visione di tipo olistico-sistemico (Bateson, 1977; Morin, 2001) del contesto educativo. Il ricorso alla metodologia dello studio di caso, condotto nella cornice teorico-metodologica di Stake (1995), ha permesso di individuare una molteplicità di temi ricorrenti utili a inquadrare il fenomeno dal punto di vista di docenti e studenti.

Nel disegno di ricerca, sono stati coinvolti 24 partecipanti (15 docenti e 9 studenti della scuola secondaria di primo e secondo grado) che hanno preso parte a interviste semi-strutturate online, con l'obiettivo di raccogliere esperienze e vissuti utili a comprendere il fenomeno indagato. I dati raccolti attraverso tali interviste sono stati analizzati mediante l'Analisi Tematica Riflessiva (Braun & Clarke, 2006; 2019) e successivamente interpretati per formulare una risposta alle domande di ricerca. L'analisi tematica ha consentito di identificare sette temi principali: da un lato, il potenziale del feedback descrittivo come risorsa per la progettazione didattica, l'autovalutazione e i piani di miglioramento delle istituzioni; dall'altro, le criticità come la comunicazione istituzionale limitata, l'uso di linguaggio tecnico nei descrittori, le tempistiche di restituzione e la necessità di un'intensa

mediazione da parte del docente. I risultati suggeriscono che la descrizione dei livelli possa assumere significato nel contesto scolastico soprattutto quando è condivisa, discussa tempestivamente e integrata nell'attività didattica; per la maggior parte degli studenti, intervistati prima di ricevere il proprio esito, si tratta di valutazioni espresse a partire dalla lettura dei descrittori. Ne derivano possibili implicazioni sia per l'ente erogatore, che dovrebbe adottare azioni proattive per superare alcuni ostacoli all'effettiva fruizione di tale feedback all'interno del contesto scolastico rispetto alle potenzialità evidenziate dagli esiti descrittivi individuali, sia per le scuole, che hanno la possibilità di sviluppare interventi pedagogici capaci di attribuire un senso a ogni attività nel contesto educativo scolastico, inclusa l'integrazione della valutazione esterna come parte del processo educativo degli studenti.

Parole chiave: feedback descrittivo, percezioni insegnanti, percezioni studenti, INVALSI, test standardizzati.

Since 2018, Italy has provided individual descriptive reports on the level achieved in standardized reading comprehension tests at the end of the school year in lower and upper secondary schools (Brown, 2022; Desimoni, 2018; Presidential Decree No. 80 of 28 March 2013 – Article 6, paragraph 3). This shift from providing data exclusively at the system level (national, school, or class) to providing feedback to individual students represents a change in perspective, placing students at the heart of the system. This is because the evaluation of the education system uses their results not only as an indicator of the system's 'well-being', but also to inform students directly of their level of reading comprehension achievement in relation to their school year and specific areas of knowledge and skills, in accordance with national laws.

This study aimed to explore the implications of such individualized feedback for teaching and school practice. It began with two research questions: *1. How do pupils and teachers perceive the individual descriptive results of national standardized reading comprehension tests? 2. Based on the experiences of those involved, both pupils and teachers, how might the descriptive results of national standardized reading comprehension tests influence teaching and learning processes?* An exploratory qualitative study was developed and situated within a constructivist paradigm (Lincoln & Guba, 1985; Mortari, 2007) and a holistic-systemic view of the educational context (Bateson, 1977; Morin, 2001) (Creswell & Poth, 2024; Denzin & Lincoln, 1994). Using the case study methodology within Stake's (1995) theoretical and methodological framework made it possible to identify a variety of recurring themes that were useful for framing the phenomenon from the perspectives of teachers and students.

The research design involved 24 participants (15 teachers and 9 students from lower and upper secondary schools) who took part in semi-structured online interviews to share their experiences and insights to help understand the phenomenon under investigation. The data collected through these interviews were analyzed using reflexive thematic analysis (Braun & Clarke, 2006; 2019) and then interpreted to answer the research questions. Thematic analysis revealed seven key themes. Firstly, the potential of descriptive feedback as a resource for instructional design, self-assessment, and institutional improvement plans was identified. Secondly, critical issues such as limited institutional communication, technical language in descriptors, feedback timing, and intensive teacher mediation were highlighted. The results suggest that descriptions of levels may become meaningful in a school context especially when they are shared, discussed promptly and integrated into teaching activities; most students, who were interviewed before receiving their own results, expressed their views on the basis of the descriptors presented during the interview. Possible implications therefore emerge for both the examination board, which should proactively remove obstacles to the effective use of such feedback within the school context, and for schools, which can develop pedagogical interventions that give meaning to all activities within the school educational context, including integrating external assessment into students' educational processes.

Keywords: descriptive feedback, teachers’ perception, students’ perception, INVALSI, standardized test.

SUL PERCHÉ DI QUESTA RICERCA

Introduzione

Negli ultimi anni, la riflessione sulla valutazione nazionale standardizzata degli apprendimenti è evoluta, passando da una visione puramente focalizzata sulla rendicontazione sociale dei risultati a una prospettiva che la considera parte da integrare nei processi educativi. In questo contesto, la questione non riguarda soltanto gli strumenti utilizzati o la loro affidabilità, ma anche il significato che le pratiche di valutazione assumono nel processo di apprendimento e la loro interpretazione da parte dei soggetti coinvolti.

Questa prospettiva si basa sull'idea che la valutazione sia un aspetto specifico dell'azione educativa che contribuisce a dare forma all'esperienza scolastica. Secondo Dewey (1938) e Bruner (1996), infatti, per l'individuo un'esperienza è considerata educativa quando riesce ad assumere un significato all'interno di un percorso coerente e su cui è stato possibile riflettere. Questo principio è particolarmente rilevante nella valutazione nei contesti educativi, perché anche pratiche ben strutturate potrebbero risultare inefficaci dal punto di vista educativo se non vengono percepite come significative da chi le vive. I processi di insegnamento, apprendimento e valutazione non sono isolati, ma fanno parte di un sistema in cui ogni azione ha senso in relazione alle altre e al contesto. Bateson (1977) sottolinea che ogni processo educativo è collegato alle modalità di comunicazione e ai processi di attribuzione di significato, in modo che ciò che si apprende dipende anche dalle cornici interpretative che inquadrano le informazioni. Morin (2001) evidenzia che bisogna concepire l'educazione come un sistema complesso, in cui comprendere i fenomeni richiede l'integrazione tra le parti e il tutto, senza ridurli l'uno all'altro. In ambito valutativo, questa visione mostra che il valore del feedback non risiede solo nella sua struttura informativa, ma anche nella sua capacità di influenzare le pratiche didattiche, le rappresentazioni di chi è coinvolto e gli obiettivi educativi. Secondo questa prospettiva, Dominici (2011) afferma che, in sistemi complessi, il senso attribuito alle azioni è fondamentale affinché producano effetti educativi, evidenziando l'importanza di sviluppare competenze interpretative capaci di operare tra livelli e molteplici relazioni.

Nel contesto italiano, l'introduzione di forme di restituzione descrittiva individuale degli esiti delle prove dell'INVALSI si inserisce proprio nel tentativo di andare oltre l'impostazione della valutazione esterna come mera azione di *accountability*, aprendo la possibilità di riconoscerle anche una funzione all'interno del contesto educativo scolastico. Affiancare alla tradizionale restituzione degli esiti, basata su punteggi medi a livello di sistema, una restituzione individuale basata su descrittori del livello raggiunto nei compiti presenti nelle prove nazionali standardizzate rappresenta un

cambiamento importante, perché consente una lettura più approfondita degli esiti di apprendimento e stimola un utilizzo diverso dei dati derivati dalla valutazione esterna.

Restituire al singolo studente informazioni più dettagliate non garantisce automaticamente un processo di integrazione nelle pratiche educative. Il punto fondamentale è se queste informazioni vengano considerate, interpretate, condivise e utilizzate all'interno della scuola in modo allineato agli obiettivi di apprendimento. Da queste riflessioni muove l'esigenza di comprendere quanto e come la valutazione esterna si collega alla valutazione interna, non come la sommatoria di strumenti diversi, ma come la creazione di un sistema valutativo globale che accompagna l'azione didattica e favorisce la riflessione degli studenti sul proprio percorso di apprendimento. Le riflessioni di Looney (2011) forniscono un quadro teorico importante in questa direzione, sottolineando il fatto che la valutazione ha effetti educativi significativi solo quando è inserita in un sistema coerente di pratiche e i dati derivati dalle valutazioni esterne vengono reinterpretati e utilizzati per migliorare i processi di insegnamento e di apprendimento. Questa prospettiva focalizza l'attenzione non solo sugli strumenti di valutazione che producono tali esiti, ma anche sulle condizioni che favoriscono l'uso efficace di tali valutazioni e dei relativi esiti nei contesti scolastici.

Anche il paradigma dell'*assessment for learning* aiuta a riflettere sul ruolo del feedback individuale, sottolineando che esso è un elemento importante per supportare i processi di autoregolazione solo se lo studente è in grado di capirlo e di usarlo consapevolmente (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). In questo senso, capire il feedback non è un'operazione che avviene semplicemente perché è disponibile, perché viene restituito, ma è una competenza che dovrebbe essere acquisita attraverso azioni didattiche che partono dal chiarimento dei criteri di valutazione e incoraggiano la riflessione. Diversi contributi della tradizione pedagogica italiana hanno evidenziato il ruolo della valutazione come strumento fondamentale per comprendere i processi di apprendimento e, allo stesso tempo, utile alla regolazione dell'azione didattica (Visalberghi, 1984; Asquini et al., 2017; Corsini, 2018; Trinchero, 2014), sottolineando come essa richieda una mediazione professionale da parte del docente e una cultura valutativa condivisa.

Partendo da queste premesse, la presente ricerca mira a esplorare come studenti e docenti percepiscono e utilizzano la restituzione descrittiva individuale degli esiti delle prove dell'INVALSI e in che modo essa possa integrarsi nei processi di insegnamento e apprendimento. La decisione di intervistare docenti e studenti deriva quindi dal fatto che entrambi sono direttamente implicati nella costruzione del significato delle pratiche valutative nel contesto scolastico. Comprendere le loro percezioni e i loro vissuti è essenziale per determinare se e in che modo il feedback descrittivo (esito descrittivo) delle prove dell'INVALSI viene percepito come parte integrante dell'esperienza educativa o come un elemento esterno e senza rilevanza. La ricerca si colloca quindi tra l'analisi di

azioni determinate da policy valutative (restituzione degli esiti ai singoli studenti) e lo studio delle ricadute di tali azioni di policy sulle pratiche educative, con l'intento di comprendere le condizioni che consentono di trasformare la valutazione esterna in una risorsa per l'apprendimento. L'occasione di realizzare questo studio all'interno di un percorso formativo come quello dottorale nasce dall'opportunità di aver potuto dedicare parte della mia attività di ricerca a questo tipo di percorso. L'Istituto INVALSI, in accordo con l'Università di Verona, ha deciso di finanziare, tramite la stipula di una convenzione ad hoc, alcuni percorsi formativi di alto livello, con l'obiettivo di consolidare e approfondire aspetti metodologici utili allo sviluppo continuo delle attività affidate all'ente. Nel caso specifico, l'aspetto più importante ha riguardato l'approfondimento dell'uso di metodologie qualitative nello studio di fenomeni legati al contesto scolastico, con particolare attenzione alle questioni relative alla comprensione del testo. Il superamento della selezione mi ha permesso di sviluppare un progetto di ricerca che, raffinato nel tempo, ha raggiunto l'obiettivo di studiare un fenomeno da prospettive qualitative. È importante sottolineare la questione del tempo per inquadrare al meglio anche alcuni cambiamenti intercorsi nello sviluppo della ricerca, come ad esempio la selezione di un numero inferiore di partecipanti rispetto a quanto ipotizzato in fase di progettazione, poiché tale elemento ha rappresentato un vincolo alla realizzazione di determinate azioni. Svolgere, infatti, un percorso di dottorato in contemporanea allo svolgimento dell'attività lavorativa, pur stando entro una cornice di tempo codificata all'interno della convenzione, ha determinato la necessità di riorganizzare completamente i miei processi di lavoro entro l'ente, dato che ricopro ruoli di responsabilità, come ad esempio quello della costruzione, per l'appunto, delle prove standardizzate di Italiano per tutti i gradi scolastici. Nonostante ciò, grazie anche al supporto di chi, all'Università di Verona e all'interno dell'ente, aveva il compito di sostenermi scientificamente in questo percorso, è stato possibile realizzare questa ricerca esplorativa su un fenomeno di interesse personale, ma anche per il mio ente e, possibilmente, per la comunità educativa nel complesso.

PARTE I: QUADRO TEORICO DI RIFERIMENTO DELLA RICERCA

In questa prima parte del lavoro verranno affrontati temi relativi alla valutazione scolastica, a partire da come la valutazione esterna standardizzata svolga diverse funzioni, fino alla valutazione nella scuola, per comprendere confini e possibili punti di collaborazione tra questi due sistemi di valutazione, entrambi focalizzati sull'apprendimento dello studente.
In particolare, si farà riferimento alle principali indagini internazionali che coinvolgono gli studenti in obbligo scolastico, ma si accennerà anche alle indagini sulla popolazione adulta al fine di mostrare come le informazioni provenienti dalle indagini su larga scala possano informare i sistemi di istruzione e le politiche educative nell'ambito della valutazione comparativa.

CAPITOLO 1: FUNZIONI E FORME DI VALUTAZIONE *SULLA* SCUOLA

1.1 La valutazione tra *assessment*, *evaluation* e *accountability*

Quando si parla di "valutazione" in ambito educativo, spesso si intendono cose molto diverse, nel senso che riguardano piani diversi, come la valutazione che avviene in classe, nel rapporto quotidiano tra insegnamento e apprendimento, la valutazione che riguarda le scuole come organizzazioni e, infine, la valutazione che interroga la qualità di un intero sistema educativo. In merito a queste ultime due dimensioni della valutazione scolastica, negli ultimi decenni, nelle riflessioni sulla valutazione scolastica (Bunda, 1979: Linn, 2000; Teltemann & Jude, 2019; Wilson, 2004) si è assistito all'ingresso di termini di derivazione anglosassone come "*evaluation*", "*assessment*" e "*accountability*", che non presentano una perfetta corrispondenza con l'uso italiano di "valutazione" e che, proprio per questo, contribuiscono a rendere visibile una tensione di fondo insita nel termine, ovvero la stessa parola valutazione che può indicare un atto educativo orientato al miglioramento oppure un dispositivo di controllo pubblico e di regolazione delle istituzioni (Corsini, 2018).
A partire dal Novecento, la valutazione scolastica amplia progressivamente il proprio campo di riferimento. Accanto alla valutazione degli apprendimenti degli studenti, che resta una dimensione centrale della vita scolastica, si afferma infatti una prospettiva più ampia, orientata a interrogare anche il funzionamento della scuola come organizzazione e la capacità del sistema educativo di raggiungere gli obiettivi dichiarati. In questo passaggio, la valutazione non riguarda più soltanto ciò che avviene *nella* scuola, ma anche ciò che può essere osservato e compreso *sulla* scuola: la qualità dei processi di insegnamento, l'efficacia delle scelte organizzative, la coerenza tra obiettivi, pratiche e risultati, e la capacità delle istituzioni scolastiche di attivare percorsi di miglioramento. Questa evoluzione si

inserisce nel più generale sviluppo della cultura valutativa, che ha progressivamente assunto una funzione non solo tecnica, ma anche educativa, sociale e politica. La valutazione diventa così uno strumento di conoscenza, regolazione e responsabilizzazione, utile tanto per sostenere i processi decisionali interni alle scuole quanto per rendere conto del funzionamento dei sistemi educativi nel loro complesso (Stame, 2001; Castoldi, 2012; Mathison, 2005; Stockmann & Meyer, 2014). Questi nuovi punti di osservazione, che in Italia vengono presi in carico in anni successivi a esperienze già sperimentate in diversi contesti nazionali europei, diventano un terreno di ampio dibattito culturale poiché sono sostenuti da filosofie diverse (Trinchero, 2014; OECD, 2013).

In contesti di *assessment*, l'obiettivo principale è raccogliere evidenze sugli apprendimenti dello studente e sui processi che li supportano con l'obiettivo di capire come intervenire sui percorsi didattici e promuovere il miglioramento individuale (Castoldi, 2012; Weeden, Winter & Broadfoot, 2002). Con *evaluation* ci si riferisce invece a una prospettiva più ampia, che considera la valutazione come un processo di interpretazione e di giudizio finalizzato a comprendere la qualità di un intervento, di un programma o di un sistema. È un concetto che ha radici nella tradizione dell'*evaluation research* nordamericana (Scriven, 1967, 1991; Chelimsky, 2012) e che lega la valutazione a una funzione conoscitiva e decisionale. In questa prospettiva si analizzano l'efficacia, l'efficienza e la coerenza di un'azione educativa, adottando una postura interpretativa che va oltre la semplice misurazione degli esiti. Per quanto riguarda l'*accountability*, rientrano in questo ambito gli aspetti connessi alla rendicontazione sociale, nell'ambito della quale la diffusione pubblica dei risultati delle valutazioni esterne e delle decisioni conseguenti tende a trasformare tali esiti in parametri, talvolta espliciti ma spesso impliciti, per l'attribuzione di premi, sanzioni e meccanismi reputazionali (Kogan, 2022).

È importante, quindi, utilizzare i termini in modo appropriato per contestualizzare la funzione che la valutazione intende assumere, così da comprendere bene che cosa si valuta, per chi, con quali conseguenze e, soprattutto, con quale idea di scuola e per quali scopi. Nella riflessione di Corsini (2018) sulla cultura della valutazione, valutare significa sempre esprimere un giudizio di valore sulla distanza tra una situazione desiderata e una situazione osservata, con l'obiettivo di orientare le decisioni per ridurla. La valutazione in sé, infatti, non è mai un'azione neutra, perché non si limita a descrivere un esito, ma orienta i comportamenti futuri e definisce le priorità (Freire, 2014). Proprio per questo, comprendere, nell'ambito dei processi di valutazione, quale funzione svolga la valutazione in termini di *evaluation*, *assessment* e *accountability*, diventa un modo per non lasciare nell'implicito il problema delle finalità.

"Nel 2009, l'OCSE lancia il programma *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes* (OECD, 2013) per rispondere all'interesse suscitato dalle politiche educative proprio sull'*assessment* e sull'*evaluation* di studenti, insegnanti, dirigenti scolastici, scuole

e sistemi educativi. Dal *report* pubblicato nel 2013 (OECD, 2013) emerge che molti Paesi si sforzano di trasformare i propri sistemi educativi per preparare i giovani a vivere in società in rapido cambiamento.

Alcune tendenze politiche comuni possono essere osservate in una forma o nell'altra nella maggior parte dei Paesi OCSE, come ad esempio il decentramento e l'autonomia scolastica che creano una maggiore necessità di *evaluation* delle scuole, dei dirigenti scolastici e degli insegnanti, d'altro canto è in esame anche lo sviluppo e l'analisi di *assessments* degli studenti su larga scala e approcci di *assessment* individualizzati. I risultati derivanti da *evaluation* e *assessment* sono importanti per capire se un sistema scolastico sta funzionando e per ottenere un *feedback* per un eventuale ulteriore sviluppo. Tuttavia, i Paesi incontrano spesso difficoltà nell'attuazione delle procedure di *evaluation* e *assessment*, dovute a un'inadeguata progettazione delle politiche, alla mancanza di analisi delle conseguenze indesiderate, alla scarsa capacità degli agenti educativi di mettere in pratica le procedure, alla mancanza di una cultura della valutazione o a una scarsa utilizzazione dei risultati della valutazione." (De Santis, 2022, p. 138). Data la parte giocata da organizzazioni internazionali, come ad esempio l'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico- *Organisation for Economic Cooperation and Development*- OECD), nei processi di *evaluation* e *assessment*, in particolare attraverso le indagini su larga scala, si descriveranno di seguito tali indagini, al fine di contestualizzare quei processi valutativi nell'ambito dei cosiddetti LSA (*large scale assessment*).

1.1.1 Le indagini internazionali IEA e OCSE

Le indagini internazionali si collocano in modo diverso rispetto a questi tre approcci di valutazione. L'OCSE, attraverso PISA (*Program for International Student Assessment*), si colloca prevalentemente nell'ambito dell'*evaluation* e dell'*accountability*, poiché l'obiettivo non è comprendere lo sviluppo degli apprendimenti dei singoli studenti, bensì offrire ai diversi paesi partecipanti all'indagine una lettura comparata dell'efficacia dei propri sistemi educativi, contribuendo alla definizione delle politiche pubbliche. Come rilevano Grek (2009) e Ozga (2016), il PISA assume una funzione quasi "diagnostica" sui sistemi nel loro complesso, producendo indicatori che alimentano il dibattito politico e che possono innescare meccanismi di responsabilizzazione a livello macro. Le indagini IEA (*International Association for the Evaluation of Educational Achievement*), invece, sviluppano una prospettiva più vicina all'*assessment* e all'*evaluation* in senso pedagogico poiché nascono da un impianto *curriculum based.* Queste indagini comparative contribuiscono a comprendere in che misura i curricoli nazionali producono gli apprendimenti attesi. La loro funzione principale non è tanto l'*accountability* quanto il sostegno alla riflessione sui processi educativi attraverso dati comparabili e radicati nei contenuti disciplinari. Le

indagini promosse dalla IEA, come TIMSS e PIRLS, non si limitano a rilevare i risultati degli studenti, ma raccolgono anche informazioni sui curricoli, sui contesti scolastici e sulle pratiche di insegnamento. Per questa ragione, possono offrire indicazioni utili sia alle amministrazioni sia alle scuole per leggere i percorsi di apprendimento, individuare le aree di criticità e orientare eventuali interventi di miglioramento (Mullis & Martin, 2012, 2017). In questo senso, il loro valore non risiede soltanto nella comparazione internazionale, ma anche nella possibilità di mettere in relazione gli esiti degli studenti con le condizioni educative e organizzative in cui tali esiti si producono. Per comprendere meglio il contributo offerto dalle indagini comparative internazionali alle riflessioni sulla valutazione, si ripercorrono brevemente alcune tappe che hanno portato alle indagini dell'OCSE e dell'IEA come le conosciamo oggi.

A partire dagli anni '50, nel panorama internazionale degli studi pedagogici, psicologici e sociali sulla scuola, si costituisce un'associazione denominata IEA (*International Association for the Evaluation of Educational Achievement*) che segna un punto di svolta fondamentale nella storia della pedagogia moderna, trasformando l'istruzione da una questione puramente nazionale a un oggetto di studio scientifico a livello globale. In questo contesto emergono i primi quesiti relativi all'individuazione di caratteristiche in grado di spiegare le differenze tra i sistemi scolastici in termini di efficacia. Successivamente l'OCSE (*Organizzazione per la Cooperazione e lo Sviluppo Economico*) ha istituzionalizzato e ampliato questo approccio, trasformando l'indagine comparativa in un dispositivo stabile di governance internazionale dell'istruzione.

A partire dagli anni '90 i governi iniziano a richiedere dati affidabili, regolari e, soprattutto, comparabili a livello internazionale per orientare le decisioni politiche relative ai sistemi di istruzione su basi empiriche solide. Ed è negli ultimi decenni che il panorama educativo globale è stato radicalmente trasformato dall'affermarsi di una vera e propria cultura della valutazione comparativa, che diventa un elemento chiave nel processo di governance pubblica. I risultati delle rilevazioni non vengono utilizzati solo da ricercatori e studiosi dei fenomeni relativi all'educazione scolastica, ma vengono discussi nei parlamenti, nelle agenzie ministeriali e dai media. Questo attribuisce loro una funzione di rendicontazione internazionale, perché rendono visibile il funzionamento di un sistema educativo e lo mettono a confronto con il funzionamento degli altri (Damiani, Losito, & Sanzo, 2018). Un ulteriore elemento caratterizzante e di rilievo riguarda il fatto che gli esiti di queste rilevazioni comparative mettono in relazione i risultati degli studenti con altre caratteristiche socioeconomiche e culturali. L'individuazione di indicatori strettamente correlati alle performance degli studenti ha evidenziato la necessità di studiare azioni di contrasto a potenziali condizioni di disuguaglianza in partenza, da sviluppare direttamente nell'ambito del contesto scolastico. Nel panorama globale, le indagini internazionali hanno quindi assunto un ruolo crescente nell'identificare modelli di "qualità"

educativa, producendo nuove conoscenze sui processi formativi, con l'obiettivo di individuare le caratteristiche determinanti dei sistemi di istruzione più efficaci.

1.1.2 Le indagini IEA

Le indagini promosse dall'IEA si sono occupate di una pluralità sempre più ampia di discipline e di popolazioni scolastiche, dalla scuola primaria fino ai livelli più avanzati dell'istruzione secondaria e, in alcuni casi, anche della formazione degli insegnanti.

Per comprendere meglio le indagini condotte dalla IEA, è necessario considerare sia la loro evoluzione storica sia la loro struttura metodologica, in quanto tale relazione ne evidenzia le peculiarità nel panorama delle valutazioni su larga scala (Rutkowski, von Davier & Rutkowski, 2014). Fin dalle origini, lo scopo di queste rilevazioni non è stato solo misurare le competenze degli studenti, ma anche analizzare sistematicamente e in modo comparativo come queste siano distribuite nei diversi sistemi educativi e come variabili scolastiche, sociali e culturali influenzino il loro sviluppo. Verso la fine degli anni Cinquanta ad Amburgo, si decise di avviare un "esperimento naturale" condiviso, in cui i sistemi educativi nazionali costituivano delle varianti in un laboratorio diffuso in cui, con strumenti comuni, era possibile osservare gli effetti delle diverse scelte curricolari, organizzative e culturali (IEA, 2019; Martin & Mullis, 2012). Le prime indagini pilota e lo studio *First International Mathematics Study* (FIMS) introducono quindi anche un approccio alla ricerca in cui la misurazione degli apprendimenti è sempre accompagnata da un'attenzione ai contesti in cui tali apprendimenti si verificano, che caratterizza ancora oggi le indagini IEA (Purves, 1987). Questo aspetto è alla base della dimensione ricorsiva di tali indagini, che non si riferisce a una semplice ripetizione nel tempo, ma a un riprendere in modo intenzionale i medesimi costrutti, come la lettura, il pensiero matematico-scientifico, la cittadinanza e, più recentemente, la competenza digitale, per osservarli nel tempo e valutarne la comparabilità tra paesi (Mullis & Martin, 2019).

Questa impostazione risulta chiaramente evidente nelle indagini attuali. Ad esempio, nell'indagine *Trends in International Mathematics and Science Study* (TIMSS), la scelta di concentrarsi su matematica e scienze al quarto e all'ottavo anno non è casuale, poiché questi momenti segnano tappe chiave del percorso scolastico durante le quali si osservano sia l'acquisizione delle conoscenze di base sia il consolidamento delle competenze disciplinari. Le prove cognitive, progettate per riflettere i programmi nazionali, valutano direttamente ciò che gli studenti hanno studiato, includendo processi cognitivi complessi come il conoscere, l'applicare e il ragionare, basati su modelli teorici consolidati nel campo della valutazione educativa (Martin, Mullis, Foy & Hooper, 2016). Tuttavia, concentrarsi esclusivamente sui risultati delle prove comporterebbe la perdita di una contestualizzazione più ampia dei dati. I questionari di contesto somministrati agli studenti, agli insegnanti e ai dirigenti scolastici

forniscono dati sulle pratiche didattiche, sull'organizzazione del tempo, sulle risorse disponibili e sul contesto socioeconomico, consentendo analisi che mettono in relazione input, processi e risultati del sistema educativo (Foy et al., 2020). È proprio combinando questi livelli che i dati TIMSS risultano importanti per comprendere la situazione di ciascun paese.

Lo stesso discorso vale per l'indagine *Progress in International Reading Literacy Study* (PIRLS), che si concentra sulla competenza di lettura al quarto anno di scuola primaria, un momento di passaggio particolarmente delicato. Le prove cognitive hanno l'obiettivo di valutare la comprensione del testo attraverso diversi processi, tra cui il recupero di informazioni esplicite, la formulazione di inferenze dirette, l'interpretazione e l'integrazione delle informazioni e la valutazione critica del testo. Questa impostazione si colloca all'interno dei modelli internazionali di reading literacy, nei quali la lettura è intesa come un processo attivo di costruzione del significato e non come semplice decodifica (Mullis, Martin, Foy & Hooper, 2017). I questionari di contesto, invece, consentono di raccogliere informazioni sulle condizioni familiari, scolastiche e didattiche che possono sostenere o ostacolare lo sviluppo della competenza di lettura, come le pratiche di insegnamento, la disponibilità di materiali, il clima scolastico e le abitudini di lettura in famiglia. In questo modo, la lettura viene considerata un fenomeno situato, in cui interagiscono il lettore, il testo, l'attività di comprensione e il contesto socioculturale in cui tale attività prende forma (RAND Reading Study Group, 2002; Snow, 2002).

Anche l'analisi di indagini come l'*International Civic and Citizenship Education Study* (ICCS) mostra chiaramente che è importante combinare prove e questionari. Le conoscenze civiche e le capacità interpretative vengono misurate attraverso prove cognitive, mentre gli atteggiamenti, i valori e le forme di partecipazione vengono esplorati tramite questionari, in linea con i modelli teorici dell'educazione civica e della cittadinanza (Schulz et al., 2016). In questo caso, l'obiettivo è comprendere il ruolo della scuola nella formazione del cittadino. Questo rende i risultati particolarmente utili per le politiche educative orientate alla partecipazione democratica (Schulz, Ainley, Fraillon, Losito & Agrusti, 2018).

Una situazione simile si ritrova anche nello studio ICILS (International Computer and Information Literacy Study), dove le prove, somministrate in formato digitale, mirano a valutare le competenze degli studenti nell'uso critico delle tecnologie e nella gestione delle informazioni riguardanti vari aspetti della digitalizzazione, mentre i questionari raccolgono dati sull'accesso, sull'uso e sul contesto di apprendimento delle ICT (*Information and Communication Technology*) (Fraillon et al., 2020). In questo contesto, l'integrazione tra dati cognitivi e contestuali si fonda su una teoria che vede la competenza digitale come il risultato di opportunità di apprendimento disponibili sia nel contesto scolastico che nella vita di tutti i giorni.

Infine, studi come TEDS-M (*Teacher Education and Development Study in Mathematics*), SITES (*Second Information Technology in Education Study*) e REDS (*Responses to Educational Disruption Survey*) estendono ulteriormente il campo di studio. Nell'indagine TEDS-M, le prove cognitive hanno il compito di misurare le conoscenze disciplinari e pedagogiche dei futuri insegnanti, mentre i questionari di contesto raccolgono informazioni sui loro percorsi formativi. L'obiettivo di questa doppia raccolta di dati è stato analizzato da Tatto e altri autori nel 2012, i quali hanno evidenziato che la formazione iniziale influisce sulla qualità dell'insegnamento.

Altre indagini, come SITES (*Second Information Technology in Education Study*) e REDS (*Responses to Educational Disruption Survey*), utilizzano strumenti di tipo *survey* per raccogliere informazioni sui contesti scolastici e sulle pratiche adottate a livello istituzionale. L'attenzione in queste indagini non si concentra quindi solo sui risultati degli studenti, ma si estende anche ai processi organizzativi delle scuole e alle modalità con cui i sistemi educativi reagiscono a situazioni di cambiamento o di crisi. Attraverso questo tipo di rilevazioni è possibile osservare come le scuole gestiscono le risorse, organizzano l'attività didattica e rispondono a sfide emergenti (IEA, 2021).

Tutte queste indagini si svolgono all'interno di una tradizione consolidata di valutazione su larga scala che considera i dati come sistemi complessi il cui significato emerge solo se si tiene conto della relazione tra ciò che viene misurato e il contesto in cui i dati si producono (Rutkowski et al., 2014). I risultati delle prove cognitive forniscono un'infrastruttura confrontabile, mentre i questionari consentono di sviluppare modelli di interpretazione dei dati più elaborati. La loro integrazione consente un uso multidisciplinare delle informazioni, utile sia a supportare le politiche educative sia alla ricerca empirica. Questa dinamica di raccolta dei dati, ripetuta in ogni ciclo successivo, costituisce la costruzione di un sapere comparativo sui sistemi scolastici.

Le ricerche della IEA hanno come obiettivo primario verificare in che misura gli studenti abbiano appreso ciò che i curricoli nazionali prevedono, adottando dunque un disegno fortemente *curriculum based*. Questa scelta implica la costruzione di prove basate sui riferimenti curricolari condivisi dai Paesi partecipanti e consente di leggere gli esiti in relazione alla coerenza tra ciò che si insegna e ciò che viene effettivamente appreso, nel rispetto della specificità dei sistemi formativi adottati dai Paesi partecipanti all'indagine. Questa prospettiva di indagine, come sottolineato da diversi autori (Benvenuto, 2020; Mullis & Martin, 2017), valorizza la dimensione pedagogica dei curricoli nazionali e fornisce informazioni più direttamente utilizzabili dalle scuole e dagli insegnanti, in quanto basate sui contenuti disciplinari e sui processi didattici effettivamente adottati.

1.1.3 Le indagini dell'OCSE

Le indagini promosse dall'Organizzazione per la cooperazione e lo sviluppo economico (OCSE) rappresentano, insieme a quelle dell'IEA, il secondo grande pilastro della valutazione comparativa internazionale, ma si distinguono per il fatto che, pur condividendo l'impianto comparativo, orientano la misurazione in una direzione in parte diversa. L'impegno dell'OCSE nella valutazione degli apprendimenti arriva con forza più tardi rispetto alle indagini IEA, tra la fine degli anni Novanta e l'inizio del nuovo millennio, periodo in cui i dati iniziano a trovare spazio nella pubblicazione annuale sugli indicatori di *Education at a Glance*, avviata nel 1992 (Bottani & Tuijnman, 1994). Divenne chiaro, infatti, che non era più sufficiente sapere quanto si investisse o come fossero organizzate le scuole, ma che servissero anche ulteriori dati per completare le informazioni sui sistemi educativi relativi agli apprendimenti degli studenti che permettessero di comprendere cosa gli studenti fossero effettivamente in grado di fare con ciò che avevano appreso. Nasce così l'idea di misurare in modo autonomo non solo quanto gli individui sanno, ma anche quanto sono pronti ad affrontare le sfide della vita adulta, un obiettivo più grande.

È da questi presupposti che, negli anni 2000, nasce il *Programme for International Student Assessment* (PISA), uno degli studi più grandi e influenti nel campo dell'istruzione comparata, che viene ripetuto, nella stessa logica delle indagini IEA, ogni tre anni. La scelta di concentrarsi sugli studenti di quindici anni, a prescindere dal grado scolastico che frequentano, è stata pensata per focalizzare l'attenzione dello studio su un preciso momento dell'istruzione, vicino alla fine dell'obbligo scolastico nella maggior parte dei Paesi. Le prove cognitive, in queste rilevazioni, non sono pensate per valutare l'aderenza ai programmi nazionali, spesso eterogenei e complessi da confrontare, ma per misurare competenze considerate fondamentali, quali la lettura, la matematica e le scienze, con l'aggiunta, successivamente, anche di altri ambiti collegati a competenze trasversali. La definizione di competenza proposta da questa indagine si fonda sul concetto di *literacy*, inteso come capacità di utilizzare conoscenze e abilità in contesti reali, di interpretare informazioni complesse e di prendere decisioni in contesti non scolastici (OECD, 2019; UNESCO, 2006). Come sottolineano autori quali Schleicher (2019) e Lundahl & Waldow (2009), questa impostazione si lega all'intento dell'OCSE di leggere l'istruzione in relazione ai processi economici e sociali globali, producendo, come già detto, indicatori che consentano ai decisori politici di orientare le riforme strutturali verso lo sviluppo di competenze ritenute fondamentali per la partecipazione attiva alla società della conoscenza.

Questo approccio comporta una modifica sostanziale anche nella progettazione delle prove cognitive. I compiti proposti agli studenti mirano a osservare la capacità degli studenti di operare in situazioni che simulano problemi della vita quotidiana, più che a verificare l'apprendimento di contenuti

disciplinari specifici. Le prove valutative sono, infatti, metodologicamente orientate alla valutazione di competenze trasversali per affrontare compiti complessi anche in contesti non scolastici. Anche nelle indagini OCSE, oltre alle prove cognitive, vengono somministrati questionari rivolti principalmente agli studenti e ai dirigenti scolastici e, in alcune edizioni, anche agli insegnanti e alle famiglie. Questi questionari sono somministrati per raccogliere diverse informazioni relative al background socioeconomico e culturale, al clima scolastico, alle pratiche didattiche, alle aspettative e al benessere degli studenti. I risultati delle prove cognitive, contestualizzati all'interno di questo quadro interpretativo più ampio, forniscono informazioni più specifiche che consentono analisi approfondite su temi come le disuguaglianze, l'identificazione dei fattori associati alle prestazioni e la costruzione di modelli esplicativi utili a comprendere il funzionamento dei sistemi educativi.

Oltre all'indagine PISA, l'OCSE ha sviluppato ulteriori indagini internazionali che coinvolgono anche la popolazione adulta, contribuendo a delineare un quadro dell'apprendimento lungo tutto l'arco della vita. Il *Programme for the International Assessment of Adult Competencies* (PIAAC), in precedenza *International Adult Literacy Survey* (IALS) e *Adult Literacy and Life Skills Survey* ALL, si rivolge agli adulti (16 – 65 anni) e misura competenze fondamentali di tipo funzionale quali la capacità di lettura, di calcolo e di risoluzione dei problemi, mettendo in relazione le abilità possedute dagli individui con i percorsi formativi (formali, non formali e informali), lavorativi e le pratiche di vita quotidiana (Schleicher, 2008).

Un'ulteriore indagine che non riguarda la popolazione studentesca, ma resta nell'ambito dei contesti scolastici, è il *Teaching and Learning International Survey* (TALIS), che si concentra su insegnanti e dirigenti scolastici. In questa indagine vengono somministrati questionari per raccogliere informazioni sulle pratiche didattiche, sullo sviluppo professionale e sulle condizioni di lavoro, spostando l'attenzione sulla modalità di realizzazione dell'insegnamento.

Tutte le indagini OCSE, in generale, mostrano un approccio che pone l'accento sulla dimensione funzionale dell'apprendimento e sul rapporto tra educazione, partecipazione sociale e sviluppo economico. Le amministrazioni dei vari paesi partecipanti alle indagini utilizzano i dati raccolti per orientare le politiche pubbliche, soprattutto quando i risultati segnalano problemi inattesi. In questo senso, quindi, le indagini OCSE, oltre a descrivere i sistemi educativi, contribuiscono attivamente alla loro trasformazione, promuovendo modelli di *policy* educative basati su evidenze empiriche e rafforzando le pratiche di *accountability*, e contribuiscono a ridefinire le priorità e le strategie dei sistemi scolastici. Nel caso di PISA, l'attenzione non è rivolta soltanto alla padronanza di contenuti disciplinari, ma alla capacità degli studenti di utilizzare conoscenze e abilità in situazioni realistiche e non necessariamente coincidenti con i contesti scolastici ordinari. L'impianto dell'indagine, infatti, mira a verificare in che misura gli studenti quindicenni siano in grado di trasferire quanto appreso a

problemi e situazioni nuove, dentro e fuori la scuola (OECD, 2019). Questa impostazione rafforza una concezione della competenza come capacità d'uso e di adattamento delle conoscenze, più che come semplice possesso di saperi disciplinari. In prospettiva critica, Mattei e Vertecchi (2016) osservano che, nelle grandi indagini comparative, la centralità attribuita alle competenze possa comportare il rischio di indebolire il riferimento ai contenuti culturali e ai canoni disciplinari dell'educazione. Si vedrà di seguito come, nel sistema scolastico italiano, le indagini dell'INVALSI si inseriscano nel panorama delle indagini condotte sugli studenti italiani.

1.1.4 Le indagini nazionali condotte dall'INVALSI

La costruzione di un sistema nazionale di valutazione in Italia non è frutto di un atto improvviso, ma si inserisce in un processo più ampio di trasformazione del sistema educativo avviato alla fine degli anni Novanta. Negli stessi anni in cui prende forma l'autonomia scolastica, con il DPR 275/1999, cresce la consapevolezza che, a fronte di maggiore autonomia, sono necessari strumenti di monitoraggio in grado di rendere visibili i risultati e di orientare le politiche pubbliche. In questo quadro si colloca il progressivo ripensamento del ruolo del CEDE, istituito negli anni Settanta con funzioni prevalentemente di ricerca, documentazione e supporto allo sviluppo del sistema scolastico. Con il Decreto legislativo 20 luglio 1999, n. 258, il Centro Europeo dell'educazione viene trasformato nell'Istituto nazionale per la valutazione del sistema dell'istruzione, segnando un passaggio importante verso la costruzione di una struttura nazionale dedicata alla valutazione del sistema scolastico. Il decreto attribuisce infatti al nuovo Istituto il compito di valutare l'efficienza e l'efficacia del sistema di istruzione, anche con riferimento alle singole istituzioni scolastiche, e di collocare la valutazione nazionale nel più ampio contesto internazionale.

Tale processo è ulteriormente definito dal Decreto legislativo 19 novembre 2004, n. 286, che istituisce il Servizio nazionale di valutazione del sistema educativo, dell'istruzione e della formazione e riordina l'INVALSI. Il provvedimento pone le basi per un servizio valutativo nazionale orientato al progressivo miglioramento e all'armonizzazione della qualità del sistema educativo, con l'obiettivo di valutarne l'efficienza e l'efficacia. Il decreto affida all'INVALSI il compito di predisporre prove standardizzate per la valutazione degli apprendimenti e di contribuire alla valutazione complessiva del sistema educativo, con l'intento di fornire informazioni comparabili sia a livello di sistema scolastico in generale sia a livello di singola scuola. In particolare, gli esiti restituiti alle singole scuole si prefiggono l'obiettivo di integrare le valutazioni interne e di supportare le decisioni di istituto basate su evidenze empiriche, volte a azioni di miglioramento. Questa linea si inserisce nella crescente importanza, a livello internazionale, delle indagini comparative come quelle dell'OCSE e dell'IEA, che negli stessi anni influenzano in modo significativo le politiche nazionali (Hopmann, 2008;

Breakspear, 2012). L'avvento delle prove standardizzate nazionali per la valutazione dei contesti scolastici diventa un tema di dibattito nazionale, in particolare in merito al ruolo che viene loro attribuito. Trinchero (2014) sottolinea che le prove dell'INVALSI sono state create non per classificare, ma come strumenti per comprendere la posizione di una classe o di una scuola in un quadro più ampio, sempre rispetto a contesti simili. In questa direzione si collocano anche le riflessioni di Lucisano e Corsini (2015) e di Corsini (2018), che richiamano la necessità di non ridurre la valutazione delle scuole alla sola misurazione dei risultati. La qualità di un sistema valutativo, infatti, non dipende esclusivamente dall'accuratezza tecnica delle prove, ma anche dal modo in cui i dati vengono letti, discussi e utilizzati all'interno delle istituzioni scolastiche. Quando gli esiti delle rilevazioni sono interpretati come classifiche o come indicatori immediati dell'efficacia dell'insegnamento, aumenta il rischio di produrre letture semplificate dei processi educativi e di orientare la didattica verso ciò che è più facilmente misurabile. In questi casi, le prove possono favorire forme di adattamento strategico, come la riduzione del curricolo ai contenuti oggetto di rilevazione o attività finalizzate prevalentemente al superamento dei test.

Un uso più formativo dei dati richiede, invece, che i risultati siano inseriti in pratiche collegiali di analisi e interpretazione. In questa prospettiva, le informazioni ricavate dalle prove standardizzate non sostituiscono il giudizio professionale dei docenti, ma possono contribuire ad alimentare percorsi di ricerca didattica, di confronto interno e di revisione delle pratiche educative.

Successivamente, con l'istituzione del Sistema Nazionale di Valutazione, attraverso il DPR 28 marzo 2013, n. 80, si propone un modello di valutazione dei sistemi scolastici che si prefigge di tenere assieme sia la dimensione dell'autovalutazione dei singoli istituti scolastici (tramite il RAV – Rapporto di Autovalutazione), sia la dimensione della valutazione esterna, attraverso osservazioni dirette da parte del corpo ispettivo ministeriale nelle singole istituzioni scolastiche, sia il supporto alle azioni di miglioramento e alla rendicontazione sociale. In questa prospettiva, l'INVALSI assume un ruolo più ampio rispetto alla sola produzione di prove, diventando uno dei pilastri del sistema valutativo nazionale, assieme all'INDIRE e al personale ispettivo ministeriale. La normativa crea nuove opportunità, ma anche preoccupazioni. Da un lato, la valutazione del sistema aiuta a verificare che le scuole lavorino in modo equo e a trovare una definizione comune su cosa significa qualità educativa. Dall'altro lato, però, c'è il rischio che la valutazione venga percepita come un controllo dall'alto, riducendo la rendicontazione a una semplice formalità e diminuendo il valore dell'autovalutazione. De Santis e Asquini (2020) evidenziano che il RAV diventa uno strumento realmente formativo solo quando le scuole sviluppano competenze interpretative e adottano un approccio riflessivo, capace di collegare i risultati delle rilevazioni nazionali standardizzate alla vita quotidiana della didattica.

All’interno del Sistema Nazionale di Valutazione, l’INVALSI svolge un ruolo centrale, occupandosi della progettazione e della realizzazione delle rilevazioni standardizzate degli apprendimenti, che costituiscono il principale strumento per la raccolta di dati empirici sul funzionamento del sistema educativo. Tali rilevazioni si prefiggono di misurare alcuni aspetti legati agli apprendimenti attraverso prove cognitive costruite sulla base di quanto indicato nei Quadri di riferimento predisposti per ciascuna prova (Mattei et al, 2016; De Santis et al, 2024). Questi ultimi definiscono in modo sistematico gli oggetti di valutazione, le competenze attese e i processi cognitivi coinvolti. Le prove riguardano gli apprendimenti scolastici relativi all’italiano (in termini di padronanza linguistica relativa alla comprensione di testi scritti), alla matematica e alla lingua inglese e sono somministrate con cadenza annuale all’intera popolazione studentesca di specifici gradi scolari, in particolare nelle classi II e V della scuola primaria (inglese solo in V primaria), nella classe III della scuola secondaria di primo grado e nelle classi II e V della scuola secondaria di secondo grado, garantendo una copertura censuaria e una comparabilità sistematica dei risultati nel tempo e tra contesti territoriali (De Santis et al, 2024).

Le rilevazioni si svolgono ogni anno in un arco temporale compreso tra marzo e maggio e, grazie alla stabilità metodologica del sistema, consentono di costruire serie storiche e analizzare l’evoluzione degli apprendimenti nel tempo. Gli esiti delle prove vengono restituiti secondo una scansione temporale differenziata che riflette le diverse finalità del sistema valutativo. Una prima restituzione a livello nazionale e generale avviene generalmente nel mese di luglio, con la pubblicazione dei principali risultati (basati su un campione nazionale rappresentativo a cui sono state somministrate le prove da personale non afferente alla scuola ma individuato e formato da INVALSI) e indicatori utili per l’analisi del sistema educativo. Successivamente, a partire dal mese di settembre, le scuole ricevono gli esiti dettagliati relativi alla propria situazione scolastica e ulteriori elementi di approfondimento (posizionamento rispetto a scuole con le medesime caratteristiche territoriali, esiti relativi al valore aggiunto prodotto dalla scuola, confrontati con scuole simili per contesto socioeconomico) che possono essere utilizzati nei processi di autovalutazione e di progettazione di azioni di miglioramento.

Anche le rilevazioni nazionali sono strutturate con somministrazioni di prove cognitive e di questionari per la raccolta dei dati di contesto con gli stessi obiettivi già menzionati nella trattazione delle indagini internazionali. In particolare, le informazioni di contesto per la scuola primaria sono raccolte prevalentemente attraverso le segreterie scolastiche, che si occupano di far compilare i questionari alle famiglie (non in modo obbligatorio), e solo poche altre informazioni sono raccolte direttamente dagli studenti durante la somministrazione delle prove cognitive. Nel caso dei livelli scolastici superiori gran parte delle informazioni di contesto sono raccolte tramite la

somministrazione contestuale dei questionari durante la fase di somministrazione delle prove cognitive (https://www.invalsi.it/rilevazioni-nazionali/). I dati raccolti sul background socioeconomico e culturale degli studenti permettono appunto di integrare dati cognitivi e dati di contesto e consentono di costruire indicatori complessi che permettono di analizzare non solo i livelli di apprendimento, ma anche i fattori che li influenzano. Tra questi, assumono particolare rilievo gli indicatori di equità, che mettono in relazione i risultati degli studenti con il loro background socioeconomico o con elementi di variabilità entro la scuola. La disponibilità di dati sistematici consente inoltre di adottare modelli di analisi più sofisticati, tra cui quelli basati sulla stima del valore aggiunto delle istituzioni scolastiche, che stimano il contributo delle scuole agli apprendimenti tenendo conto delle condizioni di partenza. Tale approccio consente di distinguere tra il contributo del contesto e quello della scuola ai risultati degli studenti, offrendo una rappresentazione più articolata della qualità educativa. La letteratura nazionale sul tema dell'equità scolastica ha evidenziato la necessità di leggere i risultati degli studenti tenendo conto delle disuguaglianze di partenza e dei diversi contesti in cui le scuole operano (Benadusi & Giancola, 2014). In questo contesto si inserisce anche l'uso di indicatori di valore aggiunto, che mirano a fornire un'interpretazione più articolata e meno immediatamente classificatoria degli esiti scolastici. La stima del valore aggiunto consente infatti di distinguere, almeno sul piano analitico, ciò che può essere ricondotto alle condizioni iniziali degli studenti da ciò che può essere attribuito al contributo specifico della scuola. Tali indicatori consentono quindi di andare oltre il semplice confronto tra punteggi grezzi, poiché considerano fattori esterni all'azione educativa, come il contesto socioeconomico e le caratteristiche della popolazione scolastica. Allo stesso tempo, il loro impiego richiede cautela: si tratta di misure fondate su modelli statistici e su assunzioni metodologiche che devono essere rese esplicite e interpretate con attenzione, evitando letture automatiche o classificatorie delle scuole (Ricci, 2008; Martini, 2018). La restituzione dei dati INVALSI può assumere una funzione diagnostico-formativa quando viene utilizzata dalle scuole come occasione per la lettura interna dei risultati e per la riflessione sulle pratiche didattiche. In questo senso, i dati non andrebbero considerati misure definitive della qualità scolastica, ma elementi informativi da interpretare alla luce del contesto, della composizione degli studenti e delle scelte organizzative e didattiche dell'istituto. I materiali INVALSI dedicati alla lettura dei risultati insistono proprio su questa funzione di supporto all'autovalutazione e al miglioramento, precisando che i dati non devono essere letti in chiave punitiva o classificatoria (INVALSI, 2024b; Ricci, 2020).

Questa esigenza di cautela è particolarmente rilevante nel caso degli indicatori di valore aggiunto. Ricci (2008) evidenzia come tali modelli possano offrire informazioni utili sull'effetto della scuola, ma solo a condizione che siano chiari i presupposti metodologici su cui si fondano e i limiti della loro

interpretazione. Se utilizzati in modo meccanico, infatti, essi possono produrre rappresentazioni parziali o distorte dei risultati scolastici e incoraggiare confronti poco fondati tra scuole diverse.
Sulla stessa linea si inseriscono gli studi di Grimaldi e Serpieri (2016), per cui le metriche, come il valore aggiunto, presentano anche criticità, soprattutto quando entrano nel dibattito pubblico scolastico. Il rischio principale è usare una misura complessa come se fosse un indicatore diretto e trasparente della qualità scolastica, pur trattandosi in realtà di un costrutto statistico che richiede competenze interpretative, conoscenza dei modelli di regressione multilivello e una chiara consapevolezza dei suoi limiti. Corsini (2018) sottolinea l'importanza di questo aspetto: affinché il valore aggiunto possa effettivamente favorire il miglioramento, deve essere supportato da una cultura valutativa robusta, che lo utilizzi come strumento di riflessione interna piuttosto che come un meccanismo di sorveglianza.
Le prove cognitive costituiscono il cuore della valutazione degli apprendimenti e sono progettate per misurarli in una prospettiva che integra conoscenze disciplinari e processi cognitivi. Nel caso della prova di Italiano, la valutazione si concentra sulla competenza di lettura, intesa come capacità di comprendere, interpretare e riflettere su testi scritti, articolata in diversi livelli che includono la comprensione lessicale, la capacità inferenziale e la riflessione sulla lingua e sulla forma e contenuto del testo (De Santis & Mastrogiovanni, 2024; De Santis et al, 2024). Per la matematica, la valutazione integra il dominio dei contenuti e la capacità di ragionamento, proponendo compiti che richiedono di interpretare situazioni, risolvere problemi e utilizzare concetti matematici in contesti applicativi. La prova di inglese, costruita in riferimento al Quadro Comune Europeo di Riferimento per le Lingue, misura prevalentemente le competenze ricettive di comprensione del testo e di ascolto, in relazione a livelli di competenza standardizzati (https://www.invalsi.it/rilevazioni-nazionali/).
L'ultimo inquadramento normativo relativo al sistema di valutazione esterna delle scuole si riferisce al Decreto legislativo 13 aprile 2017, n. 62, in cui si introduce l'obbligatorietà dello svolgimento della prova INVALSI per gli studenti delle classi dell'ultimo anno della scuola secondaria di primo e secondo grado ai fini dello svolgimento dell'Esame di Stato e le restituzioni individuali degli esiti ai singoli studenti sotto forma i livelli descrittivi. La restituzione individuale è prevista solo per alcuni gradi scolastici, alla fine della scuola secondaria di primo e secondo grado, e si integra con i percorsi di certificazione delle competenze e con il curriculum dello studente (De Santis, Mastrogiovanni, Resio, & Vendramin, 2024). Un elemento distintivo dell'impianto valutativo strutturato dall'Istituto INVALSI è l'attenzione alla restituzione degli esiti individuali, articolati in livelli descrittivi empirici relativi alla difficoltà dei compiti proposti, che hanno consentito la costruzione di scale in grado di rappresentare in modo graduato i livelli di possesso dei contenuti oggetto della prova (Desimoni et

al., 2020). In questa prospettiva, le prove non si limitano a verificare il possesso di conoscenze, ma mirano a rilevare la capacità di utilizzarle in contesti significativi.

I dati e gli indicatori raccolti dall'INVALSI sono fondamentali per gli strumenti operativi del Sistema Nazionale di Valutazione, poiché supportano la stesura del Rapporto di Autovalutazione, la pianificazione dei Piani di Miglioramento e le attività di rendicontazione sociale (De Santis & Asquini, 2020). In questo modo, le rilevazioni sugli apprendimenti diventano parte integrante di un sistema di governance basato su evidenze, dove la valutazione svolge sia una funzione descrittiva che diagnostica e regolativa, mirando a migliorare qualità e equità del sistema educativo.

Un disegno valutativo così strutturato necessita di prove standardizzate che, in quanto tali, permettono di raccogliere dati comparabili, semplificando però al tempo stesso i contenuti dell'apprendimento, favorendo quegli aspetti che sono più facilmente misurabili rispetto alla complessità dei contenuti ritenuti rilevanti dal punto di vista pedagogico. Una nota critica alle prove standardizzate nazionali nasce dal confronto con una caratteristica delle prove delle indagini internazionali IEA e OCSE (che prevedono rilevazioni campionarie e a cadenza triennale o quadriennale) che includono molte domande aperte, a differenza delle prove INVALSI, in particolare quelle di italiano, che tendono a usare domande prevalentemente a risposta chiusa riducendo quindi la possibilità di misurare processi cognitivi elevati (Corsini, 2018; Trinchero, 2014; Le Rose & Ricciardi, 2021). La natura censuaria e la cadenza annuale delle somministrazioni, purtroppo, limitano l'utilizzo di tali formati di domande, che richiedono processi di correzione lunghi e articolati (Yamamoto et al., 2017; McCaffrey et al., 2021). Proprio perché in presenza di questa condizione strutturale, l'ente dedica molto tempo alla formazione per la costruzione di test e domande standardizzate di tipo chiuso, cercando di coniugare gli aspetti di metodo con le necessità di contenuto (De Santis & Mastrogiovanni, 2024). Autori come Haladyna, Downing e Rodríguez (2002) hanno evidenziato che un item a scelta multipla può stimolare abilità di livello superiore, come il ragionamento inferenziale o la comprensione profonda di un testo, purché sia costruito secondo criteri metodologici rigorosi. Anche nel contesto nazionale, autori come Vertecchi (2010) e Calvani (2009) hanno sostenuto che i quesiti a risposta chiusa sono in grado di misurare gli apprendimenti e, se ben costruiti dal punto di vista metodologico, sono capaci di attivare processi di selezione, confronto, analisi e discriminazione, processi cognitivamente sofisticati. Altri autori, come Pellegrino, Chudowsky e Glaser (2001), affermano nella loro revisione per il *National Research Council* che la validità cognitiva di un test dipende dall'allineamento tra il modello teorico dell'apprendimento e il modo in cui l'item elicita la prestazione. Un item a risposta chiusa può, ad esempio, essere progettato per valutare abilità interpretative molto sofisticate, soprattutto quando presenta testi, grafici, scenari o problemi complessi che richiedono integrazione di informazioni o inferenze multilivello. Le prove INVALSI di italiano e matematica degli ultimi

anni, come evidenziato dalle analisi (INVALSI, 2022), hanno progressivamente aumentato la presenza di item che richiedono processi cognitivi complessi, mantenendo comunque il formato chiuso necessario per una correzione oggettiva delle prove in tempi ristretti rispetto ai modelli di valutazione delle indagini comparative internazionali.

Un rischio evidente nell'uso di prove standardizzate per valutare gli apprendimenti, come osservato da Carbrey (2024), è che se da un lato forniscono dati utili, tendono anche a contribuire a un restringimento del curricolo e a semplificare i processi cognitivi valutati, con effetti negativi sugli apprendimenti più complessi e meno misurabili. Quando la pressione valutativa aumenta, il rischio è che l'insegnamento si focalizzi maggiormente sui contenuti più facilmente valutabili. In ogni caso, Shepard (2000) sottolinea che la standardizzazione può offrire una rappresentazione parziale dei processi cognitivi, ma anche che la creazione di un quadro di riferimento chiaro, che definisca i limiti dello strumento e faciliti l'interpretazione del costrutto nel sistema valutativo più ampio, può ridurre tali rischi. Il framework di una prova standardizzata delimita le aree di misurazione, evita interpretazioni esagerate, definisce chiaramente gli obiettivi cognitivi, garantisce trasparenza alle scuole e consente una valutazione più equilibrata dei risultati.

Nel nostro contesto nazionale, i test standardizzati rappresentano un supporto fondamentale per un sistema valutativo a più dimensioni e con più obiettivi. La costruzione di una cultura della valutazione rappresenta quindi una condizione indispensabile per valorizzare le potenzialità degli esiti resi disponibili e per evitare letture semplificate del funzionamento del sistema educativo.

1.1.5 Il Sistema Nazionale di Valutazione

Nel nostro Paese, il passaggio a una valutazione di sistema attraverso la creazione del Sistema Nazionale di Valutazione (SNV), prevede che venga attivato un processo complesso in quattro fasi strettamente collegate: autovalutazione (RAV), valutazione esterna (prove nazionali standardizzate), azioni di miglioramento e rendicontazione sociale (Scuola in Chiaro). Queste quattro fasi rappresentano un modello di valutazione in cui gli esiti degli apprendimenti, le interpretazioni dei dati e degli indicatori, le decisioni prese e da prendere per procedere verso il miglioramento e la responsabilità pubblica si influenzano reciprocamente. Bottani e Checchi (2012) hanno visto in questo passaggio una ridefinizione del ruolo della valutazione nazionale degli apprendimenti, improntata a diventare un elemento strutturale della governance del sistema educativo, in termini di infrastruttura stabile di raccolta e analisi dei dati. Questo sistema di valutazione non si limita a produrre informazioni, ma promuove l'attivazione di processi riflessivi e contribuisce nel sostenere l'assunzione di decisioni informate da parte dei diversi attori istituzionali.

Ma lo sviluppo di un sistema di valutazione nazionale consente di aprire uno spazio a temi importanti, come quello dell'equità. Secondo Checchi (2006) le disuguaglianze educative sono espressione di

meccanismi strutturali che legano i risultati scolastici alle condizioni socioeconomiche di partenza questo significa che i risultati delle prove di valutazione degli apprendimenti non misurano esclusivamente le abilità individuali, ma riflettono anche i vantaggi o gli svantaggi legati alle condizioni di partenza socioeconomico e culturali. Il sistema educativo allora può o intervenire, ad esempio, per individuare le situazioni di maggiore fragilità e orientare interventi migliorativi, facendo attenzione, al contrario, a non riprodurre tali condizioni, contribuendo a etichettare scuole o studenti e a consolidare le disuguaglianze esistenti (Checchi, 2006). In questo senso, quindi, si parla di equità intendendo la capacità del sistema educativo di ridurre il peso del background sugli esiti e la valutazione del sistema educativo assume una funzione conoscitiva indispensabile, in quanto consente di rendere visibili tali disuguaglianze e di orientare interventi mirati. Bottani (2006) rafforza questa prospettiva mettendo in luce il contributo delle indagini comparative internazionali nel rendere evidenti le differenze tra sistemi educativi. Le analisi dei dati evidenziano la persistenza di disuguaglianze anche in contesti caratterizzati da elevata scolarizzazione, contribuendo a riportare al centro del dibattito il tema della giustizia educativa (Bottani, 2006). Le rilevazioni nazionali dell'INVALSI rappresentano uno strumento fondamentale per la costruzione di un quadro conoscitivo del sistema educativo, poiché consentono di raccogliere informazioni comparabili nel tempo e tra diversi contesti territoriali. La regolarità delle rilevazioni e la stabilità degli ambiti indagati costituiscono elementi essenziali per garantire la solidità e l'affidabilità delle informazioni prodotte, contribuendo alla costruzione di un'infrastruttura informativa a supporto delle politiche educative.

In questa architettura, la valutazione nazionale degli apprendimenti non coincide con "la valutazione", ma costituisce uno dei nodi del funzionamento dell'ecosistema valutativo. Produce dati comparabili a livello nazionale sugli apprendimenti e offre elementi utili alla lettura dell'efficienza e dell'equità del sistema, ma tali dati vengono integrati con informazioni provenienti da altre fonti, come la documentazione interna, i dati ministeriali e gli indicatori statistici nazionali. La scelta di coniugare prospettive interne ed esterne è coerente anche con quanto emerge dalla letteratura internazionale sulla progettazione di sistemi valutativi equilibrati, che tengono conto del rischio che la sola autovalutazione crei una condizione di autoreferenzialità, ma al contempo richiedono attenzione a non creare un sistema totalmente dipendente dagli indicatori standardizzati (Marion, Pellegrino & Berman, 2024; Mifsud & Wilkins, 2025).

Nel lavoro di De Santis e Asquini (2020), la ricostruzione della logica che sostiene l'SNV mostra come il RAV sia diventato il punto di convergenza tra le analisi valutative interne e le indicazioni esterne relative alle valutazioni degli apprendimenti, contestualizzate dal punto di vista territoriale e socioeconomico. La rendicontazione sociale, attraverso la pubblicazione dei dati relativi ai risultati delle prove e alle strategie adottate, completa il ciclo. Un'impostazione di questo tipo riflette la

prospettiva sostenuta da Benadusi e Giancola (2014), secondo cui una valutazione di sistema deve coniugare autonomia professionale e responsabilità pubblica, evitando che si riduca a un atto burocratico o si trasformi in una forma di controllo dall'alto. Il RAV si è progressivamente configurato come un dispositivo di raccordo tra *evaluation* e *accountability*, in quanto, da un lato, sostiene l'autodiagnosi e la progettazione del miglioramento, dall'altro, è pubblicamente accessibile tramite il portale "Scuola in Chiaro". Come notano Grimaldi e Serpieri (2016), la pubblicazione dei documenti valutativi e autovalutativi da una parte favorisce la trasparenza, la partecipazione e il dialogo con la comunità, dall'altra può creare delle condizioni di protezione dell'immagine propria o dell'istituzione, evitando una riflessione autentica e profonda sui problemi reali. Questo rischio si accentua quando l'*accountability* non è adeguatamente integrata con strumenti interpretativi robusti. Lo sviluppo dello SNV è stato seguito da una serie di sperimentazioni (vedi i progetti VALSIS - *Valutazione del sistema scolastico e delle scuole* - e VALES - *Valutazione del sistema scolastico e delle scuole* - https://www.invalsi.it/innovazione-sviluppo/) e da revisioni periodiche del formato del RAV, con l'obiettivo di renderlo più sostenibile per le scuole e più efficace come strumento di governo. In questa direzione si collocano anche le riflessioni di Previtali (2023), che sottolineano come il RAV sia ormai ampiamente diffuso come attività scolastica e che la rendicontazione sociale abbia acquisito una collocazione stabile nel ciclo della valutazione, confermando che il SNV rappresenta oggi un'infrastruttura istituzionale consolidata della *governance* scolastica.

In questo quadro, la rendicontazione sociale costituisce il momento in cui l'*accountability* diventa visibile. È l'ultima fase del ciclo e prevede la restituzione pubblica dei risultati, non come forma di controllo sanzionatorio, ma come dispositivo di dialogo con la comunità e di orientamento delle strategie future. È pensata come un momento di dialogo con il territorio e non come un dispositivo di esposizione competitiva, coerentemente con quanto sottolineato da Benadusi e Giancola (2020) sulla necessità di fondare la responsabilità su una lettura contestualizzata degli esiti. Il principio normativo che collega autonomia e responsabilità non chiede alle scuole di misurarsi su *ranking* impliciti, ma di rendere comprensibili obiettivi, vincoli, risultati e interpretazioni, favorendo un rapporto più maturo e consapevole con la comunità.

1.2 Le indagini nazionali e internazionali a confronto

Se quindi le rilevazioni OCSE tendono a collocare gli studenti all'interno di un quadro di competenze utili alla cittadinanza globale e al lavoro nel mondo contemporaneo, le indagini dell'IEA mettono a fuoco la qualità dei percorsi scolastici e la coerenza tra curricolo, insegnamento e apprendimento. Come osserva Hopmann (2008), questi due aspetti metodologici sottendono due diverse visioni dell'educazione: una più orientata alla comprensione delle capacità degli individui (sia studenti che

adulti) di affrontare problemi nuovi e complessi, e l'altra centrata sulla padronanza dei contenuti curricolari. Queste due prospettive offrono interpretazioni complementari del funzionamento dei sistemi educativi. Con il sistema di valutazione nazionale degli apprendimenti, poi, condividono alcuni obiettivi generali, come il monitoraggio dell'efficacia dei sistemi formativi attraverso la produzione di dati comparabili, ma lo fanno adottando impostazioni metodologiche in parte diverse e con ricadute anche molto differenti sul dibattito educativo e sulla governance scolastica. Il punto di contatto più evidente tra questi sistemi di produzione di dati informativi sulla scuola è l'idea che la qualità del sistema educativo non possa essere letta solo attraverso valutazioni interne e locali, perché rischiano di essere eccessivamente arbitrarie, e che tale variabile possa essere controllata attraverso l'integrazione di dati provenienti anche da prove esterne standardizzate.

Rispetto invece all'impostazione metodologica delle prove cognitive ci sono sia punti di incontro che specificità di ciascuna di queste rilevazioni.

Le prove nazionali sono costruite a partire dal quadro normativo italiano e dai traguardi previsti dalle Indicazioni Nazionali (le attuali si rifanno alle Indicazioni del 2012, ma è in corso la revisione del Quadro di Riferimento delle prove per i gradi scolastici interessati dalle nuove Indicazioni del 2025) e dalle Linee Guida che orientano gli obiettivi di apprendimento a cui la scuola dovrebbe puntare. In questo senso, le prove nazionali standardizzate presentano un'impostazione tendenzialmente orientata al curriculum, simile a quella delle indagini IEA, con attenzione anche all'inclusione di elementi specifici del curricolo nazionale, non misurabili nei contesti delle indagini comparative internazionali, come, ad esempio, la riflessione sulla lingua (lessico, formazione delle parole, sintassi, ecc.), che si occupa espressamente di aspetti linguistici indicati nella normativa e misurabili attraverso prove standardizzate (https://www.invalsi.it/rilevazioni-nazionali/quadri-riferimento-prove-invalsi/). Autori come Benadusi e Giancola (2014) sottolineano l'importanza di interpretare i risultati scolastici in un'ottica attenta all'equità e alle differenze tra i contesti educativi, evitando interpretazioni puramente classificatorie o indipendenti dalle condizioni in cui le scuole operano. In questa prospettiva, l'uso dei dati valutativi non riguarda soltanto la misurazione del livello di padronanza raggiunto dagli allievi, ma può anche contribuire a mettere in discussione il rapporto tra risultati, opportunità formative e obiettivi del sistema educativo. Tale lettura trova ulteriore riferimento nei Quadri di Riferimento delle prove INVALSI, che esplicitano i presupposti teorici, disciplinari e didattici alla base della loro costruzione. Essi rendono più chiaro il rapporto tra le competenze attese, i contenuti oggetto di valutazione e i criteri che orientano la rilevazione degli apprendimenti (Mattei et al., 2016). In questo senso, i risultati delle prove possono diventare un'occasione di riflessione interna per le scuole sulla congruenza tra pratiche didattiche, traguardi formativi e apprendimenti effettivamente osservati, purché siano interpretati in modo contestualizzato e non come semplici

indicatori di rendimento. Il rapporto, invece, tra le prove nazionali standardizzate e quelle dell'OCSE, in particolare quelle di PISA, è sicuramente più complesso. Un obiettivo specifico delle rilevazioni sugli studenti dell'OCSE è valutare la capacità degli studenti di trasferire le conoscenze in contesti non scolastici, a differenza delle prove nazionali standardizzate, che si concentrano maggiormente sulla rilevazione della padronanza di alcuni apprendimenti descritti nel curricolo scolastico e che, come osservano Carretta e Fontana (2018), rappresentano una sorta di "zoccolo duro" in termini di obiettivi di apprendimento nel percorso scolastico italiano. È un modo di concepire la misurazione che, pur rimanendo all'interno di contesti psicometrici simili che prevedono l'utilizzo del modello di Rasch o la costruzione di *proficiency scale*, si mantiene, nei contenuti, più radicato nelle specificità del contesto nazionale.

La natura censuaria delle rilevazioni INVALSI, rispetto a quella campionaria delle rilevazioni internazionali, si riflette direttamente sul tipo di *accountability* che ne consegue. Le rilevazioni nell'ambito dell'OCSE e dell'IEA producono informazioni di tipo "macro", orientate ai decisori politici, mentre le rilevazioni dell'INVALSI producono un'*accountability* "territorializzata", perché le scuole ricevono dati propri, li interpretano e, attraverso il RAV, li condividono nel contesto della rendicontazione sociale.

Il sistema nazionale di valutazione standardizzata degli apprendimenti non svolge soltanto una funzione di monitoraggio della qualità del sistema educativo, ma fornisce anche alle scuole elementi utili per leggere i propri risultati in modo più analitico. Attraverso la restituzione dei dati, le istituzioni scolastiche possono confrontare gli esiti delle proprie classi e della scuola con riferimenti territoriali e con contesti comparabili, ricavando indicazioni da utilizzare nei processi di autovalutazione e di miglioramento. Nei materiali INVALSI dedicati alla lettura dei risultati si precisa, infatti, che tali dati sono pensati per sostenere la riflessione di dirigenti e docenti sui risultati conseguiti dagli studenti, sulle pratiche didattiche e sulla programmazione delle attività scolastiche, in una logica non punitiva ma orientata al miglioramento (INVALSI, 2024a; INVALSI, 2024b). In questo senso, l'impiego dei dati INVALSI presenta una ricaduta più direttamente operativa rispetto alle grandi indagini comparative internazionali, poiché restituisce alle singole scuole informazioni riferite al proprio contesto e quindi più facilmente collegabili alle scelte organizzative e didattiche interne. Le rilevazioni nazionali, infatti, consentono alle scuole di verificare con maggiore precisione le criticità specifiche dei propri contesti in termini di efficacia, efficienza ed equità, soprattutto quando si considerano aspetti quali la variabilità all'interno delle classi o il ruolo del contesto socioeconomico.

In generale, come evidenziato da Hopmann (2008) e Sjøberg (2015) nelle valutazioni standardizzate, come quelle internazionali, il sistema di misurazione può produrre effetti indesiderati se mancano interpretazioni consapevoli e contestualizzate dei dati restituiti; questo aspetto è ancor più marcato

nel contesto nazionale, perché le prove standardizzate sono presenti nel quotidiano scolastico rispetto alle indagini internazionali e hanno un impatto sul sistema scolastico più diretto. È in questo senso che comprendere a fondo la finalità di un sistema di rilevazione esterno in termini di *assessment*, *evaluation* e *accountability* diventa cruciale sul piano operativo, perché, come evidenziato da Grimaldi e Serpieri (2016), questa contestualizzazione fornisce alle scuole strumenti utili per interpretare gli esiti in modo più consapevole e per individuare le direzioni di miglioramento da intraprendere.

1.2.1 Valutazione esterna e valutazione interna nel contesto scolastico

Nella scuola contemporanea, la valutazione di sistema è una questione delicata, perché deve affrontare esigenze diverse che spesso si intrecciano. Da una parte dovrebbe aiutare a riflettere e a migliorare, mentre dall'altra può facilmente creare dinamiche ambivalenti, determinate dai confronti tra le scuole e dalla rendicontazione pubblica della qualità degli istituti.

Il problema reale, nella scuola contemporanea, è che questi due piani spesso si sovrappongono: si pretende che i dati servano "per migliorare", ma li si usa (o li si percepisce) come strumenti di controllo.

Ma è proprio all'interno di ciò che viene percepito che si costruiscono significati ambivalenti, riconducibili alle valutazioni di sistema. Questa ambivalenza è ben descritta anche dall'OCSE (2013), che nelle proprie analisi sui temi di *evaluation* and *assessment* insiste sulla necessità di integrare i diversi livelli (*student assessment*, *teacher appraisal*, *school evaluation*, *system evaluation*) in un impianto coerente, capace di sostenere la qualità e l'equità senza generare effetti indesiderati. In sostanza, l'integrazione è auspicabile perché evita la frammentazione (prove che "dicono" una cosa, pratiche scolastiche che ne suggeriscono un'altra, rendicontazioni che ne impongono una terza); diventa però fragile quando i dati, nati per supportare le decisioni professionali e il miglioramento, vengono utilizzati all'interno di logiche reputazionali o sanzionatorie.

Lo studio di Looney (2011), commissionato dall'OCSE nell'ambito della *Review on Evaluation and Assessment Frameworks for Improving School Outcomes*, mira a gettare le basi per un sistema che metta in dialogo le valutazioni interne ed esterne, in modo che i dati delle valutazioni esterne, utilizzati per il monitoraggio del sistema, possano essere impiegati anche per elaborare strategie di apprendimento e metodologie didattiche. In questa prospettiva, l'integrazione è intesa come la costruzione di un sistema informativo in cui gli esiti delle diverse forme di valutazione siano leggibili, comparabili entro limiti dichiarati e, soprattutto, utilizzabili in modo coerente con lo scopo per cui sono stati raccolti. Proprio per questo motivo, Looney sottolinea che "attualmente esistono importanti ostacoli tecnici" che impediscono una piena integrazione (Looney, 2011, p. 3). Gli ostacoli non riguardano solo la tecnologia, ma anche il fatto che i dati utili a livello di sistema (aggregati, periodici,

comparabili) non coincidono con quelli utili a livello di classe (frequenti, diagnostici, sensibili ai micro-progressi). Confondere gli usi, per esempio usando indicatori sommativi per giudicare direttamente le pratiche didattiche, spesso porta a semplificazioni che riguardano il lavoro degli insegnanti (insegnare "in funzione della prova", ridurre il programma, irrigidire la valutazione). Looney conclude infatti che una migliore integrazione richiede investimenti non soltanto in strumenti, ma anche nella formazione e nello sviluppo professionale degli insegnanti.

La riflessione sviluppata dall'OCSE sul rapporto tra *l'evaluation e l'assessment* chiarisce perché il tema della coerenza interna ai sistemi di valutazione s*ia* cruciale. In *Synergies for Better Learning* l'OCSE sostiene che un sistema di valutazione può dirsi efficace soltanto quando vi è un allineamento stabile tra obiettivi, interpretazioni dei risultati e conseguenze legate all'uso dei dati (OECD, 2013). Questa visione considera la valutazione come un'infrastruttura di *governance* che richiede una regia chiara e condivisa, capace di evitare sovrapposizioni, incoerenze e duplicazioni che producono dispersione di risorse, sfiducia negli strumenti e incentivi a comportamenti opportunistici. Non è un caso che il rapporto insista sulla necessità di un "quadro unitario", in cui valutazioni diverse, formative, di scuola e di sistema, possano convivere senza entrare in competizione né generare pressioni che snaturano la loro funzione originaria. In questo contesto il tema della coerenza diventa un nodo eminentemente politico, perché riguarda chi utilizza i dati, per quali decisioni, con quale competenza interpretativa e con quali protezioni rispetto agli usi impropri che possono indebolire la dimensione formativa.

La letteratura internazionale ha ampiamente discusso gli effetti dei sistemi di *accountability* quando questi non sono accompagnati da adeguate forme di supporto professionale e da un'interpretazione contestualizzata dei dati. In particolare, gli studi sulla *performance-based accountability* mostrano che l'uso dei risultati dei test come indicatori immediati della qualità delle scuole o dell'efficacia degli insegnanti può produrre effetti non intenzionali, tra cui la pressione reputazionale, gli adattamenti strategici e la riduzione della valutazione a strumento di controllo (Skedsmo & Huber, 2024; Levatino et al., 2024).

In questi casi, il problema non riguarda soltanto la presenza di prove standardizzate, ma anche il modo in cui i loro risultati vengono integrati nei meccanismi di governance. Quando i punteggi sono direttamente associati a giudizi pubblici, incentivi o conseguenze simboliche sulla reputazione di scuole e docenti, si crea uno squilibrio tra le responsabilità richieste, le condizioni effettive di lavoro e la possibilità di utilizzare la valutazione in chiave formativa. La valutazione rischia così di perdere la sua funzione di supporto al miglioramento e di essere percepita prevalentemente come un meccanismo di sorveglianza o di classificazione. L'attenzione si sposta dalla comprensione degli errori alla gestione del rischio reputazionale, un cambiamento che impoverisce il senso stesso della

valutazione come strumento di apprendimento. È un equilibrio fragile che, se non governato, tende a sbilanciare gli scopi verso l'*accountability high stake*, con effetti negativi documentati anche in sistemi educativi molto diversi tra loro (Looney, 2011; Lingard & Sellar, 2013).

Negli ultimi tempi, vari contributi hanno ripreso queste tematiche utilizzando un linguaggio moderno, pur mantenendo una preoccupazione comune. In questi studi (Marion et al., 2019; Marion et al., 2024) i ricercatori propongono la costruzione di *balanced assessment system*, ossia sistemi valutativi che sostengono contemporaneamente l'apprendimento degli studenti, la responsabilità pubblica e il miglioramento professionale. In questa prospettiva, il bilanciamento non consiste in una semplice sommatoria di prove diverse, bensì nella costruzione intenzionale di un "ecosistema coerente" in cui le valutazioni formative, intermedie e sommative collaborano a fini distinti ma complementari. L'idea di fondo è che nessun indicatore, da solo, possa sostenere un intero sistema educativo senza generare effetti collaterali indesiderati.

La letteratura recente mostra come questo equilibrio sia difficile da realizzare quando l'*accountability* esercita una pressione eccessiva. Uno studio (Polikoff & Hutt, 2020) ha confermato che i sistemi in cui i risultati delle prove esterne hanno un peso predominante tendono a limitare il curriculum, a ridurre la difficoltà delle attività didattiche e a promuovere metodi di addestramento al test che contrastano con l'obiettivo di un effettivo miglioramento. Allo stesso modo, la letteratura sulle politiche di *accountability* (Skedsmo & Huber, 2024) mette in evidenza gli effetti non intenzionali delle politiche valutative *high stake*, mostrando come l'enfasi sui risultati esterni possa trasformare la cultura professionale degli insegnanti e modificare la percezione del loro lavoro. Sono trasformazioni che incidono sulla qualità della valutazione di classe e, più in generale, sul senso pedagogico dell'insegnamento.

Proprio per rispondere a queste questioni sta crescendo l'attenzione verso l'*assessment literacy*, cioè la capacità professionale degli insegnanti di comprendere il significato dei dati, riconoscerne i limiti e utilizzarli in modo coerente con gli obiettivi didattici. La letteratura internazionale (De Luca et al., 2019; Gotch & French, 2019) sottolinea che la tenuta dei sistemi bilanciati dipende dalla capacità degli attori di considerare la valutazione non come un obbligo esterno, bensì come una risorsa per orientare la didattica. In termini operativi, questo implica investire in percorsi di sviluppo professionale focalizzati sulle competenze valutative e sulla *data literacy*, così da favorire un uso consapevole e pedagogicamente fondato delle informazioni disponibili. Richiede inoltre una progettazione istituzionale che espliciti, a ogni livello del sistema, gli scopi e le routine decisionali connesse alle diverse forme di valutazione, dalla classe fino ai livelli di governo nazionale.

Se la guardiamo in questo modo, quello che l'OCSE indica come obiettivo di integrazione non è soltanto un semplice aggiustamento tecnico degli strumenti di valutazione, ma va inteso come un

processo di costruzione istituzionale. È una forma di equilibrio che riguarda l'interazione tra attori, scopi e significati della valutazione, ed è proprio questa dimensione relazionale a determinare la capacità di un sistema di sostenere, al contempo, equità, qualità e crescita professionale. Uno schema di valutazione veramente integrato richiede dunque un investimento nella cultura valutativa del sistema, e non soltanto nella riorganizzazione degli strumenti.

L'equilibrio tra *evaluation*, *assessment* e *accountability* rispecchia quanto suggerito anche dall'OCSE in *Synergies for Better Learning* (OECD, 2013), ossia che *l'accountability* può sostenere il miglioramento solo se guidata da una cultura valutativa solida e integrata con dispositivi capaci di favorire la comprensione e non soltanto il controllo. È una posizione coerente con quella di Vertecchi, che richiama la necessità di distinguere tra le diverse funzioni della valutazione per evitare derive riduzionistiche. Se l'*accountability* opera senza *evaluation* rischia di trasformarsi in un controllo cieco; se l'*evaluation* opera senza *accountability*, può diventare autoreferenziale; se l'*assessment* è privo di una cultura valutativa condivisa, produce numeri che non riescono a parlare. La sfida, allora, non è scegliere tra queste tre dimensioni, ma governarne la relazione. La scuola ha bisogno di *assessment* robusti per disporre di evidenze comparabili, di *evaluation* per interpretare i processi e orientare le strategie, e di *accountability* per rendere conto alla comunità in modo trasparente e comprensibile. Ciascuna di queste funzioni possiede una propria dignità epistemologica e deve essere mantenuta entro confini che ne tutelino le finalità. Solo una loro integrazione equilibrata consente al SNV di operare come un sistema al servizio del miglioramento e dell'equità, evitando gli eccessi di una governance sbilanciata, che la letteratura internazionale e nazionale mette da tempo in guardia (Skedsmo & Huber, 2024; Levatino et al., 2024).

1.3 Gli aspetti etici della valutazione standardizzata nei contesti educativi

Rispetto ai temi discussi finora, è importante evidenziare alcuni aspetti cruciali, soprattutto se si considerano nelle questioni etiche relative ai sistemi di valutazione esterna.

La valutazione esterna, sia a livello nazionale sia internazionale, si è affermata come strumento fondamentale per fornire informazioni affidabili e comparabili, contribuendo all'equità educativa. Contribuisce a monitorare il funzionamento dei sistemi scolastici, a evidenziare le differenze territoriali e sociali e a guidare politiche basate su dati empirici. Si configura come uno strumento di trasparenza delle scuole e per le scuole, aiutando a mettere in luce disuguaglianze spesso difficili da individuare con le sole pratiche locali.

Un primo elemento etico da considerare riguarda la capacità della valutazione esterna di rendere il sistema di istruzione più leggibile e governabile. Questa funzione permette di andare oltre le percezioni soggettive o gli indicatori indiretti ed è il focus dei principali rapporti comparativi, che

mettono in relazione misurazione, equità e politiche formative (OECD, 2024a; European Commission, 2025) per identificare squilibri, definire priorità di azione e valutare l'efficacia delle riforme. Tuttavia, questa stessa funzione può diventare problematica quando la valutazione viene ridotta a un'unica metrica o a classifiche che non colgono appieno la complessità dei contesti.

Un altro tema etico riguarda l'uso dei risultati. Le prove esterne possono generare criticità, soprattutto quando i dati vengono interpretati in modo eccessivamente semplice, attribuendo responsabilità a scuole o insegnanti anche per fattori strutturali, o collegando la qualità educativa esclusivamente ai risultati ottenuti. La letteratura ha documentato fenomeni quali la riduzione del curricolo, la diffusione di pratiche orientate all'addestramento al test, l'esclusione di studenti considerati più fragili e un clima di pressione che rischia di indebolire la cultura professionale (Stobart, 2008; Berliner, 2011; Klenowski, 2012). Studi recenti mostrano che gli effetti dei sistemi di *accountability* non dipendono soltanto dai risultati delle prove standardizzate, ma anche dalle percezioni di studenti e docenti e dalle modalità con cui i processi valutativi vengono interpretati e sostenuti nelle scuole (Darling-Hammond, 2004; Fullan, 2011). La letteratura internazionale sugli effetti dei test standardizzati evidenzia che benefici e criticità possano coesistere e che il loro impatto dipenda in larga misura dal modello di *accountability* in cui tali strumenti vengono impiegati. In particolare, le conseguenze delle valutazioni standardizzate non sono riconducibili ai test in sé, ma alle finalità, alle modalità di utilizzo dei risultati e alle forme di supporto offerte alle scuole (Darling-Hammond, 2004; Fullan, 2011). Un elemento particolarmente rilevante riguarda il tipo di responsabilità associato ai risultati delle prove. Nei sistemi caratterizzati da forme di *accountability* ad alta posta in gioco (*high-stakes accountability*), nelle quali i risultati sono collegati a premi, sanzioni, classifiche pubbliche o conseguenze reputazionali, si osservano più frequentemente fenomeni di restringimento del curricolo, di insegnamento orientato al test (*teaching to the test*) e di comportamenti strategici finalizzati al miglioramento degli indicatori piuttosto che ai processi educativi (Au, 2007; Darling-Hammond, 2004). Al contrario, nei modelli di *accountability* orientati al miglioramento (*low-stakes o soft accountability*), i dati vengono utilizzati prevalentemente come strumenti diagnostici e di riflessione professionale. In questi contesti, l'interpretazione contestualizzata dei risultati e il sostegno ai processi di sviluppo scolastico tendono a ridurre gli effetti distorsivi e a favorire un impiego più efficace delle evidenze valutative nei processi di miglioramento dell'insegnamento e dell'apprendimento (Fullan, 2011; Darling-Hammond, 2004).

Un terzo aspetto etico riguarda la qualità tecnica delle prove, che può essere considerata un vero e proprio dovere morale. I risultati delle valutazioni esterne influenzano decisioni rilevanti, dalla reputazione degli istituti alla distribuzione delle risorse, fino alle narrazioni pubbliche sui territori.

La metodologia adottata rappresenta un aspetto etico fondamentale, poiché stabilisce le condizioni di affidabilità e giustizia necessarie per l'interpretazione e l'uso dei risultati della valutazione esterna. Gli *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) non propongono semplicemente un insieme di regole tecniche, ma delineano un vero e proprio quadro normativo che responsabilizza chi progetta, somministra e interpreta le prove. Una documentazione accurata dei processi non è solo un requisito formale, ma permette di ricostruire passo passo come sono state ottenute le misure, quali assunzioni le sostengono e quali limiti influenzano la loro interpretazione. La trasparenza metodologica diventa quindi un elemento chiave affinché decisori, docenti e ricercatori possano valutare la solidità dei risultati e le condizioni di applicabilità. Un altro pilastro indicato dagli standard riguarda il controllo dei *bias*, cioè le procedure per verificare che gli strumenti non favoriscano né svantaggino determinati gruppi di studenti. Questo comporta analisi statistiche dedicate, revisioni qualitative degli item, verifiche linguistiche e culturali, e processi di pre-test ripetuti. L'obiettivo non è ottenere test "neutri", una condizione irrealizzabile, ma ridurre al minimo gli effetti ingiustificati di fattori estranei alla competenza che si intende misurare. La correttezza delle inferenze è un'altra dimensione cruciale: secondo gli standard, ogni risultato derivato da test standardizzati deve essere accompagnato da indicazioni chiare su quali conclusioni sia lecito trarre e quali invece richiedano cautela. Una misura valida in un determinato contesto o per una certa popolazione potrebbe non esserlo in altri contesti o per altre popolazioni; ignorare questa distinzione comporta il rischio di generalizzazioni indebite che possono influenzare in modo scorretto il giudizio su scuole, territori o gruppi sociali. Per questo motivo, l'OECD enfatizza la necessità di descrivere in modo dettagliato le scelte metodologiche, i criteri di costruzione degli strumenti, i passaggi di traduzione e di adattamento culturale, nonché i processi di standardizzazione e di scaling. Rendere espliciti i limiti della comparazione, siano essi statistici, linguistici o legati alle differenze di curricolo, rappresenta una forma di responsabilità nei confronti dell'intera comunità scientifica e politica che utilizza i dati per orientare riforme e decisioni pubbliche. Anche la tutela della privacy e le procedure di adattamento della prova costituiscono aspetti eticamente sensibili. Il primo aspetto riguarda la protezione dei dati personali e scolastici, soprattutto quando i risultati vengono uniti ad altre informazioni o utilizzati per analisi longitudinali. Il secondo aspetto, invece, riguarda le procedure di adattamento della prova, che garantiscono condizioni eque, senza alterare la natura della competenza valutata.

Un ultimo aspetto etico riguarda la relazione tra la valutazione esterna, i dati digitali e le nuove tecnologie. L'aumento dell'uso di piattaforme online, dei tracciamenti e delle analisi automatizzate solleva domande sulla privacy, sulla sicurezza, sulla minimizzazione dei dati e sui rischi di violazione dei dati. L'introduzione di processi automatizzati e di intelligenza artificiale nelle valutazioni solleva

ulteriormente queste problematiche, rendendo necessaria una maggiore attenzione alla trasparenza degli algoritmi, alla loro spiegabilità e ai potenziali *bias*. Studi recenti evidenziano che studenti e insegnanti giudicano la legittimità di questi sistemi anche in base alla loro percezione di equità e al livello di controllo che hanno sulle informazioni personali (Karimov et al., 2024). Anche lo studio di Karimov et al. (2024) ha evidenziato che la fiducia degli studenti nelle *learning analytics* è moderata, poiché si fonda su due punti chiave: da un lato, la percezione che l'uso dei dati non comporti ingiustizie o trattamenti diseguali tra i vari gruppi di studenti; dall'altro, la possibilità di conoscere quali informazioni vengono raccolte e di avere un controllo su come vengono conservate e utilizzate. L'aumento della complessità tecnica mette quindi in evidenza quanto sia importante definire responsabilità chiare e rafforzare le competenze necessarie per interpretare i dati e i sistemi digitali.

CAPITOLO 2: FUNZIONI E FORME DI VALUTAZIONE *NELLA* SCUOLA

Parlare di valutazione nel contesto scolastico significa muoversi in un ambito in cui le decisioni quotidiane e le scelte politico-istituzionali si intersecano. La valutazione a scuola ha diverse funzioni, sia esplicite che implicite. Da una parte, infatti, serve a verificare se gli studenti hanno raggiunto gli obiettivi di apprendimento, dall'altra, serve anche a orientare i processi di apprendimento. Inoltre, la valutazione a scuola incide su dimensioni quali l'autonomia e l'autostima degli studenti. Da un punto di vista sociale, la valutazione scolastica, specialmente al termine dei percorsi obbligatori certificati dagli esami di Stato, ha l'obiettivo di indicare il raggiungimento di quelle competenze prefissate per la formazione di cittadini consapevoli, pronti a contribuire allo sviluppo socioeconomico del Paese. Rappresenta anche, in questo senso, l'indicazione di un'opportunità di mobilità sociale per coloro che provengono da contesti svantaggiati. Osservare come la storia della valutazione nella scuola italiana abbia progressivamente codificato criteri, strumenti e responsabilità permette di riflettere sul perché, in alcuni momenti storici, la valutazione sia stata concepita soprattutto come selezione e, in altri, come garanzia di equità, accompagnamento dei processi e promozione del successo formativo. Ma la valutazione nella scuola non si limita ai momenti finali dei passaggi tra i vari livelli scolastici. La valutazione scolastica riguarda sia chi ha il compito di produrla sia chi la riceve. Professori e studenti sono strettamente collegati in una relazione inscindibile, poiché entrambi sono coinvolti nel lungo percorso dell'apprendimento. In questo capitolo analizzerò il fenomeno dell'apprendimento, considerando la valutazione come un feedback formativo per docenti e studenti.

2.1 Dalla pedagogia per obiettivi alla centralità dello studente

Il rapporto tra valutazione e insegnamento subisce un profondo cambiamento quando la scuola non considera più l'apprendimento come un semplice insieme di contenuti da trasmettere, ma lo concepisce come un percorso con obiettivi chiari e definiti. La pedagogia per obiettivi nasce, quindi, da questa volontà di rendere pubbliche e verificabili le finalità educative, andando oltre la semplice tradizione professionale o disciplinare (Calenda & Milito, 2020). In questa prospettiva, l'obiettivo è uno strumento di trasparenza che definisce in anticipo che cosa si considera apprendimento, quali prove possono dimostrarlo e su quali criteri si valuta la bontà del giudizio.

Il contributo di Tyler (1949) si colloca tra i riferimenti più influenti della ricerca educativa del Novecento, soprattutto per il legame che stabilisce tra progettazione curricolare, definizione degli obiettivi e valutazione. La sua proposta, maturata anche nel contesto dell'*Eight-Year Study*, interpreta la valutazione non come semplice misurazione dei risultati, ma come verifica della coerenza tra finalità educative, esperienze di apprendimento e cambiamenti osservabili negli studenti. In questo

senso, il lavoro di Tyler ha contribuito a consolidare un'idea di valutazione strettamente connessa alla progettazione didattica e curricolare, come evidenziato anche nelle ricostruzioni storiche della pedagogia sperimentale (Tyler, 1949; de Landsheere, 1988).Questa riflessione è spesso considerata un punto di svolta perché pone al centro l'idea che la valutazione non può limitarsi a registrare solo i risultati finali, ma deve essere coerente con gli obiettivi dichiarati dalla scuola. La proposta di Tyler, infatti, struttura la progettazione scolastica attorno a quattro domande: 1) Quali obiettivi educativi dovrebbe perseguire la scuola? 2) Quali esperienze educative possono essere proposte per raggiungere tali obiettivi? 3) In che modo si possono organizzare queste esperienze educative in modo efficace? 4) Come possiamo verificare se tali obiettivi sono stati raggiunti? (Tyler, 1949). Un aspetto rilevante di questo approccio è lo spostamento dell'attenzione dal giudizio al patto formativo. Quando gli obiettivi sono definiti chiaramente, anche la valutazione diventa più facile da discutere e da motivare, il che ne facilita la giustificazione in termini di giustizia educativa. È in questo passaggio che emerge l'influenza della tassonomia di Bloom (Bloom et al., 1956), che, nel tentativo di creare un linguaggio condiviso tra docenti ed esaminatori, propone una classificazione degli obiettivi cognitivi che consente di distinguere diversi livelli di complessità, evitando che ogni verifica si limiti a una semplice richiesta di richiamare informazioni.

Mager, da parte sua, sottolinea l'importanza della controllabilità, introducendo obiettivi formulati come risultati osservabili e verificabili, per minimizzare fraintendimenti e interpretazioni soggettive. Egli definisce gli obiettivi "in termini di risultati attesi come comportamenti osservabili, riconoscibili e dunque valutabili, con basse possibilità di fraintendimenti o sottoposti alla soggettività di chi valuta" (Mager, 1962/1987). In questo modo, la pedagogia degli obiettivi presenta un modello di valutazione più equo non perché oggettivo in senso assoluto, ma perché i criteri sono resi espliciti, riducendo così il rischio di arbitrarietà.

Nel dibattito italiano, la tradizione della pedagogia sperimentale ha riadottato questo approccio collegandolo alla riflessione sulla ricerca valutativa e sui suoi sviluppi, anche in relazione alle indagini comparative internazionali (Vertecchi, Agrusti, & Losito, 2010). Se la pedagogia per obiettivi mira a definire ciò che si desidera raggiungere, l'approccio per competenze si concentra sul raggiungimento di tali obiettivi attraverso un uso significativo e consapevole del sapere. La competenza non è solo la somma di conoscenze e abilità, ma implica anche la capacità di mobilitarle in contesti diversi, di scegliere le strategie, di adattarsi e di riflettere. La cornice internazionale che ha sostenuto questa idea è quella sviluppata dall'OCSE con il progetto DeSeCo, in cui le competenze chiave sono considerate risorse (cognitive e non solo) utili per affrontare compiti complessi in contesti reali (OECD, 2005). In Europa, la Raccomandazione del Consiglio del 22 maggio 2018 rafforza ulteriormente questa prospettiva proponendo un quadro di otto competenze chiave per

l'apprendimento permanente (*Council of the European Union*, 2018). Questa impostazione è stata recepita e si è consolidata anche nella normativa nazionale e nei documenti di indirizzo, attraverso un progressivo rafforzamento del legame tra curricolo, competenze chiave e certificazione delle competenze (MIUR, 2012; MIUR, 2018a).

Questo cambiamento ha necessariamente comportato conseguenze dirette anche sulla valutazione. Se l'apprendimento è inteso come sviluppo di competenze, non è più sufficiente verificare le conoscenze, ma occorre valutare anche ulteriori dimensioni relative alla capacità di trasferirle in contesti differenti. Un approccio basato sulle competenze rende lo studente un "co-produttore" delle esperienze formative (De Santis, 2022, p. 154). La centralità dello studente, in questo contesto, diventa un criterio per la costruzione della valutazione, poiché prevede l'inserimento di dimensioni quali l'autovalutazione, la riflessione condivisa sui criteri dei momenti di verifica, la revisione del lavoro sulla base delle indicazioni fornite in sede di valutazione e l'interazione dello studente con il feedback. La letteratura sull'*assessment for learning* (Black & Wiliam, 1998; Wiliam, 2011), che verrà approfondita nel paragrafo successivo, aiuta a mettere a fuoco il punto. Gli studi empirici indicano che queste strategie aiutano a sviluppare competenze metacognitive che, a loro volta, migliorano la qualità dell'apprendimento (Nicol & Macfarlane-Dick, 2006). In questo modo, la valutazione si trasforma in un'opportunità di apprendimento anziché in uno strumento di controllo esterno. Anche il contributo di Corsini (2018), ad esempio, evidenzia che la valutazione per l'apprendimento comprende attività che forniscono informazioni "per modificare il lavoro didattico e i processi di apprendimento" (Black & Wiliam, 1998, come citato in Corsini, 2018, p. 117). In questa prospettiva, la valutazione assume un ruolo più pedagogico che non neutro o tecnico. Essa aiuta a creare un ambiente di apprendimento in cui l'errore è visto come una risorsa informativa, anziché come una mancanza da punire. Questo cambiamento culturale ha importanti ripercussioni sulla motivazione degli studenti, come evidenziato da diversi studi che mostrano che le pratiche di *assessment for learning* sono collegate a un aumento della percezione di autoefficacia e a un maggiore focus sull'apprendimento piuttosto che sulla prestazione (Perrenoud, 1998; Stiggins, 2005).

2.2 La valutazione nella legislazione scolastica italiana tra primo e secondo ciclo d'istruzione

La valutazione scolastica è sempre stata al centro di un dibattito che coinvolge questioni di politica e cultura riguardanti temi quali la soggettività e l'obiettività, l'equità nel sistema scolastico nazionale e la legittimazione dei passaggi di livello tramite titoli ufficiali. La Legge Casati del 1859 costituisce il primo ordinamento organico dell'istruzione pubblica e organizza il sistema scolastico in ordini e gradi. In tale impianto, la valutazione è strettamente legata alla verifica delle conoscenze e alla regolazione dei passaggi tra un percorso e l'altro. Pur non essendo ancora presente l'esame di Stato

nella forma che sarà introdotta con la Riforma Gentile, la legge attribuisce agli esami e ai titoli una funzione centrale nella selezione scolastica e nell'accesso agli studi successivi. Con la Legge Coppino del 1877 l'obbligo dell'istruzione elementare viene rafforzato e collegato a forme più esplicite di controllo pubblico sulla frequenza e sugli apprendimenti di base. L'obbligo riguardava il corso elementare inferiore, previsto di norma fino ai nove anni, e comprendeva le prime nozioni dei doveri dell'uomo e del cittadino, la lettura, la calligrafia, i rudimenti della lingua italiana, dell'aritmetica e del sistema metrico. La legge prevedeva inoltre che l'obbligo potesse cessare anticipatamente solo nel caso in cui il fanciullo superasse un "esperimento" sulle materie indicate, svolto nella scuola o davanti al delegato scolastico, alla presenza dei genitori o di altri parenti. Qualora tale prova non avesse avuto esito positivo, l'obbligo si sarebbe protratto fino al compimento del decimo anno di età. In questo senso, la norma introduceva una prima forma di accertamento pubblico degli apprendimenti elementari, più che una certificazione in senso moderno.

Un passaggio successivo avviene con la Riforma Gentile del 1923, realizzata attraverso una serie di provvedimenti normativi che ridefiniscono in modo organico l'ordinamento scolastico italiano. In particolare, il Regio decreto 6 maggio 1923, n. 1054, relativo all'istruzione media, attribuisce agli esami un ruolo centrale nella regolamentazione dei percorsi scolastici e nel riconoscimento dei titoli di studio. L'esame di Stato assume così una funzione selettiva e certificativa: non si limita a verificare il completamento di un percorso, ma contribuisce a stabilire il valore pubblico del titolo conseguito e la possibilità di accedere ai gradi successivi dell'istruzione o a percorsi professionali per i quali è richiesto un titolo legalmente riconosciuto. In questa prospettiva, la valutazione diventa anche uno strumento di ordinamento istituzionale e di riconoscimento sociale dei percorsi scolastici.

Una svolta significativa si ha con la Legge 4 agosto 1977, n. 517, che contribuisce a modificare il modo di intendere la valutazione nella scuola dell'obbligo. L'attenzione non è più rivolta soltanto all'esito finale, ma anche al percorso compiuto dall'alunno nel corso dell'esperienza scolastica. Per la scuola elementare, l'articolo 1 stabilisce che l'esame di licenza costituisce il momento conclusivo dell'attività educativa e deve tener conto delle osservazioni sistematiche operate dall'insegnante o dagli insegnanti di classe. In questo modo, la prova conclusiva non è considerata un momento separato dal percorso, ma parte di un processo più ampio di osservazione e valutazione.

La stessa legge introduce inoltre la scheda personale dell'alunno, che l'insegnante è tenuto a compilare e aggiornare. Tale documento raccoglie informazioni sulla partecipazione alla vita scolastica, sul processo di apprendimento e sui livelli di maturazione raggiunti, costituendo la base per una valutazione periodica più informativa e meno riducibile al solo voto numerico. La valutazione assume così una funzione più descrittiva e formativa, orientata a rappresentare il percorso dell'alunno e non soltanto il risultato finale di una prova. Con il D.P.R. 8 marzo 1999, n. 275, l'autonomia

scolastica viene regolamentata e assume una forma più definita sul piano organizzativo e didattico. Le istituzioni scolastiche sono chiamate a progettare e realizzare la propria offerta formativa tenendo conto degli obiettivi nazionali, ma anche delle esigenze del contesto, delle famiglie e delle caratteristiche degli alunni. In questo quadro, la valutazione non può essere considerata un atto separato dalla progettazione, ma diventa parte del modo in cui la scuola rende esplicite le proprie scelte educative e didattiche.

L'autonomia didattica, infatti, consente alle scuole di organizzare tempi, percorsi, attività e forme di flessibilità in relazione ai ritmi di apprendimento degli studenti, con l'obiettivo di valorizzare le diversità e promuovere il successo formativo. Proprio per questo, la maggiore libertà progettuale richiede anche criteri di valutazione chiari e condivisi: senza un quadro comune di riferimento, il rischio è che l'autonomia produca pratiche valutative molto disomogenee e difficilmente giustificabili.

Da questa prospettiva emerge un'idea di valutazione meno burocratica e più legata alla responsabilità professionale dei docenti. La valutazione diventa uno strumento per leggere i percorsi degli studenti, sostenere le scelte didattiche e rendere trasparente il rapporto tra obiettivi, attività e risultati. Ne deriva anche la necessità di un confronto collegiale tra gli insegnanti, finalizzato a definire criteri comuni e modalità condivise per l'attribuzione dei giudizi. Per quanto riguarda gli esami finali del secondo ciclo, la legge 10 dicembre 1997, n. 425, ha riformato l'esame di Stato e ha stabilito regole più precise in un sistema che, per tradizione, è stato tra i più esposti alla variabilità delle valutazioni. È fondamentale che la valutazione sia considerata credibile dalla società, poiché da essa dipendono l'accesso all'università e al mondo del lavoro.

Nell'ultimo decennio, la riforma che ha avuto un impatto significativo è rappresentata dal decreto legislativo del 13 aprile 2017, n. 62, che ha ridefinito i criteri di valutazione del primo ciclo, le certificazioni delle competenze e l'esame di Stato, con l'obiettivo di creare una maggiore coerenza tra la valutazione interna, la certificazione e le prove finali. La norma, perlomeno nelle intenzioni, ha come obiettivo l'esigenza di conciliare da una parte, il riconoscimento della responsabilità professionale degli insegnanti, e dall'altra favorire visioni comuni per accrescere la trasparenza e la comparabilità.

Con l'ordinanza ministeriale 172/2020, la normativa si occupa ancora una volta delle modalità di valutazione nella scuola primaria introducendo giudizi descrittivi articolati per livelli, con l'obiettivo di rendere la valutazione più informativa e meno ridotta a un semplice numero.

Successivamente, la legge 1° ottobre 2024, n. 150, rivede nuovamente le norme sulla valutazione introducendo, dall'anno scolastico 2024/2025, i giudizi sintetici nella scuola primaria.

Tutti questi cambiamenti legislativi mettono in luce un aspetto molto importante che riguarda il modo in cui viene strutturata la valutazione, che influisce sul tipo di informazione che la scuola fornisce agli studenti e alle famiglie e, di conseguenza, sul modo in cui l'apprendimento viene interpretato culturalmente.

2.3 Elementi di docimologia: le riflessioni sulle modalità di valutazione

Nel contesto scolastico, la valutazione ha sempre svolto un ruolo fondamentale nei processi di insegnamento e apprendimento, essendo lo spazio in cui insegnante e studente interagiscono e negoziano il significato dell'esperienza educativa. Sin dalle prime riflessioni pedagogiche del Novecento, è stata concepita come un'attività complessa che include misurazione, verifica e interpretazione, andando oltre la semplice registrazione delle prestazioni e contribuendo a indirizzare il percorso formativo nel suo insieme. Visalberghi (1955) contribuisce a chiarire il rapporto tra misurazione e valutazione nel processo educativo, sottolineando che la valutazione non può essere ridotta a un semplice atto tecnico o alla mera registrazione di un risultato. Essa implica sempre un giudizio interpretativo, che richiede criteri, strumenti e attenzione al significato pedagogico delle informazioni raccolte. In questa prospettiva, la valutazione assume un ruolo rilevante nei processi di insegnamento e apprendimento, poiché consente di tracciare l'andamento del percorso formativo e di orientare le decisioni didattiche successive. In continuità con questa impostazione, de Landsheere (1973) colloca la valutazione all'interno della tradizione docimologica, evidenziandone le diverse funzioni. Una prima funzione è di tipo prognostico, poiché consente di raccogliere elementi utili a prevedere le possibilità dello studente all'interno di un determinato percorso formativo. Una seconda funzione è diagnostica, in quanto consente di individuare difficoltà, bisogni e condizioni di partenza. Una terza funzione riguarda il controllo e la regolazione del processo, attraverso l'osservazione dei progressi compiuti e l'utilizzo di informazioni valutative per adattare l'azione didattica. L'atto valutativo assume diverse forme a seconda degli obiettivi e del momento in cui viene effettuato (Visalberghi, 1955), evidenziando che la valutazione è un processo strutturato che accompagna l'intero percorso scolastico. Storicamente, la responsabilità della valutazione spetta all'insegnante, che crea prove, sviluppa strumenti, interpreta i risultati e li comunica allo studente e alla comunità educativa. Questa responsabilità si inserisce in un quadro normativo nazionale che stabilisce obiettivi, traguardi e criteri condivisi, affinché l'analisi dei risultati sia chiara e condivisa (Bloom, 1971; Domenici, 2007). Al termine dei cicli scolastici, il sistema educativo riassume il percorso attraverso esami finali che conferiscono un valore certificativo (Di Rienzo, 2017), basato sulle verifiche svolte e valutate da insegnanti o commissioni. Nei primi decenni del Novecento, gli studi docimologici si sono concentrati sugli esami finali, evidenziando criticità legate alla soggettività nelle valutazioni. Le

ricerche indicano che l'influenza del correttore può incidere in modo significativo sull'esito, creando disuguaglianze tra studenti con prestazioni simili. Benvenuto (2008) e Vertecchi (1991) evidenziano che tali distorsioni non sono dovute a cattiva volontà, ma alla mancanza di condizioni tecniche e organizzative che garantiscano coerenza intersoggettiva.

La questione centrale sollevata dalla docimologia riguarda la vulnerabilità delle decisioni valutative, soprattutto quando incidono sul futuro degli studenti. Ogni volta che un esame assegna titoli, consente l'accesso a percorsi successivi o influisce sulle opportunità di carriera, la valutazione non dovrebbe dipendere troppo dal corretto di chi corregge. Se ciò avviene, la scuola rischia di trasformare l'impegno per l'equità in un sistema in cui non conta solo ciò che uno studente ha effettivamente dimostrato, ma anche chi corregge e le sue aspettative o idee sul "buon compito", sulla "risposta adeguata" o, più in generale, sullo "studente meritevole".

Un contributo importante in questa direzione è stato dato dagli studi francesi di Piéron sul *baccalauréat*, in cui gli stessi elaborati venivano valutati da diversi gruppi di insegnanti: le differenze, talvolta anche considerevoli, evidenziavano la mancanza di criteri stabili e condivisi (de Landsheere, 1973). Da questi risultati emerge un punto centrale: il voto non è una misura neutra dell'apprendimento, poiché dipende dalle rappresentazioni, dalle aspettative e dallo stile di valutazione del correttore. La variabilità incontrollata del giudizio può così diventare un elemento di ingiustizia sistemica. Di fronte a queste evidenze, la dottrina della docimologia ha spostato l'attenzione dalla responsabilità individuale degli insegnanti al funzionamento complessivo del sistema valutativo. Il problema non è nel comportamento morale del correttore, ma nella mancanza di condizioni che rendano affidabile la valutazione intersoggettiva. Da questa prospettiva, è essenziale definire chiaramente il concetto di oggettività. Non si tratta di eliminare il giudizio professionale, che sarebbe sia impossibile che indesiderabile nelle prove aperte, ma di creare condizioni che rendano tale giudizio controllabile, coerente e basato su criteri dichiarati. Come osserva Bottani (1986), l'oggettività nella valutazione scolastica non coincide con l'assenza di interpretazione, ma richiede criteri e procedure espliciti e controllabili. Il contributo della docimologia ha inoltre mostrato che la variabilità dei giudizi valutativi può essere ridotta mediante strumenti adeguati e criteri condivisi (Vertecchi, 2003). A tal fine, sono state sviluppate pratiche come la doppia correzione, la definizione di scale di valutazione condivise, l'adozione di criteri condivisi, la formazione dei commissari e, quando necessario, l'uso di prove strutturate, tutte strategie finalizzate a ridurre la variabilità non necessaria. La tradizione italiana, soprattutto dagli anni Cinquanta, ha affrontato queste riflessioni in un contesto storico segnato dall'espansione dell'istruzione e dall'esigenza di rendere più chiare e giustificabili le decisioni valutative. Visalberghi (1955) afferma che il problema non sta nella natura tecnica o meno delle prove, ma nella necessità di formulare giudizi "giusti e obiettivi quanto

possibile", riconoscendo che ogni valutazione combina sempre elementi di misura e di giudizio. In continuità con questa visione, Notti (2014) ha riabilitato la docimologia come disciplina essenziale per la professionalità docente. Egli sostiene che una valutazione equa deriva dal riconoscimento dei fattori soggettivi che influenzano l'esaminatore e dalla capacità di gestirli attraverso strumenti progettati con attenzione. L'approccio si distingue per la capacità di collegare tre livelli spesso separati nella pratica scolastica: quello tecnico (metodi di costruzione delle prove e delle rubriche), quello psicopedagogico (influenze di stereotipi, aspettative, effetti cognitivi) e quello istituzionale (significato della certificazione e bisogno di comparabilità).

L'applicazione di questo impianto negli Esami di Stato evidenzia implicazioni fondamentali. La valutazione che assegna titoli di studio deve essere equa e comparabile, poiché si tratta di una decisione pubblica piuttosto che di un atto interno alla classe. Superare la soggettività non significa eliminarla, ma trasformarla in un giudizio argomentato e tracciabile. L'equità dell'esame non può essere garantita soltanto nella fase di correzione, ma deve essere prevista fin dalla progettazione delle prove, dalla definizione dei criteri valutativi e dalla condivisione preliminare delle modalità di giudizio tra i commissari. La letteratura sulla valutazione della scrittura conferma infatti che l'accordo tra valutatori non dipende solo dalla presenza di criteri o rubriche, ma anche dal modo in cui tali strumenti vengono interpretati e applicati. In questa direzione, Lopera-Oquendo, Lipnevich e Máñez (2024) mostrano come, nella valutazione di testi scritti, possano emergere differenze significative tra valutatori anche quando si utilizzano approcci di correzione strutturati, evidenziando il peso delle concezioni professionali, degli atteggiamenti verso il feedback e delle caratteristiche individuali dei correttori. Ne deriva che la costruzione di prove e criteri condivisi deve essere accompagnata da momenti di confronto e calibrazione tra valutatori, affinché il giudizio finale sia il più possibile trasparente, coerente e giustificabile. Analogamente, alcune ricerche sulle pratiche di attribuzione dei voti mostrano che la valutazione non può essere considerata soltanto un'operazione tecnica, ma va letta anche come una pratica professionale situata, influenzata da criteri, valori e consuetudini propri dei diversi contesti scolastici. In particolare, gli studi comparativi sulle pratiche di *grading* evidenziano che il significato attribuito ai voti, i criteri adottati e le modalità con cui gli insegnanti mirano a garantire equità e coerenza possono variare in relazione al contesto culturale e istituzionale di riferimento (Cheng et al., 2020). Anche Gamlem, Segaran e Moltudal (2024), analizzando l'uso di una scala di valutazione a 26 punti, mostrano come le scale di voto siano interpretate dagli insegnanti non solo in termini di precisione, ma anche come strumenti comunicativi e motivazionali. Ne deriva che la soggettività professionale non può essere eliminata del tutto; può però essere resa più consapevole e orientata attraverso criteri condivisi, strumenti adeguati e percorsi di formazione alla valutazione.

La docimologia ci ricorda che l'equità non dipende solo dalla buona volontà dei docenti, ma dalla creazione di un sistema di criteri, strumenti e procedure che minimizzino l'effetto del caso. Per l'Esame di Stato, che mira a essere un certificato equo, è fondamentale rispondere a questa domanda: in che misura il risultato è influenzato dalla prestazione dello studente e in che misura invece da come si svolge la valutazione? Solo un approccio sistematico, basato sulla progettazione, sulla formazione e sulla consapevolezza professionale, può rendere il giudizio educativo credibile e socialmente accettabile.

2.4 Il rapporto tra misurazione e valutazione

La valutazione nei sistemi educativi contemporanei è uno degli aspetti più complessi e delicati dell'esperienza scolastica, poiché coinvolge aspetti pedagogici, responsabilità professionali e legittimazione istituzionale. Non si limita alla semplice misurazione delle prestazioni o a una procedura amministrativa: è un processo culturale e relazionale che aiuta a definire cosa una comunità consideri "apprendimento", "merito", "successo" e "fallimento". Per questo motivo, la valutazione non può essere affrontata solo dal punto di vista tecnico, ma richiede una riflessione approfondita sulla sua natura formativa, sul legame tra evidenze e giudizi, e sulle condizioni che ne garantiscono il significato.

Nella pratica quotidiana, "misurare" una prestazione e "valutare" uno studente tendono a coincidere, poiché entrambi portano a un risultato visibile: un voto, un giudizio, un livello. Tuttavia, questa apparente sovrapposizione rappresenta uno degli aspetti più delicati della cultura valutativa: confondere misurazione e valutazione porta a dare ai numeri un potere che non possiedono, o a considerare i giudizi come certi quando in realtà non lo sono. Discutere del rapporto tra misurazione e valutazione serve quindi a chiarire un problema che è allo stesso tempo tecnico e politico: tecnico, perché riguarda la qualità delle evidenze su cui si basano decisioni educative; politico, perché queste decisioni hanno effetti concreti sui percorsi degli studenti e sulla distribuzione delle opportunità. Per misurare correttamente, si intende descrivere una prestazione attraverso una procedura esplicita, replicabile e soggetta a controlli di affidabilità; valutare invece significa interpretare tale descrizione in base a finalità formative, criteri di qualità, contesti e conseguenze. Questa distinzione non è astratta, ma fondamentale per evitare due errori opposti. Il primo consiste nel confondere la misurazione con la valutazione, come se il punteggio fosse un giudizio intrinseco; il secondo, nello scambiare la valutazione con semplice impressione, come se l'esperienza del docente fosse sufficiente a garantire equità e coerenza. Visalberghi (1955) insiste proprio su questo punto, sottolineando che il valore della misurazione non sta nel numero in sé, ma nell'uso che se ne fa: "il problema non è tanto la misurazione in sé, quanto il significato che le si attribuisce" (Visalberghi, 1955, p. 11). Nello stesso

contesto, propone una distinzione che, se presa sul serio, aiuta a evitare molti equivoci contemporanei: “il test come tale non valuta più, misura” (Visalberghi, 1955, p. 18).

Solo dopo aver compreso questa differenza è possibile capire perché, nella storia della valutazione scolastica, la questione della soggettività sia stata affrontata non solo come un problema di “giustizia” individuale, ma anche come un problema legato alle modalità di valutazione e alle garanzie. In questo quadro, il paragrafo precedente, che tratta degli studi docimologici, fornisce il contesto essenziale: la critica docimologica non cercava di eliminare il giudizio, ma di ridurre le componenti di casualità e arbitrarietà, che emergono quando il giudizio non è supportato da criteri condivisi e strumenti affidabili.

Ma gli strumenti progettati per controllare l'arbitrarietà possono, se usati senza una corretta cultura valutativa, diventare una forma di semplificazione. Visalberghi lo esprime con un’immagine che è sia epistemologica che morale: “non c’è nessuna ragione di fondo perché la valutazione debba limitarsi ad accertare i bianchi e i neri, ma sia incapace di percepire gradazioni di grigio” (Visalberghi, 1955, p. 16). La misura è utile quando aumenta la trasparenza (che cosa si osserva? con quali criteri? con quale livello di precisione?) e quando permette confronti più equi; diventa dannosa quando viene trattata come sostituto del giudizio educativo. Qui si colloca un secondo avvertimento, altrettanto attuale: “il test come tale non valuta più, misura” (Visalberghi, 1955, p. 18). Con questa frase Visalberghi non intende rifiutare i test, ma suggerisce di inserirli nel contesto appropriato: sono strumenti di raccolta di evidenze, non strumenti di decisione. Avere dei numeri può far credere di avere automaticamente oggettività. In realtà, anche i numeri possono trasmettere valori tanto quanto le parole: scegliere cosa misurare, con quali elementi, in quali domini, utilizzando quale modello di scala e con quali soglie di livello, implica già una rappresentazione di apprendimento e di scuola. Per questo motivo, nella letteratura internazionale sulla validità (Messick, 1989; Kane, 2006; AERA, APA, & NCME, 2014) si sottolinea l’importanza di considerare la validità non come un semplice “timbro” tecnico, bensì come un’argomentazione pubblica sull’interpretazione e sull'uso dei punteggi.

A questo punto il nodo centrale diventa più chiaro: misurazione e valutazione non vanno né confuse né separate. Se le si confonde, i punteggi diventano “verità” e i criteri spariscono nell’ombra; se le si separa, la misura resta muta (un dato senza senso) e la valutazione torna a essere un atto opaco, non discutibile e quindi poco giustificabile. La tradizione italiana, quando è più rigorosa, insiste proprio su questa dialettica. Da un lato, Visalberghi ricorda che la misurazione nasce dalla valutazione e nella valutazione confluisce; dall’altro, precisa che la valutazione non è il “dopo” della misurazione, ma attraversa anche la fase di costruzione degli strumenti, perché scegliere “cosa” chiedere e “perché” è già un atto di valore (Visalberghi, 1955, pp. 18–19).

Questo tema emerge nuovamente in modo significativo anche tra autori contemporanei che hanno contribuito a modernizzare la riflessione sulla valutazione. Ad esempio, Castoldi (2015) propone di distinguere nel processo di valutazione "due fasi [...]: la fase istruttoria [...] e la fase di espressione del giudizio" (Castoldi, 2015, p. 7). In poche parole, distingue tra misurazione come fase descrittiva (che comprende raccolta, documentazione e controllo della qualità dei dati) e valutazione come fase interpretativa (che assegna un valore basato su criteri dichiarati). Tuttavia, c'è anche un avvertimento metodologico che si trasforma in un principio etico: ridurre il giudizio a una semplice somma o a una media di misure rende il processo meno rigoroso e, al contempo, meno trasparente. Castoldi smonta questa falsa credenza diffusa nella scuola, quella del "mito dell'oggettività": la credenza che il numero possa fungere da protezione, come un camice bianco nascosto dietro cui si celano ansie e responsabilità (Castoldi, 2015). Misurare, se fatto giudiziosamente, significa rendere esplicite le regole con cui si trasforma una prestazione in un'informazione: quali compiti, quali criteri di correzione, quale scala, quale margine di incertezza. Valutare, se fatto giudiziosamente, significa esplicitare i criteri con cui quell'informazione viene interpretata: quale standard di padronanza, quali finalità (formative, sommative, orientative), quali conseguenze. Non esiste una misura "neutra" poiché ogni misura è costruita, e non c'è una valutazione "pura" perché ogni valutazione necessita di evidenze, altrimenti rischia di essere solo un'impressione. Un altro modo efficace di chiarire questa relazione è ricordare che misurare non significa semplicemente "quantificare", ma stabilire una corrispondenza controllata tra le osservazioni e un sistema di riferimento. Questo principio si applica sia alle misure più semplici, come il conteggio delle risposte corrette, sia a quelle più complesse, come le scale sviluppate tramite modelli psicometrici. In questo contesto, la definizione di Trinchero si rivela utile perché evidenzia la natura "comparativa" della valutazione, intesa come "un processo in cui una situazione osservata viene messa a confronto con una situazione attesa" (Trinchero, 2024, p. 13). La misurazione si occupa della parte "osservata", ossia della descrizione controllata della performance, mentre la valutazione stabilisce la relazione tra ciò che si osserva e ciò che si considera desiderabile o sufficiente. In ogni caso, la misurazione implica sempre l'idea di errore: non esiste un punteggio senza incertezze, e nessuna decisione può essere presa senza un certo rischio. Se la scuola non considera l'errore di misura, finisce per trattare il voto come un dato assoluto e naturale, anziché come un indicatore. È anche per questo motivo che la letteratura docimologica e psicometrica sottolinea l'importanza dell'affidabilità e della validità come condizioni minime per parlare di equità: senza verifiche di affidabilità, non possiamo essere certi di osservare qualcosa di stabile; senza una validazione dell'interpretazione, non possiamo essere sicuri di usare correttamente le informazioni ottenute (AERA, APA, & NCME, 2014; Kane, 2006). La docimologia, in questa prospettiva, non si limita alla storia dei test; rappresenta una pedagogia della responsabilità valutativa. Notti (2010, 2017)

sottolinea che la valutazione è una funzione educativa, non semplicemente un atto finale di certificazione, e che la qualità della valutazione dipende dalla capacità di integrare scopi, criteri e strumenti, evitando sia una visione troppo tecnicistica sia un arbitrio basato sull'impressione (Notti, 2010; Notti, 2017; Coggi & Notti, 2002). In altri termini, la misurazione è parte integrante della valutazione, ma non la definisce completamente e la valutazione implica scelte di valore, pur rimanendo obbligata a motivare, documentare e discutere le proprie decisioni. È proprio questa possibilità di discussione a segnare il passaggio da una valutazione come potere a una valutazione come pratica professionale. Se la valutazione rappresenta un giudizio educativo, deve essere motivabile; se la misurazione fornisce dati, deve poter essere verificata. La regola ricordata da Domenici "per ogni funzione e obiettivo specifico della valutazione va impiegato uno strumento di verifica omologo e congruente con quella funzione e con quell'obiettivo" (Domenici, 1993, p. 62) va letta esattamente in questo senso: non esistono strumenti "migliori" in assoluto, esistono strumenti coerenti con una finalità e con un'idea di apprendimento.

Arriviamo così al punto più delicato del rapporto tra misurazione e valutazione: non basta chiedere se uno strumento è "oggettivo"; bisogna chiedersi quale tipo di oggettività promette e quale tipo di soggettività richiede. Una prova standardizzata può rafforzare l'equità riducendo le fonti di distorsione come criteri impliciti, severità variabile e fattori esterni. Tuttavia, può anche introdurre nuove distorsioni se limita il curricolo, viene usata come scorciatoia decisionale o tenta di misurare aspetti non definiti chiaramente. Per questo motivo, la letteratura internazionale sottolinea l'importanza della validità: non si tratta di definire se un test sia "valido" in modo astratto, ma di verificare la ragionevolezza delle inferenze e degli usi che si ne fanno (Messick, 1989; Kane, 2006).

Se si considerano le pratiche scolastiche, il problema non riguarda semplicemente "più misurazione" o "meno misurazione", ma la qualità della relazione tra evidenza e giudizio. La misura diventa giusta quando rende evidenti le regole e permette il confronto; la valutazione diventa giusta quando evidenzia i criteri e favorisce la motivazione. Quando ciò avviene, la distinzione tra misurazione e valutazione non separa: le organizza. È lo stesso principio che Visalberghi esprimeva chiaramente, invitando non a scegliere prove "più calde", ma prove utili "per un giudizio impegnativo e motivato [...] che deve perciò esser il più possibile giusto e obiettivo" (Visalberghi, 1955, p. 16). In questa frase si percepisce già una visione della valutazione in cui l'oggettività non mira a eliminare il giudizio, ma a rafforzarne la validità.

2.5 Tra forme e funzioni valutative

Nel contesto scolastico, la valutazione assume funzioni diverse a seconda delle attività e degli obiettivi specifici. Per sintetizzare differenze e funzioni, la valutazione sommativa serve a rendere conto di un risultato per prendere decisioni come l'ammissione, la certificazione o la selezione. La valutazione diagnostica identifica ostacoli e prerequisiti, non per classificare, ma per guidare l'azione didattica. La valutazione formativa, infine, riguarda il modo in cui le evidenze vengono rielaborate nel processo di apprendimento.

La letteratura internazionale ha contribuito a chiarire questa distinzione anche attraverso la triade *assessment of learning*, *assessment for learning*, *assessment as learning*. Il primo richiama la dimensione certificativa; il secondo identifica pratiche valutative progettate per sostenere l'apprendimento; il terzo enfatizza il ruolo dello studente che, attraverso criteri e feedback, diventa progressivamente capace di autoregolare il proprio lavoro. *L'assessment for learning* sottolinea che la valutazione ha senso quando guida l'insegnamento e l'apprendimento in itinere (Black & Wiliam, 1998; Wiliam, 2011). Mentre l'approccio *assessment as learning* va oltre, poiché collega la valutazione allo sviluppo di competenze metacognitive e autoregolative, cioè alla capacità dello studente di comprendere i criteri, controllare le strategie e utilizzare il feedback per migliorare (Panadero, 2017).

Il mio approccio sistemico-relazionale considera ogni esperienza scolastica come il risultato di molte relazioni che si intrecciano e si evolvono nel tempo. Per questo motivo, considero la valutazione non come un atto neutro o puramente tecnico, ma come un evento plasmato dalle interazioni quotidiane tra studenti, insegnanti, famiglie e organizzazioni scolastiche (Bateson, 1977; Bronfenbrenner, 1979). Approcciando la valutazione in quest'ottica, si riconosce che un voto o un commento può influenzare l'autopercezione dello studente e il clima della classe, poiché ciò che avviene nella valutazione fa parte di un sistema complesso di relazioni scolastiche. In questo quadro epistemologico, i contributi sull'*assessment for learning* e sull'*assessment as learning* aiutano a concepire la valutazione come un processo che supporta l'apprendimento, e non semplicemente lo misura. Queste prospettive evidenziano come il feedback, se costruito collettivamente e spiegato chiaramente, possa aiutare gli studenti a comprendere cosa stanno imparando e quali passi possono compiere per migliorare (Black & Wiliam, 1998; Corsini, 2018; Corsini, 2017; Earl, 2013). Quando la valutazione si trasforma in un dialogo, gli studenti iniziano a riflettere sulle proprie strategie, a riconoscere negli errori un'opportunità di crescita e a sentirsi più coinvolti nel percorso scolastico.

L'approccio formativo della valutazione si manifesta, ad esempio, nelle pratiche di restituzione narrativa, nell'uso di rubriche co-costruite e nelle dinamiche di peer-feedback, che consentono agli studenti di confrontarsi sulle strategie cognitive adottate e di riconoscere come errori e difficoltà

possano trasformarsi in opportunità di apprendimento significativo. Questo metodo trova riscontro negli studi italiani che sottolineano la valenza formativa, partecipativa e relazionale della valutazione, la chiarezza dei criteri, la negoziazione degli obiettivi, la mediazione didattica e il miglioramento dei processi scolastici attraverso forme di valutazione incentrate sulla qualità. (Castoldi 2009; Castoldi 2021; Corsini, 2018; Damiano, 2013; Faggioli & Allulli, 2014).

Negli ultimi anni, accanto a queste distinzioni, si è sviluppata una maggiore attenzione verso le forme di *assessment* inclusive. Qui, l'obiettivo non è solo "adattare" strumenti ai bisogni specifici, ma sviluppare pratiche di valutazione che permettano a tutti gli studenti di partecipare, riducendo le barriere e rendendo i risultati più equi e facilmente interpretabili. Una definizione efficace afferma che l'*assessment* inclusivo "considera i bisogni di tutti gli studenti, assicurando che tutti partecipino alle procedure di valutazione" (Kefallinou & Donnelly, 2016, p. 210). Coerentemente, l'UNESCO ha sottolineato come le pratiche di *assessment* che classificano e gerarchizzano rigidamente possano creare esclusione, mentre il monitoraggio continuo dell'apprendimento costituisce una componente della buona didattica e può favorire l'inclusione.

2.6 Il feedback educativo: definizioni e indicazioni operative

Parlare di feedback significa entrare nel nucleo del rapporto tra misurazione e valutazione, affrontando la sua dimensione più concreta. Il feedback rappresenta sia informazione che azione: non basta comunicare un risultato, bisogna anche renderlo utile per migliorare. Secondo la letteratura di riferimento, un contributo chiave è quello del modello di Hattie e Timperley, che distingue vari livelli di feedback (sul compito, sul processo, sull'autoregolazione e sulla persona) dimostrando come la loro efficacia possa variare. Hattie e Timperley offrono una definizione ormai classica che chiarisce bene la questione: il feedback è "information provided by an agent […] regarding aspects of one's performance or understanding" (Hattie & Timperley, 2007, p. 81).

Questa definizione è utile perché evidenzia subito due aspetti principali. Innanzitutto, il feedback è un tipo di informazione, quindi richiede chiarezza e pertinenza. In secondo luogo, non esiste un feedback "neutro": l'informazione viene letta, interpretata, accettata o rifiutata, e la sua efficacia dipende dal contesto e dalla fase dell'apprendimento. Hattie e Timperley (2007) sottolineano che il feedback non ha effetto "nel vuoto", ma funziona quando riesce a ridurre la distanza tra la comprensione attuale e l'obiettivo di apprendimento. Inoltre, una tradizione importante legata al *formative assessment* (Sadler, 1989) definisce la condizione minima affinché il feedback sia davvero formativo: lo studente deve avere un'idea dello standard, confrontare la propria prestazione con tale standard, e intraprendere un'azione per ridurre lo scarto. Come evidenziato da Hattie e Clarke, "L'apprendente deve a) aver familiarità con il concetto di standard [...], b) confrontare il livello

effettivo [...] di performance con lo standard, e c) adottare le azioni appropriate [...]" (Sadler, 1989, come citato in Hattie & Clarke, 2019, p. 4). Da questa base emergono alcune indicazioni operative, supportate in modo coerente dalla ricerca. Innanzitutto, il feedback risulta più efficace se si concentra sul compito, sul processo e sulle strategie, anziché sulla persona. Questa impostazione è importante anche per garantire equità, poiché riduce l'effetto stigmatizzante dei giudizi globali e aiuta lo studente a capire cosa fare in seguito. Secondo la rassegna di Harks e colleghi (2014), la differenza tra feedback orientato al processo e feedback orientato al voto mostra effetti sistematici sulla percezione dell'utilità del feedback e, indirettamente, su apprendimento e interesse: il feedback processuale è percepito come più utile, e questa percezione si collega a miglioramenti in *achievement* e interesse.

In secondo luogo, la qualità del feedback dipende dalla chiarezza dei criteri adottati. Se i criteri sono impliciti, lo studente riceve informazioni ma non ha una vera "mappa" su come usarle. Per questa ragione, la valutazione formativa è inseparabile dal processo di definizione condivisa di standard, rubriche ed esempi di prestazione. Sadler (1989) sostiene che il feedback è efficace quando lo studente conosce lo standard, confronta la propria prestazione con esso e possiede strategie per ridurre la distanza tra la propria prestazione e lo standard. La responsabilità non ricade solo sull'insegnante che fornisce i riscontri, ma anche sull'ambiente di apprendimento che consente di agire per migliorare. È per questa ragione che Nicol e Macfarlane-Dick collegano strettamente feedback e autoregolazione, proponendo principi di buona pratica che prevedono occasioni per utilizzare il feedback, il dialogo e lo sviluppo della capacità di autovalutazione (Nicol & Macfarlane-Dick, 2006).

Uno dei motivi principali della valutazione esterna è il richiamo al principio dell'opportunity to learn: non si dovrebbero confrontare semplicemente gli esiti senza considerare il curriculum effettivo, le opportunità di apprendimento e la coerenza tra le prove e l'insegnamento (Lichtner, 2004, p. 295).

Terzo, il feedback dovrebbe tutelare la motivazione non con "lodi" generiche, ma con indicazioni precise di miglioramento e con il riconoscimento dei progressi nel tempo. La ricerca mostra chiaramente che il confronto sociale può ridurre l'efficacia formativa del feedback, mentre riferimenti critici o individuali risultano più efficaci (Harks et al., 2014). Recentemente, gli studi si sono spostati dal focus sul feedback "dato" a quello "usato". Carless e Boud (2018) introducono il concetto di feedback literacy, ossia l'insieme di capacità e attitudini necessarie per interpretare e utilizzare le informazioni al fine di migliorare. Questo spiega perché, anche con feedback scritto accurato, gli esiti variano molto in base alla preparazione degli studenti nel comprenderlo, accettarlo e trasformarlo in azioni.

Ai fini di questa ricerca è utile distinguere tra l'informazione restituita e il suo uso. Con "descrizione dei livelli", "esito descrittivo" o "restituzione descrittiva" si indica il documento con cui l'INVALSI comunica allo studente il livello raggiunto e i compiti che lo caratterizzano. Con "feedback

formativo" si intende invece l'uso di questa informazione nei processi di apprendimento: nella prospettiva richiamata in questo paragrafo, un'informazione diventa feedback formativo solo quando lo studente può comprenderla, confrontarla con uno standard e usarla per orientare le azioni successive (Sadler, 1989; Hattie & Timperley, 2007; Carless & Boud, 2018). L'espressione "feedback descrittivo", diffusa nella letteratura, è mantenuta quando si riportano le parole dei partecipanti o le denominazioni degli strumenti di ricerca; nel resto del testo si usano i termini riferiti alla restituzione.

2.7 Le prove INVALSI: dalla valutazione di sistema alla restituzione degli esiti al singolo studente

Nel dibattito contemporaneo sulla valutazione educativa si è progressivamente affermata una concezione della valutazione non limitata alla misurazione finale degli apprendimenti, ma strettamente connessa alla regolazione dei processi di insegnamento e apprendimento. In questa prospettiva, le informazioni prodotte dalla valutazione assumono valore nella misura in cui vengono utilizzate per orientare le scelte didattiche, sostenere il feedback e favorire una maggiore consapevolezza degli studenti del proprio percorso formativo (Black & Wiliam, 1998; Andrade et al., 2021; Hattie & Clarke, 2018).

Accanto a questa dimensione più vicina alla pratica didattica quotidiana, le rilevazioni standardizzate su larga scala svolgono una funzione diversa, ma complementare. Esse permettono di raccogliere dati comparabili sugli apprendimenti e sul funzionamento dei sistemi di istruzione, offrendo elementi utili per il monitoraggio, la rendicontazione e l'elaborazione di politiche educative. Il loro contributo al miglioramento, tuttavia, non dipende soltanto dalla qualità tecnica degli strumenti, ma anche dal modo in cui i risultati vengono interpretati, discussi e ricondotti ai contesti in cui vengono prodotti (Kellaghan et al., 2009; Brink, 2020).

Tali funzioni non sono però neutre, ma possono generare tensioni strutturali legate sia al livello di analisi sia alle modalità di restituzione dei risultati (Biesta, 2010). In particolare, si osserva una duplice funzione dei dati: da un lato, l'uso aggregato è funzionale al governo del sistema, dall'altro, la restituzione individuale è orientata agli studenti e quindi funzionale al supporto dei processi di apprendimento (OECD, 2013; Looney, 2011). In questo contesto, il caso italiano delle prove INVALSI rappresenta un esempio significativo di un'evoluzione dalla logica esclusivamente sistemica a una progressiva attenzione al singolo studente.

Le prove INVALSI, nella loro configurazione originaria, sono state progettate come strumenti di valutazione del sistema educativo, in linea con l'impostazione del Sistema Nazionale di Valutazione. I risultati venivano presentati in forma aggregata, permettendo analisi comparative tra classi, scuole

e territori, e collocando le performance degli studenti in un contesto più ampio, come le macro-aree e il contesto nazionale. Questa decisione riflette un modello tipico delle valutazioni su larga scala, finalizzate a fornire dati affidabili per il monitoraggio complessivo del sistema, senza influenzare direttamente i percorsi individuali degli studenti (van Rijn et al., 2024). In linea con questa impostazione, tali valutazioni sono state interpretate nella letteratura come strumenti assimilabili a "termometri" del sistema educativo, capaci di rilevarne lo stato e l'evoluzione nel tempo, ma non orientati alla certificazione delle competenze individuali. Un primo elemento di discontinuità è stato introdotto con il decreto legislativo 13 aprile 2017, n. 62, che ha ridefinito il quadro normativo della valutazione nel sistema scolastico italiano. Il decreto stabilisce che le rilevazioni INVALSI sono volte a verificare i livelli di apprendimento delle studentesse e degli studenti, introducendo esplicitamente un riferimento all'individuo. Nonostante la funzione principale delle prove continui a essere quella di monitoraggio sistemico, questa formulazione rappresenta un passaggio significativo, in quanto pone l'accento sulla certificazione degli apprendimenti individuali, misurati attraverso le prove nazionali standardizzate, prefigurando un possibile uso personalizzato dei risultati. Inoltre, il provvedimento rafforza la posizione delle prove INVALSI all'interno del percorso scolastico, prevedendone l'obbligatorietà ai fini dell'ammissione agli esami di Stato e consolidando il legame tra valutazione standardizzata e carriera scolastica. Un ulteriore cambiamento significativo si è concretizzato con l'introduzione, a partire dal 2018, dei livelli descrittivi di competenza. Tali livelli rappresentano un cambiamento rilevante nel modo in cui vengono restituiti gli esiti, in quanto descrivono in termini qualitativi ciò che gli studenti hanno dimostrato di saper fare nelle diverse prove, superando la logica dei punteggi numerici. Questa scelta si inserisce nel paradigma dei sistemi standard based, che si fonda sull'idea che la valutazione non debba essere interpretata come una semplice misurazione quantitativa, ma come una rappresentazione dei livelli di competenza degli studenti rispetto a standard esplicitamente definiti e condivisi (nel caso delle prove standardizzate INVALSI non ci si riferisce a dei veri e propri standard ma si fa riferimento alla normativa che definisce obiettivi di apprendimento in termini di Indicazioni nazionali e Linee guida). In questa prospettiva, il dato valutativo acquista significato solo se ancorato a descrizioni qualitative di ciò che lo studente sa e sa fare, consentendo di utilizzare i risultati per migliorare l'apprendimento (Looney, 2011). Allo stesso tempo, questa modalità di restituzione degli esiti si allinea con le indicazioni presenti nella letteratura sulla valutazione formativa, che evidenzia l'importanza di fornire feedback chiari e rilevanti per gli studenti, capaci di orientare i processi di apprendimento e di autoregolazione (Black & Wiliam, 1998). In questa prospettiva, i livelli INVALSI non costituiscono solo un dispositivo di monitoraggio, ma anche uno strumento potenzialmente utilizzabile in chiave pedagogica.

Il passaggio a una restituzione pienamente individuale si consolida con la normativa più recente, in particolare con il decreto-legge n. 127 del 2025 e il decreto ministeriale n. 2 del 2026, che prevedono l'inserimento dei risultati delle prove INVALSI nel curriculum dello studente, allegato al diploma finale. Nel documento rilasciato, i livelli di apprendimento sono riportati in forma descrittiva per ciascuna disciplina e riferiti esplicitamente allo studente, configurandosi come una rappresentazione delle competenze acquisite nel corso del percorso scolastico. La normativa chiarisce che tali informazioni devono essere inserite successivamente alla conclusione dell'esame di Stato, al fine di evitare qualsiasi interferenza con la valutazione finale, e ne attribuisce una funzione prevalentemente orientativa e informativa. Questo passaggio è molto importante perché cambia il modo in cui si valuta, perché unisce la valutazione di sistema e quella individuale, tenendole insieme in un unico processo. Il quadro internazionale evidenzia come questa evoluzione si inserisca in una pluralità di modelli di restituzione dei risultati delle valutazioni standardizzate. Nei sistemi anglosassoni, la restituzione individuale rappresenta una pratica consolidata e strutturata: gli studenti ricevono report dettagliati che includono punteggi, livelli di competenza e, in molti casi, indicazioni operative per il miglioramento. Tali sistemi si fondano su una forte enfasi sugli standard di apprendimento e sulla responsabilità delle istituzioni scolastiche rispetto ai risultati conseguiti (Looney, 2011). Nei Paesi nordici, come la Finlandia, la Svezia e la Norvegia, le valutazioni standardizzate tendono a essere integrate nelle pratiche didattiche e interpretate in chiave formativa, con l'obiettivo di sostenere i processi di apprendimento piuttosto che di esercitare funzioni di controllo o selezione. Tale impostazione riflette sistemi educativi caratterizzati da un'elevata fiducia nella professionalità degli insegnanti e da un uso limitato delle evidenze valutative a fini di responsabilità (Looney, 2011; OCSE, 2013). Diversamente, in alcuni Paesi dell'Europa continentale, come la Francia e la Germania, la valutazione si basa su esami nazionali o regionali, i cui risultati hanno valore certificativo e influenzano direttamente le carriere scolastiche e professionali degli studenti. In questi contesti, le prove assumono una funzione selettiva e regolativa, configurandosi come strumenti di valutazione ad alta posta in gioco (high stakes), in grado di determinare le opportunità educative e sociali degli studenti (OCSE, 2013; Eurydice, 2020). Guardando a queste realtà, si evince che il modello italiano si trova oggi in una posizione intermedia. Questa impostazione riflette una scelta strategica volta a valorizzare il potenziale informativo delle prove senza trasformarle in strumenti di selezione. Il processo descritto può essere interpretato alla luce del quadro teorico proposto da Looney (2011), secondo il quale i sistemi educativi stanno progressivamente cercando di integrare valutazioni sommative e formative in un sistema coerente. Tuttavia, tale integrazione presenta criticità. La letteratura evidenzia che i dati prodotti dalle valutazioni standardizzate, pur garantendo elevati livelli di comparabilità, richiedono una mediazione pedagogica per essere interpretati in modo significativo

e utilizzati a supporto dell'apprendimento. In questo processo, il ruolo centrale spetta ai docenti, chiamati a tradurre i contenuti dei risultati in azioni educative concrete (Shepard, 2000; Andrade & Brookhart, 2020). In questa prospettiva, il docente non si limita a "trasmettere" il dato valutativo, ma svolge una funzione interpretativa e traduttiva, trasformando l'informazione proveniente dalle prove standardizzate in indicazioni significative per lo sviluppo e il miglioramento degli apprendimenti. Come sottolineano Moss e Brookhart (2019), la valutazione ha un impatto sull'apprendimento solo quando i dati vengono contestualizzati, discussi e resi comprensibili agli studenti all'interno delle attività didattiche, implicando un uso professionale e riflessivo delle evidenze da parte degli insegnanti. In Italia, diversi studi hanno evidenziato come le prove INVALSI siano oggetto di percezioni e atteggiamenti contrastanti da parte dei docenti, oscillanti tra il riconoscimento della loro utilità informativa e la critica riguardo al loro impatto sulle pratiche didattiche e sull'autonomia professionale (Benadusi & Giancola, 2020). In particolare, la letteratura sottolinea come una parte degli insegnanti percepisca le rilevazioni standardizzate come strumenti esterni e poco integrati con il lavoro quotidiano in classe, evidenziando resistenze culturali e professionali che possono limitarne l'uso a fini pedagogici (Benadusi & Giancola, 2020). Questa distanza è stata interpretata come il risultato di una storica separazione tra la valutazione interna, legata alla pratica didattica, e la valutazione esterna, finalizzata al monitoraggio del sistema. Tale separazione può generare difficoltà nell'appropriazione e nell'utilizzo dei dati da parte degli insegnanti (OECD, 2013). Inoltre, come sottolineato da Benadusi e Giancola (2020), le prove standardizzate possono essere percepite da alcuni docenti come strumenti di controllo esterno potenzialmente in contrasto con l'autonomia professionale e con la complessità delle pratiche educative, favorendo atteggiamenti di resistenza o un utilizzo meramente burocratico e formale dei risultati. Alcuni studi sulla cultura valutativa delle scuole hanno mostrato che l'effettivo impiego dei dati delle rilevazioni su larga scala dipende in larga misura dalle rappresentazioni che gli insegnanti hanno della valutazione standardizzata stessa e dalla loro disponibilità a riconoscerne l'importanza per una valutazione dell'azione didattica (Datnow & Park, 2014). In questa prospettiva, la difficoltà di integrare la valutazione standardizzata e le pratiche di classe non è solo una questione formale, ma culturale e professionale, che influenza l'uso di questi strumenti per la formazione e il cambiamento. La restituzione individuale degli esiti delle prove standardizzate può, quindi, assumere un valore significativo in termini di opportunità per attivare processi riflessivi nello studente, favorendo una maggiore consapevolezza del proprio apprendimento. Il potenziale di un feedback, infatti, non risiede tanto nell'informazione in sé, ma nella capacità del soggetto di saperla usare per orientare le proprie azioni future, trasformandola in risorsa per il miglioramento (Hattie & Clarke, 2018). Anche i dati raccolti in contesti di valutazione standardizzata possono contribuire alla qualità dell'apprendimento, a condizione che siano resi

significativi all'interno di un processo educativo più vasto. Tuttavia, bisogna ancora capire se e come questi strumenti vengano davvero usati nelle scuole, per pensare a soluzioni che portino a un miglioramento del processo di valutazione complessivo che ruota attorno al sistema scolastico.

PARTE II: LA RICERCA EMPIRICA

CAPITOLO 3: IL CONTESTO DELLA RICERCA

Il presente capitolo introduce il progetto di ricerca, soffermandosi sulla realtà della prova INVALSI di italiano e, in particolare, sulla natura del costrutto valutato e sul contributo dei descrittori dei livelli restituiti in esito alla prova. In questa prospettiva, dunque, l'analisi non si limita a una descrizione tecnica dello strumento, ma intende evidenziare il valore interpretativo dei livelli come dispositivi di restituzione capaci di tradurre l'esito della prova in informazioni leggibili e significative per lo studente rispetto al suo apprendimento. I descrittori, infatti, svolgono un ruolo fondamentale nel rendere espliciti i processi di comprensione e le abilità linguistiche sottese alla prova, fungendo da ponte tra la misurazione standardizzata e la dimensione formativa della valutazione. In tal senso, i descrittori non sono considerati solo come esiti classificatori, ma anche come elementi chiave per sostenere la riflessione didattica e orientare percorsi di apprendimento consapevoli.

3.1 Il costrutto della prova di Italiano INVALSI

Una prova standardizzata di comprensione del testo ha l'obiettivo di misurare empiricamente un costrutto latente. La comprensione, infatti, non può essere osservata direttamente attraverso una prova standardizzata, ma può essere inferita attraverso le prestazioni degli studenti nel rispondere a compiti costruiti sulla base di un modello teorico di riferimento (Messick, 1989; Kane, 2006; Kintsch, 1998). Per costruire quindi una prova è necessario innanzitutto definire come la lettura viene considerata, quali dimensioni è possibile osservare attraverso quella prova e, soprattutto, sapere che ogni scelta comporta inevitabilmente una semplificazione di ciò che si vuole misurare, perché non è possibile valutare la comprensione del testo in modo assoluto, ma solo aspetti specifici di essa tra quelli considerati rilevanti e misurabili.

Sulla base di questa premessa, in questa parte di capitolo verrà descritto il modello teorico di riferimento che sottende al costrutto della comprensione del testo verificata dalla prova INVALSI di Italiano. Il presupposto di partenza riguarda le indicazioni normative che definiscono i confini entro cui andare a delineare tale costrutto. Le prove INVALSI, come ampiamente descritto nella parte teorica, hanno l'obiettivo di misurare gli apprendimenti scolastici, i cui riferimenti sono declinati nei documenti istituzionali quali le Indicazioni Nazionali e le Linee Guida. Per quanto riguarda la prova di Italiano, come ulteriore riferimento, sono stati presi in considerazione i Quadri di riferimento delle prove comparative internazionali che misurano la cosiddetta reading comprehension. In questi

contesti si sottolinea la centralità della comprensione del testo, in quanto competenza chiave per gli individui in formazione scolastica nei diversi paesi. Questa competenza è sostenuta anche nei documenti normativi che regolamentano la scuola italiana, pur non essendo specificate le diverse componenti che compongono un costrutto come la comprensione del testo. In questi testi normativi si esprime la necessità, progressiva nei diversi gradi scolastici, di sviluppare un'adeguata capacità di comprendere testi di vario tipo e genere, costruiti con scopi specifici e scritti con un lessico sempre più ampio e di registri diversi. Una parte più dettagliata di questi documenti riguarda gli aspetti dell'uso della lingua e, in particolare, del lessico, ai quali è stato dedicato uno spazio nella definizione complessiva del costrutto della prova nazionale standardizzata. I diversi ambiti grammaticali oggetto di apprendimento scolastico vengono quindi integrati nel Quadro di Riferimento (QdR) della prova di Italiano come parte di riflessione sulla lingua e lessico, funzionali alla comprensione del testo. A partire quindi da una definizione di padronanza linguistica intesa come possesso ben strutturato di una lingua e capacità di servirsene per vari scopi, tra le varie competenze racchiuse in questo contenitore quali, ascolto, produzione e interazione orale, lettura e comprensione, scrittura, lessico, e riflessione sulla lingua, sono state individuate quelle parti misurabili attraverso una prova nazionale standardizzata inquadrata all'interno di un preciso contesto valutativo.

Per quanto riguarda gli aspetti più specifici della comprensione del testo, si è scelto di declinarla secondo caratteristiche derivate dalla letteratura di riferimento e dalle rilevazioni presenti nelle indagini standardizzate internazionali, i cui costrutti sottendono modelli di derivazione prevalentemente cognitiva. Tra gli approcci teorici più fondati si ritrova il modello costruttivo integrativo di Kintsch (1988) dove la comprensione viene descritta come un processo dinamico che si articola in due fasi interdipendenti: una fase di *costruzione* e una di *integrazione*. Inizialmente, il lettore crea una vasta rete di significati a partire dagli elementi linguistici del testo, utilizzando sia le informazioni esplicite sia le conoscenze pregresse. Questa fase è caratterizzata da una certa ridondanza e apertura, poiché si generano molteplici connessioni di significato potenziali. Nella fase successiva interviene un processo di *selezione* e *stabilizzazione* che, attraverso meccanismi di *attivazione* e *inibizione*, mantiene le informazioni coerenti e scarta quelle incompatibili, portando alla costruzione di una rappresentazione organizzata del testo (Kintsch, 1998). Il risultato di questo processo è la formazione di diversi livelli di rappresentazione del significato attribuito ad un testo. Accanto a una comprensione superficiale e a una microstruttura proposizionale, assume particolare rilevanza il *modello situazionale*, ossia una rappresentazione mentale che integra le informazioni testuali con le conoscenze del lettore e consente di comprendere il significato globale del testo (van Dijk & Kintsch, 1983). In questa prospettiva, la comprensione non consiste nel semplice recupero di

informazioni, ma nella costruzione attiva di un modello coerente che permette di interpretare e riorganizzare i contenuti.

Un elemento fondamentale di questo processo è il ruolo delle conoscenze pregresse, spesso descritte come schemi cognitivi, secondo cui la comprensione è intesa come un processo interattivo in cui il testo fornisce stimoli che attivano strutture di conoscenza già presenti, che orientano la selezione e l'integrazione delle informazioni (Rumelhart, 1980). Questi schemi permettono di attribuire significato anche a informazioni incomplete, o implicite, guidando le aspettative del lettore e rendendo possibile la costruzione di coerenza anche al di là di quello che è stato esplicitamente detto. I processi cosiddetti inferenziali ricoprono, infatti, un ruolo centrale nella comprensione, che richiede una continua produzione di inferenze, necessarie per colmare le lacune del testo, laddove presenti, e garantire la coerenza sia a livello locale sia a livello globale. Le inferenze si attivano quando il testo non chiarisce completamente le relazioni tra le informazioni. Anche gli studi di Rice (1986) hanno evidenziato che la capacità di fare inferenze è collegata alla crescita della competenza linguistica e narrativa, e contribuisce a creare rappresentazioni integrate del testo. Secondo l'autrice, le inferenze permettono di stabilire relazioni causali, temporali e referenziali e agiscono durante tutto il processo di lettura, contribuendo a mantenere la coerenza.

Questi approcci teorici consentono quindi di descrivere la comprensione come un'attività complessa che coinvolge processi distinti ma coordinati, tra cui il riconoscimento delle informazioni, l'attivazione di conoscenze pregresse, l'elaborazione inferenziale, l'integrazione semantica e il monitoraggio della coerenza. Proprio questi processi possono essere distinti sul piano analitico e resi osservabili mediante compiti strutturati.

Il QdR della prova INVALSI di Italiano traduce il suo costrutto in diversi compiti, sintetizzati in due tabelle: una relativa a tre macro-aspetti, che possono essere letti come una traduzione operativa (operazionalizzazione) delle diverse dimensioni della comprensione, e un'altra relativa agli ambiti di riflessione sulla lingua, che ripercorrono i contenuti linguistici oggetto di studio nel percorso di apprendimento degli studenti.

Il primo macro-aspetto, Localizzare e trovare informazioni nel testo, riguarda i processi di accesso alla superficie testuale e alla microstruttura, che implicano il riconoscimento e il recupero di informazioni esplicite. Sul piano cognitivo, questo rappresenta un primo momento in cui il lettore inizia a comprendere il significato di ciò che legge. Questo macro-aspetto verifica le capacità del lettore di muoversi nel testo per individuare gli elementi più importanti su cui poi lavorerà per comprenderlo (Kintsch, 1998).

Il secondo macro-aspetto, Ricostruire il significato del testo, sia a livello locale sia globale, si colloca invece al centro del processo di integrazione. I compiti di questo macro-aspetto riguardano i

meccanismi inferenziali descritti in letteratura, attraverso i quali il lettore collega informazioni anche distanti, costruisce relazioni logiche e temporali e attiva schemi pertinenti per colmare gli impliciti (Lumbelli, 1996; Rice, 1986). Questo livello corrisponde alla costruzione del modello situazionale e rappresenta il nucleo della comprensione profonda (van Dijk & Kintsch, 1983).

Il terzo macro-aspetto, Riflettere sul contenuto o sulla forma del testo, sia a livello locale sia globale, e valutarlo, considera prevalentemente aspetti maggiormente legati a una dimensione metacognitiva e riflessiva poiché si occupa di verificare la capacità di analizzare e valutare il testo, prendendo distanza dal contenuto e interrogandosi sulle scelte linguistiche e comunicative e sulle intenzioni dell'autore. Tale attività richiede un controllo consapevole dei processi di comprensione e il ricorso a schemi più complessi legati ai tipi e ai generi testuali e anche ai contesti d'uso, ampliando il modello situazionale verso una dimensione critica.

A ciascuno di questi macro-aspetti sono associati degli indicatori che rappresentano in maggior dettaglio i compiti proposti nella prova standardizzata INVALSI. Durante la fase di costruzione delle prove, tali indicatori definiscono le specificità di ciascuna domanda, fornendo ulteriori dettagli sul compito proposto (De Santis et al., 2024). Nei contributi di De Santis e Mastrogiovanni (2024) viene descritto in particolare il processo di costruzione della prova, per cui, a partire dalla selezione di testi secondo criteri specifici, indicati sempre nel QdR, si procede alla produzione di domande standardizzate secondo criteri psicometrici codificati e all'attenzione alla copertura del costrutto oggetto di misurazione. Per ogni domanda viene prodotta una descrizione che spieghi il compito che si prefigge di verificare, a partire dall'individuazione dei processi cognitivi coinvolti, riferibili alla tassonomia di Bloom (Bloom et al., 1956; Anderson & Krathwohl, 2001), e che poi descriva le informazioni necessarie al loro inquadramento, in coerenza con i macro-aspetti oggetto della prova e con le specificità legate al testo su cui il compito viene costruito, soprattutto in termini di distrattori plausibili. Questa parte di specificazione è fondamentale per esplicitare gli aspetti linguistici ed eventualmente inferenziali sostenuti da uno specifico testo, o da parti di esso, che concorrono a definire la difficoltà di un compito e, di conseguenza, di una domanda.

I tre macro-aspetti presentati nel QdR della prova di Italiano INVALSI non sono, quindi, solo una cornice teorica, ma anche un criterio di progettazione delle domande della prova standardizzata. La prova, infatti, deve rappresentare, attraverso le domande proposte, le diverse operazioni descritte nel QdR, cercando di evitare che la misurazione della comprensione del testo si riduca a sole alcune componenti tra quelle misurabili. Nella logica dei test standardizzati, la copertura delle diverse componenti dichiarate come oggetto di misurazione costituisce una condizione strutturale del test legata alla sua validità di contenuto, poiché la misura finale (punteggio o livello) dovrebbe rappresentare l'insieme delle operazioni che definiscono il costrutto e non una sola parte di esso.

Sicuramente il passaggio a prove basate su banche di domande ha permesso di raffinare questi aspetti, sia di contenuto sia metodologici (Desimoni et al. 2020).

3.2 Attendibilità e validità della prova: dalla teoria psicometrica alla discussione critica

Se il costrutto indica "cosa" si vuole misurare, l'attendibilità e la validità rispondono rispettivamente a "quanto bene" lo si misura e a "quale legittimità" i risultati hanno. Sebbene questa distinzione sia classica, continua a verificarsi confusione. Lucidi (2014) la spiega chiaramente, richiamando i principi della psicometria secondo cui la validità rappresenta "il grado in cui uno strumento misura ciò che si propone di misurare", mentre l'attendibilità rappresenta "la precisione con cui lo fa" (Lucidi, 2014, p. 2). Questa distinzione è importante perché chiarisce che un test può essere molto preciso nel produrre punteggi stabili, ma debole nel giustificarne il significato e le azioni che autorizza.

La struttura metodologica delle prove INVALSI si fonda sui modelli psicometrici basati sull'*Item Response Theory* (IRT) come ampiamente descritto nei diversi rapporti tecnici pubblicati nel corso degli anni. In una prima fase, il costrutto viene definito con accuratezza e operazionalizzato come "variabile latente continua", costituendo il riferimento teorico per l'intero processo di misurazione. In seguito, gli item vengono progettati per coprire l'intero continuum del tratto latente, garantendo una distribuzione adeguata dei livelli di difficoltà, condizione fondamentale per assicurare un'elevata informatività del test. Infine, la calibrazione degli item, attraverso procedure di stima di parametri specifici, consente di individuare e rimuovere quelli non adeguati, migliorando progressivamente sia la precisione della misura sia la coerenza interna del test (Chiorri, 2020).

Secondo questo approccio, l'affidabilità e la validità di un test non sono caratteristiche fisse, ma dipendono da com'è costruito, dal modello con cui viene analizzato e dal gruppo di individui a cui viene applicato, quindi cambiano in base alla situazione, ma in questo modo è possibile costruire strumenti più precisi e teoricamente più coerenti con l'intero costrutto latente (Chiorri, 2020).

L'attendibilità, in questa prospettiva, è data dalla precisione della stima del tratto latente, che varia in funzione del livello del costrutto posseduto dal soggetto, ed è direttamente collegata alle caratteristiche degli item, quali la loro difficoltà e la capacità discriminativa. Pertanto, un test può essere molto preciso nell'individuare differenze tra soggetti a determinati livelli di abilità, ma meno efficace in altri, mostrando che l'attendibilità è locale e non uniforme (Chiorri, 2020). Questo approccio consente dunque di progettare strumenti standardizzati, specifici per determinate popolazioni o con obiettivi diagnostici specifici, ottimizzando la qualità della misurazione.

Nei modelli IRT, l'unidimensionalità del test, cioè il fatto che le risposte agli item siano spiegate da un unico tratto latente, e l'indipendenza locale, cioè il fatto che, a parità di tratto latente, le risposte

ai singoli item siano indipendenti tra loro, sono assunti necessari per l'applicazione del modello (Barbaranelli & D'Olimpio, 2007). La loro verifica fornisce evidenze sulla struttura interna del test e contribuisce a escludere l'intervento di fattori estranei che renderebbero multidimensionale la misura; non costituisce però, di per sé, una garanzia di validità. Nella prospettiva richiamata più avanti in questo paragrafo, la validità riguarda le interpretazioni e gli usi dei punteggi e richiede di integrare evidenze di natura diversa, relative anche al contenuto, ai processi di risposta e alle conseguenze (Messick, 1989; Kane, 2006; AERA, APA, & NCME, 2014). Questo significa che ogni test richiede una fase di *pretest* per verificare che le proprietà indicate siano rispettate. Inoltre, l'analisi dei parametri dei singoli item consente di valutare la coerenza tra il funzionamento empirico del test e il modello teorico di riferimento, rafforzando ulteriormente la validità dello strumento.

Un'ulteriore procedura di analisi psicometrica, importante per verificare la robustezza di un test standardizzato, è l'analisi del funzionamento differenziale degli item (DIF), che consente di rilevare l'assenza di *bias* sistematici nelle risposte tra diversi gruppi di soggetti. L'individuazione di eventuali item che funzionano in modo diverso a parità di livello del tratto latente consente di migliorare l'equità dello strumento e di rafforzarne la validità complessiva (Barbaranelli & D'Olimpio, 2007).

Queste procedure sono previste nel processo di costruzione delle prove INVALSI come descritto nei diversi rapporti tecnici.

I documenti tecnici dell'INVALSI sulla transizione alle prove Computer Based Test (CBT) sono fondamentali perché evidenziano gli aspetti metodologici adottati. Nella parte dedicata a questi aspetti delle prove CBT, sono descritte le banche di domande delle prove INVALSI, composte da gruppi di domande i cui parametri derivano da un campione rappresentativo nazionale e grazie alla calibrazione, le prove somministrate in momenti e con modalità diversi possono essere confrontate (INVALSI, 2019). Lo stesso Quadro di Riferimento di Italiano specifica che gli studenti affronteranno prove diverse, ma formulate per avere "la medesima composizione e un livello di difficoltà equivalente" (INVALSI, 2018, p. 26). Questa affermazione è importante perché affronta una delle questioni storiche della docimologia relativa alla comparabilità delle valutazioni quando cambiano le prove o i valutatori. Tuttavia, il punto è che la robustezza tecnica, sebbene possa apparire come una soluzione, non risolve il problema, ma lo sposta solo altrove.

In questo contesto si colloca una parte rilevante della riflessione critica italiana relativamente alla valutazione standardizzata. In un contributo dedicato alla validità di contenuto delle prove di comprensione della lettura, Corsini (2014) evidenzia come la duplicità delle finalità attribuite alle rilevazioni INVALSI, orientate contemporaneamente alla valutazione diagnostica degli apprendimenti degli studenti e alla funzione di *accountability* del sistema, generi una tensione strutturale che può incidere sulla qualità degli strumenti. In particolare, tale duplicità può influenzare

la costruzione delle prove in merito alla selezione dei contenuti e orientarne le caratteristiche in modo implicito, riducendo la capacità degli esiti di fornire informazioni realmente utili in un'ottica educativa e formativa (Corsini, 2014). Un'analisi più sistematica di queste criticità è sviluppata nel contributo di Corsini, Scierri e Scionti (2017) dove l'attenzione è posta sulla validità delle prove INVALSI di comprensione della lettura nel periodo 2010–2016, evidenziando alcune criticità relative alla validità di contenuto in particolare e alla rappresentatività del costrutto. Secondo gli autori, quando una stessa prova è chiamata a svolgere contemporaneamente le funzioni di comparazione e di diagnosi, si riduce la possibilità di fornire informazioni analitiche sui processi di comprensione, con conseguente semplificazione del costrutto.

Nel contributo di Lucidi (2013), la validità è considerata una questione che riguarda l'intero processo di ricerca, dalla definizione del costrutto alla produzione di inferenze sugli esiti ottenuti. In questa prospettiva, la validità è rappresentata dalla verifica di più dimensioni che riguardano la coerenza teorica, la qualità delle operazionalizzazioni e le condizioni di uso delle prove. Lucidi (2013) sottolinea anche la necessità di sviluppare una cultura condivisa della valutazione, resa necessaria dall'aumentata diffusione delle prove standardizzate, con l'obiettivo di garantire un uso consapevole dei risultati e di evitare quindi interpretazioni semplificate o improprie. Queste riflessioni trovano un punto di convergenza nella tradizione internazionale sulla validità, che la intende come un processo argomentativo che mette in relazione teoria, evidenze empiriche e l'uso dei punteggi. In questa prospettiva, la validità non riguarda il test in sé, ma le inferenze che a partire da esso vengono costruite, le quali devono essere giustificate attraverso un insieme coerente di evidenze (Messick, 1989). Messick sottolinea inoltre il carattere unitario della validità, che integra sia la rappresentazione del costrutto sia le conseguenze dell'uso dei risultati. In continuità con questa impostazione, il modello argomentativo proposto da Kane interpreta la validazione come un processo strutturato volto a sostenere la plausibilità delle interpretazioni dei punteggi attraverso una catena esplicita di inferenze da giustificare e da controllare empiricamente (Kane, 2006).

Alcune di queste criticità, in particolare quelle legate alla comparabilità e alla precisione della misura, sono affrontate dall'introduzione delle prove *computer-based*; la tensione tra funzione diagnostica e funzione di comparazione segnalata da Corsini (2014) riguarda invece gli usi dei risultati più che le proprietà psicometriche dello strumento, e non è risolta dal solo cambiamento di formato. Per la secondaria di primo grado, ad esempio, il documento tecnico chiarisce che una singola prova, in modalità *computer-based*, è composta da item selezionati da un repertorio più ampio e costruiti per garantire comparabilità tra forme diverse (INVALSI, 2018, p. 26). La costruzione di una banca di domande per la valutazione degli apprendimenti nella scuola secondaria non viene presentata come un semplice ampliamento numerico di quesiti, ma come una scelta di disegno metodologico utile a

tenere in considerazione tre esigenze: 1) la sicurezza del test in una finestra di somministrazione più lunga e flessibile; 2) la comparabilità dei risultati tra studenti che svolgono prove diverse; 3) la precisione della misura lungo il continuum dell'abilità (in particolare nei punti che poi diventeranno soglie per i livelli). Proprio perché il CBT aumenta i tempi di somministrazione, data l'impossibilità di garantire, in un'unica giornata, per tutti gli studenti la dotazione informatica necessaria allo svolgimento della prova, la soluzione è stata individuata in un disegno metodologico basato su banche di domande calibrate e quindi su forme multiple di test da somministrare contemporaneamente. Nel rapporto tecnico INVALSI (2018) viene specificato che le rilevazioni CBT si fondano sul modello psicometrico di Rasch (1980) per ciascuna disciplina e grado scolastico di scuola secondaria, che il processo di costruzione della banca di domande si fonda su due fasi di calibrazione (una preliminare in fase di *pretest* e una sulla rilevazione principale) e su un assemblaggio automatizzato delle forme denominato *Automated Test Assembly* (ATA) per costruire prove "*weakly parallel*" in senso IRT, mantenendo allo stesso tempo il controllo dei vincoli di contenuto, precisione attesa e dell'esposizione degli item (Desimoni et al., 2018). Realizzare una banca di domande significa non solo raggiungere un certo numero di quesiti, ma anche progettare una distribuzione equilibrata delle difficoltà e garantire la copertura di diversi aspetti del contenuto (costrutto). Per la prova di Italiano INVALSI, le banche sono composte da un numero elevato di domande (oltre 300), con una difficoltà distribuita su un intervallo ampio, da valori molto facili a più difficili lungo la scala Rasch, e con un adattamento degli item al modello entro un range stabilito secondo le indicazioni della letteratura, a garanzia della robustezza dello strumento. La realizzazione di un modello per la costruzione di prove standardizzate basate su banche di domande ha permesso di raggiungere un importante obiettivo, ovvero la definizione di scale descrittive della progressione della difficoltà empirica del costrutto misurato attraverso la prova INVALSI. I livelli descrittivi rendono trasparente il significato dei valori numerici desunti dalle prove e ne supportano l'interpretazione, garantendo una maggiore comunicabilità degli esiti, in particolare verso la scuola. Nel lavoro di Desimoni (2018) emerge con chiarezza l'idea derivata dal modello di Rasch, per cui la variabile del tratto latente (costrutto) può essere rappresentata come una posizione lungo un continuum e il singolo punteggio colloca lo studente su quella linea, ma da solo fornisce poche informazioni su cosa sia stato in grado di fare chi ha ottenuto un certo valore. Per questo motivo, si è andati verso un obiettivo che era già stato sviluppato nelle indagini nazionali e internazionali (NAEP, PISA, TIMSS, PIRLS), ovvero quello di suddividere il continuum dei punteggi del tratto latente in livelli descrittivi (*descriptive proficiency scales*), in cui sono stabilite delle soglie che rappresentano i punti della scala che delimitano il passaggio da un livello al successivo e che sono supportate da descrizioni articolate di tutti i compiti relativi al costrutto che sono rappresentati all'interno di quei livelli.

La chiarezza di alcuni aspetti critici permette di stabilire quali inferenze siano ammissibili. Un livello ottenuto alle prove standardizzate dell'INVALSI, anche se tecnicamente rigoroso, non rappresenta una fotografia completa dell'abilità di lettura e comprensione del testo di uno studente né una misura esaustiva della qualità di una scuola. È un indicatore, costruito secondo criteri tecnici e rigorosi dal punto di vista scientifico, che va integrato con altre valutazioni nei percorsi individuali di apprendimento degli studenti.

3.3 Dal punteggio numerico ai livelli descrittivi

La decisione di affiancare ai punteggi numerici delle descrizioni del contenuto che rappresenta ciascun livello raggiunto rappresenta una svolta culturale più che tecnica. Mentre il punteggio offre un confronto immediato, ha un significato rispetto agli apprendimenti misurati scarso se non viene tradotto in contenuti che lo studente può effettivamente dimostrare rispondendo alle domande della prova INVALSI. I livelli descrittivi hanno proprio questo obiettivo, cioè contestualizzare la misura ottenuta attraverso la prova standardizzata all'interno di contenuti oggetto di apprendimento. Non è quindi una semplificazione della valutazione standardizzata, ma un passo verso una maggiore comprensibilità all'interno di un certo grado di complessità. La transizione dal punteggio numerico ai livelli descrittivi ha voluto essere una risposta a un'esigenza forte proveniente dal contesto educativo ovvero rendere trasparente un esito valutativo. La scuola ha bisogno di un linguaggio che possa trasformare le misure rilasciate dalle prove INVALSI in informazioni utili, in grado di guidare interpretazioni didattiche e informazioni più trasparenti per famiglie e studenti. Le scale di livello svolgono questa funzione interpretativa e comunicativa, a condizione che non diventino semplici etichette, ma restino strettamente collegate al costrutto dichiarato nel QdR e alle operazioni che lo definiscono. Nel lavoro di Desimoni (2018) viene descritta con chiarezza l'impostazione adottata da INVALSI nella costruzione delle scale: "Il primo passo nella costruzione delle scale descrittive INVALSI è stato quello di approfondire quale dovesse essere la tipologia di livelli INVALSI" (Desimoni, 2018, p. 6). Questa affermazione chiarisce sin da subito che sono state effettuate scelte metodologiche da parte dell'INVALSI, anche in relazione alla definizione di scale descrittive di livello, perché il livello non nasce come una semplice "soglia" numerica. Infatti, è stato determinante stabilire a priori il modello di livelli ancorati a standard, oppure quello di livelli intesi come descrittori della progressione di un tratto latente lungo un continuum (*descriptive proficiency scales*). Desimoni (2018) colloca esplicitamente la scelta INVALSI nel panorama internazionale e chiarisce che la descrizione di ciascun livello si basa sugli aspetti e sugli ambiti descritti nel QdR, e che le descrizioni sono costruite per restare coerenti con il costrutto dichiarato (INVALSI, 2019a; INVALSI, 2019b). La lettura degli esiti in questo senso significa che un livello non è un'etichetta rispetto ad uno standard (come nel caso invece della prova INVALSI di Inglese) ma una sintesi di prestazioni tipiche su base

empirica rispetto a compiti definiti nel QdR INVALSI. Le descrizioni analitiche dei livelli rilasciati al grado 8 (ultimo anno della scuola secondaria di primo grado), rappresentano esattamente questo modello di restituzione. Nei livelli più bassi, la comprensione del testo, dimostrata empiricamente dalle prove, appare legata soprattutto alla gestione di informazioni esplicite e a inferenze semplici, ossia a operazioni prevalentemente tipiche del primo macro-aspetto, con un accesso ancora parziale al secondo. Proseguendo verso i livelli intermedi, la comprensione del testo è maggiormente rappresentata da compiti in cui gli studenti dimostrano capacità di integrare parti del testo, distinguere informazioni pertinenti da quelle concorrenti, riconoscere elementi di struttura dei testi e saper usare indizi linguistici per ricostruire significati lasciati impliciti nel testo. In questi livelli intermedi, oltre ai compiti relativi al primo macro-aspetto, ma caratterizzati da una difficoltà superiore a quella riscontrata nei livelli precedenti, si individuano molti compiti afferenti al secondo macro-aspetto e si riconoscono anche diversi contenuti specifici del terzo macro-aspetto, come la capacità di cogliere intenzioni comunicative e la comprensione di scelte espressive. Ai livelli più alti, le descrizioni richiamano esplicitamente capacità di riorganizzazione autonoma, di sintesi dei contenuti letti e di riconoscimento del tono e dello stile di un testo. Anche nelle descrizioni analitiche di grado 13 (ultimo anno della scuola secondaria di secondo grado), la progressione descrittiva include la caratterizzazione dei testi su cui è costruita la prova che sono via via più densi e non lineari e, soprattutto, la caratterizzazione di compiti che richiedono la ricostruzione di logiche argomentative, la parafrasi non letterale, l'individuazione degli impliciti su testi più complessi e il riconoscimento delle strategie comunicative. Anche in queste descrizioni il passaggio da un livello all'altro non è solo quantitativo, ma attraversa processi via via più complessi che vanno dall'individuare al costruire e poi al valutare, anche se, in questo grado scolastico, ogni macro-aspetto è sufficientemente rappresentato all'interno dei diversi livelli. Questo collegamento tra livelli e macro-aspetti diventa ancora più significativo se lo si affianca alle Indicazioni nazionali (2012; 2025) e alle linee guida (2012) di riferimento per l'Italiano, che attribuiscono alla lettura un ruolo che parte dal saper individuare informazioni fino a giungere a una capacità di lettura interpretativa, alla comprensione critica, alla capacità di riconoscere scopi, tipi e generi testuali, all'uso consapevole del testo per apprendere e per partecipare alla vita culturale. In questa prospettiva, l'ancoraggio dei livelli ai macro-aspetti consente di leggere la scala INVALSI non come un dispositivo esterno al curricolo, bensì come un tentativo di descrivere, con un linguaggio misurabile, una progressione che la normativa indica nei propri documenti ufficiali. Le Indicazioni nazionali includono più dimensioni, come il piacere della lettura, la costruzione dell'identità culturale e la discussione interpretativa in comunità, che una prova standardizzata non può misurare direttamente.

Osservando le *proficiency scales* internazionali, è possibile individuare analogie e differenze determinate dagli obiettivi differenti di ciascuna di queste *large-scale assessments*. PISA, ad esempio, definisce livelli di competenza in lettura (e in altri domini) per descrivere ciò che gli studenti quindicenni sanno fare nelle loro prove, con descrittori associati a soglie di punteggio, analogamente a quanto realizzato anche dall'INVALSI. Le descrizioni pubblicate indicano le abilità associate a ciascun livello e descrivono la progressione del costrutto indicato nel loro *Framework*; tali descrizioni permettono di comprendere in maniera più esplicita gli esiti ottenuti per la comparazione internazionale e le analisi di sistema (OECD, 2023b). I livelli di PIRLS utilizzano benchmark internazionali (*Low, Intermediate, High, Advanced*) e forniscono descrizioni delle strategie e delle abilità osservate, ancorate a testi e compiti del quarto anno di scolarità, con un'enfasi molto chiara sul legame tra il tipo di testo, i processi di comprensione e le prestazioni (Mullis et al., 2023; PIRLS 2021). Il NAEP americano, invece, adotta *achievement levels* (*Basic, Proficient, Advanced*) che sono standard di prestazione definiti da un processo di *standard setting* e accompagnati da descrizioni di ciò che gli studenti dovrebbero dimostrare; gli stessi enti statunitensi sottolineano che questi livelli vanno interpretati con cautela, proprio perché "*proficient*" non coincide con un generico "adeguato", ma indica un livello elevato di padronanza (NCES, 2022; NAGB, 2022). In tutti questi casi, la funzione è la stessa ovvero garantire la comunicabilità del dato rilevato. Le scale descrittive INVALSI sono state costruite per comunicare in maniera più chiara e utile gli esiti alle scuole, al singolo studente e alle famiglie per permettere di comprendere meglio il significato della valutazione esterna standardizzata all'interno di un sistema che mantiene un collegamento con le indicazioni e i traguardi di apprendimento nazionali. Le indagini internazionali PISA e PIRLS, pur attenti ai *framework* disciplinari soprattutto nel caso del PIRLS, sono progettate per le comparazioni internazionali e, quindi, privilegiano descrittori che restano in ogni caso "trasversali" ai curricoli nazionali. Il NAEP, invece, proponendo un modello di restituzione dei livelli basato su *standard* rientra in un contesto di maggior dibattito pubblico nazionale collegato alla definizione di prestazione attese, più o meno facili, in riferimento a degli standard indicati a priori.

La specificità di INVALSI nella costruzione dei livelli descrittivi risiede nell'ancoraggio al quadro nazionale collegato ai diversi gradi scolastici, che rende le scale potenzialmente utili alle scuole, purché si eviti di interpretarli come giudizi definitivi e si considerino entro i confini delineati da quanto dichiarato nel QdR. Ogni scala descrittiva, infatti, rischia di trasformarsi in una struttura rigida se viene usata come dispositivo di classificazione anziché come strumento per comprendere ciò che ciascuno studente è stato in grado di fare e ciò che ancora potrebbe fare. In questo senso torna nuovamente in evidenza quanto la letteratura sulla validità dei test sottolinea in merito alla validità delle prove standardizzate, ovvero l'interpretazione e l'uso responsabile dei risultati, che possono

portare a semplificazioni riduttive in termini di insegnamenti focalizzati sui contenuti dei test e quindi alla riduzione del curricolo (AERA, APA, & NCME, 2014; Messick, 1989; Corsini, 2013).
La scelta di procedere con una restituzione degli esiti attraverso livelli descrittivi può essere letta anche come un tentativo di ricostruire il rapporto tra misurazione e valutazione, come discusso nei capitoli precedenti. Un punteggio, da solo, rischia di essere confuso con un giudizio, mentre un'indicazione di livello descrittivo rende l'esito maggiormente comprensibile e discutibile.
È in questa zona intermedia, in cui la misura non si configura come una valutazione definitiva e la valutazione si fonda ancora su evidenze empiriche, che la prova INVALSI di Italiano si colloca in modo più convincente.

3.4 I metodi di lavoro del gruppo di italiano per la costruzione delle prove nazionali

Come presentato nel *working paper* di De Santis, Mastrogiovanni, Resio e Vendramin (2024), disponibile sul sito dell'INVALSI nella sezione Pubblicazioni, si è svolta la sperimentazione che ha riguardato la costruzione di prove *computer-based* per la scuola primaria. Nell'avviare la sperimentazione, che ha comportato la selezione dei testi per costruire le prove per la classe quinta primaria, che potranno costituire la banca di item, come accade per gli altri gradi scolastici, è stato messo a punto un programma di lavoro articolato in diverse fasi che, per mantenere fedeltà alla fonte, si riportano di seguito come presentate nel working paper.
«Il Gruppo di Lavoro (GdL) dell'INVALSI per la costruzione delle prove di italiano in formato *computer based* (CBT - *Computer Based Test*) per la classe quinta di scuola primaria è composto da ricercatrici e C-Ter (Collaboratori Tecnici Enti di Ricerca) interni all'ente e da assegnisti e insegnanti di scuola primaria che, attraverso i bandi pubblici a cui hanno partecipato, sono entrati a far parte del GdL. Inoltre, il GdL si avvale della consulenza di esperti esterni che prendono parte non solo alla riflessione e alla realizzazione delle prove, ma che insieme al gruppo interno contribuiscono all'attuazione di momenti formativi per l'intero GdL. Riflettere sul formato CBT delle prove di italiano per la classe quinta della scuola primaria chiama sia a mettere in primo piano ciò che attiene alla comprensione del testo, come espresso nel paragrafo sopra dedicato, sia a confrontarsi con l'occasione e le sfide offerte dal digitale, accogliendo cambiamenti sostenibili (Asquini, 2018). Pensare le prove di italiano per la quinta primaria CBT ha portato a un programma di formazione per il GdL sviluppato in tre incontri online teorico-metodologici, per giungere a un incontro operativo (composto da tre giornate) in presenza, per la costruzione delle prove. Negli incontri teorici si sono ripresi i costrutti che compongono la comprensione del testo e la riflessione sulla lingua perché, come precisato anche nel framework di PIRLS 2026 (von Davier & Kennedy, 2024), la prova resta una prova di comprensione e dunque non una prova che misuri le abilità digitali degli studenti e delle

studentesse; tanto meno, tale formato, dovrà costituire un possibile "ostacolo" durante la compilazione in formato digitale. Il primo incontro formativo si è tenuto a marzo 2024, condotto da uno degli esperti esterni per l'intero GdL, e centrato sulla comprensione del testo. Si è riflettuto sulla complessità del testo per gli aspetti di leggibilità e comprensione (Vena, 2022) in relazione al QdR delle prove di italiano; si è tornati a ragionare sulle modalità di lettura (Ferraboschi & Meini, 1993) che un lettore può attuare di fronte a un testo. Ci si è confrontati con le definizioni internazionali di *reading literacy* (OECD, 2019), per giungere poi agli elementi che bisogna considerare quando ci si trova ad avere a che fare con i testi e cioè coesione, coerenza, intenzioni comunicative, intertestualità e le attese del lettore rispetto al testo (Pratesi, 2000). Le tipologie testuali (Lavinio, 2000) sono state discusse anche in relazione alle testualità derivanti dal Web interattivo (Montani, 2020). Nelle considerazioni sul vocabolario di base (De Mauro, 1980, 2016) cioè 7000 parole che hanno maggiore frequenza nella nostra lingua, parlando di frequenza, si sono considerati gli indici di leggibilità, in particolare quello di GULPEASE (Vena, 2022), come altri elementi di complessità su cui riflettere. Il secondo incontro, svolto ad aprile 2024 da un altro esperto esterno, ha approfondito i temi della comprensione del testo, concentrandosi particolarmente sui processi cognitivi. Nello specifico, a partire dalla concettualizzazione di *reading literacy* nel quadro dell'OCSE-PISA (OECD, 2019) e dell'IEA-PIRLS (Mullis & Martin, 2019), si sono analizzati i processi di individuare informazioni, comprendere un testo, valutare i testi e riflettere su di essi, come descritti nel paragrafo sopra dedicato alla comprensione. Ci si è soffermati sulla comprensione del testo come processo di costruzione di una rappresentazione mentale del testo, che è un'attività interattiva e costruttiva che richiede un'interazione tra lettore e testo (Lumbelli, 2009; Siniscalco et al., 2007). Il focus dell'incontro si è basato sul processo inferenziale che serve a creare connessioni per una rappresentazione coerente del testo (Soto et al., 2019). Il terzo incontro, a maggio 2024, è stato organizzato dal gruppo interno all'ente come una giornata di *workshop online*, con una prima parte a carattere seminariale e una seconda con gruppi di lavoro. La parte seminariale ha riguardato un affondo sulla tassonomia degli obiettivi educativi di Bloom nella revisione di Anderson e Krathwohl (2001), rilevando come la descrizione degli obiettivi educativi attraverso i processi cognitivi e i contenuti di conoscenza (Anderson & Krathwohl, 2001) costituisca un utile connubio per la descrizione degli item delle prove di italiano, operando una sorta di sintesi con i temi dei primi due incontri. La parte di lavoro in gruppi ha avuto un carattere prettamente operativo. Divisi in due gruppi si è lavorato alla costruzione e alla descrizione di item di una prova tipo, utilizzando non solo i riferimenti del QdR per i macro-aspetti, ma anche la descrizione dei processi cognitivi sottostanti in base alla tassonomia degli obiettivi educativi di Anderson e Krathwohl (2001). A conclusione dei lavori dei due gruppi, in plenaria, ci si è confrontati sulle descrizioni degli item costruiti e sui processi cognitivi descritti. Al termine del

workshop è stato inviato un breve questionario a risposte aperte ai partecipanti per avere un riscontro sia sull'uso della tassonomia come strumento per la descrizione degli item, sia sulle modalità operative di lavoro, per replicare la struttura in incontri futuri. Il riscontro derivante dalle risposte testimonia l'apprezzamento (Dewey, 1939/1960) sia delle modalità di lavoro, sia dell'uso della tassonomia. Durante il trimestre che ha visto l'avvicendarsi degli incontri formativi, parallelamente si è proceduto alla ricerca e alla selezione dei testi per costruire le prove CBT di quinta primaria. Dopo i tre incontri, in sottogruppi, si è proceduto a una prima stesura delle domande sui testi selezionati, lavorando online sia in modalità asincrona sia sincrona. La scelta dei testi e la prima stesura delle domande hanno coinvolto sia gli insegnanti sia il gruppo interno di italiano. I tre incontri sopra descritti e i processi a essi collegati hanno costituito dei momenti formativi propedeutici all'incontro in presenza, avvenuto a luglio 2024 nella sede dell'ente, composto da tre giornate, durante le quali ci si è concentrati sulla costruzione e sulla revisione degli item dei testi fino a quel momento lavorati in sottogruppi. La struttura degli incontri sopra descritti mira a mostrare, in questo scritto, le modalità di lavoro che la "comunità di pratica" (Lave & Wenger, 2012) ovvero il GdL di italiano ha attuato, in particolare per la costruzione delle prove CBT di quinta primaria. In un altro scritto (De Santis & Mastrogiovanni, 2024), si sono approfonditi gli aspetti che riguardano il lavoro del GdL di italiano come comunità di pratiche, non solo in funzione della costruzione delle prove, ma anche per ciò che concerne la condivisione degli obiettivi, come la *mission* e la *vision* dell'ente, le dinamiche relazionali e valoriali." (De Santis, Mastrogiovanni, Resio & Vendramin, 2024, pp. 13-15).

La stessa struttura organizzativa si replica, tendenzialmente, anche nei gruppi di lavoro dei gradi scolastici della scuola secondaria. A volte sono necessari approfondimenti su aspetti sia teorici sia metodologici, perché i nuovi insegnanti esterni che entrano a far parte del GdL spesso non hanno avuto modo di fare esperienza pratica di tutti i processi e degli strumenti utilizzati nel GdL per la costruzione delle prove di comprensione del testo. Ad esempio, alcuni approfondimenti riguardano l'uso della tassonomia degli obiettivi di Anderson e Krathwohl (2001) in relazione ai macro-aspetti e ai compiti cognitivi-linguistici del QdR per costruire determinati item delle prove. Inoltre, si sono messi a punto degli strumenti che supportano gli insegnanti nella selezione dei testi per le prove per stabilirne l'adeguatezza rispetto al grado scolastico e rispetto agli elementi cognitivi e linguistici considerati nella scheda di valutazione dei testi presentata nel *working paper*, a cui si rimanda. Oltre alla scheda, è stata messa a punto una griglia di valutazione dei testi da utilizzare nel momento in cui l'intero GdL valuta i testi che entreranno a far parte delle prove. Nel tempo, tali strumenti sono stati affinati e messi alla prova con i GdL dei diversi gradi scolastici.

CAPITOLO 4: LA CORNICE METODOLOGICA DELLA RICERCA

Il presente capitolo delinea la cornice metodologica della presente ricerca, esplicitando le scelte teoriche e operative che ne hanno orientato l'impianto complessivo. In primo luogo, vengono definite le domande e gli obiettivi della ricerca, collocandoli all'interno di un paradigma interpretativo di matrice costruttivista e all'interno di una visione sistemica dell'azione didattica e valutativa. Successivamente, si illustra la scelta della metodologia dello studio di caso e si descrivono il disegno della ricerca, nonché i criteri di selezione dei partecipanti.

In seguito, il capitolo descrive le strategie di raccolta dei dati, che includono la revisione sistematica della letteratura, l'analisi documentale e l'intervista semi-strutturata, e approfondisce le procedure di analisi, con un focus particolare sull'Analisi Tematica Riflessiva e sugli strumenti per la visualizzazione dei risultati. Viene infine presa in esame anche la dimensione etica della ricerca, elemento imprescindibile per garantire rigore, trasparenza e responsabilità nel processo di ricerca.

4.1 Domande e obiettivi di ricerca

Il presente capitolo illustra gli obiettivi e le domande di ricerca che orientano lo studio, nonché le scelte metodologiche adottate per la raccolta e l'interpretazione di dati significativi rispetto alle questioni indagate. In coerenza con le finalità della ricerca, è stato adottato un approccio qualitativo che consente di analizzare il fenomeno oggetto di studio nel suo contesto reale, valorizzandone la complessità e le molteplici dimensioni. L'attenzione è rivolta a casi circoscritti e chiaramente definiti, al fine di comprenderne in profondità le caratteristiche e i processi sottostanti. Questo approccio si inserisce in una visione olistica dell'educazione e attribuisce alla valutazione una funzione formativa, considerando l'apprendimento non solo come un risultato, ma anche come parte integrante del processo valutativo stesso (*assessment for learning*). Con questa ricerca si intende comprendere se, ed eventualmente in che modo gli esiti descrittivi della valutazione esterna standardizzata siano entrati nell'esperienza educativa scolastica. L'introduzione delle restituzioni individuali degli esiti, tramite livelli descrittivi nelle prove standardizzate nazionali, ha rappresentato un tentativo strategico di creare un collegamento tra la valutazione interna ed esterna, trasformando il dato valutativo sommativo della valutazione standardizzata in un momento di comprensione e apprendimento, soprattutto per lo studente, ma non solo. Sviluppare azioni scolastiche che attribuiscano significato e valore all'azione valutativa standardizzata significa enfatizzare il ruolo formativo di ogni intervento valutativo, sia interno sia esterno al percorso scolastico. L'esito descrittivo può diventare un'occasione per promuovere pratiche riflessive e azioni educative mirate, purché sia interpretato e utilizzato in modo consapevole e integrato all'interno dei contesti educativi. Le prove standardizzate

utilizzate per la valutazione esterna degli apprendimenti sono costruite tenendo conto delle indicazioni contenute nei documenti normativi nazionali. Questo collegamento con la normativa nazionale rappresenta un primo livello di integrazione tra il sistema di valutazione esterno e quello interno della scuola.

Partendo da questa premessa, l'obiettivo principale di questo studio è quello di comprendere la percezione di studenti e docenti in merito alla restituzione degli esiti delle prove standardizzate nazionali sulla comprensione del testo. Allo stesso tempo, sempre in un'ottica esplorativa, si è voluto comprendere in che modo tali esiti descrittivi potessero avere ricadute sull'azione educativa nel contesto scolastico.

Per raggiungere tali obiettivi, sono state formulate due domande di ricerca:

1. *Quali sono le percezioni degli studenti e dei docenti riguardo agli esiti descrittivi individuali delle prove nazionali standardizzate sulla comprensione del testo*?

2. *In che modo, a partire dall'esperienza dei soggetti coinvolti, studenti e docenti, gli esiti descrittivi delle prove standardizzate nazionali relativi ai livelli raggiunti nella prova di comprensione del testo possono incidere sui processi di insegnamento e apprendimento?*

La prima domanda di ricerca si concentra sulla raccolta delle percezioni degli studenti e dei docenti e, in particolare, per i docenti, sulla loro idea di valore della restituzione individuale agli studenti di esiti descrittivi rispetto alla prova esterna standardizzata nazionale; e per gli studenti, sulla loro capacità di comprendere, interpretare e utilizzare l'esito individuale che riceveranno, considerata fondamentale per trasformare i dati valutativi in strumenti di crescita. L'esito di una prova standardizzata rappresenta un momento educativo nel contesto scolastico, poiché è parte integrante dei processi che vi si svolgono; dare senso alle azioni e alle procedure che avvengono all'interno del contesto scolastico significa capire che ogni cosa che avviene a scuola comunica qualcosa, anche se non in modo esplicito. Anche attraverso queste azioni di costruzione del significato relative alle attività svolte a scuola si contribuisce al processo di apprendimento (Luhmann, 1995).

Infatti, l'efficacia di un esito valutativo sul processo di apprendimento dipende non solo dalla qualità del feedback, ma anche dalla capacità dello studente di comprenderlo e di impiegarlo in modo consapevole e strategico. Raccogliere informazioni sulla percezione degli studenti in merito a tale fenomeno è quindi cruciale per capire quanto il feedback sia percepito come parte integrante dell'ecosistema educativo oppure come un semplice giudizio o, nel peggiore dei casi, considerato come privo di significato.

La seconda domanda ha l'obiettivo di comprendere quale potrebbe essere l'effetto trasformativo della restituzione descrittiva della prova standardizzata nazionale sulla pratica didattica, considerando tutti

gli aspetti collegati alla valutazione dello studente nell'ambito dell'*assessment for learning*. La ricerca internazionale, infatti, sottolinea che gli esiti prodotti dalle valutazioni standardizzate assumono un peso significativo quando i dati vengono integrati con le valutazioni formative e sommative degli insegnanti (Looney, 2011). È in questo spazio che i risultati descrittivi individuali possono essere visti come strumenti utili per creare un collegamento con l'azione didattica in aula. Il presente studio si propone di comprendere se e in che modo docenti e studenti percepiscono gli esiti descrittivi individuali rilasciati da INVALSI e come tali esiti possano essere integrati nei processi di insegnamento e apprendimento. In particolare, l'attenzione è rivolta alla possibilità che la restituzione descrittiva non rimanga un'informazione certificativa isolata, ma diventi una risorsa per progettare azioni didattiche mirate e per promuovere competenze di *feedback literacy*, intese come capacità di comprendere, interpretare e utilizzare il feedback in funzione del miglioramento dell'apprendimento (Bevilacqua, 2023; Carless & Boud, 2018).

4.2 Il paradigma della ricerca: la visione costruttivista e la visione sistemica dell'azione didattica e valutativa

Ogni progetto di ricerca è guidato da assunti, espliciti o impliciti, sulla natura del mondo sociale e sulle modalità con cui esso può essere indagato. Come sostengono Burrell e Morgan (1979), ogni ricercatore si avvicina al proprio oggetto di studio attraverso una lente di presupposti che ne modellano l'indagine. Questa lente è il paradigma di ricerca, definito da Guba (1990) come un "insieme di credenze che guidano l'azione" e da Hughes (2010) come una "maniera di vedere il mondo" che inquadra il tema di studio. La scelta di un paradigma non è una questione secondaria, ma la base su cui si costruisce l'intera architettura di uno studio, poiché le assunzioni implicite del ricercatore sulla natura del mondo sociale determinano il modo in cui esso può essere indagato. Un paradigma di ricerca si fonda su tre concetti filosofici chiave che operano in una relazione gerarchica: *Ontologia, Epistemologia e Metodologia*. La loro interdipendenza è stata ben descritta da Hitchcock e Hughes (1995), che affermano che "le assunzioni ontologiche daranno origine ad assunzioni epistemologiche che hanno implicazioni metodologiche". In termini più semplici, la nostra visione della realtà (ontologia: *Qual è la natura della realtà?*) determina come pensiamo di poterla conoscere (epistemologia: *Come conosciamo ciò che conosciamo? Qual è la relazione tra chi conosce e ciò che viene conosciuto?*) e, a sua volta, questo definisce come dovremmo procedere per acquisire tale conoscenza (metodologia: *Come dovremmo studiare il mondo per acquisire conoscenza?*). In questo capitolo viene presentato il paradigma che sottende la ricerca condotta durante la formazione di dottorato, attraverso la descrizione delle sue fondamenta filosofiche, in termini di ontologia ed epistemologia, e delle implicazioni metodologiche che ne derivano. La comprensione di questo

quadro epistemologico è pertanto essenziale per valutare la coerenza interna e il rigore del percorso di ricerca che verrà successivamente presentato. I fenomeni educativi si collocano nel "mondo della vita" (Dahlberg, Dahlberg, Nyström, 2008), e rappresentano un tessuto di esperienze, relazioni e significati che non può essere compreso appieno attraverso procedure "di tipo laboratoriale e sperimentale". Per questo motivo, la ricerca pedagogica, come quella qui presentata, si orienta verso un paradigma epistemologico naturalistico, anche detto costruttivista o interpretativo (Guba & Lincoln, 1989). Dal punto di vista di Mortari (2007), la ricerca pedagogica richiede un "paradigma di tipo ecologico" che enfatizzi l'interrelazione tra tutti gli elementi del contesto. La catena logica che collega le assunzioni filosofiche del paradigma costruttivista-naturalistico alle scelte metodologiche della ricerca può essere ricostruita a partire dal rapporto tra ontologia, epistemologia e metodologia. Come evidenzia Kamal (2019), il paradigma costruttivista non riguarda soltanto una preferenza per le tecniche qualitative, ma implica una precisa concezione della realtà, della conoscenza e del modo in cui essa può essere indagata. In modo analogo, Hitchcock e Hughes (1995) sottolineano che le assunzioni ontologiche ed epistemologiche orientano le scelte metodologiche e, di conseguenza, le tecniche di raccolta e interpretazione dei dati. Questa impostazione è coerente con la prospettiva costruttivista di Lincoln e Guba (2013), con il costruzionismo di Crotty (1998) e con l'attenzione, richiamata da Mortari (2007), alla specificità dei fenomeni educativi e alla necessità di una ricerca capace di restituirne la complessità.

Nel presente studio, tale orientamento si traduce in una concezione della realtà come molteplice, situata e socialmente costruita. La ricerca non si propone quindi di individuare una verità unica e indipendente dai soggetti, ma di comprendere i significati che i partecipanti attribuiscono alla propria esperienza. La conoscenza viene intesa come esito di un processo relazionale e interpretativo, nel quale ricercatrice e partecipanti contribuiscono alla costruzione del dato.

È in questa cornice che si colloca il mio posizionamento come ricercatrice. La conoscenza prodotta non è considerata esterna o neutrale rispetto al processo di ricerca, ma prende forma attraverso l'interazione con il contesto, il dialogo con i partecipanti e la riflessione continua sulle scelte compiute durante l'indagine. Il mio ruolo, pertanto, non è quello di osservatrice esterna, ma di soggetto coinvolto in un processo interpretativo, nel quale la presenza, le domande e le categorie di lettura della ricercatrice contribuiscono alla costruzione del sapere prodotto.

Come sottolinea Mortari

> …lo strumento di indagine non venga individuato nelle tecniche tradizionali di raccolta dei dati, ma nel ricercatore stesso, ossia lo strumento è l'essere umano che compie l'indagine, poiché è il solo capace di adattarsi alle situazioni indeterminate che si presenteranno nel corso della ricerca (Mortari, 2007, p. 73).

Questa centralità del ricercatore come strumento privilegiato di co-costruzione della conoscenza si manifesta in tre dimensioni: la presenza, la gestione della soggettività e la co-costruzione della comprensione.

La *presenza*, ovvero il tempo dedicato a comprendere un fenomeno "dall'interno", nel contesto in cui si sviluppa. Lincoln e Guba (1985) definiscono questo approccio "impegno prolungato" (*prolonged engagement*) nel senso di "abitare" il contesto di ricerca, sviluppando una familiarità con esso che consente di cogliere sfumature altrimenti invisibili; l'impegno non si riferisce essenzialmente alla permanenza nel tempo nel contesto di ricerca, ma alla *qualità dell'attenzione*. L'"abitare" un contesto richiede una sensibilità e una consapevolezza costanti, basate su solide capacità di pratiche riflessive che rappresentano il cuore della competenza del ricercatore. Nel caso della ricerca di dottorato qui proposta, l'aver "abitato" il contesto scolastico nei diversi anni di attività svolta presso INVALSI mi ha consentito di mantenere l'attenzione sui particolari del fenomeno di studio, in quanto io stessa posso essere considerata un'interlocutrice privilegiata del fenomeno osservato. Aver contribuito alla definizione metodologica della restituzione degli esiti delle prove standardizzate INVALSI ai singoli studenti mi permette sicuramente di aver chiari alcuni passaggi, ma allo stesso tempo mi mette in una condizione di allerta rispetto alla creazione di potenziali pregiudizi e anticipazioni sullo studio di tale fenomeno. Ma è proprio all'interno di queste dinamiche che la pratica riflessiva del ricercatore assume la sua maggiore forza, in particolare quando entra in contatto con le riflessioni dei partecipanti e dà inizio al processo di co-costruzione della realtà.

Se in un approccio positivista la soggettività del ricercatore è considerata un *bias* da eliminare, all'interno di un paradigma costruttivista essa diventa una risorsa da gestire con consapevolezza. Il ricercatore non può (e non deve) spogliarsi delle proprie idee e della propria storia, ma deve imparare a metterle al servizio della ricerca. Questo processo avviene attraverso tre passaggi fondamentali:

1. *Riconoscere l'interesse iniziale*: ogni ricerca nasce da una curiosità personale. Secondo van Manen (2016), la ricerca fenomenologica richiede di esplicitare l'interesse che orienta il ricercatore verso un'esperienza vissuta e di trasformarlo in una domanda capace di mantenere aperto il senso di meraviglia e interrogazione nei confronti del fenomeno. Riconoscere perché un tema ci sta a cuore aiuta a mantenere la ricerca focalizzata, riflessiva e onesta.
2. Esplorare *il "pre-understanding*" significa che ciascun individuo ha delle pre-comprensioni e delle cornici di riferimento implicite. Moustakas (1990) sottolinea l'importanza per un ricercatore di un dialogo interiore per portare alla luce le conoscenze tacite. Se non vengono riconosciute, il rischio è quello di orientare in modo non consapevole la costruzione del percorso di ricerca.

3. *Costruire il significato* significa che la conoscenza avviene nell'interazione. Come afferma Crotty (1998), il significato dei fenomeni oggetto di studio è generato socialmente. La relazione tra ricercatore e partecipanti è il luogo stesso in cui si genera la comprensione.

La soggettività è quindi la lente attraverso cui si osserva il mondo, mentre la riflessività è il processo rigoroso attraverso cui si mette a fuoco quella lente e si trasforma un potenziale *bias* nel fondamento stesso della comprensione. La riflessività è quindi fondamentale per chi sceglie di fare ricerca in ambito pedagogico, perché, come sostiene Mortari (2003, p. 27), è "quella postura mentale che mette a fuoco situazioni di incertezza". Ma come si manifesta, in pratica, questa postura riflessiva? Quali sono le qualità che definiscono il ricercatore come uno strumento ben "calibrato"? Un ricercatore che coltiva questa postura sviluppa una serie di qualità essenziali, descritte da Mortari (2007, pp. 74-75):

- *Responsività*: la capacità di ascoltare attivamente il contesto, entrando in "risonanza ricettiva" con quanto accade.
- *Adattabilità*: la flessibilità di adeguare il proprio percorso di ricerca a ciò che emerge dalla realtà studiata, senza rimanere rigidamente ancorati al piano iniziale.
- *Estensione epistemica*: la capacità di utilizzare diverse forme di conoscenza, non solo il pensiero logico, ma anche le emozioni e le intuizioni, tenendole sotto il "vigile controllo del pensiero riflessivo".
- *Immediatezza processuale*: l'abilità di generare nuove ipotesi "sul momento" per rispondere alla mutevolezza del contesto.
- *Empatia*: la capacità di entrare in risonanza emotiva e cognitiva con i partecipanti, comprendendo il loro punto di vista.

Nella vita di tutti i giorni viviamo le cose in modo diretto e immediato, e questo flusso di eventi è chiamato, da Mortari (2003), vissuto. L'esperienza, invece, "prende forma quando il vissuto diventa oggetto di riflessione e il soggetto se ne appropria consapevolmente per comprenderne il senso" (Mortari, 2003). La riflessività trasforma il vissuto in esperienza e l'esperienza in apprendimento. Uno strumento concreto per esercitare questa competenza è rappresentato dal diario di ricerca, uno strumento in cui annotare non solo dati, ma anche pensieri, dubbi, intuizioni e decisioni metodologiche, rendendo trasparente il proprio processo conoscitivo. Questa postura riflessiva è oggi più che mai strategica per "abitare" i contesti di ricerca in modo critico e consapevole, contrapponendosi alle logiche della contemporaneità che enfatizzano la velocità, la produttività e la prestazione istantanea (Gambacorti-Passerini & Palmieri, 2021; Boarelli, 2019). Come ricercatrice, ho compreso che la ricerca richiede tempo, lentezza e disponibilità alla riflessione. La riflessività mi ha aiutata a vivere questa lentezza come una risorsa, e non come un ostacolo. Nel corso dello studio

ho avuto più volte bisogno di fermarmi, ripensare alcune scelte, ridefinire i confini dell'indagine e rimettere in dialogo le domande di ricerca con i dati raccolti e con il contesto scolastico osservato. Questa postura flessibile e riflessiva mi ha permesso di accompagnare il processo di ricerca in modo più consapevole e di dare maggiore profondità all'interpretazione dei risultati.

La mia postura epistemica di ricercatrice, in generale, oltre a quanto già descritto, è caratterizzata da uno sguardo sistemico-relazionale, poiché considero il contesto educativo, e non solo, un sistema complesso, dinamico e interconnesso, in cui ogni componente influenza e viene influenzata dalle altre componenti (Bronfenbrenner, 1979; Fullan, 2006). In questa prospettiva, le azioni di formazione e valutazione sono viste come una rete di relazioni tra studenti, insegnanti, dirigenti scolastici, curricula, politiche educative e influenze socioculturali esterne, che si influenzano a vicenda. Come ricordato da Lamanauskas (2024), riprendendo Vaitkevičius (1988), l'educazione può essere interpretata come un sistema complesso, costituito da molteplici elementi interrelati e condizionato da un insieme ancora più ampio di fattori esterni. Questa prospettiva consente di leggere i processi educativi non come sequenze lineari di azioni, ma come configurazioni dinamiche in cui interagiscono soggetti, contesti, pratiche, vincoli istituzionali e condizioni culturali. Il pensiero di Edgar Morin (1993) si inserisce in questa prospettiva, definendo la "complessità" come "ciò che è tessuto insieme" (complexus). L'approccio sistemico, dunque, suggerisce di non disfare il tessuto per analizzare singolarmente ogni filo, uno a uno, ma di comprendere come questi fili siano intrecciati e si compongono per formare un'unità complessa e inscindibile. Il pensiero sistemico richiede la capacità di considerare simultaneamente molteplici dimensioni di un problema, cogliendone le interrelazioni e interrogando in modo riflessivo le assunzioni che orientano l'interpretazione della situazione (Senge, 2006; Richmond, 1994).

L'approccio sistemico, ispirato alle teorie di Gregory Bateson, permette al ricercatore di osservare il contesto educativo come un sistema interconnesso, in cui l'apprendimento e la valutazione non possono essere compresi isolatamente, ma vanno analizzati come parte di un sistema complesso di relazioni.

Questo approccio si integra con il pensiero complesso di Morin (2001), che considera ogni sistema vivente e cognitivo caratterizzato da interdipendenza, ricorsività e auto-eco-organizzazione. Morin osserva che la conoscenza non progredisce in modo lineare, ma si forma attraverso interazioni e retroazioni che collegano il soggetto al suo ambiente. Nella prospettiva di Bateson, il feedback assume un ruolo centrale nei processi di apprendimento, perché consente al soggetto di cogliere informazioni significative e di riorganizzare il proprio modo di comprendere e di agire. Il feedback, infatti, introduce "differenze che fanno la differenza" (Bateson, 1977, p. 459), ossia informazioni che, se riconosciute come rilevanti, possono attivare processi di adattamento, cambiamento e

apprendimento. Ed è in questo contesto che la valutazione viene concepita non solo come misura delle performance, ma anche come un processo riflessivo e trasformativo, diventando così uno strumento che sviluppa la consapevolezza, stimola il dialogo e favorisce il cambiamento, contribuendo all'evoluzione complessiva del sistema educativo.

Anche negli studi di Black e Wiliam (1998) sull'*assessment for learning* si rafforza l'idea che la valutazione formativa sia uno dei principali motori del miglioramento educativo. Questo approccio alla valutazione nei percorsi scolastici permette agli studenti di valutare il proprio livello di competenza, di vedere gli errori come opportunità di crescita, di sviluppare abilità metacognitive e di guidare consapevolmente il proprio percorso di apprendimento. Ulteriori studi osservano che una valutazione efficace è un processo che coinvolge attivamente gli studenti nella comprensione dei feedback e nella responsabilizzazione per il proprio percorso di studio (Heritage, 2018). Anche Wiliam (2018) insiste sul concetto di "*responsive teaching*", ossia un modello di insegnamento che si adatta continuamente alle esigenze degli studenti in base ai risultati ottenuti nelle fasi di valutazione. In tal senso, l'attività valutativa si configura come un sistema autoregolativo che alimenta in tempo reale il processo educativo. La teoria socioculturale dell'apprendimento, riconducibile al pensiero di Vygotskij, sottolinea che lo sviluppo cognitivo non può essere compreso separatamente dalle relazioni sociali, dagli strumenti culturali e dai contesti in cui l'apprendimento prende forma. In questa prospettiva, le funzioni cognitive superiori si sviluppano inizialmente sul piano interpersonale, attraverso l'interazione con adulti, pari o figure più esperte, e vengono poi progressivamente interiorizzate dal soggetto (Vygotsky, 1978; Vygotskij, 1990).

Applicata alla valutazione, questa prospettiva consente di superare una concezione puramente individuale e misurativa dell'*assessment*. La valutazione può infatti essere interpretata come una pratica mediata e situata, nella quale il feedback non si limita a comunicare un esito, ma contribuisce a sostenere la comprensione, la riflessione e la regolazione dell'apprendimento. In tale direzione si collocano anche gli sviluppi della *dynamic assessment*, che, a partire dalla zona di sviluppo prossimale, considera valutazione e insegnamento come attività integrate, orientate non solo a rilevare ciò che lo studente sa fare autonomamente, ma anche a comprendere ciò che può realizzare con un supporto adeguato (Lantolf & Poehner, 2004; Poehner, 2008)

Questa prospettiva si distingue chiaramente dall'idea tradizionale di insegnamento, in cui il docente si limita a trasmettere contenuti a studenti passivi, attribuendo maggiore importanza al ruolo attivo dello studente, che viene coinvolto direttamente nella costruzione delle proprie conoscenze. L'apprendimento avviene attraverso esperienze significative, strettamente collegate al contesto e alla realtà, che permettono ai discenti di attribuire senso a ciò che apprendono e di rielaborarlo in modo personale e consapevole, come sottolineato da Lucisano e Corsini (2015).

Il docente, in questo contesto, agisce come “agente del cambiamento”, capace di gestire situazioni complesse e ambigue, trasformando la “palude” della pratica quotidiana in un terreno fertile per l’innovazione (Schön, 1983). Le sue decisioni non sono mai neutre, ma sono influenzate da paradigmi epistemologici che determinano il tipo di sapere proposto e le modalità di costruzione delle situazioni di apprendimento. L’insegnante non si limita a seguire teorie, ma condivide con gli studenti il ruolo di protagonista nella creazione di ambienti favorevoli all’apprendimento.

In questo scenario ogni azione promossa in classe ha ricadute sui processi di apprendimento dello studente. In questo senso, allora, anche lo svolgimento delle prove di valutazione standardizzata esterna costituisce un momento educativo a cui attribuire un significato all’interno dell’azione didattica in classe. Le prove INVALSI, spesso viste come strumenti distanti dalla quotidianità dell’azione didattica, possono svolgere un ruolo diverso se integrate in essa in modo consapevole e significativo. Alcuni studi di ricerca educativa hanno sottolineato come queste prove, se utilizzate in una prospettiva che va oltre la sola funzione di misurazione e di rendicontazione, possano diventare dispositivi informativi, capaci di fornire spunti per la costruzione di processi di apprendimento e di miglioramento sia per gli studenti sia per i docenti (Lucisano & Corsini, 2015; Freddano & Pastore, 2020). Le prove INVALSI, attraverso la restituzione individuale degli esiti, possono diventare strumenti di riflessione funzionali agli apprendimenti nel contesto scolastico, offrendo agli studenti l’opportunità di analizzare le strategie cognitive usate per rispondere alle domande, di riconoscere gli errori e, di conseguenza, di imparare da essi, in accordo con quanto sostenuto dalle teorie della pedagogia dell’errore (Freddano & Pastore, 2020). Per favorire questo processo, è fondamentale creare un senso condiviso delle prove tra docenti e studenti, affinché comprendano le finalità e le inseriscano in un contesto di significato che vada oltre il semplice insegnamento orientato ai test. Adottare un approccio sistemico nell’osservazione dei fenomeni che riguardano il contesto scolastico significa riconoscere che anche i risultati di una prova nazionale standardizzata, se riferiti al singolo studente, hanno un’influenza su di lui, poiché fanno parte di un'azione educativa più ampia. Come già sottolineato, in questo approccio epistemologico il feedback non si limita a correggere gli errori, ma modifica il sistema cognitivo dell’individuo, creando nuove strutture mentali. Una visione sistemica della valutazione propone di considerare tutte le forme di valutazione come parte di un unico ecosistema formativo, capace di sostenere lo sviluppo dell'apprendimento degli studenti e la riflessività professionale dei docenti.

4.3 La metodologia della ricerca: lo studio di caso

L'approccio qualitativo adottato in questa ricerca mira a esplorare nel dettaglio le percezioni, le esperienze e le rappresentazioni di docenti e studenti in merito alla valutazione esterna standardizzata, in particolare alle prove INVALSI di comprensione del testo. Questo approccio, in linea con l'obiettivo della ricerca, consente di cogliere la complessità dei vissuti educativi e di analizzare le implicazioni formative del feedback descrittivo. Come sottolinea Berg (2001), la ricerca qualitativa "permette ai ricercatori di condividere le comprensioni e le percezioni degli altri e di esplorare come le persone strutturano e danno significato alla loro vita quotidiana" (p. 7). Questo è particolarmente importante in un contesto educativo, in cui la valutazione non è soltanto una misurazione, ma anche un'esperienza di apprendimento. Questo approccio si caratterizza come un processo principalmente induttivo, che organizza i dati in categorie e identifica schemi (relazioni) tra di essi, lasciando che i dati e i significati emergano in modo naturale dal contesto della ricerca. La rigorosità nella ricerca qualitativa non manca, ma si manifesta attraverso un metodo sistematico, trasparente e riflessivo, volto a ottenere una comprensione profonda e verificabile dei fenomeni sociali, spesso generando nuove teorie radicate nei dati. È importante, quindi, rispettare criteri quali credibilità, trasferibilità, dipendenza e confermabilità, che si ottengono attraverso pratiche riflessive, trasparenza e coerenza metodologica (Ahmed, 2024; Noble & Smith, 2025; Cain, 2023; Tariq, 2025; Daniele, 2025; Lincoln & Guba, 1985).

Per la realizzazione di questo lavoro di ricerca esplorativo si è scelto di utilizzare lo studio di caso per analizzare in profondità il fenomeno nel suo contesto reale, focalizzandosi su alcuni casi ben definiti, rappresentati da studenti e docenti, e utilizzando diverse fonti di dati, quali interviste e documenti relativi al fenomeno oggetto di studio (Creswell & Poth, 2024). Questo metodo ha il vantaggio di comprendere le dinamiche complesse, così come si presentano nei contesti reali, spesso trascurate dagli approcci quantitativi o sperimentali. In generale, questo metodo considera il caso come un'unità dotata di un significato proprio, da studiare per ciò che è e per ciò che consente di comprendere grazie alla sua unicità.

Nello specifico di questa ricerca, ci si è orientati verso un approccio strumentale, come delineato nei lavori di Stake (1995). Per Stake, come descritto da Yazan (2015), il caso è "specifico, complesso e funzionante, è un sistema integrato che ha un confine e parti operative, ha uno scopo" (De Santis, 2022, p. 191). Quindi, per Stake, lo studio di caso intende cogliere al contempo la particolarità e la complessità di un singolo caso, per comprenderne gli aspetti caratterizzanti. Nella sua prospettiva epistemica, lo studio di caso può presentarsi come olistico, empirico, interpretativo ed empatico, differenziandosi in base al contesto di ricerca. Per questo, Stake intende il disegno di ricerca come flessibile, ossia un processo che consente ai ricercatori di riorganizzare le azioni progettuali durante

il percorso di ricerca. Tra gli strumenti utilizzabili nel contesto dello studio di caso per raccogliere i dati, Stake elenca l'osservazione, le interviste e la revisione di documenti, la cui analisi mira a costruire significato partendo dai processi di categorizzazione fino alla loro interpretazione e contestualizzazione. I dati qualitativi così raccolti vengono valutati attraverso strategie di triangolazione: fonti di dati; sperimentatore; teorica; metodologica. Secondo questo modello, l'obiettivo principale dello studio non è concentrarsi esclusivamente sul caso in sé, ma usarlo come strumento per esplorare questioni più ampie, sia di natura teorica sia pratica.

Nella presente ricerca, il caso consiste nella restituzione descrittiva individuale degli esiti della prova INVALSI di Italiano agli studenti dell'ultimo anno della scuola secondaria di primo grado (grado 8) e di secondo grado (grado 13), così come viene letta e interpretata da docenti di Italiano e da studenti di questi due gradi. I confini del caso sono definiti da tre elementi: l'oggetto (il documento con le descrizioni dei livelli e i materiali che lo accompagnano), gli attori coinvolti (15 docenti e 9 studenti, par. 4.5) e il periodo di raccolta dei dati (da maggio 2024 a gennaio 2025, par. 4.6.3). Oggetto di indagine sono le percezioni e le pratiche riferite da questi attori, non gli effetti della restituzione sugli apprendimenti. La scelta di questo studio di caso è stata guidata dall'interesse ad approfondire la comprensione di tale fenomeno nel contesto scolastico. Il disegno della ricerca, descritto successivamente (par. 4.4), è stato costruito in linea con le indicazioni di flessibilità di Stake, attraverso una focalizzazione progressiva, a partire dalla selezione dei partecipanti, che originariamente seguiva un criterio di individuazione volto al raggiungimento di una numerosità più consistente di quella definitiva, con l'obiettivo di costruire un sistema di significati a confronto tra docenti e soprattutto studenti con conoscenze già consolidate delle documentazioni di rilascio degli esiti INVALSI e docenti e studenti che non lo fossero, che ha portato alla riflessione del fenomeno su un piano più generale con la necessità di introdurre nello studio anche un'approfondita analisi documentale con l'obiettivo di verificare l'influenza dei contenuti presentati all'interno di documentazione normativa e istitutiva sulla percezione del fenomeno stesso .

Gli strumenti, quindi, utilizzati in questo studio di caso, descritti di seguito (par. 4.6), sono stati: interviste semi-strutturate, fonti documentali e una revisione sistematica della letteratura.

Come sostiene Stake (1995), lo studio di caso mira a cogliere la particolarità del fenomeno nel suo contesto naturale, valorizzando la pluralità delle prospettive attraverso cui esso può essere compreso. Attraverso la triangolazione delle fonti, il fenomeno è stato indagato con dati di natura diversa, evitando così che l'interpretazione si fondi su un'unica fonte informativa e permettendo, invece, di mettere in relazione livelli differenti di conoscenza. La triangolazione può essere intesa come una modalità di osservazione dello stesso oggetto di ricerca da prospettive differenti, con l'obiettivo di produrre una comprensione più articolata del fenomeno indagato (Caillaud & Flick, 2021). L'analisi

dei documenti normativi e istituzionali relativi all'INVALSI ha permesso di ricostruire il quadro entro cui si colloca la restituzione degli esiti, sia a livello di normativa generale sia di atti istitutivi e di programmazione delle attività che definiscono le modalità di svolgimento dei compiti istituzionali. La revisione della letteratura ha consentito di confrontare il caso con gli studi che hanno già indagato la percezione e l'uso dei dati valutativi standardizzati, anche in termini di esiti individuali, nei contesti educativi. Le interviste, infine, hanno permesso di accedere alle rappresentazioni e alle esperienze di chi vive quotidianamente la scuola ed è protagonista di questo processo di valutazione esterna. La comprensione del significato attribuito agli esiti individuali INVALSI richiede infatti di considerare il modo in cui tali risultati vengono comunicati, letti, discussi e tradotti nelle pratiche scolastiche. La descrizione accurata del contesto ha quindi reso possibile comprendere in che modo docenti e studenti interpretano gli esiti delle prove e quale valore attribuiscono a tali informazioni nell'esperienza scolastica. In questa direzione, il riferimento alla descrizione densa proposta da Geertz (1973) risulta particolarmente significativo, poiché richiama la necessità di restituire non solo ciò che accade, ma anche il sistema di significati entro cui le azioni degli attori acquistano senso. In questo senso, la triangolazione assume una funzione interpretativa, poiché consente di confrontare convergenze, questioni problematiche e differenze tra le fonti, rafforzando la solidità dell'analisi interpretativa (Denzin, 1978; Stake, 1995).

Anche nel processo di analisi dei dati le categorie interpretative non sono state assunte come schemi rigidi da applicare ai dati, ma sono state costruite e progressivamente affinate attraverso un confronto continuo con le evidenze raccolte. Il confronto costante tra dati raccolti da fonti diverse, la letteratura di riferimento e l'interpretazione consente di costruire un'analisi più aderente al materiale empirico e di evitare letture affrettate o eccessivamente distanti dalle evidenze raccolte. Merriam e Tisdell (2016) sottolineano che, nella ricerca qualitativa, la qualità dell'interpretazione dipende anche dalla capacità del ricercatore di tornare più volte sui dati, rivedere le proprie ipotesi e rendere esplicito il percorso attraverso cui attribuisce significato alle evidenze.

Infine, la validità è stata garantita dalla trasparenza del processo di ricerca. Sono state documentate le scelte relative alla selezione delle fonti, alla costruzione degli strumenti di raccolta dei dati, alla conduzione delle interviste e alle procedure di analisi. In primo luogo, tutte le interviste sono state registrate e trascritte integralmente, e le trascrizioni automatiche sono state verificate mediante il riascolto delle registrazioni (par. 5.3). Successivamente è stato costruito un database di codifica anonimizzato, in formato Excel, in cui ogni estratto è identificato dal codice del partecipante e da un numero progressivo ed è collegato all'etichetta di codifica, alla Categoria Tematica e al Tema Generale: questo consente di ricostruire a ritroso ogni passaggio interpretativo (*audit trail*). La

codifica delle interviste dei docenti ha prodotto 159 etichette, di cui 10 escluse per non pertinenti rispetto alle domande di ricerca; la codifica delle interviste degli studenti ha prodotto 51 etichette.
Inoltre, per tutta la durata della ricerca sono state tenute annotazioni di lavoro, in cui sono state registrate le osservazioni emerse durante la pianificazione dell'attività di ricerca, l'analisi della letteratura, le interviste e l'analisi e le principali decisioni metodologiche.
Le scelte di codifica e la definizione dei temi sono state discusse nel corso dell'analisi con la tutor e la co-tutor, mentre l'analisi è stata condotta dalla sola ricercatrice: il rigore è stato perseguito attraverso la riflessività e la trasparenza delle scelte, piuttosto che attraverso la convergenza tra più codificatori (Braun & Clarke, 2019).
Infine, il Capitolo 5 riporta, per ciascuna Categoria Tematica, estratti esemplificativi ed esempi del percorso che va dall'estratto al Tema Generale.
Tali attenzioni consentono di rendere tracciabile il lavoro svolto e di mostrare come le interpretazioni siano state elaborate a partire dai materiali raccolti. Lincoln e Guba (1985) riconducono questa dimensione alla credibilità e alla confermabilità della ricerca qualitativa, evidenziando l'importanza di un percorso metodologico esplicito, argomentato e coerente con gli obiettivi dello studio.

4.3.1 Posizionamento della ricercatrice

Nel presente studio di caso, la posizionalità della ricercatrice ha costituito una dimensione trasversale all'intero processo di ricerca. Essa non ha riguardato soltanto la fase di conduzione delle interviste, ma anche la costruzione del quadro teorico, la selezione e l'interpretazione della letteratura, l'analisi documentale e la lettura integrata dei dati empirici. L'indagine ha avuto come oggetto la percezione di docenti e studenti rispetto agli esiti descrittivi individuali rilasciati agli studenti dell'ultimo anno del primo e del secondo ciclo di istruzione nell'ambito delle prove INVALSI. Tale oggetto di studio si colloca in un ambito istituzionale a me vicino, poiché lavoro presso l'INVALSI occupandomi, in particolare, del processo di sviluppo delle prove di Italiano relative alla comprensione del testo e dei diversi strumenti e documenti ad esse collegati. Questa collocazione ha configurato una posizione di prossimità rispetto al fenomeno studiato, che ha rappresentato, al tempo stesso, una risorsa conoscitiva e una possibile fonte di distorsione interpretativa. In questa prospettiva, la conoscenza prodotta dalla ricerca non può essere considerata separata dalla posizione da cui il ricercatore osserva e interpreta il fenomeno. Le esperienze professionali, il ruolo istituzionale e il contesto in cui il ricercatore si colloca influenzano, infatti, il modo in cui i dati vengono selezionati, letti e compresi (Haraway, 1988). Per lo sviluppo di questa ricerca non ho quindi cercato di neutralizzare tale condizione, ma di assumerla attraverso una postura riflessiva costante, secondo una prospettiva in cui il rigore della ricerca qualitativa non dipende dall'assenza di soggettività, bensì dalla capacità di

renderne visibili le implicazioni nel processo conoscitivo (Bourdieu & Wacquant, 1992; Finlay, 2002).

L'essere interna all'INVALSI ha comportato alcuni vantaggi, in particolare, ha consentito una comprensione più approfondita del contesto istituzionale in cui l'azione è stata progettata, del linguaggio tecnico impiegato, delle finalità dichiarate dell'intervento e delle implicazioni che esso poteva avere sul rapporto con le scuole, i docenti e gli studenti. Questa conoscenza interna ha favorito una lettura più articolata dei documenti, delle scelte progettuali e delle narrazioni emerse dalle interviste, consentendo di riconoscere sfumature che, a uno sguardo esterno, sarebbero potute risultare meno visibili. In questo senso, la posizione della ricercatrice può essere collocata in quello spazio intermedio tra appartenenza e distanza descritto da Dwyer e Buckle (2009), nel quale il ricercatore non è né pienamente interno né pienamente esterno al campo, ma si muove continuamente tra comprensione ravvicinata e necessità di presa di distanza analitica. Tale condizione ha reso possibile una conoscenza più ampia del fenomeno, ma ha richiesto anche una vigilanza costante, per evitare che la familiarità con l'ente e con l'azione studiata producesse letture eccessivamente orientate dalle intenzioni progettuali o dalle categorie istituzionali già disponibili.

Un primo potenziale *bias* ha riguardato la costruzione stessa dell'oggetto di ricerca e del quadro interpretativo. La revisione della letteratura e l'analisi documentale, infatti, non possono essere considerate operazioni neutrali, poiché implicano sempre scelte di selezione, organizzazione e attribuzione di rilevanza ai materiali disponibili. Come osservano Alvesson e Sköldberg (2018), ogni processo interpretativo dipende da assunzioni teoriche e da precomprensioni che orientano il modo in cui il ricercatore guarda il fenomeno. Nel caso specifico, la ricerca poteva essere esposta al rischio di privilegiare riferimenti teorici, documenti o interpretazioni maggiormente coerenti con la logica dell'azione progettata da INVALSI, oppure di leggere la letteratura e i documenti prevalentemente alla luce di categorie già presenti nella progettazione dell'intervento. Tale rischio può essere ricondotto al *bias* di conferma, inteso come tendenza a cercare, riconoscere o valorizzare con maggiore facilità gli elementi che confermano ipotesi, aspettative o cornici interpretative già possedute (Nickerson, 1998). In una ricerca qualitativa, questo rischio non riguarda soltanto la fase di analisi dei dati empirici, ma può manifestarsi già nella definizione del problema di ricerca, nella scelta dei riferimenti teorici e nel modo in cui vengono messi in relazione letteratura, documenti e dati raccolti.

Per contenere questo rischio, la costruzione degli studi presenti in letteratura e l'analisi documentale sono state trattate come passaggi aperti e progressivi, non come strumenti finalizzati a confermare una lettura già definita. Questo ha significato utilizzare la letteratura non solo per sostenere l'impianto della ricerca, ma anche per mettere in discussione alcune premesse dell'azione studiata. I riferimenti

teorici sono stati quindi riletti più volte, cercando di distinguere ciò che aiutava effettivamente a comprendere il fenomeno da ciò che poteva soltanto rafforzare una prospettiva già vicina a quella dell'ente. Anche i documenti sono stati considerati materiali da interrogare criticamente, prestando attenzione sia a ciò che dichiaravano esplicitamente sia alle categorie e alle rappresentazioni che proponevano. Questa scelta è coerente con l'idea che i documenti, nella ricerca qualitativa, non siano semplici contenitori di informazioni, ma testi prodotti in specifici contesti istituzionali e quindi portatori di finalità, linguaggi e prospettive particolari (Atkinson & Coffey, 2011; Bowen, 2009).

Una strategia importante di contenimento è stata quella di mantenere distinti, almeno in una prima fase, i dati emersi dalle diverse fonti. La revisione della letteratura, l'analisi documentale e le interviste sono state quindi considerate come materiali dotati di una propria specificità, prima di essere messi in relazione tra loro. Questa scelta ha permesso di evitare che le percezioni di docenti e studenti fossero lette automaticamente come conferme o disconferme dell'azione costruita da INVALSI. In particolare, ha consentito di distinguere il piano dell'azione progettata, ricostruibile attraverso i documenti e il quadro teorico, da quello dell'azione percepita, che emergeva invece dalle narrazioni dei partecipanti. In questo modo, le interviste non sono state interpretate soltanto chiedendosi se gli esiti descrittivi individuali fossero stati compresi secondo le intenzioni dell'ente, ma cercando di ricostruire come docenti e studenti li avessero effettivamente percepiti, rielaborati e collocati nella propria esperienza scolastica.

Un secondo potenziale *bias* ha riguardato la familiarità con il contesto istituzionale, particolarmente rilevante quando il ricercatore analizza materiali prodotti in un ambiente che conosce dall'interno. La familiarità ha certamente facilitato l'accesso al significato dei documenti individuati, poiché ha consentito di comprendere termini, riferimenti e logiche organizzative che a uno sguardo esterno potrebbero risultare meno immediati. Allo stesso tempo, però, essa poteva ridurre la capacità di interrogare ciò che appariva già noto. In questo senso, la conoscenza ravvicinata del campo richiede una costante vigilanza riflessiva, poiché il rischio non è soltanto quello di interpretare eccessivamente i dati, ma anche quello di non vedere più ciò che meriterebbe di essere analizzato proprio perché familiare (Bourdieu & Wacquant, 1992; Finlay, 2002).

Per contenere questo *bias*, l'analisi dei documenti è stata condotta cercando di assumere uno sguardo distanziato. I documenti sono stati riletti interrogandosi su ciò che fosse effettivamente presente nel testo e su ciò che, invece, apparteneva alle mie conoscenze pregresse nel mio contesto di lavoro. Questo passaggio è stato importante per evitare di completare automaticamente eventuali impliciti documentali con informazioni già note a chi lavora all'interno dell'ente. Le parole, le espressioni e le categorie utilizzate nei documenti sono state quindi trattate come dati da analizzare, non come

significati già acquisiti. In questo modo, anche scelte linguistiche apparentemente ordinarie sono state considerate come possibili indicatori del modo in cui l'azione veniva rappresentata istituzionalmente. Un'ulteriore strategia di contenimento è consistita nel tornare più volte sui materiali, lasciando trascorrere del tempo tra una prima lettura e le riletture successive. Questo ha consentito di prendere distanza dalle prime impressioni interpretative e di verificare se esse fossero effettivamente sostenute dai dati. Le prime interpretazioni sono state quindi considerate provvisorie e non definitive. Quando una lettura appariva troppo immediata o troppo coerente con ciò che già conoscevo dell'azione studiata, è stata sospesa e ripresa in un secondo momento, cercando elementi che potessero confermarla, precisarla o metterla in discussione. Tale procedimento ha permesso di trasformare la riflessività in una pratica concreta di controllo dell'interpretazione, in linea con Finlay (2002), che la considera una risorsa metodologica per interrogare il ruolo del ricercatore nel processo di produzione del sapere.

Il contenimento dei *bias* è stato inoltre sostenuto dal confronto tra fonti diverse. La revisione della letteratura, l'analisi documentale e le interviste non sono state considerate materiali da far convergere necessariamente verso un'unica interpretazione, ma fonti capaci di mettere in luce aspetti diversi del fenomeno. In particolare, eventuali differenze tra ciò che i documenti dichiaravano, ciò che la letteratura suggeriva e ciò che docenti e studenti esprimevano sono state trattate non come incoerenze da eliminare, ma come elementi rilevanti per comprendere la complessità dello studio di caso. Questa attenzione ha permesso di evitare una lettura lineare dell'azione studiata e di mantenere aperta l'interpretazione dei dati, lasciando spazio all'emersione di significati non previsti, ambivalenti o parzialmente distanti dalle intenzioni progettuali iniziali.

Un terzo ambito di possibile distorsione ha riguardato la conduzione delle interviste e la relazione con i partecipanti. Come indicato nella sezione dedicata allo svolgimento delle interviste, il mio coinvolgimento diretto con l'oggetto della ricerca poteva incidere sia sulla formulazione delle domande e dei rilanci, sia sulla costruzione delle risposte da parte degli intervistati. In particolare, docenti e studenti avrebbero potuto orientare le proprie dichiarazioni verso posizioni più favorevoli, prudenti o socialmente desiderabili nei confronti dell'azione oggetto di studio, anche in ragione del fatto che l'intervista era condotta da una ricercatrice interna all'ente. Per contenere tale rischio, la mia posizione è stata dichiarata nel consenso informato e ribadita nella fase iniziale dell'intervista, chiarendo il ruolo ricoperto, le finalità della ricerca, riconducibili al percorso formativo di dottorato, e il valore conoscitivo attribuito alle diverse prospettive dei partecipanti. Questa esplicitazione ha avuto lo scopo di rendere trasparente una possibile fonte di condizionamento e, al tempo stesso, di chiarire che l'interesse della ricerca non era ottenere risposte aderenti alle intenzioni dell'ente, ma comprendere il modo in cui l'azione era stata percepita e interpretata dai soggetti coinvolti.

La strategia fondamentale adottata per contenere i diversi *bias* è stata, quindi, il ricorso a una riflessività sistematica e ricorsiva attraverso un'attenzione costante a rivedere quanto emergeva dall'analisi, tornando più volte sui materiali e sospendendo le prime interpretazioni quando apparivano troppo immediate o eccessivamente allineate alle aspettative iniziali. In questa prospettiva, il mio ruolo interno all'ente non è stato considerato un limite da rimuovere, ma una componente costitutiva della ricerca. Tale posizione è stata assunta come una risorsa conoscitiva da utilizzare con cautela e come una condizione da considerare responsabilmente nella costruzione dell'analisi, attraverso un lavoro continuo di esplicitazione, distanziamento e controllo riflessivo.

4.4 Il disegno della ricerca

Il disegno di ricerca adottato per lo sviluppo di questa ricerca è stato sviluppato in una cornice qualitativa volta alla comprensione profonda del fenomeno oggetto dello studio (Creswell & Poth, 2018) e si è configurato come un processo graduale e riflessivo, in cui le diverse fasi si sono intrecciate fino ad arrivare a una prima articolazione conoscitiva del fenomeno studiato.

Questa sezione si propone, dunque, di ripercorrere tale processo, illustrando le scelte metodologiche e le fasi che hanno portato alla realizzazione concreta della ricerca, dalla definizione del quadro teorico fino alla rielaborazione interpretativa dei dati raccolti.

La prima fase della ricerca è stata dedicata alla contestualizzazione del fenomeno attraverso la rassegna della letteratura e l'individuazione della documentazione utile a reperire ulteriori informazioni significative. Questo primo passaggio ha rappresentato il punto di partenza per la definizione del quadro teorico e la formulazione delle domande di ricerca. In linea con l'approccio di Stake, l'esplorazione delle fonti ha consentito di individuare concetti chiave connessi al fenomeno, modelli interpretativi e nodi critici rilevanti per l'analisi del caso. La letteratura individuata è stata utilizzata per approfondire diversi aspetti del fenomeno, non in modo rigido, ma in modo flessibile, per costruire possibili categorie di analisi (Stake, 1995).

Una fase importante del disegno della ricerca ha riguardato l'analisi documentale del contesto, in linea con l'attenzione che Stake riserva alla comprensione situata del caso. Sono stati raccolti e analizzati documenti rilevanti per il contesto oggetto di studio, tra cui i materiali relativi alle prove INVALSI, dal loro inquadramento normativo e istitutivo fino ai documenti relativi all'uso delle certificazioni descrittive. L'analisi documentale ha costituito una fonte rilevante di dati secondari, fondamentale per comprendere le specificità del contesto e per supportare il processo di triangolazione (Yin, 2018). Ha anche contribuito a creare gli strumenti di ricerca e a comprendere meglio i risultati.

Successivamente, anche grazie agli approfondimenti derivati dall'analisi della letteratura, è stato definito l'impianto metodologico e predisposte le tracce delle interviste, progettate come strumenti

flessibili e aperti, coerentemente con l'approccio qualitativo e con la logica dello studio di caso. Le tracce delle interviste semi-strutturate sono state infatti costruite a partire dai nuclei tematici emersi dalla letteratura e dall'analisi documentale, mantenendo tuttavia un margine di adattabilità che permettesse di seguire le traiettorie emerse durante l'interazione con i partecipanti (Kvale & Brinkmann, 2009).

Dopo aver costruito gli strumenti per la raccolta dei dati, è stata predisposta la domanda da sottoporre al Comitato Etico dell'Università di Verona, passaggio essenziale per garantire la conformità della ricerca ai principi etici. In questa fase sono stati esplicitati gli obiettivi dello studio, le metodologie, le modalità di coinvolgimento dei partecipanti e le strategie per tutelare la loro riservatezza. In linea con la sensibilità etica propria della ricerca qualitativa, particolare attenzione è stata dedicata alla relazione con i partecipanti e al rispetto della loro autonomia (Creswell & Poth, 2018).

Successivamente all'autorizzazione ricevuta dal Comitato Etico, si è proceduto alla selezione e al contatto dei partecipanti. I partecipanti sono stati individuati anche sulla base delle informazioni emerse dall'analisi documentale, in particolare nel contesto delle prove INVALSI, e tramite reti personali e contatti istituzionali. In linea con l'approccio di Stake, la selezione non è stata finalizzata alla rappresentatività statistica, ma alla rilevanza informativa (Stake, 2005).

La raccolta dei dati è avvenuta principalmente tramite interviste semi-strutturate svolte a distanza, e registrate previo consenso firmato

Terminata la raccolta dei dati, è stata avviata la fase di analisi delle interviste che ha previsto la trascrizione dei materiali e una lettura approfondita e ripetuta. L'analisi è stata condotta secondo un approccio tematico-riflessivo, mediante processi di codifica e di individuazione di categorie e di successivi temi generali (Braun & Clarke, 2006). In questa fase, l'attenzione è stata rivolta sia alle ricorrenze sia alle specificità che caratterizzano il fenomeno.

Le analisi dei dati raccolti hanno incluso anche il confronto tra interviste e materiali documentali, con l'obiettivo di triangolare fonti diverse per rafforzare la validità interpretativa dello studio (Yin, 2018), consentendo di cogliere connessioni, discrepanze e diversi livelli di significato. La riflessione ha riguardato sia i contenuti espliciti sia gli elementi impliciti e di contesto, con l'obiettivo di restituire la complessità del caso.

In questa fase, il riferimento all'approccio allo studio di caso di Stake ha guidato la costruzione di una narrazione interpretativa del caso, intesa come descrizione approfondita (*thick description*) delle osservazioni fatte, in cui i dati vengono intrecciati in modo coerente con il contesto, la letteratura di riferimento e le esperienze dei partecipanti (Stake, 1995).

Infine, la fase di stesura della tesi ha rappresentato il momento in cui l'intero percorso di ricerca è stato sintetizzato e rielaborato in modo costantemente riflessivo. La scrittura ha consentito di

sintetizzare il quadro teorico, l'analisi documentale e i risultati empirici, costruendo un'argomentazione coerente e solida. In linea con l'approccio qualitativo, il testo cerca di offrire una narrazione analitica che restituisca la complessità del fenomeno e ne evidenzi le implicazioni teoriche e pratiche.

4.5 La selezione dei partecipanti alla ricerca

La selezione dei partecipanti è una fase strategica di grande rilievo in ogni progetto di ricerca. Nella ricerca qualitativa, tale processo non è volto a garantire una rappresentatività statistica, quanto piuttosto a favorire la rilevanza e la ricchezza informativa dei dati raccolti (Patton, 2015). L'obiettivo principale è infatti quello di comprendere in profondità il fenomeno analizzato, piuttosto che produrre generalizzazioni su larga scala (Denzin & Lincoln, 1994). Pertanto, i campioni qualitativi tendono ad avere dimensioni contenute e vengono costruiti in modo intenzionale sulla base di criteri teorici e pratici collegati alla domanda di ricerca (Pope & Mays, 2006). Come sottolinea Stake (1995), gli individui coinvolti nello studio di caso sono scelti in quanto "informatori privilegiati", in grado di fornire spunti rilevanti e di chiarire aspetti cruciali delle dinamiche osservate.

Nell'ambito della ricerca qualitativa, una delle strategie di selezione dei partecipanti maggiormente diffusa è il campionamento intenzionale (*purposeful sampling*), che consiste nell'individuazione di partecipanti alla ricerca in grado di offrire informazioni particolarmente rilevanti riguardo al fenomeno oggetto di studio (Patton, 2015). Diverse altre strategie di selezione dei partecipanti sono ricavabili in approcci nell'ambito di metodologie di ricerca strutturate come il campionamento teorico tipico della *Grounded Theory*, che consiste nella selezione graduale dei partecipanti in base alle categorie emergenti durante l'analisi (Denzin & Lincoln, 1994), per cui gli interlocutori privilegiati per lo studio di un fenomeno in questa prospettiva possono aumentare di tipologia a seconda della necessità di introdurre voci da parte di attori differenti rispetto a quelli teorizzati all'inizio.

Laddove, invece, selezionare i partecipanti risulta più complicato, si ricorre spesso al campionamento a catena (*snowball sampling*), che si basa sulla selezione di partecipanti provenienti dalle reti relazionali dei primi partecipanti selezionati. Altre strategie includono il campionamento per massima variazione, utile per esplorare la pluralità delle esperienze, e il campionamento di convenienza, che si adotta in presenza di vincoli pratici (Pope & Mays, 2006).

Nel contesto dello studio di caso, la strategia individuata non esclude la possibilità di utilizzarne anche altre se necessario, in linea con la natura evolutiva della ricerca qualitativa (Stake, 2005).

Questa flessibilità riflette la natura iterativa del processo conoscitivo, in cui la raccolta e l'analisi dei dati procedono in parallelo e orientano progressivamente le scelte metodologiche. Inoltre, è

importante essere trasparenti su come si scelgono e reclutano i partecipanti, perché ciò garantisce la credibilità e la rigorosità della ricerca (Denzin & Lincoln, 1994).

Per la selezione dei partecipanti della presente ricerca, è stata adottata, in prima istanza, una strategia di tipo *purposeful sampling* prevalentemente per la selezione dei partecipanti docenti; un secondo gruppo di docenti, senza esperienze di collaborazione con l'INVALSI, è stato individuato tramite reti personali e contatti istituzionali legati all'INVALSI, secondo una logica di campionamento di convenienza; gli studenti, infine, sono stati individuati dai docenti intervistati, con una selezione mediata dagli insegnanti. Nello specifico, nel processo di selezione dei partecipanti a questa ricerca è stato individuato un primo gruppo di docenti di italiano che, nel corso degli anni, avevano collaborato con l'INVALSI alla costruzione delle prove. La scelta di partire dai docenti che hanno collaborato con INVALSI è stata motivata dal fatto che loro avessero familiarità con gli strumenti rilasciati in esito alle prove in quanto facenti parte dei gruppi di lavoro per la costruzione delle prove di Italiano. Il loro coinvolgimento nella ricerca ha avuto lo scopo di comprendere come un punto di vista informato sulla struttura e sull'esito delle prove percepisse tali restituzioni degli esiti all'interno del loro contesto scolastico di classe. Parallelamente, sono stati individuati altri docenti tramite reti personali e contatti istituzionali legati all'INVALSI.

Tabella 1 - Docenti partecipanti per grado scolastico e collaborazione con l'INVALSI

Docenti	**Numero docenti secondaria di primo grado**	**Numero docenti secondaria di secondo grado**	**Totale docenti**
Docenti che hanno lavorato con INVALSI	2	5	7
Docenti che non hanno lavorato con INVALSI	5	3	8
Totale docenti	7	8	15

Questa selezione ha permesso di osservare il fenomeno da prospettive diverse, arricchendone la comprensione attraverso il confronto di esperienze eterogenee. L'attenzione alla varietà delle posizioni sul fenomeno oggetto di studio rispecchia l'interesse tipico dello studio di caso per le molteplici interpretazioni che contribuiscono a definire la realtà studiata (Stake, 2005). La scelta di coinvolgere esclusivamente docenti di italiano è stata motivata dall'esigenza di concentrare l'analisi su una categoria di insegnanti particolarmente coinvolta nelle prove di valutazione esterna sulla comprensione del testo. Inizialmente era stato immaginato un campione più ampio di studenti e docenti, ma in questa prima fase di ricerca esplorativa si è preferito ridurre la numerosità, anche in

considerazione dell'organizzazione dei tempi della ricerca. In totale, sono stati coinvolti 15 insegnanti: 7 appartenenti al gruppo di docenti che avevano già collaborato con l'INVALSI e 8 individuati tramite reti personali e contatti istituzionali legati all'INVALSI, ma che non avevano mai collaborato in alcun modo con l'istituto. Inoltre, coerentemente con l'attenzione di Stake (2005) alla molteplicità dei punti di vista, si è scelto di integrare le prospettive degli insegnanti con quelle degli studenti. A ogni insegnante disponibile è stato quindi chiesto di scegliere due studenti della propria classe dell'ultimo anno del rispettivo ciclo da intervistare sui temi in questione. Nove studenti hanno accettato di partecipare alla ricerca: 3 dell'ultimo anno della scuola secondaria di primo grado e 6 dell'ultimo anno della scuola secondaria di secondo grado.

Tabella 2 - Studenti partecipanti per grado scolastico e per collaborazione con l'INVALSI del docente di riferimento

Studenti	**Numero studenti secondaria di primo grado**	**Numero studenti secondaria di secondo grado**	**Totale studenti**
Studenti di docenti che hanno lavorato con INVALSI	3	4	7
Studenti di docenti che non hanno lavorato con INVALSI		2	2
Totale studenti	3	6	9

È tuttavia importante sottolineare che non tutti i docenti hanno indicato i nominativi degli studenti e che sette studenti su nove sono stati segnalati da docenti che avevano collaborato con l'INVALSI alla costruzione delle prove. Questa composizione ha implicazioni per l'interpretazione dei risultati: è plausibile che questi studenti abbiano avuto un contatto con i materiali e con le descrizioni dei livelli più frequente della media. Le implicazioni sono discusse nel paragrafo 7.1.

Tutte le interviste si sono svolte online, sulla piattaforma Zoom, tra maggio 2024 e gennaio 2025, con una durata media di circa 40 minuti (par. 4.6.3). In coerenza con gli impegni di riservatezza assunti nel consenso informato, e considerando che una parte dei docenti appartiene a una comunità professionale ristretta come quella dei gruppi di lavoro INVALSI, le caratteristiche dei partecipanti sono riportate solo in forma aggregata, così da evitare che l'incrocio di più informazioni renda riconoscibili i singoli. Per gli studenti, l'informazione più rilevante ai fini dell'interpretazione, cioè il grado di conoscenza delle descrizioni dei livelli prima dell'intervista, è richiamata nel Capitolo 5 insieme agli estratti.

Dal percorso di ricerca è emersa l'utilità di includere, in eventuali sviluppi successivi dello studio, ulteriori tipologie di partecipanti. Oltre a docenti e studenti, il coinvolgimento di dirigenti scolastici e famiglie avrebbe potuto arricchire la comprensione del fenomeno, offrendo ulteriori piani di lettura sulla restituzione descrittiva degli esiti INVALSI e sulle condizioni che ne favoriscono l'uso nei processi educativi.

Nel complesso, la strategia di selezione adottata ha consentito di costruire un corpus di dati adeguato a un'analisi esplorativa del fenomeno, con i limiti di composizione indicati sopra e ripresi nel paragrafo 7.1.

4.6 La metodologia della raccolta dati

Il presente paragrafo illustra la metodologia adottata per la raccolta dei dati, delineando gli strumenti e le procedure impiegate per costruire un quadro conoscitivo coerente con gli obiettivi della ricerca. La raccolta dei dati è stata condotta mediante una pluralità di strategie qualitative, scelte per consentire un'analisi approfondita del fenomeno indagato.

In particolare, verranno presentate tre modalità principali di raccolta: la revisione sistematica della letteratura, finalizzata a ricostruire il quadro teorico e concettuale di riferimento; l'analisi documentale, orientata all'esame di materiali, documenti e fonti istituzionali rilevanti per il contesto della ricerca; e l'intervista semi-strutturata, descritta attraverso la costruzione delle tracce d'intervista e il loro ruolo nell'approfondimento delle prospettive dei soggetti coinvolti. Questi strumenti, considerati nel loro insieme, hanno permesso di raccogliere dati eterogenei ma complementari, contribuendo così a una comprensione sistemica dell'oggetto di studio.

4.6.1 La revisione sistematica della letteratura

La revisione sistematica della letteratura ha costituito una delle fonti conoscitive dello studio ed è stata utilizzata per ricostruire, secondo criteri espliciti, trasparenti e controllabili, lo stato della ricerca sul fenomeno indagato. A differenza di una rassegna narrativa, che consente una ricostruzione più libera e interpretativa del dibattito scientifico, la revisione sistematica prevede una procedura definita di ricerca, selezione, analisi e sintesi degli studi, fondata su criteri dichiarati di inclusione ed esclusione. Come evidenziano Gough, Oliver e Thomas (2017), la revisione sistematica non consiste soltanto nel raccogliere contributi esistenti, ma nel costruire un percorso metodologicamente esplicito attraverso cui individuare, organizzare e interpretare le conoscenze disponibili su un determinato tema. In modo analogo, Grant e Booth (2009) sottolineano che le diverse tipologie di revisione si distinguono anche per il grado di sistematicità delle procedure adottate e per il modo in cui vengono

condotte le fasi di ricerca, di analisi e di sintesi della letteratura. Nel presente studio, la revisione sistematica è stata quindi impiegata come una fonte autonoma di dati conoscitivi, utile a individuare concetti ricorrenti, orientamenti teorici, risultati consolidati, eventuali aree di incertezza e questioni ancora poco esplorate dalla letteratura esistente. Essa ha permesso di costruire una base teorica solida e di ridurre il rischio di una selezione arbitraria dei contributi, rendendo più chiaro il percorso attraverso cui la letteratura è stata utilizzata nella ricerca.

Nel contesto del presente studio qualitativo, la revisione sistematica della letteratura è stata impiegata come una delle tre fonti da mettere in relazione nell'interpretazione finale del fenomeno. I risultati della revisione sono stati utilizzati come riferimento per comprendere in che modo le evidenze emerse nel caso studiato si collocassero rispetto alle conoscenze già prodotte, svolgendo una funzione orientativa e interpretativa soprattutto quando sono poi integrati con altre fonti qualitative nel quadro complessivo della ricerca (Gough et al., 2017).

Sul piano operativo, la revisione è stata condotta sulla banca dati Scopus, con la stringa di ricerca e lo schema PCC (*Population*, *Concept*, *Context*) riportati nel paragrafo 5.1, considerando i contributi pubblicati fino al 2026. Sono stati inclusi gli studi che: a) riguardavano studenti della scuola secondaria di primo e secondo grado e/o i loro insegnanti; b) si riferivano a valutazioni standardizzate nazionali o a rilevazioni internazionali su larga scala, con particolare attenzione alla comprensione della lettura; c) prevedevano una restituzione dei risultati ai partecipanti; d) riportavano evidenze empiriche su percezioni, esperienze o usi di tale restituzione, con disegno qualitativo, quantitativo o misto. Sono stati esclusi i contributi che non riguardavano la scuola secondaria di primo o di secondo grado e quelli che non si riferivano a prove o rilevazioni standardizzate. L'ultima interrogazione della banca dati è stata effettuata a maggio 2026. Lo screening di titoli, abstract e testi completi è stato svolto con il software Rayyan dalla sola autrice, senza un secondo revisore indipendente. Poiché gli studi pienamente centrati sulla restituzione descrittiva individuale sono risultati molto scarsi, sono stati inclusi anche contributi parzialmente pertinenti; per ciascuno studio incluso, il paragrafo 5.1 e l'Appendice 1 indicano il grado di pertinenza rispetto a questo oggetto. Il ricorso a una sola banca dati e lo screening condotto da una sola persona costituiscono limiti della revisione, discussi nel paragrafo 7.1.

4.6.2 L'analisi documentale

L'analisi documentale ha avuto lo scopo di esaminare i materiali e le fonti istituzionali rilevanti, al fine di comprendere il contesto in cui il fenomeno studiato si colloca. In una ricerca qualitativa, i documenti sono rappresentati da testi prodotti in specifici contesti sociali e istituzionali, portatori di linguaggi, finalità, rappresentazioni e assunzioni implicite. Bowen (2009) definisce l'analisi

documentale come una procedura sistematica di revisione e valutazione dei documenti, attraverso la quale i materiali vengono esaminati e interpretati per ricavare significati e sviluppare una comprensione empirica dei fenomeni. Analizzare i documenti significa quindi interrogare non soltanto ciò che essi dichiarano esplicitamente, ma anche il modo in cui viene costruito l'oggetto di cui parlano, le categorie che utilizzano, i destinatari a cui si rivolgono e le logiche istituzionali che ne rendono l'oggetto visibile. Nel presente studio, l'analisi ha previsto una lettura attenta e progressiva dei materiali, orientata a individuare temi ricorrenti, riferimenti concettuali, impliciti e modalità di rappresentazione dell'azione. Tali elementi sono stati successivamente messi in relazione con la revisione sistematica della letteratura e con le interviste, per distinguere il piano del discorso istituzionale da quello delle percezioni e delle interpretazioni dei soggetti coinvolti.

4.6.3 L'intervista semi strutturata

L'intervista è una delle tecniche di ricerca qualitativa più utilizzate e importanti, considerata come una "finestra sul mondo dell'esperienza per ricercatori e professionisti" (Gubrium & Holstein, 2002). Questo strumento di raccolta dati è utile quando l'obiettivo è comprendere la complessità delle interpretazioni soggettive nei fenomeni sociali. L'intervista non è semplicemente un mezzo per raccogliere informazioni, ma uno spazio di interazione in cui i significati emergono attraverso il dialogo e la narrazione, riflettendo il carattere soggettivo e costruito dell'esperienza.

Nonostante la sua apparente semplicità, questa metodologia solleva questioni concettuali e metodologiche complesse. Come sottolineato, infatti, nel volume *Handbook of Interview Research* (Gubrium et al., 2012), la scelta di una specifica tipologia di intervista da parte del ricercatore dipende da diverse questioni su cui l'intervistatore stesso deve innanzitutto riflettere, ponendosi alcuni interrogativi fondamentali, come: "Chi è l'intervistatore agli occhi dell'intervistato? Da dove ottengono gli intervistatori le domande e gli intervistati le risposte che comunicano?". La scelta del tipo di intervista da utilizzare per la raccolta dei dati non è mai casuale poiché riflette la posizione epistemologica del ricercatore e si collega strettamente agli obiettivi della ricerca (Semeraro, 2011). L'*intervista libera* o non direttiva, ad esempio, si inserisce in un paradigma interpretativo o fenomenologico, il cui obiettivo è raccogliere materiale spontaneo attraverso domande molto generali. In questo contesto, l'analisi dei dati sarà principalmente qualitativa, concentrandosi sulle modalità con cui gli intervistati spiegano, selezionano e interpretano fatti per descrivere realtà diverse. Diversamente, un modello di *intervista* completamente *strutturata* si inserisce nell'ambito di un paradigma di tipo positivista. In questo metodo di intervista, il ricercatore costruisce uno strumento composto da domande prefissate, molto specifiche e tendenzialmente a risposta chiusa, per standardizzare la raccolta dei dati e facilitarne la generalizzazione. Pur avvicinandosi al questionario,

questa metodologia si differenzia in quanto "le risposte dell'intervistato hanno, in questo caso, maggiori gradi di libertà" (Semeraro, 2011). I dati raccolti con questo tipo di strumento sono trattati mediante analisi quantitative (frequenze, correlazioni, regolarità, fatti oggettivi, ecc.).

L'*intervista semi-strutturata* emerge come sintesi metodologica tra questi due estremi. In questa ricerca, si è scelto di utilizzare questo metodo per esplorare in profondità le esperienze e le percezioni di docenti e studenti. La costruzione di questo strumento si basa sulla preparazione di una serie di domande o argomenti chiave (*topic guide*) a cui il ricercatore cerca di ottenere risposte. A differenza degli strumenti quantitativi, nelle interviste semi-strutturate l'intervistatore può modificare l'ordine delle domande, adattarne la formulazione e, soprattutto, aggiungerne di approfondimento (*probing*) non incluse nella traccia originale. Questo sistema flessibile consente di orientare il colloquio intorno ad alcuni temi centrali, senza però escludere l'opportunità di approfondire aspetti inattesi o particolarmente rilevanti per la comprensione del fenomeno. Per questa ragione, l'intervista semi-strutturata si configura come uno strumento particolarmente adatto alla raccolta di dati qualitativi ricchi e contestualizzati dal momento che la presenza di una traccia permette di mantenere il focus della ricerca, mentre la relazione dialogica con l'intervistato consente di lasciare spazio alla narrazione, all'approfondimento e alla costruzione progressiva del significato (Natali, 2024). Un elemento metodologico importante da considerare nelle interviste per la raccolta di dati qualitativi è che l'azione di intervistare non è un processo neutro, ma è il risultato dell'interazione tra intervistatore e intervistato. Diversa letteratura metodologica (Creswell & Poth, 2024; Gubrium et al., 2012) ha evidenziato il rischio che, durante lo svolgimento delle interviste, le risposte dell'intervistato possano essere orientate dalle aspettative percepite, soprattutto quando l'intervistatore è direttamente coinvolto nei temi oggetto dell'indagine. Questa tendenza è nota come "*bias di desiderabilità sociale*", ossia il fatto che tendiamo a fornire risposte che l'altra persona potrebbe ritenere più accettabili, legittime o gradite, anche se ciò comporta fornire informazioni meno precise (Paulhus, 1991). Questo fenomeno è spesso collegato a un effetto più ampio, legato all'influenza dell'intervistatore, che deriva dalle caratteristiche, dal ruolo e dalla posizione simbolica del ricercatore e può incidere sulle risposte fornite. Corbetta (2014) evidenzia come il coinvolgimento dell'intervistatore, che sia di natura tematica, valoriale o professionale, possa influenzare il modo in cui l'intervistato interpreta la situazione comunicativa, portandolo ad adattarsi o a soddisfare le aspettative del discorso. Nel caso specifico di questa ricerca, la relazione tra l'intervistatrice e gli intervistati potrebbe essere stata influenzata da questo *bias*, dato il coinvolgimento diretto con l'INVALSI e, di conseguenza, con i temi oggetto dell'intervista. In altre parole, il fatto di aver condotto direttamente le interviste potrebbe aver introdotto alcune distorsioni che riguardano, da un lato, il modo in cui le ho condotte, potenzialmente influenzato dal fornire indicazioni sulle risposte,

e, dall'altro, il fatto di aver ricevuto risposte da parte dell'intervistato/a maggiormente orientate a mostrare atteggiamenti positivi nei confronti dei temi oggetto dell'intervista. Per contenere questo *bias*, una prima strategia è stata quella di dichiarare immediatamente tale condizione, sia nella stesura del consenso informato sia all'inizio dell'intervista. Attraverso questa spiegazione si è cercato di sottolineare in modo positivo come tale condizione fosse utile alla ricerca, poiché avrebbe permesso di comprendere meglio il fenomeno oggetto di studio, in quanto chi se ne stava occupando, essendo direttamente coinvolta, era in grado di individuare maggiori sfumature collegate alle informazioni fornite, ma anche di individuare indicazioni utili per il miglioramento di alcuni aspetti collegati alle prove nazionali standardizzate. In questa prospettiva, la riflessività (*reflexivity statement*) assume una funzione epistemologica più che meramente etica (Finlay, 2002). Durante la fase di intervista, in diverse occasioni sono state proposte domande di rilancio principalmente come spunti per chiarire meglio le informazioni fornite dagli intervistati (Corbetta, 2014). Altri elementi fondamentali per la conduzione di interviste rigorose riguardano l'atteggiamento, che dovrebbe essere sostanzialmente non giudicante, con ascolto attivo costante, nonché empatico e amichevole durante l'intero processo (Natali, 2024). Durante le interviste della presente ricerca, nella fase di presentazione e introduzione, dopo aver spiegato le modalità di svolgimento dell'intervista, veniva indicato uno spazio specifico in chiusura in cui avrebbero potuto rivolgere una domanda all'intervistatrice, relativa a una curiosità o ad altro sulla ricerca o sull'INVALSI in generale. Con questo sistema si è cercato di instaurare un rapporto interlocutorio con l'intervistato, con l'obiettivo di trasformare il potenziale *bias* determinato dal coinvolgimento dell'intervistatrice con l'oggetto dell'intervista in una risorsa.

Per quanto riguarda il setting delle interviste, come già indicato, nella presente ricerca si è scelto di condurle online per garantire la massima partecipazione dei partecipanti individuati (docenti e studenti). La scelta del setting è un aspetto rilevante nella conduzione dell'intervista qualitativa, poiché incide sulle condizioni di comfort, riservatezza e libertà espressiva dei partecipanti. Per questo motivo, è opportuno individuare un ambiente tranquillo, privo di distrazioni e percepito come sicuro, in grado di favorire la concentrazione e la disponibilità alla narrazione (Jacob & Furgerson, 2012; Elwood & Martin, 2000). Nel caso delle interviste svolte a distanza, tali condizioni assumono ulteriore importanza, poiché il contesto domestico e lo strumento tecnologico possono influire sulla qualità dell'interazione e sulla gestione della relazione tra intervistatore e intervistato (Samuk Carignani & Burchi, 2022). Nel presente studio, ogni intervistato/a si trovava in un ambiente adeguato allo svolgimento dell'intervista: i docenti erano, in molte occasioni, a scuola, in un'aula vuota, oppure presso la propria abitazione; gli studenti, invece, si trovavano tutti a casa.

La durata dell'intervista è stata un altro elemento preso in considerazione nella fase di progettazione metodologica. Non esiste una durata standard in ogni situazione, in quanto bisogna tenere in

considerazione diversi criteri, come la necessità di approfondimento e la capacità di mantenere alta l'attenzione. Interviste troppo lunghe possono provocare stanchezza, riducendo la qualità delle risposte e, soprattutto, facendo perdere coerenza al racconto. In questi casi, il ricercatore deve essere in grado di modulare il ritmo dell'interazione, lasciando spazio al racconto, ma anche, quando necessario, orientarlo verso i temi di ricerca. Le interviste hanno avuto una durata media di circa 40 minuti, tempo che ha consentito di discutere ogni traccia dell'intervista semi-strutturata.

Anche la scelta del tono è fondamentale per instaurare un dialogo rispettoso e favorire un clima di fiducia, essenziale per far emergere narrazioni ricche di dettagli e di contesto, soprattutto considerando il ruolo dell'intervistatore. Per le interviste svolte in questa ricerca è stato adottato un tono informale utilizzando il "tu" con i docenti già conosciuti, mentre con i docenti non conosciuti si è mantenuto l'uso del "lei". Per le interviste agli studenti è stato adottato un tono informale, utilizzando il "tu". Per tutte le interviste è stato possibile creare un clima sereno, di riflessione, in alcuni casi argomentata e di rispetto reciproco.

Le interviste sono state condotte per la maggior parte tra maggio e dicembre 2024, e le ultime tre di soli studenti a gennaio 2025, contattando i partecipanti via mail e concordando la data e l'orario.

Per l'interpretazione dei risultati è rilevante la collocazione delle interviste rispetto alla prova e alla restituzione degli esiti. Sette studenti su nove sono stati intervistati durante l'ultimo anno del rispettivo ciclo, prima dello svolgimento della prova e quindi prima di ricevere la restituzione individuale del proprio esito, che avviene al termine del percorso: le loro risposte non si riferiscono a un esito già ricevuto, ma alla conoscenza che avevano del modello di restituzione e alla lettura dei descrittori proposta durante l'intervista (par. 4.6.4), e alcuni di loro hanno incontrato le descrizioni dei livelli per la prima volta proprio in quella occasione. Due studenti, invece, sono stati intervistati dopo aver concluso la scuola secondaria di secondo grado e avevano già ricevuto la propria certificazione: uno dichiara di averla consultata e di aver usato i livelli ottenuti per ripassare alcuni argomenti in vista di un test di ingresso all'università, l'altro di non averla mai aperta. I docenti, invece, hanno potuto fare riferimento anche alle restituzioni degli anni precedenti. Per questa ragione, nel Capitolo 5 i commenti degli studenti basati su esperienze precedenti sono distinti da quelli generati dalla presentazione dei descrittori.

4.6.4 Le tracce delle interviste

In questa sezione verranno illustrate le tracce delle interviste e le dimensioni considerate nella formulazione delle domande. Lo sviluppo di queste tracce si è basato sugli obiettivi generali della ricerca e sulle domande specifiche. Il processo di operazionalizzazione, nel contesto della ricerca qualitativa, consiste nel trasformare concetti astratti e domande generali in domande specifiche e in

aree di indagine ben definite. In questo paragrafo vengono quindi presentate le domande utilizzate durante le interviste semi strutturate, con l'obiettivo di verificare in che modo le informazioni elicitate dalle domande siano collegate ai presupposti generali di questa ricerca. Questo procedimento è essenziale per garantire che i dati raccolti siano rilevanti, completi e affidabili, per raggiungere gli obiettivi dello studio. Innanzitutto sono state riprese le due domande di ricerca e stabilite le domande dell'intervista che hanno cercato di raccogliere le informazioni utili per rispondere ai due quesiti generali.

Per rispondere alla prima domanda di ricerca, "Quali sono le percezioni degli studenti e dei docenti riguardo agli esiti descrittivi individuali delle prove nazionali standardizzate sulla comprensione del testo?", sono state formulate domande per esplorare le opinioni, le aspettative e le interpretazioni degli studenti in merito a questo nuovo strumento di feedback (il certificato rilasciato come esito delle prove standardizzate). Con la domanda "*Cosa pensi del fatto che adesso riceverai un feedback individuale sulla prova INVALSI, che è una prova standardizzata a livello nazionale e quindi esterna alle valutazioni scolastiche?*" si è cercato di raccogliere informazioni sulla percezione degli studenti del processo di valutazione esterna, con l'obiettivo di comprendere in che modo tale feedback viene vissuto. Con la domanda "*Quali vantaggi e quali svantaggi (o punti di forza e debolezza) vedi in questo feedback?*" si è tentato di indagare alcuni aspetti connessi alla *feedback literacy*, che, secondo la letteratura (Bevilacqua, 2023; Carless & Boud, 2018), consiste nella capacità di interpretare criticamente le valutazioni e di immaginarne i possibili utilizzi. Inoltre, attraverso le domande "*Nella sua esperienza di insegnante, che idea hanno i suoi studenti della valutazione INVALSI in generale e delle prove come strumento di valutazione?"* e "*Rispetto al risultato che riceveranno e ai contenuti descrittivi del certificato?"* rivolte ai docenti, si è cercato di verificare la percezione degli studenti della prova nazionale standardizzata attraverso lo sguardo dei docenti in modo da poter incrociare le informazioni su questo stesso tema raccolte da due punti di vista differenti.

Per rispondere alla seconda domanda di ricerca, "In che modo, a partire dall'esperienza dei soggetti coinvolti, studenti e docenti, gli esiti descrittivi delle prove standardizzate nazionali relativi ai livelli raggiunti nella prova di comprensione del testo possono incidere sui processi di insegnamento e apprendimento?", sono state formulate domande per comprendere quali fossero possibili gli effetti delle restituzioni descrittive delle prove INVALSI all'interno del contesto scolastico. Con le domande *"Quali sono le informazioni che le sembrano utili nella sua azione didattica funzionale allo svolgimento delle prove INVALSI o anche al di là di esse?"* (rivolta ai docenti) o *"Mi sapresti dire quali contenuti, tra quelli presenti nelle descrizioni dei livelli, sono familiari a te nella tua esperienza a scuola?"* e *"E quali ritieni siano più utili per te per capire i tuoi apprendimenti (quello che hai imparato) rispetto alla comprensione del testo, riflettendo anche sul senso di una prova*

standardizzata come quella di INVALSI?" (rivolta agli studenti) si è tentato di comprendere come il processo di valutazione esterna e la restituzione dei suoi esiti si inseriscano nelle dinamiche della classe e nell'azione didattica quotidiana.

Nella tabella seguente sono riportate le domande delle interviste nell'ordine in cui sono state presentate.

Tabella 3 - Tracce delle interviste.

Traccia per intervista semi strutturata DOCENTI	**Traccia per intervista semi strutturata STUDENTI**
Verifica della percezione generale degli strumenti/materiali INVALSI - *Mi può raccontare, per la sua esperienza, quali tra i materiali messi a disposizione da INVALSI le sono sembrati utili e perché?* - *Quali materiali utilizza nella sua attività didattica funzionali allo svolgimento delle prove INVALSI?*	**Verifica della percezione generale degli strumenti/materiali INVALSI** - *Nella tua esperienza, quali materiali/strumenti/informazioni ti sono stati utili per capire/comprendere che cosa valuta la prova INVALSI di Italiano? Che cosa viene valutato con la prova di italiano* - *E, in particolare, quali materiali ti sono stati utili per comprendere i risultati che ti verranno restituiti dopo averla svolta?*
Verifica della percezione delle descrizioni dei livelli di comprensione del testo restituiti dalle prove - *Rileggendo le descrizioni dei livelli di comprensione del testo (in particolare nella versione analitica), cosa ne pensa di questo modello di restituzione degli esiti?* - *Quali sono le informazioni che le sembrano utili nella sua azione didattica funzionale allo svolgimento delle prove INVALSI o anche al di là di esse?*	**Verifica della percezione delle descrizioni dei livelli di comprensione del testo restituiti dalle prove** - *Mi sapresti dire quali contenuti, tra quelli presenti nelle descrizioni dei livelli, sono familiari a te nella tua esperienza a scuola?* - *E quali ritieni siano più utili per te per capire i tuoi apprendimenti (quello che hai imparato) rispetto alla comprensione del testo, riflettendo anche sul senso di una prova standardizzata come quella di INVALSI?*
Verifica della percezione nel contesto classe del processo di valutazione standardizzata - *Nella sua esperienza di insegnante, che idea hanno i suoi studenti della valutazione INVALSI in generale e delle prove come strumento di valutazione?* - *Rispetto al risultato che riceveranno e ai contenuti descrittivi del certificato?*	

Traccia per intervista semi strutturata DOCENTI	Traccia per intervista semi strutturata STUDENTI
Metariflessione sul feedback descrittivo INVALSI - *Riassumendo, quali sono, secondo lei, i punti di forza e i punti di debolezza di un modello di restituzione degli esiti della prova INVALSI attraverso un sistema di descrittori di abilità progressivo?*	**Verifica della percezione del feedback e metariflessione sul feedback descrittivo INVALSI** - *Cosa pensi del fatto che adesso riceverai un feedback individuale rispetto alla prova INVALSI che è una prova standardizzata a livello nazionale quindi esterna rispetto alle valutazioni scolastiche* - *Quali vantaggi e quali svantaggi (o punti di forza e debolezza) vedi in questo feedback*

È importante sottolineare che durante lo svolgimento delle interviste, per le domande relative alla dimensione sulla verifica della percezione delle descrizioni dei livelli descrittivi restituiti dalle prove, è stato condiviso lo schermo affinché studenti e docenti potessero vedere e leggere o rileggere il documento con le descrizioni dei livelli che viene rilasciato agli studenti e che, sul sito INVALSI, viene proposto anche in una versione analitica ovvero in una versione in cui le descrizioni riportano maggiori dettagli relativi ai contenuti dei compiti che rappresentano un particolare livello di competenza. Successivamente a tale lettura o rilettura venivano poste le domande sui contenuti del documento di certificazione dei livelli di competenza.

4.6.5 La trascrizione delle interviste

Tutte le interviste, come già descritto, sono state condotte in un ambiente neutrale e confidenziale, online, tramite la piattaforma Zoom, e registrate, previo consenso informato, per poi essere trascritte integralmente. Per la trascrizione si è scelto un metodo denominato "naturalizzato", che consiste nell'eliminazione di elementi idiosincratici, come balbettii e ripetizioni, per ottenere una riproduzione che rispetti complessivamente le convenzioni del linguaggio scritto. Questa scelta ha permesso di creare un corpus più leggibile, facilitando la lettura interpretativa incentrata sul contenuto tematico (Pagani, 2020). Non sono stati considerati, ai fini della trascrizione delle interviste, aspetti non verbali o del contesto, se non attraverso la rilettura, durante la trascrizione, di potenziali atteggiamenti determinati dal coinvolgimento dell'intervistatrice nei temi dell'intervista. I dati raccolti tramite le interviste semi-strutturate sono stati successivamente analizzati secondo metodologie di tipo qualitativo che includono la spiegazione, nel senso di individuare regolarità o ricorrenze all'interno dei discorsi, la selezione dei fatti, in termini di individuazione di segmenti di testo significativi, e l'interpretazione da parte degli intervistati, ovvero di come spiegano la loro esperienza.

4.7 La metodologia per l'analisi delle interviste

L'analisi qualitativa dei dati delle interviste è stata condotta mediante l'Analisi Tematica Riflessiva per identificare i principali nuclei di significato (temi, concetti, modelli) nei testi. Questo metodo adotta un approccio induttivo, che significa muoversi dai dati empirici per sviluppare interpretazioni, in linea con l'obiettivo idiografico della ricerca qualitativa, che intende comprendere il particolare e il singolo, invece di cercare leggi universali. Esistono diversi metodi di analisi dei dati qualitativi e, solitamente, sono strettamente collegati alla metodologia di ricerca adottata. Nella presente ricerca il metodo prescelto dello studio di caso non predilige un modello di analisi dei dati in particolare, per cui ci si è orientati verso approcci di Analisi Tematica delle interviste secondo il modello descritto da Braun e Clarke (2006; 2019).

4.7.1 L'Analisi Tematica Riflessiva

L'Analisi Tematica (AT) è un metodo versatile per l'analisi dei dati qualitativi che consente di indagare in profondità le percezioni, le esperienze e i significati che i partecipanti attribuiscono a un fenomeno specifico. Questo approccio consente di riconoscere e interpretare i modelli di significato, o pattern, presenti in un corpus di dati testuali, andando oltre le narrazioni individuali per individuare strutture di senso condivise (Clarke & Braun, 2016). Una delle sue caratteristiche più apprezzate è la flessibilità, poiché non è un modello vincolato a una cornice teorica o epistemologica specifica. Al contrario, il ricercatore può utilizzarla in studi in cui vi sia un quadro di riferimento definito, adattandola alle esigenze della propria ricerca (Braun, Clarke & Hayfield, 2019). Da questi presupposti ne deriva necessariamente che il ricercatore specifichi i fondamenti teorici e le scelte operative che lo guidano nell'utilizzo, per garantire il rigore della ricerca. Di seguito verranno illustrati gli aspetti metodologici che hanno caratterizzato l'impiego dell'Analisi Tematica in questa ricerca.

Prima di avviare il processo di analisi, è essenziale definire con chiarezza il quadro teorico di riferimento e le decisioni metodologiche ad esso associate. Questa fase preliminare serve a garantire rigore, coerenza e trasparenza nell'intero processo di ricerca, rendendo le scelte operate chiare e verificabili.

Nei lavori di Braun e Clarke (2016) si distingue tra l'Analisi Tematica come metodo e come metodologia. Questa distinzione evidenzia la flessibilità dell'AT, che, in quanto metodo, include un insieme di procedure per analizzare i dati, ma che deve essere collocata dal ricercatore in una cornice metodologica più ampia, che guida gli obiettivi e le modalità di applicazione.

Nel contesto dell'AT, Braun e Clarke (2019) hanno introdotto una tripartizione per differenziare i principali approcci alla codifica e all'interpretazione:

- *Coding Reliability,* si concentra sulla precisione e sull'affidabilità del processo di codifica, che vengono spesso valutate tramite indicatori di concordanza tra diversi codificatori.
- *Riflessivo,* un metodo in cui la soggettività del ricercatore non viene considerata un *bias* da eliminare, ma è riconosciuta e considerata una risorsa importante per una codifica credibile, coerente e dettagliata.
- *Code Book,* una variante che rappresenta una visione strutturata del processo di verifica, partendo da un processo analitico sviluppato dal ricercatore, cercando di armonizzare le due posizioni precedenti.

Per questa ricerca, è stato scelto l'approccio riflessivo, in linea con gli obiettivi della ricerca qualitativa, che mira a ottenere una comprensione approfondita e dettagliata dei fenomeni, riconoscendo l'interpretazione della ricercatrice come elemento chiave nel processo di produzione della conoscenza.

Le decisioni principali prese durante l'analisi sono state le seguenti:

- *Approccio alla Codifica*: l'approccio scelto per la ricerca è stato quello induttivo (o bottom-up), che non prevede un framework di codifica predefinito, ma lo sviluppo del coding a partire dai dati stessi. Le interviste dei docenti sono state codificate per prime, senza etichette predefinite. Le interviste degli studenti sono state codificate successivamente: le etichette già individuate per i docenti sono state applicate quando i contenuti erano analoghi, mentre per i contenuti non riconducibili ad esse sono state create nuove etichette (par. 5.3). L'analisi è proceduta un processo che va dal particolare (coding) al generale (temi generali). Questo approccio è stato scelto perché consente di esplorare le percezioni dei partecipanti senza imporre categorie di codifica preconcette, garantendo un'analisi strettamente ancorata alle esperienze dirette emerse durante le interviste.
- *Livello di Analisi*: il livello di analisi scelto è stato *semantico*, perché si è cercato di analizzare i significati espliciti e superficiali espressi dai partecipanti, mantenendo fede alle loro parole e alle loro prospettive. Scegliere questo livello di analisi consente di cogliere e riportare fedelmente le visioni espresse direttamente dai partecipanti, evitando un approccio latente che indaga idee e presupposti impliciti (Pagani, 2020).

L'ultimo elemento metodologico considerato è stato il ruolo del ricercatore in quanto strumento fondamentale del processo di analisi dei dati. Le interpretazioni dei dati, infatti, sono fortemente collegate a tutte le esperienze, alle conoscenze pregresse e alla sensibilità di chi svolge l'analisi, ovvero del ricercatore stesso. Il rigore metodologico associato a questo contesto consiste nella

capacità di riconoscere, gestire e rendere esplicita questa soggettività attraverso la pratica riflessiva continua.

Definiti questi aspetti metodologici iniziali ma fondamentali, si è proceduto all'analisi dei dati seguendo un processo articolato in sei fasi operative, descritte di seguito, riprese dal modello a sei fasi sviluppato da Braun e Clarke (2006; 2019). Questo modello non è un protocollo rigido e lineare, ma un'indicazione flessibile che struttura il processo analitico, assicurando che ogni fase venga eseguita con cura e attenzione.

Fase 1: Familiarizzazione con i dati

La prima fase è stata caratterizzata da una totale immersione nei dati raccolti, attraverso la lettura e la rilettura di tutto il materiale trascritto; in questo modo è stato possibile comprendere globalmente i contenuti emersi dalle interviste. Questo approccio ha permesso di formulare le prime impressioni e di individuare alcuni temi ricorrenti. Per esempio, in questa fase sono emerse alcune considerazioni preliminari:

- I test INVALSI sono percepiti come un obbligo più che come uno strumento utile nei contesti di apprendimento.
- Le informazioni sul feedback descrittivo dei test INVALSI fornito da alcuni insegnanti e studenti sono scarse.
- Il coinvolgimento degli studenti nelle prove INVALSI è descritto come debole, a causa della percezione che il feedback non abbia effetti sul profitto scolastico.
- I descrittori dei livelli sembrano utili per monitorare i progressi degli studenti dalla scuola primaria alla secondaria.
- Il valore che gli studenti attribuiscono alle prove dipende dalla percezione che i docenti hanno di esse.

Queste considerazioni iniziali non sono soltanto semplici osservazioni, ma costituiscono la base che ha guidato la successiva fase di codifica più sistematica, anticipando potenziali schemi di significato.

Fase 2: Generazione dei Codici Iniziali

In questa fase si è proceduto alla codifica, ossia all'etichettatura sistematica delle unità di analisi significative all'interno di ciascuna trascrizione. Come già indicato, è stato adottato un approccio induttivo nell'individuazione dei codici, costruiti direttamente a partire dal linguaggio e dai concetti espressi dai partecipanti.

Per verificare l'efficacia delle etichette prodotte, si è ricorso al criterio del *remove the data* (Clarke & Braun, 2013), per cui un codice etichetta è stato considerato valido se, anche in assenza del dato originale, ovvero della parte specifica della trascrizione cui si riferiva, riusciva a evocarne il significato essenziale.

Fase 3: Ricerca e generazione dei Temi Iniziali

Una volta generato l'elenco completo dei codici, i codici iniziali sono stati raggruppati nei primi temi possibili. In questa fase, le intuizioni emerse durante la familiarizzazione con i dati hanno svolto un ruolo fondamentale nell'organizzare i dati in modo più strutturato, man mano che i codici venivano aggregati in raggruppamenti tematici preliminari. Questi primi temi emersi sono stati considerati ancora provvisori e sono stati sottoposti a successiva revisione e raffinamento.

Fase 4: Revisione dei Temi

Il processo di revisione si è articolato su due livelli distinti:

1. Primo Livello: i temi provvisori sono stati confrontati con le etichette e gli estratti di testo che le compongono. L'obiettivo è stato verificare la coerenza interna di ciascun tema, per assicurare che i dati al suo interno presentino un *pattern* significativo.
2. Secondo Livello: i temi sono stati rivisti rispetto all'intero dataset. Questo passaggio ha garantito che la "mappa tematica" generata abbia colto con accuratezza la ricchezza e la complessità dei contenuti dei dati raccolti, senza tralasciare aspetti rilevanti né forzare interpretazioni.

In questa fase, sono state quindi verificate l'*omogeneità interna* (ogni tema è coerente e specifico) e l'*eterogeneità esterna* (ogni tema è chiaramente distinto dagli altri), come suggerito da Patton (2015).

Fase 5: Definizione e denominazione dei Temi

Dopo la fase di revisione, è stata elaborata una definizione per ciascun tema al fine di chiarirne il significato e i confini. Infine, a ogni tema è stato assegnato un nome definitivo.

Fase 6: Scrittura del report

L'ultima fase è stata la stesura del capitolo dedicato all'analisi dei dati della tesi che è stato costruito in modo da seguire i vari passaggi di codifica che hanno portato all'individuazione dei Temi generali e riporta, per ciascuna Categoria Tematica, il numero di etichette di codifica che vi sono confluite. Ogni etichetta si riferisce all'intervista di un singolo partecipante e può raggruppare uno o più estratti;

uno stesso partecipante, inoltre, può contribuire con più etichette alla stessa Categoria. Le numerosità indicano quindi occorrenze codificate e non il numero di partecipanti. Coerentemente con l'Analisi Tematica Riflessiva, le frequenze hanno una funzione descrittiva e non sono state usate come criterio per stabilire la rilevanza di un tema.

4.8 La dimensione etica della ricerca

La metodologia di ricerca adottata per la raccolta dei dati, pur essendo fondamentale per esplorare a fondo le percezioni e le esperienze individuali, comporta questioni etiche e legali complesse. La partecipazione di soggetti vulnerabili, in particolare i minorenni, ha richiesto un livello superiore di verifica e cautela per proteggerli, rispettare la loro autonomia (o quella di chi ne fa le veci) e ridurre al minimo eventuali danni psicologici o sociali. Per rispondere a queste esigenze, si è deciso di ottenere la validazione e l'approvazione del processo di ricerca dal Comitato Etico Universitario di Verona, non come semplice adempimento formale, ma come elemento essenziale del principio deontologico di tutto il progetto.

Tale scelta è coerente con le indicazioni della *British Educational Research Association*, secondo cui la ricerca educativa deve fondarsi su responsabilità esplicite nei confronti dei partecipanti, con particolare attenzione al consenso informato, alla trasparenza, al diritto al ritiro, alla tutela della riservatezza e alla prevenzione di possibili forme di danno (BERA, 2018). Nel caso di ricerche che coinvolgono bambini e giovani, Alderson e Morrow (2011) sottolineano inoltre la necessità di considerare l'etica non come una procedura preliminare alla raccolta dei dati, ma come una dimensione trasversale dell'intero processo di ricerca, che riguarda la progettazione, l'accesso al campo, la costruzione degli strumenti, la relazione con i partecipanti e l'uso successivo delle informazioni raccolte.

Il confronto con un organo terzo e indipendente ha consentito di perfezionare i protocolli di consenso informato (diversi per adulti e per minorenni/genitori), di garantire la riservatezza e l'anonimato dei dati e di verificare la conformità agli standard di ricerca riconosciuti a livello nazionale e internazionale. Questa azione è stata eseguita prima di avviare la selezione dei partecipanti e la richiesta di adesione.

Il parere positivo ottenuto in merito ai contenuti etici e legali della fase di raccolta dei dati ha consentito di condurre le interviste e di analizzare i dati, con la consapevolezza di operare in un contesto metodologico rigoroso, anche sul piano etico. In Appendice sono disponibili i testi del consenso informato per partecipanti maggiorenni e minorenni.

CAPITOLO 5: ANALISI E INTERPRETAZIONE DEI DATI

In questo capitolo verranno presentati i risultati emersi dalla raccolta dei dati, sia derivanti dalla revisione della letteratura, sia derivanti dall'analisi documentale, sia, infine, derivanti dalle interviste realizzate per approfondire le informazioni relative alle percezioni di docenti e studenti relativamente al fenomeno oggetto del presente studio.
Questa prima analisi e interpretazione supporteranno la discussione sugli esiti che costituirà il capitolo successivo, in cui si cercherà di rispondere alle domande di ricerca e di comprendere come i dati raccolti da diverse fonti possano produrre possibili interpretazioni del fenomeno.

5.1 Revisione sistematica della letteratura: risultati

Nel campo della ricerca educativa, la realizzazione di una revisione sistematica della letteratura consente di individuare, selezionare e sintetizzare in modo trasparente le evidenze disponibili su un determinato fenomeno, attraverso procedure esplicite, documentabili e potenzialmente replicabili. A differenza di una rassegna narrativa, la revisione sistematica si fonda sulla definizione preliminare della domanda di ricerca, sulla costruzione di una strategia di ricerca bibliografica, sull'applicazione di criteri di inclusione ed esclusione e sulla rendicontazione delle diverse fasi del processo di selezione degli studi (Petticrew & Roberts, 2006; Gough, Oliver, & Thomas, 2017).
Nel contesto della presente ricerca, la revisione sistematica della letteratura ha avuto l'obiettivo di individuare studi che avessero analizzato le percezioni e le esperienze di studenti e docenti in merito alla restituzione individuale degli esiti delle prove standardizzate, con particolare attenzione ai feedback descrittivi relativi ai contenuti oggetto di valutazione e alle loro possibili ricadute sui processi di insegnamento e apprendimento. La revisione è stata quindi orientata non soltanto alla ricostruzione dello stato dell'arte, ma anche all'identificazione di evidenze utili a comprendere in che modo l'esito descrittivo restituito dall'INVALSI possa essere interpretato, utilizzato e valorizzato nei contesti scolastici.
La ricerca è stata condotta su Scopus, piattaforma ritenuta adatta agli obiettivi della revisione, poiché offre un'ampia copertura multidisciplinare della letteratura scientifica *peer-reviewed* e include riviste internazionali rilevanti per gli ambiti educativo, valutativo e psicopedagogico. Inoltre, Scopus consente di costruire stringhe di ricerca complesse, combinando parole chiave, operatori booleani e filtri per area disciplinare, lingua, tipologia di documento e periodo di pubblicazione. La costruzione della strategia di ricerca è stata guidata dai principi di trasparenza e riproducibilità propri delle revisioni sistematiche, secondo cui le scelte relative alle fonti interrogate, ai termini utilizzati e ai criteri applicati devono essere documentate in modo chiaro (Booth, Sutton, & Papaioannou, 2016).

La piattaforma Scopus ha consentito di esportare in modo organizzato i dati bibliografici, che sono stati successivamente caricati nel software Rayyan per la fase di screening.
La revisione è stata condotta con il supporto della referente della biblioteca INVALSI, il cui contributo è stato rilevante nella costruzione e nella verifica della strategia di ricerca bibliografica, nella scelta delle banche dati e nella gestione delle fonti. Il coinvolgimento di bibliotecari o *information specialists* è considerato un elemento metodologicamente rilevante nelle revisioni sistematiche, poiché contribuisce a migliorare la qualità, la completezza e la riproducibilità delle strategie di ricerca (Rethlefsen et al., 2015). In particolare, figure esperte nella ricerca bibliografica possono ridurre il rischio di *bias* nella selezione delle fonti, migliorare la documentazione delle stringhe di ricerca e aumentare l'aderenza agli standard metodologici delle revisioni sistematiche (Koffel, 2015; Rethlefsen et al., 2015).

Metodologia e processo di revisione

A partire dalla domanda di ricerca è stato elaborato il PCC come raccomandato dal Joanna Briggs Institute, poiché consente, nelle revisioni orientate a mappare o sintetizzare le evidenze su fenomeni complessi, di specificare in modo trasparente la popolazione, il concetto e il contesto della revisione (Joanna Briggs Institute, 2024):
P (*Population*): *students*, *teachers*
C (*Concept*): *perceptions*, *experiences*, *impact of descriptive feedback on teaching and learning processes*
C (*Context*): *lower and upper secondary school*, *national standardized assessments, reading comprehension, individual descriptive feedback*

La stringa di ricerca da lanciare sulla piattaforma individuata (Scopus) è stata costruita in modo ampio rispetto al PCC, al fine di intercettare anche studi non esplicitamente centrati sulla comprensione della lettura o sul feedback descrittivo individuale, ma comunque rilevanti per comprendere il rapporto tra *assessment* standardizzato, percezioni degli studenti e l'uso formativo delle informazioni restituite.

(TITLE-ABS-KEY (feedback)) AND ((TITLE-ABS-KEY ("standardized test" OR "national assessment" OR "large-scale assessment")) AND ((TITLE-ABS-KEY ("perception*" OR "view*" OR "experience*")) AND (TITLE-ABS-KEY ("student*" OR teacher* OR "pupil*")))).

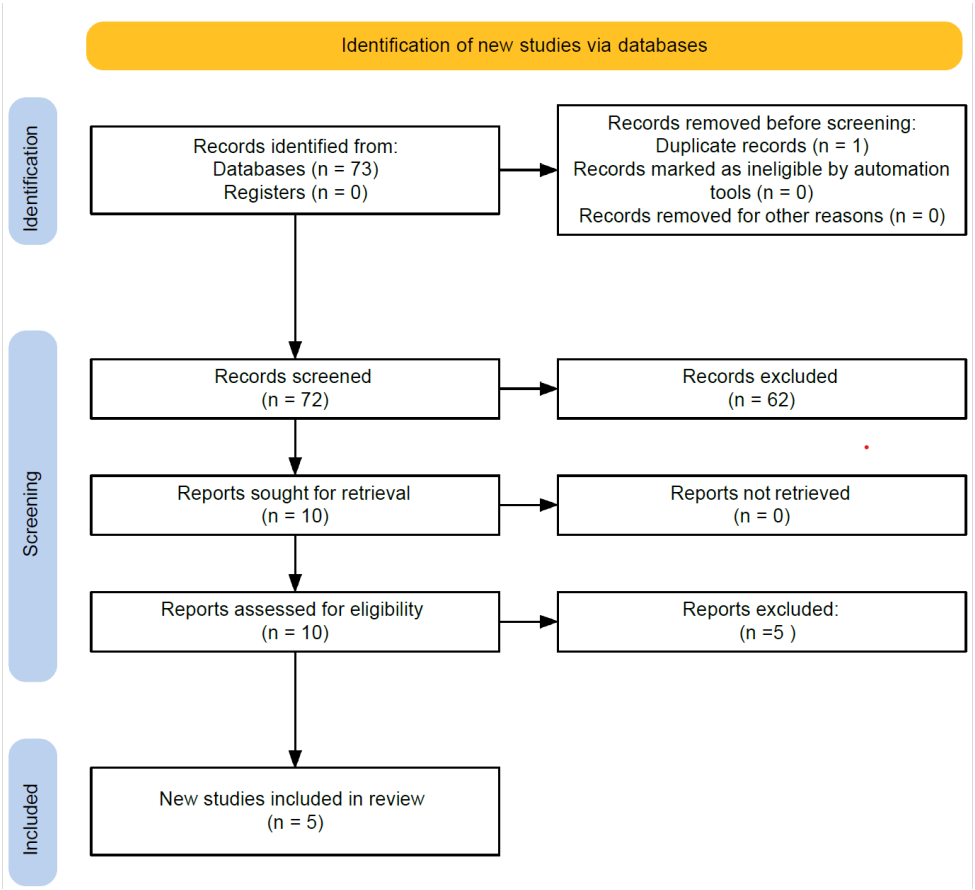
In questa revisione sono stati inclusi studi fino al 2026 che prendevano in considerazione studenti della scuola secondaria di primo e secondo grado e/o insegnanti coinvolti in processi di valutazione

standardizzata. La ricerca si è concentrata in particolare sulle percezioni, sulle esperienze e sull'utilizzo dell'esito descrittivo derivante dalle valutazioni standardizzate, con un focus specifico sulla comprensione del feedback, sulla sua utilità percepita e sui possibili effetti sui processi di insegnamento e apprendimento. Gli studi dovevano riferirsi a contesti di valutazione standardizzati a livello nazionale o a rilevazioni internazionali su larga scala, in particolare nell'ambito della comprensione della lettura, e prevedere la restituzione di feedback ai partecipanti. Sono stati presi in considerazione studi qualitativi, quantitativi e con approccio misto, studi di valutazione e report di ricerca che contengono evidenze empiriche rilevanti per gli obiettivi della revisione. Il criterio relativo alla comprensione della lettura e al feedback descrittivo individuale è stato considerato prioritario; tuttavia, data la scarsità di studi pienamente centrati su questo oggetto, sono stati inclusi anche contributi parzialmente pertinenti, utili a ricostruire dimensioni connesse alla percezione dell'*assessment*, alla motivazione, all'uso del feedback e al valore attribuito alla valutazione standardizzata.

La revisione e la presentazione dei risultati sono state condotte secondo i principi del PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), che rappresenta uno standard internazionale per la redazione delle revisioni sistematiche e fornisce una checklist e un diagramma di flusso utili a documentare in modo trasparente le fasi di identificazione, screening, eleggibilità e inclusione degli studi. L'adozione del PRISMA consente di rendere esplicite le decisioni metodologiche, di aumentare la tracciabilità del processo e di migliorare la replicabilità della revisione (Page et al., 2021).

Gli studi individuati sulla piattaforma SCOPUS sono stati esportati e successivamente caricati nel software Rayyan, utilizzato per supportare la fase di screening. Rayyan è una piattaforma sviluppata specificamente per facilitare la gestione delle revisioni sistematiche, in particolare nelle fasi di importazione delle referenze, individuazione dei duplicati, screening dei titoli e degli abstract, classificazione degli studi come inclusi, esclusi o incerti e gestione collaborativa delle decisioni di selezione (Ouzzani et al., 2016). Il software consente inoltre di assegnare etichette, motivazioni di esclusione e note ai singoli record, contribuendo a rendere più ordinato, documentabile e trasparente il processo decisionale della *review* (Ouzzani et al., 2016).

Di seguito è riportato il PRISMA della presente revisione sistematica.



La procedura di screening è stata articolata in più fasi. In primo luogo, sono stati verificati eventuali duplicati e esaminati i titoli e gli abstract in base ai criteri di inclusione precedentemente definiti. Gli studi potenzialmente pertinenti sono stati mantenuti per la fase successiva di lettura del full text. In questa seconda fase, ciascun articolo è stato valutato in coerenza con la popolazione, il concetto e il contesto della review, prestando particolare attenzione alla presenza di feedback descrittivo individuale, alla tipologia di *assessment* e alla pertinenza delle evidenze riportate rispetto ai processi di insegnamento e apprendimento.

La ricerca, aggiornata maggio 2026, ha restituito 73 record. Dopo la rimozione di un duplicato, 72 record sono stati esaminati sulla base di titolo e abstract e 62 sono stati esclusi. I 10 contributi rimanenti sono stati letti integralmente: 5 sono stati esclusi perché non riguardavano la scuola secondaria di primo o di secondo grado oppure non si riferivano a prove o rilevazioni standardizzate, mentre 5 sono stati inclusi nella revisione. Lo screening è stato condotto dalla sola ricercatrice e su una sola banca dati: entrambi gli aspetti limitano la completezza della revisione e sono ripresi nel paragrafo 7.1.

Risultati

I cinque studi inclusi analizzano il rapporto tra *assessment*, atteggiamenti degli studenti, motivazione e apprendimento nei contesti della scuola secondaria di primo e secondo grado (Appendice 1). Le ricerche considerate adottano approcci metodologici diversi. Alcuni studi esaminano interventi educativi supportati da tecnologie digitali, come nel caso di Barana, Fissore e Marchisio (2020) e di Arner, McCarthy e McNamara (2021). Altri utilizzano analisi quantitative su larga scala, come Chu, Babenko, Cui e Leighton (2014) e Happ e Förster (2018). Lo studio di Ahmed, ALZahrani e ALZahrani (2026) adotta invece un disegno mixed methods per approfondire le percezioni degli studenti rispetto a una valutazione nazionale standardizzata.

Va precisato fin da subito che nessuno dei cinque studi ha come oggetto specifico la restituzione descrittiva individuale degli esiti di una prova standardizzata, che è il fenomeno indagato in questa ricerca. Lo studio di Ahmed et al. (2026) è il più vicino: riguarda una valutazione nazionale standardizzata e documenta, dal punto di vista degli studenti, l'assenza di una restituzione individuale e la richiesta di report personalizzati. Chu et al. (2014) e Happ e Förster (2018) riguardano prove standardizzate, ma trattano il feedback in modo marginale: il primo studia le percezioni dell'ambiente di apprendimento e delle pratiche valutative in relazione ai risultati PISA, il secondo considera l'interesse a ricevere un feedback come variabile motivazionale. Barana et al. (2020) e Arner et al. (2021) riguardano forme di feedback automatico in attività costruite a partire da prove standardizzate o simili a esse, non la restituzione degli esiti di una rilevazione nazionale; lo studio di Arner et al. (2021), inoltre, è stato condotto con studenti universitari, ma è stato mantenuto, in ragione del carattere esplorativo della revisione, perché riguarda materiali e obiettivi della scuola secondaria e la preparazione a prove standardizzate di comprensione della lettura.

I cinque studi offrono quindi elementi di contesto sul rapporto tra valutazione standardizzata, percezioni, motivazione e uso del feedback, ma non evidenze dirette sul fenomeno studiato. Questa scarsità costituisce a sua volta un risultato della revisione, che conferma il carattere esplorativo della presente ricerca. La colonna «Tipo di feedback» dell'Appendice 1 sintetizza, per ciascuno studio, il modo in cui il feedback è trattato.

I campioni variano da circa cinquanta partecipanti a oltre mille studenti. Questa differenza consente di osservare il fenomeno sia in contesti più controllati, legati a specifici interventi, sia in contesti più ampi di valutazione standardizzata o nazionale.

Un primo tema emerso riguarda il ruolo delle percezioni e degli atteggiamenti degli studenti nei confronti delle pratiche valutative. Gli studi di Chu et al. (2014) e Ahmed et al. (2026) mostrano che il modo in cui gli studenti interpretano la valutazione e il significato che le attribuiscono incidono sul loro coinvolgimento e, in parte, sui risultati ottenuti.

Chu et al. (2014) rilevano una relazione positiva tra la percezione dell'ambiente di apprendimento e le prestazioni nelle prove standardizzate. Gli studenti che descrivono in modo più favorevole il proprio ambiente scolastico tendono a conseguire risultati migliori. Questa relazione risulta più evidente rispetto a quella osservata per le percezioni espresse dai dirigenti scolastici. Il dato suggerisce che l'esperienza diretta degli studenti costituisca un elemento importante per comprendere gli esiti dell'apprendimento.

Un risultato simile emerge nello studio di Ahmed et al. (2026), nel quale gli studenti manifestano atteggiamenti complessivamente negativi nei confronti del test nazionale *National Assessment for Learning Standards* (NAFS). In particolare, le dimensioni cognitive legate all'utilità del test ottengono i punteggi più bassi tra tutte le dimensioni rilevate dal questionario. Gli studenti dichiarano di non percepire un legame chiaro tra i risultati della prova e il miglioramento del proprio apprendimento. Inoltre, non riferiscono di ricevere informazioni utili per comprendere i propri punti di forza e di debolezza.

Queste evidenze indicano che il valore attribuito all'*assessment* rappresenta una componente importante nei processi di apprendimento. Quando la valutazione viene vissuta come poco significativa o distante dai bisogni degli studenti, il suo potenziale formativo tende a indebolirsi.

Un secondo tema riguarda il contributo delle tecnologie digitali ai processi valutativi. Barana et al. (2020) analizzano l'utilizzo di un sistema di valutazione automatizzata integrato con percorsi di formazione per insegnanti. I risultati mostrano che la disponibilità di informazioni immediate sulle prestazioni degli studenti favorisce pratiche di insegnamento adattivo. I docenti possono modificare le attività didattiche sulla base delle difficoltà emerse e intervenire in modo più mirato durante il percorso di apprendimento. Lo strumento viene quindi percepito come utile per monitorare i progressi degli studenti e per sostenere decisioni didattiche più personalizzate.

Anche Arner et al. (2021) prendono in esame un ambiente tecnologico di supporto all'apprendimento. In questo caso si tratta di un *Intelligent Tutoring System* orientato allo sviluppo delle strategie di comprensione della lettura. Gli studenti mostrano una buona accettazione dello strumento, ma gli effetti sulle prestazioni cognitive risultano limitati. Questo dato suggerisce che la presenza della tecnologia, da sola, non sia sufficiente a garantire miglioramenti significativi nell'apprendimento.

Considerati insieme, questi due studi mostrano che le tecnologie valutative possono offrire opportunità interessanti, ma la loro efficacia dipende dal modo in cui vengono inserite nelle pratiche didattiche. Non è lo strumento tecnologico in sé a produrre cambiamento, quanto la possibilità di trasformare le informazioni raccolte in azioni educative utili.

Un terzo tema riguarda la motivazione degli studenti. Happ e Förster (2018) evidenziano una relazione significativa tra motivazione verso il test e risultati ottenuti. Gli studenti maggiormente

motivati tendono a conseguire prestazioni più elevate rispetto a quelli con livelli inferiori di coinvolgimento.

Lo studio di Ahmed et al. (2026) rafforza questa evidenza da una prospettiva diversa. La mancanza di valore percepito della valutazione, l'assenza di feedback personalizzati e la scarsa rilevanza attribuita ai risultati contribuiscono a ridurre la partecipazione degli studenti alle attività di *assessment*. Molti studenti dichiarano di affrontare il test con scarso interesse, anche perché i risultati non sembrano avere conseguenze concrete sul loro percorso scolastico. Gli intervistati riferiscono inoltre che l'esperienza valutativa non ha modificato le loro strategie di studio né ha aumentato la loro consapevolezza del proprio apprendimento.

La motivazione appare quindi non solo come un possibile effetto dell'*assessment*, ma anche come una condizione necessaria affinché la valutazione possa contribuire realmente all'apprendimento.

Per quanto riguarda gli effetti dell'*assessment* sulla performance, le evidenze sono meno uniformi. Chu et al. (2014) riscontrano una relazione positiva tra le percezioni favorevoli dell'ambiente di apprendimento e i risultati nei test standardizzati. Happ e Förster (2018) mostrano un'associazione significativa tra motivazione e prestazioni. Arner et al. (2021), invece, rilevano miglioramenti limitati nelle competenze di comprensione della lettura dopo l'intervento tecnologico proposto.

Anche Ahmed et al. (2026) suggeriscono che bassi livelli di performance non devono essere letti soltanto come espressione di competenze insufficienti. Gli autori ipotizzano che atteggiamenti negativi nei confronti del NAFS possano aver ridotto l'impegno degli studenti durante la prova. In questo senso, la performance può riflettere non solo ciò che gli studenti sanno o sanno fare, ma anche il livello di coinvolgimento con cui affrontano la valutazione.

Gli effetti diretti dell'*assessment* sulle prestazioni appaiono quindi meno stabili rispetto a quelli osservati sulle dimensioni motivazionali e percettive.

Diversi studi richiamano inoltre l'importanza dei fattori contestuali e individuali. Chu et al. (2014) identificano effetti associati allo status socioeconomico e al genere. Happ e Förster (2018) sottolineano il ruolo delle differenze individuali nei livelli di motivazione. Ahmed et al. (2026) non osservano differenze significative tra studenti maschi e femmine, ma rilevano atteggiamenti più positivi nei confronti del NAFS tra gli studenti delle scuole private rispetto a quelli delle scuole pubbliche. Gli autori collegano questa differenza a livelli diversi di preparazione, supporto scolastico e valorizzazione dell'esame all'interno delle istituzioni educative.

Questi risultati indicano che il possibile effetto dell'*assessmen*t non può essere spiegato solo attraverso le caratteristiche della prova. Il contesto educativo in cui la valutazione viene proposta contribuisce in modo rilevante a definire l'esperienza degli studenti.

Alcune criticità emerse dagli studi consentono di individuare anche possibili limiti delle pratiche valutative analizzate. Un primo elemento riguarda la difficoltà degli studenti nel riconoscere l'utilità dell'*assessment* per il proprio percorso di apprendimento. Ahmed et al. (2026) mostrano che molti studenti non percepiscono il NAFS come uno strumento capace di sostenere il miglioramento dei risultati scolastici o di fornire indicazioni concrete sul proprio livello di preparazione. Le dimensioni legate all'utilità del test registrano infatti i punteggi più bassi dell'intera indagine. Gli studenti sembrano attribuire alla prova una funzione prevalentemente istituzionale, più che formativa.

Una seconda criticità riguarda la restituzione dei risultati. Nel caso del NAFS, lo studio di Ahmed et al. (2026) non documenta una restituzione individuale agli studenti sotto forma di punteggio dettagliato o feedback descrittivo. Gli studenti percepiscono i risultati della prova come scollegati dal proprio percorso scolastico e non utili per comprendere punti di forza, debolezze o strategie di miglioramento. In particolare, emerge una richiesta diffusa di report individualizzati che permettano di orientare lo studio e di capire quali aspetti richiedano maggiore attenzione. L'assenza di informazioni operative limita la possibilità che la valutazione sostenga processi di autoregolazione dell'apprendimento.

Un ulteriore aspetto riguarda il rapporto tra partecipazione alla prova e motivazione. Quando le valutazioni non producono conseguenze percepite come rilevanti, gli studenti tendono a investire meno impegno nella preparazione e nello svolgimento delle attività. Questo aspetto pone anche un problema di interpretazione dei risultati, perché le prestazioni potrebbero dipendere non soltanto dalle competenze possedute, ma anche dal grado di coinvolgimento degli studenti. Tale aspetto emerge in particolare negli studi di Happ e Förster (2018) e Ahmed et al. (2026).

Un'altra criticità riguarda il legame tra *assessment* e miglioramento degli apprendimenti. Gli studi inclusi mostrano che l'esistenza di una valutazione, anche quando supportata da tecnologie avanzate, non determina automaticamente un progresso nelle prestazioni. Gli esiti osservati da Arner et al. (2021) confermano questa cautela, poiché l'intervento tecnologico produce effetti limitati sulle competenze considerate.

Infine, alcuni studi sottolineano l'importanza delle condizioni organizzative in cui la valutazione viene condotta. Ahmed et al. (2026) riportano problemi legati alla pianificazione temporale delle prove, alla chiarezza delle istruzioni, alla durata dei test e, in alcuni casi, alle infrastrutture tecnologiche impiegate durante la somministrazione. Questi aspetti possono influenzare la percezione di equità della valutazione e il modo in cui gli studenti vivono l'intera esperienza.

Limiti emersi dagli studi inclusi

L'analisi degli studi selezionati evidenzia alcune criticità metodologiche da considerare nell'interpretazione dei risultati. La prima riguarda la varietà dei disegni di ricerca. Solo una parte

degli studi prevede l'implementazione di interventi educativi, mentre altri si basano su dati osservazionali o correlazionali. Questo rende difficile stabilire relazioni causali tra l'*assessment* e l'apprendimento. Gli studi di Chu et al. (2014) e Happ e Förster (2018), ad esempio, utilizzano grandi dataset e analisi statistiche avanzate, ma la loro natura non sperimentale consente di individuare associazioni tra variabili senza dimostrare che siano le pratiche valutative a produrre direttamente i risultati osservati.

Un secondo limite riguarda l'ampiezza e la rappresentatività dei campioni. Accanto a studi condotti su numeri elevati di studenti, sono presenti ricerche con campioni ridotti, come nel caso di Arner et al. (2021) e Barana et al. (2020). Questa condizione limita la possibilità di generalizzare i risultati ad altri contesti scolastici. La presenza di gruppi ristretti aumenta anche il rischio che caratteristiche specifiche del contesto abbiano inciso sugli esiti rilevati.

Un ulteriore elemento riguarda il ricorso frequente a questionari e misure self report. Gli studenti vengono spesso invitati a descrivere percezioni, motivazioni o valori attribuiti alla valutazione. Nello studio di Barana et al. (2020), inoltre, molte informazioni provengono dalle osservazioni e dalle percezioni degli insegnanti. Questi dati sono utili per comprendere l'esperienza dei partecipanti, ma possono essere influenzati dalla desiderabilità sociale, da ricordi imprecisi o da interpretazioni soggettive.

Alcuni studi presentano anche limiti legati alla durata del periodo di osservazione. Gli effetti degli interventi vengono analizzati in archi temporali relativamente brevi. Di conseguenza, non è sempre possibile capire se i cambiamenti osservati siano stabili o se rappresentino reazioni iniziali legate alla novità dell'esperienza proposta. Questo limite riguarda soprattutto gli studi di intervento, come Barana et al. (2020) e Arner et al. (2021).

Lo studio di Ahmed et al. (2026) evidenzia ulteriori limiti riconosciuti dagli stessi autori. La raccolta dei dati riguarda un'area geografica circoscritta dell'Arabia Saudita, non sono presenti misure longitudinali e non viene approfondito il ruolo di fattori culturali e socioeconomici che potrebbero influenzare gli atteggiamenti degli studenti verso le valutazioni nazionali.

Nel loro insieme, gli studi suggeriscono che l'*assessment* può assumere valore educativo quando gli studenti ne comprendono la funzione e percepiscono una chiara connessione tra valutazione e apprendimento. Le pratiche valutative risultano più promettenti quando sono integrate nei processi di insegnamento, producono informazioni utilizzabili e aiutano studenti e docenti a orientare le azioni successive.

Quando invece la valutazione viene percepita come poco utile o scollegata dal percorso formativo, emergono una riduzione della motivazione, uno scarso coinvolgimento e ricadute educative limitate. Questo elemento risulta particolarmente evidente nello studio di Ahmed et al. (2026), ma si collega

anche ai risultati di Happ e Förster (2018), che evidenziano il ruolo della motivazione nella performance.

Il contributo principale dell'*assessment* scolastico non sembra quindi risiedere nella sola misurazione delle competenze. Piuttosto, la valutazione può sostenere l'apprendimento quando favorisce la motivazione, la consapevolezza e la possibilità di intervenire sui processi di studio e di insegnamento.

5.2 L'analisi documentale: risultati

Nell'ambito dello studio di caso trattato in questa ricerca, incentrato sul ruolo dell'INVALSI e, in particolare, sui certificati di restituzione individuale degli esiti espressi in termini di livelli descrittivi, l'analisi documentale riveste un ruolo fondamentale. Questa azione consente di chiarire meglio il fenomeno dell'introduzione del sistema di restituzione degli esiti individuali nella valutazione nazionale esterna, considerando gli elementi di vincolo derivanti dagli obblighi di legge, sia attraverso l'interpretazione operativa di tali vincoli nei documenti istituzionali e nella visione pluriennale delle attività dell'Istituto, sia attraverso le modalità di utilizzo legate agli strumenti di fruizione pubblica, nei quali tali esiti assumono una loro posizione. I documenti selezionati per questa analisi documentale sono i decreti legislativi che hanno regolamentato l'istituzione e alcuni compiti dell'INVALSI, gli atti istitutivi dell'ente, in particolare lo Statuto e il Documento di Visione Decennale (DVS), e gli strumenti di diffusione pubblica di tali esiti, in particolare le piattaforme digitali ministeriali Scuola in Chiaro e la piattaforma UNICA nella sezione dell'e-portfolio dello studente, oltre al portale INVALSIopen.

I documenti sono stati selezionati secondo tre criteri: essere fonti ufficiali, normative o istituzionali, pubblicamente accessibili; disciplinare o descrivere la valutazione nazionale degli apprendimenti e, in particolare, la restituzione dei suoi esiti; appartenere a uno dei tre livelli che concorrono a definire il fenomeno studiato. Il primo è il livello normativo, costituito dai decreti che istituiscono e regolano il sistema di valutazione; il secondo è il livello istituzionale dell'ente, rappresentato dallo Statuto e dal documento di visione; il terzo è il livello della comunicazione pubblica degli esiti, costituito dalle piattaforme e dal portale attraverso cui gli esiti raggiungono scuole, studenti e famiglie.

Per l'interpretazione, ogni documento è stato letto con la stessa griglia di domande, che corrisponde alle colonne della tabella di sintesi riportata alla fine del paragrafo: quale funzione attribuisce alla valutazione, quale ruolo assegna all'INVALSI, quali implicazioni ha per la valutazione degli apprendimenti e quale spazio riserva alla restituzione degli esiti e ai livelli descrittivi. Coerentemente con quanto indicato nel paragrafo 4.3.1, i documenti sono stati trattati come testi portatori di finalità e linguaggi istituzionali, distinguendo ciò che vi è effettivamente dichiarato da ciò che deriva dalla conoscenza interna della ricercatrice.

I documenti normativi relativi all'INVALSI, istitutivi e regolativi, delineano il quadro complesso entro cui si sono orientate e costruite le azioni dell'ente. L'analisi dei principali riferimenti normativi consente di cogliere come il tema della valutazione degli apprendimenti e della loro restituzione sia progressivamente passato da una dimensione prettamente tecnica e sperimentale a un dispositivo strutturale di governo del sistema educativo.

Un primo riferimento normativo fondamentale per comprendere la configurazione del ruolo dell'INVALSI è il Decreto Legislativo 19 febbraio 2004, n. 286, con il quale è stato istituito il Servizio nazionale di valutazione del sistema educativo di istruzione e formazione. In tale provvedimento, l'INVALSI viene designato come ente di ricerca tecnico-scientifico deputato alla progettazione e alla realizzazione di verifiche periodiche e sistematiche sugli apprendimenti degli studenti, con l'obiettivo di produrre dati affidabili e comparabili a livello nazionale. In questo decreto all'Istituto viene anche affidato l'ulteriore compito di valutare l'efficienza e l'efficacia del sistema nel suo complesso, mediante la definizione di indicatori e la predisposizione di rapporti tecnici e di ricerca a supporto delle policy sull'istruzione. Questo documento normativo iniziale definisce la valutazione nazionale come uno strumento conoscitivo, necessario per descrivere e interpretare il funzionamento del sistema scolastico con l'obiettivo di individuare le condizioni che ostacolassero o promuovessero l'apprendimento delle competenze di base.

Da questa prima impostazione si è passati con il D.P.R. 28 marzo 2013, n. 80 da una prima fase di monitoraggio del sistema a una nuova fase che sviluppava e ampliava il concetto di sistema nazionale di valutazione, passando dalla sola valutazione degli apprendimenti a un sistema articolato, denominato appunto Sistema Nazionale di Valutazione. All'interno di questo sistema, l'INVALSI non è più soltanto responsabile della rilevazione degli apprendimenti, ma assume anche un ruolo di coordinamento nella costruzione e nell'organizzazione dell'intero sistema di valutazione. L'Istituto, quindi, è contemporaneamente responsabile della realizzazione della valutazione degli apprendimenti attraverso la somministrazione di prove standardizzate, i cui esiti, oltre a essere restituiti in maniera capillare alle singole scuole, convergono all'interno del Rapporto di Autovalutazione (RAV); allo stesso tempo ha il mandato di definire gli indicatori utilizzati dalle scuole nei processi di autovalutazione (ovvero di strutturare il RAV); e di fornire dati a supporto delle attività di valutazione esterna a livello di sistema attraverso la produzione di rapporti di ricerca annuali. Il ruolo così assegnato all'INVALSI è quello di collegamento tra la produzione dei dati e il loro utilizzo nei processi di analisi e di miglioramento all'interno dei RAV, in una visione della valutazione come pratica continua e integrata nella vita delle scuole.

All'interno del successivo Decreto Legislativo del 13 aprile 2017, n. 62, si presenta il tema della valutazione degli studenti e rafforzando in questo modo la presenza delle prove INVALSI all'interno

dei percorsi scolastici. Il decreto stabilisce l'obbligatorietà delle prove nazionali standardizzate di Italiano, Matematica e Inglese in momenti specifici del primo e del secondo ciclo e ne definisce le modalità di somministrazione (prove lineari alla primaria e *computer based* negli altri gradi scolastici interessati), affidando all'INVALSI il compito di garantirne la solidità tecnico-scientifica degli strumenti e la comparabilità nel tempo degli esiti delle prove standardizzate nazionali. È in questo documento normativo che viene introdotta la distinzione tra la funzione di certificazione individuale degli apprendimenti per le classi dell'ultimo anno della scuola secondaria di primo e secondo grado e quella di monitoraggio del sistema che rimane stabile rispetto a quanto già indicato nel decreto precedente. In merito alla restituzione degli esiti individuali, si chiarisce che i risultati delle prove standardizzate non influenzano direttamente il voto finale o l'esito degli esami di Stato ma sono obbligatorie per poter sostenere tali esami finali. Questo nuovo assetto delle prove nazionali standardizzate all'interno del sistema scolastico apre a un orientamento della lettura degli esiti più integrato con la comprensione dei processi di apprendimento dei singoli studenti. Ai singoli studenti di questi due gradi scolastici viene quindi restituito il cosiddetto "certificato", in cui gli esiti individuali sono espressi in termini di livelli descrittivi. Il decreto, pur non specificando in modo esplicito la struttura di questi livelli descrittivi, crea le premesse per lo sviluppo di forme di comunicazione degli esiti che siano accessibili e rilevanti per i vari destinatari. La necessità di rendere i dati interpretabili da studenti, famiglie e scuole promuove infatti un passaggio da una restituzione puramente numerica a modalità che integrano indicatori quantitativi e descrizioni qualitative. In questo senso, l'introduzione e la diffusione dei livelli descrittivi possono essere lette come un'evoluzione che segue l'idea alla base del decreto, poiché rispondono al bisogno di trasformare le informazioni statistiche in conoscenze utilizzabili nei contesti educativi.

Un dato interessante che emerge dall'analisi della normativa, e che va nella stessa direzione e quindi ne rafforza l'interpretazione, è l'attenzione, soprattutto nelle scuole primarie, alla restituzione agli studenti della certificazione delle competenze, anch'essa basata su descrizioni dei livelli di acquisizione. Questo elemento concorre a definire un quadro comune di riferimento per quanto viene rilasciato dalla scuola agli studenti e, di conseguenza, anche alle famiglie, in merito agli esiti degli apprendimenti, e contribuisce a diffondere una cultura della valutazione, integrando, in questo contesto, anche la restituzione dei risultati delle prove standardizzate.

Passando ora all'analisi dei documenti istitutivi dell'ente, lo Statuto e i documenti di programmazione strategica a medio e lungo termine, tra cui il cosiddetto documento di visione decennale che ne definisce l'evoluzione, rivestono un ruolo particolarmente importante. Mentre i decreti legislativi stabiliscono vincoli e funzioni all'interno del sistema educativo, questi altri documenti chiariscono

l'identità epistemologica e operativa dell'Istituto di ricerca, offrendo indicazioni importanti anche su come vengono raccolti e presentati i dati sugli apprendimenti.

Lo Statuto dell'INVALSI definisce l'Istituto come un ente di ricerca autonomo dal punto di vista scientifico, organizzativo e amministrativo, i cui compiti istituzionali sono relativi al monitoraggio del funzionamento del sistema educativo e alla predisposizione di strumenti per la valutazione degli apprendimenti. All'interno di questo contesto emergono chiaramente due aspetti distinti: da un lato la natura scientifica, basata sulla produzione di informazioni accurate e precise, fondate su metodologie solide dal punto di vista tecnico e scientifico, e dall'altro la natura istituzionale, connessa al ruolo che la produzione di queste informazioni ha nei processi di monitoraggio del sistema di istruzione e formazione. Per quanto riguarda in particolare la valutazione degli apprendimenti, lo Statuto sottolinea le responsabilità dell'INVALSI nella progettazione di prove standardizzate per il contesto scolastico, necessariamente accompagnate da documentazione tecnica che ne spieghi le caratteristiche di costrutto e psicometriche, ovvero i Quadri di Riferimento (QdR), e nella costruzione di indicatori che consentano di confrontare i risultati su larga scala. Se quindi la raccolta dei dati viene sostenuta attraverso strumenti rigorosamente basati su modelli statistici e psicometrici, finalizzati a garantire validità, affidabilità e comparabilità degli esiti, è evidente la necessità di individuare strategie e modalità che consentano una lettura di tali risultati accessibile e utilizzabile da diversi *stakeholder*. In quest'ottica, la produzione dei dati non si limita, quindi, solo alla fase di misurazione, ma include anche un'attenzione alla comunicazione dei dati a vari livelli, come quello delle istituzioni, della scuola e delle singole persone.

Questa visione è ulteriormente rafforzata nei documenti di pianificazione strategica, come il Documento di Visione Decennale (DVS) che orienta le attività dell'INVALSI nel lungo periodo. In questo documento, la valutazione degli apprendimenti è considerata parte di un sistema informativo complesso, in cui la qualità dei dati dipende strettamente dalla loro chiarezza e utilizzabilità, soprattutto nel contesto scolastico. La restituzione dei dati alle singole scuole impone una crescente esigenza di elaborare modelli di comunicazione delle informazioni derivate dalle rilevazioni nazionali, funzionali alla costruzione di piani di miglioramento. Si nota, in particolare, una crescente attenzione alla presentazione dei risultati, non solo per diffonderli, ma anche per sviluppare formati comunicativi adatti ai vari destinatari. Di conseguenza, la progressiva introduzione di modalità di restituzione basate su livelli descrittivi può essere letta come l'avvicinarsi al superamento di una logica esclusivamente numerica degli apprendimenti. Nei documenti istitutivi e di visione strategica dell'INVALSI si sottolinea, infatti, l'esigenza di rendere i risultati delle prove più comprensibili per studenti, famiglie e insegnanti, proprio per contribuire alla costruzione di una cultura dell'uso consapevole dei dati. Sicuramente, accompagnare la restituzione degli esiti a un modello di rilascio

di descrizioni dei compiti superati o meno consente di collegare la complessità delle analisi statistiche a linguaggi più accessibili.

In questa prospettiva, l'analisi congiunta dei documenti sulle funzioni dell'istituto permette di comprendere che l'evoluzione dell'INVALSI non riguarda solo l'ampliamento delle sue funzioni, ma anche un cambiamento nel modo in cui la valutazione viene concepita e comunicata. La crescente importanza attribuita alla restituzione, in particolare alla sua dimensione qualitativa, indica una tendenza verso metodi di valutazione più orientati alla comprensione dei risultati, pur rimanendo ancorati a un approccio metodologico fortemente standardizzato.

Dopo aver analizzato alcune componenti dei documenti istitutivi e di indirizzo dell'INVALSI, passiamo all'analisi di alcuni strumenti finalizzati alla diffusione e alla fruizione pubblica dei dati che svolgono un ruolo cruciale nell'orientare l'azione di restituzione degli esiti delle prove nazionali sugli apprendimenti scolastici.

Il portale Scuola in Chiaro è stato creato con l'obiettivo di mettere a disposizione della comunità pubblica una serie di informazioni sulle scuole derivate dai RAV, compresi gli esiti delle prove INVALSI, e di inserirsi nel contesto delle politiche di trasparenza amministrativa. Una nota interessante è relativa al fatto che la pubblicazione di tali esiti non è obbligatoria, ma è a discrezione di chi è proprietario di tali dati, ovvero della scuola stessa, nella persona della sua dirigenza scolastica. I dati raccolti dalle prove INVALSI vengono rielaborati per contribuire a costruire un'immagine della qualità della scuola, e inseriti all'interno del RAV comunicano l'andamento degli apprendimenti scolastici relativi alle prove oggetto di rilevazione in termini di punteggi medi messi a confronto con i punteggi medi nazionali. Sicuramente, nell'ambito di questa piattaforma, lo spazio dedicato agli aspetti di valutazione standardizzata degli apprendimenti è di tipo sintetico, poiché l'attenzione è maggiormente incentrata sulla pubblicazione di informazioni di contesto e di assetto organizzativo, con l'obiettivo di condurre un'analisi di autovalutazione improntata a una visione d'insieme del contesto scolastico.

All'interno, invece della Piattaforma Unica, sviluppata nell'ambito delle riforme legate al Piano Nazionale di Ripresa e Resilienza (PNRR), sono raccolti in un unico ambiente diversi strumenti e servizi per la scuola per accompagnare il percorso scolastico e orientativo a cui possono accedere studenti, famiglie e docenti.

In questo ambiente digitale troviamo l'e-portfolio, ovvero uno spazio strutturato in cui il percorso formativo dello studente è presentato in modo progressivo, in una prospettiva evolutiva. Nell'ambito della pedagogia moderna (Pellerey, 2004; Barrett, 2007), il portfolio nasce con l'idea di costituire una raccolta dei lavori dello studente, creata per documentare il suo percorso di apprendimento, i risultati ottenuti e i progressi compiuti nel tempo. In questo senso, lo strumento assume una funzione

riflessiva, poiché supporta lo studente nell'interpretazione delle proprie esperienze e nell'attribuzione di un significato ad esse, ponendo l'accento sulla costruzione dell'identità personale e formativa. Nella filosofia dell'e-portfolio, in particolare, assume centralità la funzione orientativa di questo strumento, per cui le informazioni raccolte contribuiscono a delineare interessi, capacità e aspirazioni, e il percorso scolastico è collegato anche alle scelte future in ambito formativo e professionale.

All'interno di questo ambiente digitale, i risultati delle prove INVALSI vengono inseriti, per gli studenti dell'ultimo anno della scuola secondaria di secondo grado, tra i dati di risultato, diventando parte integrante della documentazione del loro percorso formativo. Il loro valore in questo contesto non è puramente certificativo, ma assume un significato se collegato agli altri elementi, come le esperienze di apprendimento e le riflessioni personali, attraverso cui lo studente interpreta ciò che ha appreso, riconosce le difficoltà e individua le strategie adottate. Tutte queste informazioni, messe assieme, contribuiscono a rendere più completa la lettura del percorso e a sostenere la funzione orientativa dell'e-Portfolio.

È proprio in questo passaggio che assume particolare rilevanza la modalità di restituzione basata su livelli descrittivi, poiché rende gli esiti più chiari e significativi. Mentre, sulla piattaforma Scuola in Chiaro, i risultati delle prove nazionali di valutazione degli apprendimenti servono a costruire una visione sulla scuola, nella Piattaforma Unica questi valori assumono un ruolo diverso, diventando parte integrante della descrizione del percorso scolastico e formativo dello studente. È prevalentemente in questi due contesti di comunicazione pubblica dei risultati e degli esiti degli apprendimenti che si definisce la duplice finalità a cui l'INVALSI è chiamato a rispondere e che segna un'evoluzione significativa nel ruolo della valutazione esterna, che non è più solo uno strumento di analisi del sistema, ma diventa un elemento che contribuisce attivamente alla costruzione del percorso educativo del singolo studente.

Un ulteriore elemento rilevante per l'analisi documentale è rappresentato dal portale INVALSIopen, concepito come sito ufficiale dell'Area Prove nazionali dell'INVALSI, con l'obiettivo di comunicare in modo diretto, accessibile, esaustivo e trasparente il lavoro di ricerca sulla valutazione scolastica e l'impegno dell'Istituto nel misurare gli esiti di apprendimento degli studenti nella scuola italiana. Il portale si rivolge a diversi interlocutori come insegnanti, studenti, genitori, dirigenti scolastici, ricercatori, interessati a comprendere il funzionamento delle rilevazioni nazionali. Il portale INVALSIopen è accessibile dal 2019 come spazio dedicato alle Prove INVALSI, pensato per informare sul funzionamento e sull'organizzazione delle rilevazioni nazionali, sugli obiettivi e sui risultati. Il portale si propone come uno strumento utile per spiegare come nascono le prove, cosa misurano e in che modo possono aiutare a comprendere ciò che gli studenti hanno dimostrato di saper fare rispondendo alle domande presenti nelle prove.

Il sito si compone di diversi contenuti, tra cui articoli, interviste, video, testi, infografiche, approfondimenti specifici e confronti con altri Paesi, evidenziando una finalità non soltanto informativa, ma anche divulgativa e formativa, orientata alla costruzione di un ponte tra il mondo della scuola e quello del sistema nazionale di valutazione. Sono presenti anche sezioni dedicate ai dati e alla ricerca di INVALSI. Nella sezione sui dati sono raccolte le informazioni e le pubblicazioni relative agli esiti delle prove, con approfondimenti su temi quali i risultati delle rilevazioni, le differenze tra i gradi scolastici, le competenze degli studenti, il peso del contesto e le fragilità del sistema scolastico. La sezione dedicata alla ricerca raccoglie invece pubblicazioni, articoli, rapporti, libri e working papers derivanti dal lavoro dell'Istituto, rendendo disponibili contributi scientifici sull'uso dei dati INVALSI nella ricerca didattica, nella valutazione delle scuole e nell'orientamento delle politiche educative.

Tre sono le dimensioni rispetto alle finalità del portale. La prima riguarda la trasparenza, poiché vengono messe a disposizione informazioni sulle prove, sui risultati e sulle modalità di costruzione e di somministrazione delle rilevazioni. La seconda riguarda la comprensione pubblica della valutazione, poiché il portale mira a spiegare che cosa misurano le prove e perché i dati prodotti possano essere utili per leggere il sistema scolastico. La terza riguarda la costruzione di una cultura della valutazione, poiché i contenuti non si limitano alla presentazione dei risultati, ma propongono chiavi di lettura, approfondimenti, guide e materiali capaci di sostenere una comprensione più ampia del ruolo della valutazione esterna.

Nel contesto di questa analisi, INVALSIopen può quindi essere interpretato come uno strumento che contribuisce a definire le condizioni di comprensibilità della valutazione esterna. Se Scuola in Chiaro rende visibili alcuni dati a livello di istituzione scolastica e la Piattaforma Unica colloca gli esiti individuali all'interno del percorso formativo dello studente, INVALSIopen è uno spazio di spiegazione, accompagnamento e divulgazione, finalizzato a rendere più comprensibili le prove, i risultati e le ragioni della valutazione nazionale.

La tabella seguente illustra in modo sintetico il collegamento tra i vari documenti normativi e istituzionali relativi all'INVALSI e le implicazioni per la valutazione degli apprendimenti e le pratiche di restituzione.

Tabella 4 - Documenti analizzati per lo studio di caso.

Documento	Tipologia	Contesto e funzione	Ruolo dell'INVALSI	Implicazioni per la valutazione	Rilevanza per restituzione e livelli
D.Lgs. 286/2004	Decreto legislativo	Istituzione valutazione nazionale standardizzata e comparabilità sistemi	Rilevazioni sistematiche e censuarie	Misurazione standardizzata	Restituzione implicita
D.P.R. 80/2013	Regolamento	Sistema integrato valutazione	Indicatori e supporto RAV	Uso dati per miglioramento	Lettura organizzativa
D.Lgs. 62/2017	Decreto legislativo	Valutazione studenti	Doppio obiettivo valutativo: di sistema e individuale	Certificazione apprendimento	Livelli descrittivi
Statuto INVALSI	Documento istitutivo	Identità scientifica	Modelli valutativi	Validità e affidabilità della raccolta dati	Valutazione apprendimenti e valutazione sistema scolastico
DVS	Documento programmatorio	Innovazione e digitalizzazione	Sviluppo metodologico	Uso dati in termini di *data literacy* e *feedback literacy*	Sviluppo livelli descrittivi e indicatori per la valutazione scolastica
Scuola in Chiaro	Piattaforma digitale	Trasparenza	Diffusione dati a livello di singola scuola	Indicatori istituto	Restituzione di sistema
Piattaforma Unica	Piattaforma digitale	Integrazione servizi	Esiti individuali	Orientamento studente	Restituzione individuale
INVALSIopen	Sito web	Divulgazione informativa e formativa	Comunicazione	Trasparenza	Strumenti di informazione

5.3 Dalle Categorie Tematiche ai Temi Generali emergenti nelle interviste di docenti e studenti

L'obiettivo di questo paragrafo è presentare in modo sistematico i risultati dell'analisi tematica delle interviste condotte con gli studenti e i docenti. In prima battuta, verrà presentato il processo di analisi tematica delle azioni di *coding* delle interviste, condotto separatamente per docenti e studenti. Le interviste codificate di ciascun partecipante sono state analizzate e interpretate per individuare i temi emergenti. Il processo di analisi dei dati, come già indicato nel capitolo precedente, è iniziato durante la fase di trascrizione delle interviste. Durante questa fase, le trascrizioni automatiche generate dalla piattaforma Zoom sono state sistemate e corrette mediante il riascolto delle interviste registrate, previo consenso degli intervistati. Questo riascolto ha permesso di consolidare alcune impressioni emerse durante le interviste e di delineare una prima idea delle tematiche emergenti. Le trascrizioni sono state organizzate in file Word separati e anonimizzati. A ciascuna intervista è stato assegnato un codice identificativo, registrato nel database, in un file Excel predisposto in fase di selezione dei partecipanti e di calendarizzazione delle interviste.

Nella fase successiva di codifica delle interviste, i diversi contenuti emergenti sono stati etichettati con codici identificativi che comprendevano un identificativo in lettere assegnato al partecipante e un numero assegnato all'etichetta di coding relativa a uno specifico contenuto tematico. Le stringhe di conversazione relative ad uno specifico contenuto tematico sono state evidenziate nei file Word di trascrizione. Di seguito viene riportato un esempio di questo processo di attribuzione del codice a una stringa di conversazione: *"RC42_Con questi risultati, sono dei risultati che servono per la scuola. Essendo io, avendo svolto, e tuttora sono referente Rav e quindi del rapporto di autovalutazione, li analizziamo, ma avendo come obiettivo quello di un miglioramento scolastico, dell'istituto."*

Ogni volta che veniva completata la codifica di un'intervista, tutte le stringhe di conversazione e le relative etichette di coding tematico ad esse associate venivano inserite in un database generale anonimizzato. In questa fase, la rilettura costante delle diverse etichette di coding tematico ha anche consentito di procedere a una revisione continua di tali etichette alla luce delle nuove codifiche. Questa azione ha consentito di procedere a una revisione terminologica delle etichette alla luce della loro effettiva comprensibilità e della loro coerenza con i temi di ricerca.

All'interno di questo database è stato successivamente creato un nuovo foglio di lavoro in cui sono state isolate le etichette individuate (Coding Tematico), per procedere alla riflessione e alla costruzione di raggruppamenti tematici (Categorie Tematiche) più astratti, con l'individuazione di dimensioni tematiche generali (Temi Generali). Per i temi generali individuati, si è poi cercato di fornire definizioni che ne consentissero una comprensione del significato, ai fini dell'interpretazione dei dati emersi dalle interviste.

Sono state codificate prima le interviste dei docenti e successivamente quelle degli studenti. Le etichette individuate nella codifica delle interviste dei docenti sono state applicate alle interviste degli studenti quando i contenuti erano analoghi; per i contenuti non riconducibili ad esse sono state create nuove etichette, come nel caso della Categoria Tematica *L'esito individuale aumenta il valore della prova di valutazione esterna*, emersa solo dalle interviste degli studenti. Le Categorie Tematiche degli studenti sono quindi in parte condivise con quelle dei docenti e in parte specifiche, e i Temi Generali (par. 5.4) sono stati costruiti sull'insieme delle due codifiche.

5.3.1 Le Categorie Tematiche individuate nelle interviste di docenti e studenti

Questa sezione del capitolo presenta l'analisi qualitativa dei dati raccolti attraverso conversazioni con studenti e docenti sulle loro percezioni ed esperienze con i feedback descrittivi restituiti dalle prove INVALSI. Come evidenziato nei capitoli precedenti, l'obiettivo delle interviste era di approfondire il significato che studenti e docenti attribuiscono al sistema di valutazione, in particolare al feedback descrittivo ricevuto. Di seguito verranno quindi presentati prima i risultati dei docenti e poi quelli degli studenti.

5.3.1.1 Analisi delle interviste dei docenti

Le tabelle seguenti presentano le 19 Categorie Tematiche individuate e li illustrano con alcuni esempi di stringhe di conversazione degli insegnanti (Ins_). Vengono riportate anche le numerosità dei Coding Tematici che compongono la Categoria Tematica. Le numerosità indicano il numero di etichette di codifica confluite nella Categoria. Ogni etichetta si riferisce all'intervista di un singolo docente e può raggruppare più estratti; poiché uno stesso docente può contribuire con più etichette alla stessa Categoria, le numerosità indicano occorrenze codificate e non il numero di docenti.

Nell'ambito della parte dell'intervista dedicata alla *Verifica della percezione generale degli strumenti/materiali INVALSI*, sono state individuate 6 Categorie Tematiche. Per ciascuna di esse vengono riportati estratti di stringhe di conversazione significative.

La tabella 5 fornisce una prima mappatura della percezione generale, da parte dei docenti, degli strumenti e dei materiali INVALSI utili nelle fasi dell'attività didattica incentrata su questo tema. Le narrazioni mostrano come l'uso delle prove non sia circoscritto alla sola preparazione degli studenti alla loro somministrazione, ma si collochi all'interno di pratiche didattiche più articolate. La Categoria Tematica con più ricorrenze riguarda, infatti, l'utilizzo dei contenuti delle prove per progettare attività didattiche. La percezione di una distanza tra la prova INVALSI e le attività scolastiche segnala tuttavia che l'integrazione dello strumento nella didattica non è sempre immediata, ma richiede un lavoro di mediazione e ricontestualizzazione da parte dei docenti.

Tabella 5 - Categorie Tematiche relative alla Verifica della percezione generale degli strumenti/materiali INVALSI

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Ins_RC11_"*prepariamo i ragazzi alle prove INVALSI già a partire dalla prima media. Però, le esercitazioni chiaramente le svolgiamo in terza media e facciamo sempre riferimento alla vostra banca dati di esercitazioni e quindi diamo ai ragazzi le indicazioni del sito. E poi le vediamo insieme a loro.*" Ins_RS31_ "*per prepararli all'invalsi, cerchiamo di ragionare su come funziona questa prova. Avevamo studiato con loro il quadro di riferimento e avevamo provato a costruire insieme a loro su un testo scelto da noi.*" Ins_ EL13 *"vado sul sito dell'invasi e faccio con loro le prove che l'invalsi mette a disposizione. E poi ognuno le fa… faccio scrivere …perché non dando la risposta… faccio scrivere su un foglio le risposte, poi le guardiamo insieme e vediamo."*	Familiarizzazione al modello di prova valutativa	5
Ins_EL15_"*la prova la faccio soprattutto per metterli in condizione di non avere difficoltà col computer e con il sito dell'invalsi in modo che loro siano abbastanza veloci nel leggere e nel capire che cosa devono fare, come devono rispondere. Fondamentalmente faccio questo"*	Familiarizzazione all'uso della piattaforma di somministrazione	3
Ins_LP11_"*una conoscenza molto approfondita del quadro di riferimento Invalsi, quindi tendenzialmente per l'attività didattica mi baso su quello, sulla tabella dei macroaspetti o seguendo, ecco appunto il modo in cui viene inteso la comprensione del*	Utilizzo dei contenuti delle prove per progettare attività didattiche (sulla comprensione del testo o riflessione sulla lingua)	15

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
testo, per quello che riguarda insomma la prova di italiano chiaramente e cerco di creare delle mie prove ex novo tra virgolette, cioè attività didattiche tendenzialmente con domande a risposta aperta su testi che possono essere indifferentemente narrativi, quindi, di base il mio punto di riferimento è il quadro di riferimento direttamente" Ins_FD24_*"con gli studenti in passato soprattutto mi è capitato di chiarire, per esempio, per i ragazzi della quinta in prossimità anche delle prove, far vedere alcune cose, ma mai con diciamo l'idea del "prepariamoci a" quanto piuttosto* andiamo *a vedere quali sono le questioni"*		
Ins_RC31_"*È una logica molto diversa dalla valutazione scolastica e anche dal modo di affrontare i testi a scuola"*	Percezione di distanza tra lo strumento della prova INVALSI rispetto ai modelli e alle attività proposte a scuola	4
Ins_AI71_*"i dati sono utili, perché così noi sappiamo soprattutto se manteniamo la continuità didattica, diciamo che è un po' una cartina di tornasole per vedere se effettivamente quello che abbiamo fatto…e come lavorare poi nel triennio e impostando lavoro in terza e in quarta."* Ins_MG45_*"Io penso che questi strumenti debbano essere usati per aiutare anche la scuola in modo pratico a fare tutte quelle azioni che servono a portare avanti un percorso scolastico buono"*	Utilizzo dei dati delle prove INVALSI per azioni di miglioramento nel contesto scolastico	7

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
Ins_MG12_ "*Su Invalsi Open arrivano parecchi articoli approfondimenti e quando ho il tempo, io quelli li leggo..."*	Utilizzo materiali INVALSI a scopo informativo per il docente	4

Rispetto alla dimensione della *Verifica della percezione delle descrizioni dei livelli di comprensione del testo restituiti dalle prove,* la tabella 6 riporta le narrazioni dei docenti secondo cui i descrittori dei livelli di comprensione del testo, restituiti agli studenti come esito dello svolgimento della prova nazionale standardizzata, sono percepiti come strumenti potenzialmente utili per leggere in modo più analitico le competenze degli studenti. Le Categorie Tematiche con maggiore ricorrenza di Codici riguardano la programmazione di interventi didattici e l'uso dei descrittori in chiave formativa, suggerendo che il documento che riporta la restituzione descrittiva è riconosciuto non soltanto come informazione sull'esito della prova, ma anche come supporto alla progettazione e alla riorganizzazione dell'insegnamento. I docenti attribuiscono valore alla possibilità di disporre di indicazioni sui contenuti oggetto della prova più articolate rispetto al voto o al punteggio, poiché queste consentono di individuare aree di lavoro per la progettazione di attività mirate e di sostenere negli studenti una maggiore consapevolezza dei processi di apprendimento.

Tabella 6 - Categorie Tematiche relative alla Verifica della percezione delle descrizioni dei livelli di comprensione del testo restituiti dalle prove

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
Ins_RC91_*"Faccio un esempio, cioè, mi arrivano magari i descrittori della primaria della quinta di una della mia classe nuova di una prima media nuova e secondo me è un'immagine completamente diversa, un'immagine diversa di valutazione e rispetto da mettere chiaramente a confronto col percorso che loro hanno svolto nella scuola primaria e con il voto d'uscita, però da un'idea, per quanto riguarda la disciplina d'italiano, è*	Descrittori per comprendere le competenze degli studenti	5

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
più analitica anche perché è più oggettiva, provengono tutti dalla scuola elementare e hanno tutti i voti altissimi e abbiamo certe volte difficoltà nell'individuare, se non poi con le nostre prove, i test d'ingresso, che sono finalizzati proprio a questo, le vere competenze del degli alunni. Cioè dà una restituzione più oggettiva"		
Ins_EF14_*"all'inizio proprio avevo fatto una tabellina che corrispondeva…una tabellina Excel che corrispondeva ai tre… quando ancora avevamo i tre macro aspetti…. E mettevo dentro alcuni esercizi pratici e a loro facevo trovare a quale aspetto… una specie di riconoscimento dell'ambito della domanda e questo piaceva molto."* Ins_GL51_"*Secondo me sono tutti utilissimi e sarebbe proprio un lavoro da dipartimento, estrapolarli magari da questa… Perché ogni livello contiene diverse aree. Giusto, quindi l'area della comprensione, della riflessione dell'uso della lingua e del lessico, sarebbe l'ideale sarebbe proprio che ogni dipartimento per la propria scuola, ma per il tipo di lavoro che individua di dover fare per migliorare l'offerta, sarebbe interessante estrapolarli e indicarli proprio nelle programmazioni di classe"*	Descrittori per la programmazione di interventi didattici	14
Ins_CT71_*"in una modalità di valutazione formativa…gli dai una restituzione che loro capiscono e in qualche modo si rendono conto, di, diciamo, della tipologia di errore che hanno fatto. Ecco, Quindi questo loro io gliel'ho fatto. Devo dire che avevo iniziato subdolamente, non*	Descrittori per una valutazione formativa	14

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
proprio in maniera, facendogli leggere i livelli, ma perché in qualche modo io li conoscevo già, gliel'avevo proposto in terza, in quarta. Quindi se in maniera diciamo così, molto discorsiva. Poi quest'anno, quando gli ho fatto vedere che c'erano proprio i livelli descritti, allora hanno iniziato a crederci di più, ecco in qualche modo Ins_AI53_*"personalmente, per me più sono dettagliate più ho dati per comprendere me stesso il mio lavoro".*		

Rispetto alla dimensione della *Verifica della percezione nel contesto classe del processo di valutazione standardizzata,* la tabella 7 mostra il punto di vista dei docenti rispetto alla loro esperienza dei vissuti degli studenti rispetto al processo di valutazione standardizzata, che viene percepito dal loro punto di vista nel contesto classe. La Categoria Tematica con le maggiori ricorrenze riguarda il fatto che l'esito della prova INVALSI non abbia un peso diretto sulla valutazione scolastica dello studente, spiegando in questo modo la difficoltà nel costruire una motivazione condivisa attorno alla prova e alla sua certificazione. Parallelamente, emerge con chiarezza il ruolo del docente nel contribuire alla programmazione di azioni informative-formative che possano dare un senso alle prove. Le Categorie Tematiche relative alla percezione della valutazione esterna come conferma della valutazione scolastica e agli atteggiamenti percepiti degli studenti nei confronti delle prove mostrano un quadro articolato, in cui si riscontrano contemporaneamente senso di responsabilità e serenità, ma anche ansia, ambivalenza e, talvolta, uno scarso impegno. La prova INVALSI sembra quindi uno strumento il cui valore educativo dipende in larga parte dal modo in cui viene presentata, spiegata e discussa in classe.

Tabella 7 - Categorie Tematiche relative alla Verifica della percezione nel contesto classe del processo di valutazione standardizzata

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Ins_EF81_"*arriverà un certificato senza il voto, sicché non fa media con l'esame"* Ins_ES52_*"La prova devono farla, perché altrimenti non vanno all'esame, ma poi per loro è come un rito di passaggio, e il risultato non lo vanno a vedere o non lo vanno a vedere" Ins_SR71_"sono studenti di 13 anni, e sono ovviamente persone sensibilissime al tema della valutazione. Quindi danno valore se una cosa fa media. Quale ricaduta potrà avere sul mio voto. Ragionano per voti"*	L'esito della prova INVALSI non ha un peso sulla valutazione scolastica dello studente	17
Ins_FD71_*"è un'argomentazione che utilizzo anche nelle classi perché prendano molto sul serio le prove. Insomma, nella classe quinta, il dire ai ragazzi: guardate che questo certificato poi ha un valore a tutti gli effetti, per cui è importante che voi facciate le prove con attenzione, perché è qualcosa che poi vi ritorna attraverso un documento che descrive il vostro livello, che descrive il traguardo che avete raggiunto e quindi diciamo anzi, io credo che veramente la certificazione rappresenti un'ottima occasione"*	Il docente promuove azioni informative per costruire un significato per lo svolgimento delle prove INVALSI	11
Ins_FD91_*"Ecco, quindi questo fa piacere. Fa piacere anche ai ragazzi perché dicono: Beh, però è una conferma esterna"*	La valutazione esterna utile come conferma rispetto alla valutazione scolastica	4
Ins_FD121_ "*in generale la prendono con senso di responsabilità. Cioè, in generale non capita che lo studente che poi si siede a far le*	Atteggiamento responsabile e sereno degli studenti rispetto allo svolgimento della prova	3

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
prove non le faccia o le faccia diciamo in modo molto distratto, è sempre più raro questo."		
Ins_EL81_ "*alcuni si rendono conto dell'utilità, altri, secondo me, assolutamente no. Perché siamo ancora in un grado di scuola dove gli alunni sono abbastanza immaturi, quindi alcuni la fanno proprio così la affrontano senza impegno, alcuni altri, forse con troppo impegno, quindi si sentono troppo pressati, però secondo me, anche magari anche le famiglie non percepiscono proprio bene l'utilità di questo tipo di prova"*	Atteggiamenti ambivalenti e criticità percepite rispetto alla prova	3

Rispetto alla dimensione della *Metariflessione sul feedback descrittivo INVALSI* la tabella 8 raccoglie le esperienze o le riflessioni dei docenti relativamente ad aspetti di criticità sul documento che riporta le descrizioni sintetiche e analitiche dei livelli rilasciati dall'INVALSI come esito individuale dello svolgimento della prova. Le Categorie Tematiche con maggiore ricorrenza riguardano la scarsa informazione sulla certificazione e sulla descrizione dei livelli, e la restituzione tardiva dell'esito, che arriva quando il percorso scolastico è ormai concluso. Questi due elementi limitano la possibilità di avviare un confronto significativo con gli studenti e di trasformare l'esito descrittivo delle prove standardizzate nazionali in un'occasione di apprendimento. Altre criticità riguardano la struttura e i contenuti del documento, che, essendo statico, non consentono di individuare facilmente potenzialità e aree di miglioramento specifiche. Questo limita la capacità di comprendere quanto un determinato elemento influisca sulla comprensione del testo dello studente rispetto a un altro elemento, all'interno dello stesso livello. Inoltre, l'uso di terminologia specialistica nei descrittori e la scarsa adeguatezza dei materiali informativi e formativi rispetto ai contenuti delle descrizioni rendono più complessa l'organizzazione delle attività di potenziamento e di recupero. L'esito descrittivo e la struttura dei livelli sono riconosciuti come strumenti potenzialmente utili nel contesto didattico, ma la loro efficacia appare condizionata dalla chiarezza comunicativa, dalla tempestività della restituzione e dalla disponibilità di strumenti di accompagnamento per un'approfondita comprensione dei descrittori dei livelli.

Tabella 8 - Categorie Tematiche relative alla Metariflessione sul feedback descrittivo INVALSI

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Ins_RC53_*"E quindi poi loro neanche sanno come è andata, com'è andata questa prova invalsi, mentre e chiaramente con gli alunni, avevamo subito un risultato, un riscontro concreto, e quindi si poteva parlare, avremmo potuto approfondire, così attraverso anche le vostre schede, i loro livelli di competenza, adesso è lasciato tutto un po' così."*	L'esito della prova INVALSI arriva allo studente e al docente terminata la scuola	10
Ins_MG71_*"questi livelli e questo tipo di certificazione Invalsi è conosciuto pochissimo"*	Poca informazione sulla certificazione INVALSI e sulla descrizione dei livelli	15
Ins_FD151_*"sarebbe, diciamo, un po' sofisticato però interessante se la descrizione, che pure restituisce lo stesso livello di Caio e di Sempronio potesse adattarsi a seconda della tipologia delle risposte su cui poi si è coperto il livello e quelle che invece no"* Ins_SL31_*"riusciamo sia a dare un discorso generico di risultato sei a livello 3…però, la cosa che sarebbe fondamentale è quella dire guarda che sull'inferenza sei a livello 1"*	L'esito delle prove INVALSI non permette di individuare potenzialità o aree di miglioramento specifiche	5
Ins_LP22_*"Nella mia esperienza eco, perlomeno è difficile che uno va ad addentrarsi nei livelli anche perché chiaramente sono scritti con terminologia scientifica che non è non è accessibile a tutti".*	Concetti e terminologia dei descrittori specialistici	6
Ins_AI111_*"forse si potrebbe dettagliare sì, ma rendendo anche solo graficamente un po' più schematico. Se fosse anche solo fatto per punti visivamente, potrebbe essere più facile, più fruibile per i ragazzi"*	Scarsa adeguatezza dei materiali informativi e formativi	3

Estratti di stringhe di conversazione significative	Categorie Tematiche	Numerosità Coding Tematico
Ins_EL121_*"Qualcosa che potrebbe essere utile mettere su una piattaforma appunto, magari più materiale e vicino al materiale, anche per alunno, fare delle domande sui descrittori: che cosa hai fatto…hai confrontato…? Questo potrebbe essere un qualcosa che potrebbe essere fatto in itinere."*		

5.3.1.2 Analisi delle interviste degli studenti

Le analisi delle interviste degli studenti sono avvenute successivamente all'individuazione delle prime Categorie Tematiche individuate nel coding delle interviste dei docenti.

Anche per queste interviste si riportano le tabelle con le Categorie Tematiche individuate e alcuni esempi di stringhe di conversazione (Stud_).

Anche in questo caso le numerosità indicano etichette di codifica e non studenti. Nel leggere gli estratti va inoltre tenuto presente che sette studenti su nove sono stati intervistati prima di svolgere la prova dell'ultimo anno e quindi prima di ricevere il proprio esito, mentre due (AC e FG) sono stati intervistati dopo il diploma (par. 4.6.3). Per distinguere le diverse basi delle loro affermazioni, ogni estratto è contrassegnato con (E) quando si riferisce a esperienze, conoscenze od opinioni precedenti all'intervista, e con (D) quando è stato suscitato dalla lettura delle descrizioni dei livelli proposta durante l'intervista. Gli estratti della Tabella 10, raccolti nella parte dell'intervista in cui le descrizioni venivano mostrate a schermo, sono tutti di questo secondo tipo; alcuni studenti, inoltre, dichiarano di non aver conosciuto il modello di restituzione prima dell'intervista (Tabella 12), mentre i due studenti intervistati dopo il diploma si riferiscono alla certificazione già ricevuta, che uno ha consultato e l'altro non aveva aperto. Quando la sezione dell'intervista e il contenuto dell'estratto non coincidono, la classificazione segue il contenuto. Le percezioni degli studenti sui descrittori vanno quindi lette come reazioni a un documento appena presentato, più che come esperienze d'uso.

Nell'ambito della parte dell'intervista dedicata alla *Verifica della percezione generale degli strumenti/materiali INVALSI*, la tabella 9 riporta la Categoria Tematica ripresa dalle Categorie individuate per il gruppo dei docenti, relativa alla familiarizzazione con il modello di prova valutativa, e mostra come le esperienze di preparazione siano presenti, ma non sempre omogenee

nei diversi contesti scolastici. Dalle stringhe di conversazione emerge infatti una differenza tra gli studenti che non ricordano attività specifiche di preparazione e quelli che riferiscono di aver utilizzato materiali o libri dedicati per comprendere la struttura della prova. Il dato suggerisce che la familiarizzazione con il modello INVALSI dipende in modo rilevante dalle pratiche promosse dai docenti e dalla continuità con cui tali pratiche vengono integrate nel percorso didattico.

Tabella 9 - Categorie Tematiche relative alla Verifica della percezione generale degli strumenti/materiali INVALSI

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Stud_LM11_ (E)*"Allora, da quanto mi ricordo alle superiori, non mi sono stati presentati materiali particolari e non mi sembra ci siano state neanche delle preparazioni particolari in classe."* Stud_MB_f1_12_ *(E) "Abbiamo un libro, specialmente la nostra professoressa d'inglese che usa per farci vedere un po' come saranno le prove d'Invalsi."*	Familiarizzazione al modello di prova valutativa	8

Rispetto alla dimensione della *Verifica della percezione delle descrizioni dei livelli di comprensione del testo restituiti dalle prove* la tabella 10 presenta le percezioni degli studenti sulle descrizioni dei livelli di comprensione del testo restituiti dalle prove INVALSI. La Categoria Tematica con il numero maggiore di ricorrenze riguarda l'uso dei descrittori per l'autovalutazione e il miglioramento, seguita dalla possibilità di comprendere in modo più articolato le competenze dimostrate. Tale dato suggerisce che gli studenti, di fronte alle descrizioni mostrate durante l'intervista, riconoscono valore a una restituzione capace di indicare non solo un esito, ma anche ciò che è stato svolto correttamente e ciò che può essere migliorato. I descrittori sono percepiti come strumenti potenzialmente utili quando supportano la consapevolezza individuale, la preparazione e la riflessione sui propri punti di forza e sulle aree di miglioramento.

Tabella 10 - Categorie Tematiche relative alla Verifica della percezione delle descrizioni dei livelli di comprensione del testo restituiti dalle prove

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Stud_LM52_ (D) *"È proprio una valutazione di "Ok, 6 stato in grado di fare questo, questo, questo, questo questo no"* Stud_FL61_ (D) *"Sì, lo preferisco al voto è importante soprattutto in prove come quella di italiano, in cui cioè del voto poco importa, importa se sia riuscito a riconoscere e analizzare il testo e a capirlo bene, quindi io lo trovo utile."*	Descrittori per comprendere le competenze degli studenti	8
Stud_LM72_ *(D) "andare ad applicare delle versioni, magari non necessariamente identiche ma simili, di questo metro di valutazione, nella valutazione dei lavori che vengono compiuti in classe."*	Descrittori per la programmazione di interventi didattici	1
Stud_MB_f1_42_ (D) *"se qualcuno sa già quali sono gli obiettivi che bisogna raggiungere e superare in qualche modo si prepara già mentalmente al fatto che durante la prova d'italiano bisogna avere determinate competenze e quindi si mette anche nel Mindset di riuscire ad arrivare al livello 5"* Stud_GO41_ (D) *"una griglia di valutazione tale possa aiutare anche una persona a capire su cosa deve migliorare e cosa invece può portare avanti come suo punto di forza"*	Descrittori per l'autovalutazione e il miglioramento	9

Rispetto alla dimensione della *Verifica della percezione nel contesto classe del processo di valutazione standardizzata* nella tabella 11 le Categorie Tematiche più frequenti riguardano l'aumento del valore della prova quando l'esito assume una dimensione individuale e gli atteggiamenti ambivalenti e le criticità percepite rispetto alla prova. Questo dato mostra come gli studenti riconoscano una maggiore utilità della prova quando essa restituisce informazioni

personali, ma al tempo stesso ne colgano i limiti legati alla natura episodica della rilevazione e alla possibile distanza tra la prestazione del giorno della prova e le competenze effettive. Lo scarso peso dell'esito sulla valutazione scolastica può inoltre ridurre la motivazione e l'impegno, mentre le azioni informative dei docenti, anche se poco rappresentate nelle conversazioni con gli studenti, emergono come attività utili alla costruzione di un significato più chiaro dello strumento.

Tabella 11 - Categorie Tematiche relative alla Verifica della percezione nel contesto classe del processo di valutazione standardizzata

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Stud_LM61_ (E) *"uno dei problemi che forse si può attribuire alle prove invalsi è il fatto che, non essendo valutate, ed essendo noi studenti in un sistema che invece che che se ne dica mette molta enfasi sul voto, c'è il rischio che qualcuno non prenda seriamente la prova."*	L'esito della prova INVALSI non ha un peso sulla valutazione scolastica dello studente	3
Stud_AC11_ (E) *"l'aiuto dei professori. Quindi quello che ci han detto effettivamente l'ultimo anno su come era cambiato il valore, diciamo della prova Invalsi"*	Il docente promuove azioni informative per costruire un significato per lo svolgimento delle prove INVALSI	1
Stud_FG41_ (E) *"prima era una prova, diciamo serviva a fare una sorta di sondaggio... Invece ora è qualcosa di più personale e secondo me è molto, molto più utile rispetto a prima."*	L'esito individuale aumenta il valore della prova di valutazione esterna	6
Stud_FG32_ (E) *"se quel giorno una persona, ad esempio, non si sente particolarmente bene... aveva una valutazione che non lo rispecchiava."* Stud_ AC41_ *(E) "gli Invalsi le vedo molto di più come una cosa generale, cioè, che serve in generale, non solo a me."* Stud_ AC42_ *(E) "E quindi più a fini statistici che al singolo."*	Atteggiamenti ambivalenti e criticità percepite rispetto alla prova	6

Rispetto alla dimensione della *Metariflessione sul feedback descrittivo INVALSI* la tabella 12 mostra come Categoria Tematica più ricorrente la scarsa informazione sulla certificazione e sulla descrizione dei livelli, evidenziando una conoscenza ancora parziale e poco diffusa dello strumento e delle modalità di restituzione dell'esito. Accanto a questo elemento emergono criticità legate alla terminologia dei descrittori troppo complessi per un'immediata comprensione. Nel complesso, l'esito descrittivo appare potenzialmente significativo per gli studenti, ma la sua efficacia dipende dalla chiarezza delle informazioni fornite, dalla comprensibilità del linguaggio utilizzato e dalla presenza di dispositivi di accompagnamento che aiutino a leggere e interpretare i livelli in relazione al proprio percorso di apprendimento.

Tabella 12 - Categorie Tematiche relative alla Metariflessione sul feedback descrittivo INVALSI

Estratti di stringhe di conversazione significative	**Categorie Tematiche**	**Numerosità Coding Tematico**
Stud_MB_f1__22_ (E) *"Però non so in alcun modo come sarà strutturato questo risultato." Stud_ IC21_ (E) "Non lo sapevo. L'ho scoperto ora"*	Poca informazione sulla certificazione INVALSI e sulla descrizione dei livelli	7
Stud_ IC41_ (D) "*gli svantaggi, forse che non tutti gli studenti possono conoscere il significato delle parole alcune sono un po' complesse forse."*	Concetti e terminologia dei descrittori specialistici	1

5.4 Definizione dei Temi Generali

Come illustrato nel capitolo metodologico, l'individuazione dei Temi Generali costituisce una fase di astrazione successiva alla codifica e alla costruzione delle Categorie Tematiche. In coerenza con l'approccio dell'analisi tematica proposto da Braun e Clarke, il tema non è stato inteso come una semplice sintesi descrittiva delle occorrenze più frequenti, ma come un pattern di significato capace di organizzare più categorie attorno a un nucleo interpretativo rilevante rispetto alle domande di ricerca. La definizione dei temi ha quindi richiesto un processo ricorsivo di lettura, confronto e rielaborazione, finalizzato a precisare il significato di ciascun tema, i suoi confini rispetto agli altri temi e il contributo specifico che esso offre alla comprensione del fenomeno indagato.

In questa prospettiva, i sette Temi Generali non corrispondono a contenitori neutri delle Categorie Tematiche, ma costituiscono costrutti analitici attraverso i quali interpretare il rapporto tra le prove INVALSI, il feedback descrittivo, le pratiche didattiche e le percezioni di docenti e studenti.
Le definizioni che seguono sono state elaborate con l'obiettivo di esplicitare il nucleo concettuale di ciascun Tema Generale e di chiarire come le Categorie Tematiche confluiscano in una struttura interpretativa più ampia. Ogni definizione specifica il significato del tema, le categorie che lo compongono e la funzione analitica che esso assume nella comprensione del processo di restituzione degli esiti INVALSI. In questo senso, le definizioni non sono soltanto una descrizione dei contenuti emersi, ma costituiscono una formalizzazione del lavoro interpretativo svolto sui dati.

Tabella 13 - Temi generali.

Categorie Tematiche	Temi Generali
Familiarizzazione al modello di prova valutativa	**T1_Utilizzo dei contenuti delle prove INVALSI integrato nell'attività didattica**
Familiarizzazione all'uso della piattaforma di somministrazione	
Utilizzo dei contenuti delle prove per progettare attività didattiche sulla comprensione del testo o riflessione sulla lingua	
Utilizzo materiali INVALSI a scopo informativo per il docente	
Descrittori per comprendere le competenze degli studenti	**T2_Esito descrittivo per la progettazione di azioni educative**
programmazione Descrittori per la programmazione di interventi didattici	
Descrittori per una valutazione formativa	
Descrittori per l'autovalutazione e il miglioramento	
L'esito della prova INVALSI non ha un peso sulla valutazione scolastica dello studente	**T3_La certificazione INVALSI**
L'esito della prova INVALSI arriva allo studente e al docente terminata la scuola	
Poca informazione sulla certificazione INVALSI e sulla descrizione dei livelli	

Categorie Tematiche	Temi Generali
Il docente promuove azioni informative per costruire un significato per lo svolgimento delle prove INVALSI	**T4_Il ruolo di intermediazione del docente tra prova INVALSI e il contesto scolastico**
Percezione di distanza tra lo strumento della prova INVALSI rispetto ai modelli e alle attività proposte a scuola	
Atteggiamento responsabile e sereno degli studenti rispetto allo svolgimento della prova	
Atteggiamenti ambivalenti e criticità percepite rispetto alla prova	
Utilizzo dei dati delle prove INVALSI per azioni di miglioramento nel contesto scolastico	**T5_Utilizzo dei dati INVALSI per piani di miglioramento scolastico**
L'esito delle prove INVALSI non permette di individuare potenzialità o aree di miglioramento specifiche	**T6_Criticità percepite sui descrittori INVALSI per un utilizzo nella didattica**
Concetti e terminologia dei descrittori specialistici	
Scarsa adeguatezza dei materiali informativi e formativi	
La valutazione esterna utile come conferma rispetto alla valutazione scolastica	**T7_Vantaggi percepiti sulla valutazione individuale dello studente**
L'esito individuale aumenta il valore della prova di valutazione esterna	

Per rendere visibile il percorso analitico, la tabella seguente riporta alcuni esempi del passaggio dagli estratti delle interviste alle etichette di codifica, alle Categorie Tematiche e ai Temi Generali. Gli esempi riguardano sia le interviste dei docenti sia quelle degli studenti.

Tabella 14 – Esempi del percorso dall'estratto al Tema Generale

Estratto	Etichetta di codifica	Categoria Tematica	Tema Generale
Ins_RC53: «E quindi poi loro neanche sanno come è andata [...] avremmo potuto approfondire [...] i loro livelli di competenza, adesso è lasciato tutto un po' così.»	RC5. Mancanza di un lavoro sulla restituzione con gli studenti	L'esito della prova INVALSI arriva allo studente e al docente terminata la scuola	**T3. La certificazione INVALSI**

Ins_LP22: «è difficile che uno va ad addentrarsi nei livelli anche perché chiaramente sono scritti con terminologia scientifica»	LP2. Linguaggio troppo specialistico dei descrittori	Concetti e terminologia dei descrittori specialistici	**T6. Criticità percepite sui descrittori INVALSI**
Ins_GL51: «sarebbe proprio un lavoro da dipartimento [...] indicarli proprio nelle programmazioni di classe»	GL5. Riconoscimento di un potenziale utilizzo dei descrittori per la programmazione e la costruzione di azioni e strumenti didattici	Descrittori per la programmazione di interventi didattici	**T2. Esito descrittivo per la progettazione di azioni educative**
Stud_GO41 (D): «una griglia di valutazione tale possa aiutare anche una persona a capire su cosa deve migliorare»	GO4. Utilità del feedback per l'autovalutazione e il miglioramento	Descrittori per l'autovalutazione e il miglioramento	**T2. Esito descrittivo per la progettazione di azioni educative**
Stud_FG41 (E): «prima era una prova [...] una sorta di sondaggio... Invece ora è qualcosa di più personale»	FG4. Valore attribuito alla prova INVALSI	L'esito individuale aumenta il valore della prova di valutazione esterna	**T7. Vantaggi percepiti sulla valutazione individuale dello studente**

I sette temi emergenti dall'analisi tematica offrono una visione multidimensionale delle dinamiche, delle potenzialità e delle criticità nell'interazione tra il sistema di valutazione esterna e il contesto scolastico. Di seguito sono presentate definizioni e analisi di questi temi, basate sui dati disponibili.

Tema 1: Utilizzo dei contenuti delle prove INVALSI integrato nell'attività didattica

Questo Tema (T1) descrive come i materiali, i contenuti e le modalità operative delle prove INVALSI entrano, con intensità variabile, nelle pratiche ordinarie di insegnamento e apprendimento. Il tema raccoglie le categorie relative alla familiarizzazione al modello di prova, alla familiarizzazione con la piattaforma di somministrazione, all'uso dei contenuti per progettare attività didattiche e all'impiego dei materiali INVALSI come risorsa informativa per il docente. Il nucleo interpretativo del tema riguarda il passaggio da una concezione della prova come evento esterno e circoscritto a una sua possibile integrazione nel lavoro didattico quotidiano. In questa prospettiva, i contenuti delle prove non sono considerati soltanto strumenti di esercitazione in vista della somministrazione, ma possono essere utilizzati per sostenere attività, nel caso specifico di questa ricerca, sulla comprensione del testo e sulla riflessione linguistica. Il tema evidenzia quindi una distinzione tra l'uso finalizzato alla preparazione alla prova, orientato a fornire informazioni utili per ridurre eventuali incertezze, e

l'uso didattico, che prevede l'impiego dei materiali INVALSI come risorse per la progettazione e la riflessione professionale del docente.

Tema 2: Esito descrittivo per la progettazione di azioni educative

Questo tema (T2) riguarda il valore attribuito ai descrittori degli esiti INVALSI, intesi come strumenti in grado di restituire informazioni più articolate rispetto al semplice risultato numerico. Il tema comprende le Categorie Tematiche relative all'uso dei descrittori per comprendere le competenze degli studenti, programmare interventi didattici, sostenere la valutazione formativa e promuovere l'autovalutazione e il miglioramento. Il nucleo del tema è costituito dalla funzione interpretativa e progettuale dei descrittori di livello. I descrittori, infatti, consentono di leggere l'esito della prova non soltanto come certificazione di un livello raggiunto, ma anche come informazione utile a comprendere ciò che lo studente sa fare, ciò che necessita di essere consolidato e quali azioni educative possono essere progettate a partire da tali risultati. Il tema mette quindi in relazione la restituzione descrittiva con una concezione formativa della valutazione, in cui il feedback, o meglio l'esito individuale ricevuto dallo studente, assume significato solo se viene compreso, discusso e trasformato in occasione di apprendimento, di regolazione didattica e di consapevolezza individuale.

Tema 3: La certificazione INVALSI

Questo Tema (T3) definisce l'insieme delle percezioni e delle criticità connesse alla restituzione formale degli esiti delle prove nazionali standardizzate. Esso raccoglie le Categorie Tematiche relative al mancato peso dell'esito INVALSI nella valutazione scolastica dello studente, alla restituzione dell'esito quando il percorso scolastico è ormai concluso e alla scarsa informazione sulla certificazione e sulla descrizione dei livelli. Il tema evidenzia il carattere ambivalente della certificazione rilasciata allo studente come esito delle prove standardizzate nazionali. Da un lato, essa rappresenta uno strumento potenzialmente utile perché restituisce allo studente un profilo individuale rispetto alla prova svolta. D'altro canto, la sua efficacia è limitata da condizioni organizzative e comunicative che ne riducono le potenzialità educative. Il fatto che l'esito non incida direttamente sulla valutazione scolastica, che arrivi in una fase successiva alla conclusione del percorso e che non sia sempre adeguatamente spiegato contribuisce a indebolirne il significato per studenti e docenti. Il tema consente quindi di leggere la certificazione come un dispositivo formalmente presente, ma non sempre pienamente integrato nei processi di apprendimento.

Tema 4: Il ruolo di intermediazione del docente tra prova INVALSI e il contesto scolastico

Questo tema (T4) descrive la funzione svolta dagli insegnanti nel rendere la prova INVALSI comprensibile e significativa nell'ambito del contesto scolastico. Il tema include le Categorie Tematiche relative alle azioni informative promosse dai docenti, alla percezione della distanza tra la prova e le attività scolastiche, agli atteggiamenti responsabili e sereni degli studenti e agli atteggiamenti ambivalenti e alle criticità percepite rispetto alla prova. Il nucleo interpretativo del tema riguarda la mediazione tra un dispositivo valutativo esterno e il contesto interno della classe. La prova INVALSI può essere percepita dagli studenti come lontana dalle attività scolastiche ordinarie, come un adempimento formale o come un'esperienza valutativa non pienamente compresa. In questo quadro, il docente assume un ruolo decisivo nel costruire il significato attorno alla prova, chiarirne finalità e conseguenze, collegarla alle pratiche didattiche e sostenere un atteggiamento più consapevole da parte degli studenti. Il tema mostra quindi che la percezione della prova non dipende esclusivamente dalle sue caratteristiche tecniche, ma anche dal lavoro interpretativo e relazionale attraverso cui la scuola la presenta, la spiega e l'accompagna.

Tema 5: Utilizzo dei dati INVALSI per piani di miglioramento scolastico

Questo tema (T5) riguarda l'uso dei risultati delle prove come risorsa informativa per l'autovalutazione e la progettazione di azioni di miglioramento a livello dell'istituzione scolastica. Esso deriva dalla Categoria Tematica relativa all'utilizzo dei dati delle prove INVALSI per azioni di miglioramento nel contesto scolastico. Il tema definisce i dati INVALSI come strumenti per la lettura del funzionamento complessivo della scuola, più che come informazioni riferite esclusivamente alla prestazione del singolo studente. I risultati aggregati restituiti alle scuole possono contribuire a individuare aree di forza e di criticità, a orientare le scelte dei dipartimenti, a sostenere la progettazione di interventi didattici e ad alimentare i processi di autovalutazione istituzionale. Il valore del tema risiede nella possibilità di collegare la valutazione esterna a una funzione di miglioramento, a condizione che i dati siano interpretati in modo contestualizzato e inseriti in processi collegiali di riflessione e di progettazione.

Tema 6: Criticità percepite sui descrittori INVALSI per un utilizzo nella didattica

Questo tema (T6) raccoglie le difficoltà percepite da docenti e studenti nell'uso e nella comprensione del feedback descrittivo restituito dalle prove. Esso comprende le Categorie Tematiche relative alla difficoltà dell'esito individuale nel riportare informazioni ancora più specifiche per cogliere potenzialità o aree di miglioramento, alla presenza di concetti e terminologia specialistici e alla scarsa adeguatezza dei materiali informativi e formativi. Il nucleo del tema riguarda lo scarto tra il potenziale

formativo dei descrittori e le condizioni concrete che ne rendono possibile l'utilizzo. I descrittori possono offrire una restituzione più articolata rispetto al punteggio, ma rischiano di non essere pienamente efficaci quando risultano troppo generali, linguisticamente complessi o non accompagnati da strumenti di supporto alla lettura. Il tema evidenzia quindi che la qualità del feedback non dipende soltanto dalla presenza di descrizioni analitiche, ma anche dalla loro accessibilità, dalla chiarezza comunicativa e dalla possibilità di tradurle in indicazioni comprensibili e operative per studenti e docenti.

Tema 7: Vantaggi percepiti sulla valutazione individuale dello studente

Questo Tema (T7) riguarda il valore attribuito alla restituzione individuale degli esiti INVALSI come forma di riscontro esterno rispetto alla valutazione scolastica. Il tema comprende le Categorie Tematiche relative alla valutazione esterna, come la conferma rispetto alla valutazione scolastica e l'aumento del valore della prova quando l'esito assume una dimensione individuale. Il nucleo interpretativo del tema consiste nel riconoscimento della prova INVALSI come possibile fonte di informazione complementare rispetto alla valutazione svolta in classe. La restituzione individuale può essere percepita come significativa perché consente allo studente di ricevere un riscontro standardizzato, esterno e personalizzato sulle proprie competenze. Tale riscontro può rafforzare la consapevolezza individuale, confermare o problematizzare la valutazione scolastica e attribuire maggiore valore alla prova, distinguendola da una rilevazione percepita esclusivamente come statistica o istituzionale. Il tema mostra quindi che la dimensione individuale dell'esito può contribuire a rendere la valutazione esterna più rilevante per lo studente, soprattutto quando viene compresa e interpretata in relazione al proprio percorso formativo.

CAPITOLO 6: DISCUSSIONE DEI RISULTATI

In questo capitolo viene proposta una lettura interpretativa dei risultati della ricerca con l'obiettivo di comprendere come i dati dell'analisi documentale, della revisione sistematica e delle interviste possano dialogare tra loro alla luce del quadro teorico di riferimento. La discussione è organizzata attorno alle due domande di ricerca, ma adotta la triangolazione come modalità di interpretazione capace di collegare i piani istituzionali, scientifici ed esperienziali. La riflessione si sviluppa a partire dall'idea che la restituzione descrittiva individuale degli esiti delle prove dell'INVALSI non sia soltanto un documento di comunicazione del risultato ma un oggetto educativo complesso, nel quale si incontrano la funzione di monitoraggio del sistema, la certificazione individuale e le pratiche concrete attraverso cui docenti e studenti attribuiscono significato a tale valutazione.

Nella lettura dei risultati si distinguono tre piani: le percezioni espresse dai partecipanti; le pratiche che i partecipanti riferiscono di aver svolto, non osservate direttamente; le possibili implicazioni educative: ipotesi interpretative della ricercatrice. La ricerca non ha rilevato effetti della restituzione descrittiva sugli apprendimenti, che richiederebbero disegni di altro tipo. Per gli studenti, inoltre, va ricordato che sette interviste su nove si sono svolte prima della prova dell'ultimo anno e che una parte delle loro affermazioni è stata suscitata dalla lettura delle descrizioni dei livelli durante l'intervista (par. 4.6.3 e 5.3.1.2).

Per collegare in modo sistematico i risultati alle domande di ricerca e alla letteratura, la tabella seguente riporta, per ciascun Tema Generale, la domanda di ricerca a cui contribuisce, le convergenze con gli studi richiamati nel quadro teorico e nella revisione sistematica e gli elementi specifici del caso studiato.

Tabella 15 - Temi Generali, domande di ricerca e confronto con la letteratura

Tema Generale	Domanda di ricerca	Convergenze con la letteratura	Elementi specifici del caso
T1. Utilizzo dei contenuti delle prove INVALSI integrato nell'attività didattica	2 (pratiche riferite dai docenti)	I materiali e i dati delle rilevazioni diventano una risorsa didattica quando sono reinterpretati dai docenti (Looney, 2011; Freddano & Pastore, 2020)	Le pratiche riferite riguardano soprattutto i materiali e le prove, meno le descrizioni dei livelli restituite ai singoli studenti
T2. Esito descrittivo per la progettazione di azioni educative	1 e 2	Il feedback orientato al compito e al processo è percepito come più utile del voto (Harks et al., 2014) e sostiene	Per i docenti si tratta in larga parte di usi possibili; per gli studenti di valutazioni espresse

		l'autoregolazione se lo studente può usarlo (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006)	leggendo i descrittori durante l'intervista
T3. La certificazione INVALSI	1	Le prove a bassa posta in gioco riducono motivazione e impegno (Happ & Förster, 2018); gli studenti chiedono restituzioni individuali utilizzabili (Ahmed et al., 2026)	La restituzione individuale esiste, ma è poco conosciuta e arriva a percorso concluso: il problema si sposta dalla disponibilità dell'informazione alla sua appropriazione
T4. Il ruolo di intermediazione del docente tra prova INVALSI e il contesto scolastico	2	L'uso dei dati valutativi dipende dalle rappresentazioni e dalla mediazione dei docenti (Shepard, 2000; Datnow & Park, 2014); la tradizione italiana sottolinea la mediazione nella valutazione (Visalberghi, 1955; Trinchero, 2014)	La mediazione è descritta soprattutto come iniziativa del singolo docente, più che come pratica collegiale
T5. Utilizzo dei dati INVALSI per piani di miglioramento scolastico	2 (livello di scuola)	I dati aggregati alimentano i processi di autovalutazione e miglioramento delle scuole (Freddano & Pastore, 2020)	Riguarda i dati aggregati e non l'esito descrittivo individuale: contribuisce alla seconda domanda solo indirettamente
T6. Criticità percepite sui descrittori INVALSI per un utilizzo nella didattica	1	La comprensione di un feedback richiede la conoscenza dello standard di riferimento (Sadler, 1989) e competenze specifiche di chi lo riceve (Carless & Boud, 2018)	La critica alla terminologia viene da docenti che hanno collaborato con l'INVALSI; la richiesta di materiali di supporto da docenti esterni
T7. Vantaggi percepiti sulla valutazione individuale dello studente	1	Gli studenti attribuiscono valore alla valutazione quando ne riconoscono l'utilità per sé (Chu et al., 2014; Ahmed et al., 2026)	Il valore è espresso da alcuni docenti come conferma esterna della valutazione scolastica e da alcuni studenti come valore personale della prova

6.1 Lettura integrata dei risultati

La discussione dei risultati della presente ricerca prende avvio da una considerazione di fondo secondo cui la restituzione descrittiva individuale degli esiti delle prove dell'INVALSI rappresenta uno spazio in cui si intrecciano finalità istituzionali, pratiche didattiche, processi di attribuzione di significato e possibilità di uso formativo delle informazioni restituite. Questo fa sì che l'interpretazione dei dati debba tenere conto di più fonti, con l'obiettivo di giungere a una lettura più ampia del caso presentato nella ricerca.

Questa scelta è coerente con l'impostazione complessiva dello studio che vede nella triangolazione tra più fonti, il metodo per comprendere il fenomeno in modo più articolato (Denzin, 1978). In questa ricerca, la triangolazione è stata utilizzata soprattutto come confronto tra fonti dove la restituzione descrittiva individuale degli esiti delle prove dell'INVALSI è stata osservata come dispositivo istituzionale, come oggetto discusso nella letteratura sull'*assessment* e come esperienza vissuta da docenti e studenti.

La revisione sistematica ha mostrato che l'*assessment* assume valore educativo quando gli studenti ne riconoscono l'utilità, quando il feedback è comprensibile e quando le informazioni prodotte possono essere collegate ai processi di apprendimento. L'analisi documentale ha evidenziato come, nel contesto italiano, il ruolo dell'INVALSI si sia progressivamente articolato tra funzione di monitoraggio del sistema, supporto ai processi di autovalutazione delle scuole e restituzione individuale degli esiti. L'analisi delle interviste ha infine consentito di osservare come questi strumenti vengano percepiti e interpretati da docenti e studenti nella concretezza dell'esperienza scolastica.

Il punto centrale che attraversa l'intero studio riguarda quindi il passaggio dalla disponibilità del dato alla sua effettiva significatività educativa. Come sottolineano Dewey (1938) e Bruner (1996), un'esperienza diventa educativa quando può essere compresa, rielaborata e collocata in un percorso di senso. Applicata alla valutazione, questa prospettiva suggerisce che un feedback, anche quando è formalmente corretto e tecnicamente fondato, non produce necessariamente apprendimento se non viene compreso da chi lo riceve e non trova un posto riconoscibile nella storia scolastica dello studente. In questo senso, la restituzione descrittiva individuale degli esiti delle prove dell'INVALSI rappresenta una possibilità rilevante, ma non ancora una garanzia di integrazione nei processi educativi.

La prospettiva sistemica aiuta a chiarire ulteriormente questo aspetto. Bateson (1977) invita a leggere i fenomeni educativi come sistemi di relazioni e di comunicazioni, nei quali il significato di un'informazione dipende dalla cornice entro cui viene interpretata. Morin (2001) richiama

l'attenzione sulla necessità di comprendere i fenomeni complessi tenendo insieme le parti e il tutto, evitando di ridurre un processo educativo a un singolo elemento. In questa ricerca, tale impostazione consente di leggere la restituzione descrittiva non come informazione isolata, ma come elemento di un sistema che comprende la prova, il documento di restituzione, i tempi di consegna, il linguaggio dei descrittori, le piattaforme digitali, la mediazione docente e le pratiche didattiche della scuola.
La discussione dei risultati della ricerca è articolata in relazione alle due domande di ricerca. La prima riguarda le percezioni degli studenti e dei docenti sugli esiti descrittivi delle prove nazionali standardizzate di comprensione del testo. La seconda riguarda il modo in cui, a partire dall'esperienza di studenti e docenti, questi esiti possono incidere sui processi di insegnamento e apprendimento.

6.2 Le percezioni di docenti e studenti rispetto agli esiti descrittivi

La prima domanda di ricerca chiedeva quali fossero le percezioni degli studenti e dei docenti riguardo agli esiti descrittivi individuali delle prove nazionali standardizzate sulla comprensione del testo. Gli studenti, di fronte alle descrizioni presentate durante l'intervista, riconoscono una potenziale utilità della restituzione descrittiva, soprattutto quando questa consente di andare oltre il voto o il punteggio e di comprendere in modo più analitico ciò che sono stati in grado di fare. Tuttavia, questa percezione positiva appare fragile, perché dipende dalla conoscenza dello strumento, dalla chiarezza del linguaggio utilizzato e dalla possibilità di collegare l'esito ottenuto a un percorso di miglioramento.
Il tema relativo all'esito descrittivo per la progettazione di azioni educative mostra che gli studenti attribuiscono valore ai descrittori quando li percepiscono come informazioni utili per l'autovalutazione e il miglioramento. In alcune narrazioni, la descrizione dei livelli viene preferita al voto proprio perché consente di comprendere quali aspetti della comprensione del testo siano stati padroneggiati e quali richiedano ulteriore lavoro. Questo risultato si collega direttamente alla letteratura sull'*assessment for learning*, secondo cui il feedback sostiene l'apprendimento quando aiuta lo studente a ridurre la distanza tra la prestazione attuale e gli obiettivi da raggiungere (Hattie & Timperley, 2007). Allo stesso modo, Nicol e Macfarlane-Dick (2006) sottolineano che il feedback è efficace quando sostiene processi di autoregolazione, cioè quando lo studente riesce a usarlo per orientare il proprio modo di apprendere.
Nel caso degli esiti delle prove INVALSI, però, questa possibilità non sembra pienamente realizzata. Gli studenti intervistati mostrano infatti una conoscenza non sempre adeguata della certificazione e della descrizione dei livelli. Alcuni non sanno esattamente com'è strutturato il certificato finale che riceveranno, altri sembrano scoprire lo strumento soltanto nel corso dell'intervista. Questo dato è molto significativo, perché indica che la disponibilità formale dell'esito individuale non coincide con

la sua appropriazione da parte dello studente. In altri termini, l'esito individuale esiste, ma non sempre entra nell'esperienza scolastica come informazione riconosciuta, attesa e interpretabile.

Il confronto con lo studio di Ahmed et al. (2026), il più vicino al caso tra quelli individuati dalla revisione della letteratura, aiuta a precisare questo punto. Nel caso presentato nell'articolo, gli studenti lamentano l'assenza di una restituzione individuale e chiedono report personalizzati; nel caso italiano la restituzione individuale esiste, ma gli studenti intervistati la conoscono poco. Il problema si sposta quindi dalla disponibilità dell'informazione alla sua appropriazione, che la letteratura sul feedback collega alla conoscenza dei criteri e dello standard di riferimento (Sadler, 1989) e alle competenze di chi riceve il feedback (Carless & Boud, 2018).

Le percezioni dei docenti, che la prima domanda di ricerca considera insieme a quelle degli studenti, presentano un quadro in parte diverso. I docenti si esprimono a partire dall'esperienza delle restituzioni degli anni precedenti e riconoscono ai descrittori un valore soprattutto per leggere le competenze degli studenti in ingresso, per la programmazione didattica e per una valutazione di tipo formativo (Tabella 6). Allo stesso tempo, sono soprattutto i docenti a segnalare le criticità: la scarsa informazione sulla certificazione, la restituzione a percorso concluso, il mancato peso dell'esito sulla valutazione scolastica, la terminologia specialistica dei descrittori e la difficoltà di individuare, all'interno di uno stesso livello, aree di miglioramento specifiche (Tabelle 7 e 8). Le osservazioni critiche sui descrittori si distribuiscono in modo diverso tra i due gruppi di docenti. Le critiche alla terminologia specialistica provengono soltanto da due docenti che hanno collaborato con l'INVALSI, cioè da chi conosce meglio lo strumento e i suoi riferimenti teorici; la richiesta di materiali di supporto per leggere i descrittori viene invece da due docenti che non hanno collaborato con l'Istituto; il limite di un livello che non distingue aree di miglioramento specifiche è segnalato da docenti di entrambi i gruppi. La familiarità con lo strumento sembra quindi orientare il tipo di critica più che la sua presenza. Questo quadro è coerente con gli studi italiani che descrivono atteggiamenti ambivalenti dei docenti verso le rilevazioni standardizzate (Benadusi & Giancola, 2020).

Questa distanza tra disponibilità del documento e comprensione effettiva richiama una questione più ampia, emersa anche dalla revisione sistematica. Gli studi analizzati mostrano che gli studenti attribuiscono valore all'*assessment* quando ne comprendono la funzione e quando percepiscono che i risultati possono aiutarli a migliorare. Al contrario, quando la valutazione appare esterna, poco spiegata o non collegata al percorso di apprendimento, il coinvolgimento tende a ridursi. La percezione dell'utilità non è quindi un effetto automatico della restituzione del risultato, ma dipende dal modo in cui la valutazione viene inserita in una cornice educativa riconoscibile.

Un ulteriore elemento riguarda il valore attribuito alla dimensione individuale dell'esito. Gli studenti sembrano attribuire maggiore significato alla prova quando questa non è percepita soltanto come una

rilevazione statistica o istituzionale, ma come un'occasione per ricevere un riscontro personale sulle proprie competenze. Questo dato è particolarmente importante, perché segnala che la valutazione esterna non viene necessariamente rifiutata dagli studenti. Al contrario, essa può essere considerata utile quando restituisce informazioni che riguardano il singolo e che possono essere messe in relazione con il proprio percorso formativo.

Al tempo stesso, l'esito individuale genera anche ambivalenze. Alcuni studenti riconoscono che una prova svolta in un singolo momento può non rappresentare pienamente le competenze effettive, perché la prestazione può essere influenzata da condizioni contingenti, come lo stato emotivo o fisico del giorno della somministrazione. Altri evidenziano che il fatto che la prova non incida direttamente sulla valutazione scolastica può ridurre l'impegno con cui viene affrontata. Questi elementi mostrano che la percezione della restituzione descrittiva è strettamente legata alla percezione complessiva della prova, del suo valore e delle conseguenze che essa produce.

In risposta alla prima domanda di ricerca, si può quindi sostenere che docenti e studenti attribuiscono agli esiti descrittivi delle prove dell'INVALSI una potenziale utilità: i docenti soprattutto per leggere le competenze e programmare, gli studenti per comprendere meglio le proprie competenze di comprensione del testo, sulla base di un documento che diversi di loro hanno conosciuto durante l'intervista. Per entrambi i gruppi questa utilità appare condizionata dal fatto che gli esiti rischiano di essere poco conosciuti, restituiti tardivamente o spesso non accompagnati da una mediazione dei docenti che ne chiarisca il senso.

6.3 L'incidenza degli esiti descrittivi sui processi di insegnamento e apprendimento

La seconda domanda di ricerca chiedeva in che modo, a partire dall'esperienza dei soggetti coinvolti, studenti e docenti, gli esiti descrittivi delle prove standardizzate nazionali relativi ai livelli raggiunti nella prova di comprensione del testo potessero incidere sui processi di insegnamento e apprendimento in termini di significato riconosciuto a tale azione valutativa all'interno del percorso di apprendimento. I risultati suggeriscono che tale incidenza non dipende soltanto dalla presenza di descrittori, ma anche dalle condizioni che possano consentirne un uso pedagogico nel contesto scolastico.

Il primo elemento riguarda l'integrazione dei contenuti delle prove nell'attività didattica. Dalle interviste emerge che i materiali dell'INVALSI vengono utilizzati in modi diversi. In alcuni casi servono a familiarizzare gli studenti con la struttura della prova o con la piattaforma di somministrazione; in altri, invece, diventano strumenti per progettare attività didattiche volte alla comprensione del testo e alla riflessione linguistica. Questa distinzione è rilevante perché mostra due

differenti modalità di rapporto con la valutazione esterna. Da un lato emerge un uso preparatorio, volto soprattutto a ridurre l'incertezza in vista della prova, mentre dall'altro si coglie la possibilità di un uso didattico, in cui i materiali dell'INVALSI entrano nella progettazione ordinaria e diventano risorse per lavorare su competenze rilevanti anche al di fuori della prova.

Questa seconda modalità appare più coerente con la prospettiva di Looney (2011), secondo cui l'integrazione tra valutazione sommativa e formativa richiede che i dati prodotti dalle valutazioni esterne possano essere reinterpretati e utilizzati per orientare l'insegnamento e l'apprendimento. L'analisi documentale mostra che il quadro istituzionale italiano sembra muoversi in questa direzione, soprattutto nel passaggio da una valutazione centrata sulla comparabilità di sistema a forme di restituzione individuale e descrittiva. Inoltre, in alcuni casi emergono esperienze di fruizione dei materiali informativi presenti sul sito INVALSIopen, utilizzati per progettare azioni di integrazione dei contenuti delle prove nell'attività didattica. In coerenza con questo dato, dalle interviste emerge che l'integrazione effettiva dipende in larga misura dal lavoro di mediazione dei docenti e dalla capacità della scuola di trasformare i dati in opportunità per la progettazione didattica.

Va precisato che queste pratiche sono riferite dai docenti e non sono state osservate direttamente. Inoltre, riguardano soprattutto l'uso dei materiali e delle prove; i riferimenti a un uso effettivo delle descrizioni dei livelli restituite ai singoli studenti sono più rari e sono presentati in larga parte come possibilità, con alcune eccezioni: tra i docenti, Ins_CT71 (Tabella 6); tra gli studenti, uno dei due intervistati dopo il diploma, che riferisce di aver usato i livelli ricevuti per ripassare alcuni argomenti in vista di un test di ingresso universitario.

Il secondo elemento riguarda il ruolo del docente. Le analisi delle interviste hanno messo in luce l'importanza che i docenti rivestono nel costruire il significato della prova. Essi possono spiegare agli studenti la finalità della valutazione, chiarire il senso della certificazione, collegare le competenze rilevate alle attività svolte in classe e ridurre la percezione della prova come corpo estraneo rispetto alla didattica ordinaria. Questo risultato è coerente con quanto sostenuto dalla tradizione pedagogica italiana sulla necessità di una mediazione da parte del docente nella valutazione. La valutazione non si esaurisce nella misurazione, ma richiede interpretazione, attribuzione di valore e orientamento dell'azione educativa (Visalberghi, 1955; Trinchero, 2014). Anche la letteratura internazionale sull'uso dei dati valutativi attribuisce alle rappresentazioni e alla mediazione dei docenti un ruolo decisivo (Shepard, 2000; Datnow & Park, 2014). Nelle interviste, tuttavia, questa mediazione appare affidata soprattutto all'iniziativa del singolo docente, più che a pratiche collegiali.

La mediazione del docente è particolarmente importante perché la descrizione del livello, da sola, non sempre consente allo studente di comprendere come agire. Se il descrittore indica un livello raggiunto, ma non viene discusso o collegato a esempi concreti, lo studente può non riuscire a

utilizzarlo per modificare le proprie strategie di apprendimento. In questo senso, l'esito descrittivo delle prove dell'INVALSI sembra richiedere una "traduzione didattica". Non si tratta di semplificarne il contenuto, ma di costruire ponti tra il linguaggio di una prova di valutazione standardizzata e quello dell'esperienza didattica quotidiana.

Il terzo elemento riguarda la tempestività della restituzione. Uno dei risultati più rilevanti emersi dalle interviste è che l'esito delle prove INVALSI spesso arriva quando il percorso scolastico è ormai concluso. Questa condizione limita fortemente la possibilità di utilizzare gli esiti come feedback e quindi in chiave formativa. Se il docente e lo studente non hanno più uno spazio comune per leggere, discutere e interpretare l'esito, l'esito descrittivo rischia di trasformarsi in una certificazione finale priva di ricadute immediate sull'apprendimento. Questo aspetto è particolarmente critico alla luce della letteratura sul feedback, che sottolinea l'importanza di poter utilizzare l'informazione ricevuta per orientare le azioni successive (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). Nei termini di Sadler (1989), lo studente non ha più l'opportunità di utilizzare l'informazione per ridurre lo scarto tra la propria prestazione e lo standard. Questo elemento, peraltro, deriva dalle modalità di restituzione previste dalla normativa (par. 2.7) e trova conferma nelle esperienze riferite dai docenti: è quindi un punto di convergenza tra le fonti più che una scoperta delle sole interviste.

Il quarto elemento riguarda la comprensibilità dei descrittori e la disponibilità di materiali di accompagnamento. Il tema relativo alle criticità dei descrittori mostra che docenti e studenti riconoscono il potenziale informativo dei livelli, ma segnalano anche difficoltà legate alla terminologia specialistica e alla scarsa disponibilità di strumenti formativi per comprenderli. Questo aspetto è decisivo: un feedback può essere tecnicamente accurato e tuttavia poco efficace se non è accessibile ai destinatari. La qualità educativa del feedback non dipende quindi soltanto dalla precisione del descrittore, ma anche dalla possibilità che esso venga letto, compreso e trasformato in azione.

L'incidenza sui processi di insegnamento non riguarda però solo il livello individuale. Il tema relativo all'utilizzo dei dati dell'INVALSI per piani di miglioramento scolastico mostra che i docenti riconoscono anche una funzione organizzativa e collegiale dei risultati. I dati aggregati possono aiutare la scuola a individuare le aree di criticità, a orientare la riflessione dei dipartimenti e a contribuire alla progettazione di azioni di miglioramento. Questo risultato è coerente con l'analisi documentale, che ha evidenziato il ruolo dei dati dell'INVALSI all'interno del Rapporto di Autovalutazione e dei processi di miglioramento previsti dal Sistema Nazionale di Valutazione.

Tuttavia, anche su questo piano, il dato non produce cambiamento per il semplice fatto di essere disponibile. Come osserva Trinchero (2014), la valutazione può diventare agente di cambiamento solo se la scuola attribuisce ai dati un significato corretto e si dispone a utilizzarli come occasione di

apprendimento organizzativo. Questo risultato è coerente con quanto emerso dalle interviste: i dati dell'INVALSI diventano utili quando sono inseriti in processi collegiali di lettura, discussione e progettazione, mentre restano deboli quando vengono considerati come indicatori esterni o come adempimenti burocratici.

In risposta alla seconda domanda di ricerca, si può quindi ipotizzare, sulla base delle esperienze riferite dai partecipanti, che gli esiti descrittivi delle prove dell'INVALSI possano incidere sui processi di insegnamento e apprendimento a condizione che non restino confinati nel documento di certificazione. Secondo docenti e studenti, ciò avviene quando vengono integrati nelle pratiche didattiche, quando sono mediati dai docenti, quando sono presi in considerazione in tempi compatibili con un loro uso educativo e quando sono accompagnati da strumenti che ne facilitano la comprensione. In assenza di queste condizioni, la restituzione descrittiva mantiene una funzione prevalentemente certificativa e fatica a diventare una risorsa per l'apprendimento. La ricerca non consente invece di stabilire se e in che misura queste condizioni producano effetti sugli apprendimenti.

6.4 Conclusioni

L'integrazione dei risultati nell'ambito di questo studio di caso ha permesso di mettere a fuoco che la restituzione individuale dei risultati delle prove dell'INVALSI si ritrova coinvolta in tre funzioni distinte, ovvero la funzione di monitoraggio di sistema, la funzione di restituzione certificativa individuale e la potenziale funzione di restituzione formativa, in termini di possibilità di integrazione tra informazioni di valutazione esterna e il processo di apprendimento sviluppato nel contesto scolastico. Queste funzioni non sono necessariamente in contraddizione tra loro, ma richiedono condizioni specifiche per essere mantenute insieme in modo coerente.

La funzione di sistema è chiaramente rappresentata nell'analisi documentale, ma emerge anche dalle interviste, laddove l'esito delle prove viene ricondotto esclusivamente alle letture dei dati restituiti alle scuole, utili alla costruzione del Rapporto di Autovalutazione o all'individuazione di piani di miglioramento scolastici. Nell'analisi documentale, la normativa generale sottolinea in particolare che, fin dall'istituzione del Sistema Nazionale di Valutazione, l'INVALSI è chiamato a produrre dati affidabili e comparabili sugli apprendimenti, utili a comprendere il funzionamento del sistema scolastico, e tali dati, con il Sistema Nazionale di Valutazione, entrano nei processi di autovalutazione delle scuole e svolgono una funzione di supporto al miglioramento. In questa prospettiva, la valutazione esterna risponde a un'esigenza pubblica di conoscenza, trasparenza e orientamento delle politiche educative. Nei documenti istitutivi dell'Ente di ricerca gli approcci metodologici individuati per rispondere ai compiti istituzionali portano verso un ampliamento di tale funzione di valutazione

di sistema e *accountability*, poiché tende a rivolgere azioni e strumenti informativi e formativi a diversi *stakeholders*. Questa funzione è particolarmente importante rispetto al tema della restituzione descrittiva individuale degli esiti. I risultati empirici della ricerca hanno mostrato che studenti e docenti riconoscono il potenziale dei livelli descrittivi, ma segnalano criticità legate alla scarsa conoscenza della certificazione, alla scarsa comprensibilità dei descrittori e alla necessità di materiali di accompagnamento. In questo quadro, INVALSIopen può essere letto come parte dell'ecosistema documentale e comunicativo attraverso cui l'Istituto cerca di rendere più accessibile il sistema delle prove e dei risultati. Tuttavia, la presenza del portale non garantisce, di per sé, che le informazioni arrivino effettivamente agli studenti o che siano utilizzate nella pratica didattica. Il suo valore dipende dalla capacità delle scuole e dei docenti di integrare tali materiali in percorsi di informazione, mediazione e riflessione.

La funzione certificativa individuale emerge invece con maggiore forza a partire dall'introduzione dei livelli descrittivi nella normativa generale e, quindi, dalla restituzione degli esiti al singolo studente. Questo passaggio è rilevante perché modifica in parte la natura dell'esperienza valutativa. La prova non produce soltanto dati aggregati per il sistema o per la scuola, ma restituisce allo studente un'informazione personale. Tale trasformazione è riconosciuta anche dagli studenti, che attribuiscono maggiore valore alla prova quando la percepiscono come qualcosa che riguarda il proprio percorso e non soltanto come un mero esito statistico o istituzionale.

La funzione formativa, tuttavia, è quella meno automatica e più dipendente dal contesto. Le percezioni raccolte attribuiscono all'esito descrittivo un potenziale educativo, ma questo potenziale resta parzialmente inespresso quando mancano la conoscenza dello strumento, la tempestività nella restituzione, la mediazione del docente e i materiali di supporto. In questo senso, i risultati sono coerenti con l'idea che la valutazione esterna possa avvicinarsi ai processi di apprendimento, ma non diventi formativa per il solo fatto di adottare una restituzione descrittiva. Perché ciò avvenga, è necessario che l'esito della prova nazionale standardizzata entri in un sistema di pratiche in grado di sostenerne l'interpretazione e l'uso. Questa interpretazione è coerente con quanto emerso dalla revisione sistematica, dove gli studi analizzati suggeriscono che la valutazione esterna può assumere valore educativo quando gli studenti ne comprendono la funzione e percepiscono una connessione chiara tra valutazione e apprendimento. Le pratiche valutative, infatti, risultano più promettenti quando sono integrate nei processi di insegnamento, producono informazioni utilizzabili e aiutano studenti e docenti a orientare le azioni successive. Quando invece la valutazione viene percepita come poco utile o scollegata dal percorso formativo, emergono una riduzione della motivazione, uno scarso coinvolgimento e ricadute educative limitate. Questo passaggio è centrale anche per interpretare gli esiti descrittivi delle prove dell'INVALSI. I livelli descrittivi possono offrire un'informazione più

articolata rispetto al punteggio, ma diventano educativamente rilevanti solo se sono compresi dagli studenti, discussi con i docenti e collegati ad azioni didattiche o riflessive. In assenza di queste condizioni, la restituzione descrittiva rischia di rimanere una comunicazione individuale dell'esito, senza trasformarsi in una reale occasione di apprendimento.

Questa questione è particolarmente evidente se si considera il rapporto tra la valutazione esterna e quella interna. Le premesse della ricerca richiamavano la necessità di non considerare tale rapporto come una semplice sommatoria di strumenti diversi, bensì come la costruzione di un sistema valutativo globale. Le esperienze riferite suggeriscono che questa integrazione sia possibile, ma richieda un lavoro intenzionale. La valutazione esterna offre dati e descrittori, la valutazione interna può contestualizzarli, discuterli e metterli in relazione con la storia dello studente. Senza questo raccordo, i due livelli rischiano di rimanere paralleli.

Se essere interna all'Istituto mi ha permesso di osservare in modo più immediato le ragioni istituzionali e metodologiche che sostengono la restituzione descrittiva individuale, dall'altro lato le interviste mi hanno obbligata a sospendere l'idea che l'intenzione istituzionale coincida con l'effetto educativo. Questo è stato un passaggio centrale nella ricerca che mi ha permesso di leggere le criticità non come semplici mancanze informative, ma come segnali di una trasformazione ancora in corso, nella quale il sistema valutativo cerca di avvicinarsi allo studente senza avere ancora costruito tutte le condizioni per rendere quella vicinanza pienamente educativa.

CAPITOLO 7: CONCLUSIONI

Lo studio di caso presentato in questa ricerca ha cercato di superare alcune interpretazioni riduttive della valutazione esterna, spesso limitate tra accettazione senza critiche e rifiuto totale, per analizzare il valore che l'attuale restituzione individuale agli studenti, nelle classi dell'ultimo anno del primo e del secondo ciclo, assume all'interno delle pratiche scolastiche attraverso il racconto di esperienze dirette dei soggetti coinvolti.

Il contributo principale di questa ricerca consiste nell'aver documentato, dal punto di vista di un gruppo circoscritto di docenti e studenti, che la questione della restituzione descrittiva individuale degli esiti delle prove dell'INVALSI non riguarda soltanto la forma del documento o la qualità tecnica dei descrittori ma le condizioni in cui un'informazione valutativa che attiene al contesto scolastico può diventare un'esperienza educativa, in coerenza con l'impianto teorico della ricerca, secondo cui, se l'apprendimento è un processo di costruzione di significati e la valutazione è parte dell'azione educativa, allora un esito valutativo rivolto al singolo studente può essere concepito non solo come una semplice comunicazione di un risultato, ma anche come un'occasione di riflessione.

L'analisi dei sette Temi Generali emersi dalle interviste nella ricerca ha permesso di individuare alcune condizioni che, secondo i partecipanti, potrebbero favorire questa trasformazione e che andrebbero verificate in studi successivi. La prima è la possibilità di integrare i contenuti delle prove nella didattica, intesa come azione che mette in dialogo i contenuti già proposti nel contesto scolastico con quelli presentati nella prova standardizzata nazionale, il che consentirebbe di ridurre la distanza tra la prova standardizzata e le attività scolastiche. La seconda è la comprensibilità dei descrittori, un altro elemento importante che contribuirebbe a sostenere l'integrazione della valutazione esterna nel contesto scolastico, perché, senza una chiara comprensione dei contenuti, la restituzione descrittiva rischierebbe di risultare opaca. La terza è la mediazione del docente, fondamentale nella costruzione di questo ponte e, soprattutto, nella creazione di un valore significativo della valutazione esterna per lo studente. La quarta è la tempestività della restituzione, indispensabile affinché un risultato assuma il valore di un feedback, utile a attivare azioni di autovalutazione e a orientare le azioni successive. La quinta è la costruzione di una cultura valutativa condivisa, in cui i dati esterni siano interpretati non come giudizi conclusivi, ma come informazioni da discutere in itinere, a partire dalle riflessioni sui contenuti oggetto di valutazione nazionale e sugli esiti descrittivi che, al termine, verranno restituiti agli studenti e messi al servizio del miglioramento.

Da questo punto di vista, i risultati di questo studio di caso dialogano anche con la riflessione italiana sulla valutazione educativa. Visalberghi (1955) aveva già evidenziato la necessità di distinguere e, al tempo stesso, di collegare la misurazione e la valutazione. Corsini (2023) sottolinea che la valutazione

educa quando non si limita alla classificazione, ma produce riscontri in grado di sostenere l'insegnamento e l'apprendimento. Le evidenze dello studio seguono questa prospettiva, nel momento in cui vedono nell'esito descrittivo delle prove dell'INVALSI una potenziale funzione educativa quando tali esiti non vengono relegati alla sola logica certificativa, ma diventano un riscontro leggibile, significativo e utilizzabile.

Nel complesso, i risultati suggeriscono che la restituzione individuale degli esiti delle prove dell'INVALSI si trova in una fase di transizione. Sul piano normativo e istituzionale, essa segnala il tentativo di superare una visione della valutazione esterna esclusivamente orientata alla rendicontazione e di aprire uno spazio per una lettura più individuale e descrittiva degli apprendimenti. Sul piano dell'esperienza scolastica, tuttavia, questa trasformazione appare ancora parziale. Studenti e docenti riconoscono il potenziale dei livelli descrittivi, ma segnalano anche criticità che ne limitano l'uso educativo.

Rispetto alla prima domanda di ricerca, docenti e studenti attribuiscono all'esito descrittivo una potenziale utilità: i docenti soprattutto per leggere le competenze degli studenti e per programmare, gli studenti quando l'esito consente di comprendere meglio le competenze raggiunte nella comprensione del testo e quando permette di individuare aspetti da migliorare. Tuttavia, questa utilità è indebolita dalla scarsa conoscenza della certificazione, dalla complessità di alcuni descrittori e dalla mancanza di un accompagnamento stabile nella loro lettura. L'esito descrittivo non viene quindi percepito automaticamente come parte dell'esperienza educativa, ma deve essere reso riconoscibile e significativo. Per la maggior parte degli studenti si tratta di valutazioni espresse leggendo i descrittori durante l'intervista, prima di aver ricevuto il proprio esito.

Rispetto alla seconda domanda di ricerca, le esperienze riferite indicano che gli esiti descrittivi potrebbero incidere sui processi di insegnamento e apprendimento quando vengono integrati nelle pratiche didattiche, discussi con gli studenti e utilizzati dai docenti per progettare interventi. La loro incidenza dipende meno dalla sola esistenza del documento di restituzione e più dalla qualità delle pratiche attraverso cui quel documento viene interpretato. La ricerca, di tipo esplorativo e basata sulle percezioni dei partecipanti, non consente di stabilire effetti sugli apprendimenti.

La scelta di coinvolgere direttamente docenti e studenti come interlocutori, in quanto attori protagonisti nel contesto educativo scolastico, ha evidenziato che la valutazione non è solo un concetto astratto, ma una pratica reale, contestualizzata e interpretata, arricchendo la comprensione di un tema spesso affrontato solo dal punto di vista normativo o tecnico. In questo contesto, la ricerca offre spunti per ulteriori sviluppi, sia dal punto di vista empirico, approfondendo lo studio delle pratiche di restituzione degli esiti delle prove standardizzate in contesti diversi, coinvolgendo anche nuovi interlocutori privilegiati (dirigenti scolastici, il consiglio di classe), sia dal punto di vista

formativo e trasformativo, promuovendo lo sviluppo di competenze valutative che favoriscano un impiego pedagogicamente consapevole della valutazione esterna nel sistema scolastico.
La risposta complessiva alle domande di ricerca porta quindi a ipotizzare che la valutazione esterna possa diventare una risorsa per l'apprendimento soprattutto se inserita in un sistema valutativo coerente, capace di collegare dati standardizzati, la valutazione interna, la mediazione del docente e la riflessione dello studente. In assenza di queste condizioni, la restituzione descrittiva rischia di restare un dispositivo formalmente innovativo ma pedagogicamente debole. La sfida, per la scuola e per il sistema di valutazione, non è dunque soltanto restituire più informazioni, ma creare le condizioni affinché tali informazioni possano essere comprese, discusse e trasformate in apprendimento.

7.1 Limiti della ricerca e proposte future

Rispetto agli esiti mostrati, è bene non perdere di vista quali parti di questa ricerca esplorativa possano costituire potenziali limiti all'effettiva comprensione del fenomeno studiato.
Un limite della ricerca riguarda la selezione dei partecipanti per la raccolta dei dati. La decisione di concentrarsi su docenti e studenti è stata coerente con gli obiettivi e l'approccio metodologico adottato, consentendo un'analisi approfondita delle percezioni e delle pratiche di chi vive direttamente l'esperienza della valutazione scolastica. Tuttavia, i risultati indicano che una comprensione più completa del fenomeno avrebbe potuto essere arricchita coinvolgendo altri attori fondamentali del sistema educativo, come dirigenti scolastici e famiglie, che svolgono un ruolo cruciale nei processi di attribuzione di valore e di interpretazione degli esiti delle prove dell'INVALSI.
Un secondo limite, più rilevante per l'interpretazione dei risultati, riguarda le modalità di reclutamento. Tutti i docenti sono stati raggiunti attraverso reti collegate all'INVALSI: il primo gruppo è composto da docenti che avevano collaborato alla costruzione delle prove, il secondo da docenti individuati tramite reti personali e contatti istituzionali legati all'Istituto. Gli studenti sono stati segnalati dai docenti stessi e sette su nove provengono dalle classi di docenti che avevano collaborato con l'INVALSI. È plausibile che questa composizione abbia favorito una familiarità con le prove e con i materiali di restituzione superiore a quella della popolazione di riferimento, e che la selezione mediata dagli insegnanti abbia privilegiato studenti motivati o vicini alla prospettiva dei propri docenti. A ciò si aggiunge la posizione della ricercatrice, discussa nel paragrafo 4.3.1: le interviste sono state condotte da una ricercatrice interna all'INVALSI, in alcuni casi già conosciuta dai docenti, che ha presentato il proprio ruolo anche come occasione per individuare aspetti delle prove da migliorare. Queste condizioni possono aver orientato alcune risposte verso posizioni più favorevoli o più prudenti. I dati mostrano, tuttavia, che la familiarità con lo strumento non ha prodotto

soltanto valutazioni positive: anche i docenti che avevano collaborato con l'INVALSI hanno espresso critiche ai descrittori, in particolare sulla loro terminologia (par. 6.2). La composizione del gruppo sembra quindi aver inciso sia su quali aspetti sono emersi sia sul tono delle risposte.

Un terzo limite riguarda gli studenti. Sette di loro sono stati intervistati prima della prova dell'ultimo anno e quindi prima di ricevere il proprio esito; durante l'intervista, le descrizioni dei livelli venivano mostrate a schermo e alcuni le hanno conosciute in quell'occasione. Le loro percezioni riguardano quindi un documento presentato dalla ricercatrice più che un esito ricevuto e utilizzato. I due studenti intervistati dopo il diploma mostrano, già all'interno di questo piccolo gruppo, esiti opposti: uno ha usato i livelli ricevuti, l'altro non ha mai aperto la certificazione. Studi successivi potrebbero intervistare gli studenti dopo la restituzione, per osservare se e come l'esito venga effettivamente letto e usato.

La revisione sistematica, inoltre, è stata condotta su un unico database, con lo screening affidato a un solo revisore, e ha individuato pochi studi pertinenti al fenomeno indagato.

Per quanto riguarda invece gli sviluppi futuri della ricerca, questi possono essere ricondotti ad alcune direzioni principali.

Una prima, strettamente collegata alla ricerca e al suo principale limite, può consistere nella selezione di studenti e docenti appartenenti allo stesso contesto scolastico e nel confronto degli esiti delle interviste provenienti da contesti scolastici differenti, approfondendo ulteriori caratteristiche di contesto che possano ampliare il campo di comprensione del fenomeno.

Una seconda direzione, orientata maggiormente verso quella parte di informazioni ricavate dalla ricerca che va nella direzione dell'autovalutazione da parte dell'INVALSI, data la mia posizione lavorativa interna all'istituto. Potrebbero essere sperimentati strumenti standardizzati da mettere a disposizione della scuola, utilizzabili nel contesto scolastico, che offrano informazioni immediate sugli esiti degli studenti, in un'ottica di valutazione formativa totalmente gestita dai docenti. L'obiettivo principale potrebbe consistere in una maggiore integrazione tra la valutazione esterna e quella interna, attraverso modalità valutative più immediate che restituiscano esiti significativi, a supporto del processo di apprendimento e dello sviluppo di processi di autovalutazione da parte dello studente. Alcune indicazioni emerse dalle interviste, in particolare sulla tempestività della restituzione, suggeriscono di esplorare la possibilità di sviluppare da parte dell'INVALSI strumenti valutativi standardizzati, in grado di fornire esiti descrittivi contestuali alla prova. Questi strumenti potrebbero favorire una maggiore continuità valutativa e una funzione diagnostica più chiara, a supporto della didattica.

Una terza prospettiva riguarda l'INVALSI come attore di *policy* educativa, evidenziando l'importanza di adottare procedure sistematiche per valutare l'impatto delle iniziative intraprese, in particolare

nella comunicazione dei risultati agli studenti e alla comunità scolastica in generale. In questo contesto, un dialogo strutturato con le scuole e la raccolta di opinioni da parte di docenti, studenti, dirigenti e famiglie potrebbero contribuire a migliorare le modalità di comunicazione e a rafforzare il valore pedagogico dell'esito individuale.

Per concludere, la sfida, dunque, non è soltanto produrre e restituire dati più articolati sugli apprendimenti, ma anche costruire le condizioni affinché tali dati possano essere compresi, discussi e trasformati in esperienza educativa. Come ricorda Morin (2001), la comprensione dei fenomeni complessi richiede di tenere insieme le parti e il tutto; allo stesso modo, la restituzione descrittiva può acquisire valore formativo, e diventare feedback in senso proprio, solo quando è collegata alla storia dello studente, alla mediazione del docente e alla cultura valutativa della scuola.

BIBLIOGRAFIA

Ahmed, R. M., ALZahrani, A., & ALZahrani, A. (2026). Students' attitudes toward national evaluations and their impact on learning: A mixed method study of the NAFS test. International Journal of Education and Practice, 14(1), 183-201. https://doi.org/10.18488/61.v14i1.4686

Ahmed, S. K. (2024). *The pillars of trustworthiness in qualitative research*. Journal of Medicine, Surgery, and Public Health, 2, 100051. DOI: https://doi.org/10.1016/j.glmedi.2024.100051

Ajello, A. M. (2000). *Apprendimento e competenza: un nodo attuale*. La Nuova Italia.

Ajello, A. M., & Castelli, F. (2017). Valutazione in educazione: Tra misurazione e senso. Cadmo, 25(2), 7-22.

Alderson, P., & Morrow, V. (2011). *The ethics of research with children and young people: A practical handbook*. SAGE Publications.

Alvesson, M., & Sköldberg, K. (2018). Reflexive methodology: New vistas for qualitative research (3rd ed.). Sage.

Faggioli, M., & Allulli, G. (2014). *Migliorare la scuola. Autovalutazione, valutazione e miglioramento per lo sviluppo della qualità.* Roma: Anicia

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001/2022). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. L. Andrade & G. J. Cizek (Eds.), Handbook of formative assessment (pp. 90–105). Routledge.

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. Assessment in Education: Principles, Policy & Practice, 27(4), 350–372. https://doi.org/10.1080/0969594X.2019.1571992

Andrade, H., & Heritage, M. (2018). Using formative assessment to enhance learning, achievement, and academic self-regulation. Routledge.

Arner, T., McCarthy, K. S., & McNamara, D. S. (2021). iSTART StairStepper: Using comprehension strategy training to game the test. Computers, 10(4), 48. https://doi.org/10.3390/computers10040048

Andrade, H. L., Brookhart, S. M., & Yu, E. C. Y. (2021). Classroom assessment as co-regulated learning: A systematic review. Frontiers in Education, 6, Article 751168. https://doi.org/10.3389/feduc.2021.751168

Asquini, G., & Dodman, M. (2018). La sostenibilità della Ricerca-Formazione. In G. Asquini (Ed.), La Ricerca-Formazione. Temi, esperienze, prospettive (pp. 157-169). Franco Angeli.

Asquini, G., Benvenuto, G., & Cesareni, D. (2017). La funzione educativa della valutazione. Teoria e pratiche.

Atkinson, P., & Coffey, A. (2011). Analysing documentary realities. In D. Silverman (Ed.), Qualitative research (3rd ed., pp. 77-92). Sage.

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. Educational Researcher, 36(5), 258–267.

Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA - a scale for the quality assessment of narrative review articles. Research Integrity and Peer Review, 4(1), Article 5. https://doi.org/10.1186/s41073-019-0064-8

Bandura, A. (1997). Self-efficacy: The exercise of control. Freeman.

Barana, A., Fissore, C., & Marchisio, M. (2020). From standardized assessment to automatic formative assessment for adaptive teaching. In Proceedings of the 12th International Conference on Computer Supported Education (CSEDU 2020), Volume 1 (pp. 285-296). SciTePress. https://doi.org/10.5220/0009577302850296

Barbaranelli, C., & D'Olimpio, F. (2007). Analisi dei dati con SPSS. Le basi. LED.

Barrett, H. C. (2007). Researching electronic portfolios and learner engagement: The REFLECT initiative. Journal of Adolescent & Adult Literacy, 50(6), 436–449. https://doi.org/10.1598/JAAL.50.6.2

Barrett, M. (2025). Balancing the scales: How culturally relevant pedagogy mitigates the harmful impacts of standardized testing. Literature Reviews in Education and Human Services, 4(2), 55–67.

Barzanò, G., & Grimaldi, E. (2013). Valutare la scuola. Prove standardizzate e qualità dell'insegnamento. FrancoAngeli.

Bateson, G. (1977). Verso un'ecologia della mente. Adelphi.

Battistin, E., De Santis, F., & Schizzerotto, A. (2015). Merito, valutazione e incentivi nella scuola italiana. Il Mulino.

Bazeley, P. (2013). Qualitative data analysis: Practical strategies. Sage.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74.

Brink, S. (2020). Effective use of large-scale learning assessments to improve learning outcomes and education system reform. UNESCO.

Benadusi, L., & Giancola, O. (2014). Sistemi di scuola secondaria comprensivi versus selettivi: una comparazione in termini di equità. Scuola democratica, 2, 461–482. DOI: https://doi.org/10.12828/77426

Benadusi, L., & Giancola, O. (2020). Equità e merito nella scuola: Teorie, indagini empiriche, politiche. FrancoAngeli.

Benvenuto, G. (2008). La ricerca educativa. Carocci.

Benvenuto, G. (2020). Valutazione e voti nel sistema scolastico e universitario. In Valutare a scuola e all'università (pp. 9–12). Nuova Cultura.

Berg, B. L. (2001). Qualitative research methods for the social sciences (4th ed.). Allyn & Bacon.

Berliner, D. C. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. Cambridge Journal of Education, 41(3), 287–302.

Bevilacqua, A. (2023). Praticare il feedback a scuola: dalla ricerca, suggestioni per un approccio autentico e sostenibile. Pedagogia Oggi, 21(1), 209-219.

Bevilacqua, A., & Girelli, C. (2020). Valutazione sostenibile e feedback literacy in ambito universitario. Una proposta didattica fra complessità e dilemmi. Formazione & insegnamento, 18(2), 154-168.

Biesta, G. J. J. (2010). Good education in an age of measurement: Ethics, politics, democracy. Paradigm Publishers.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. Longmans.

Bloom, B.S. (1971). Mastery learning and its Implications for Curriculum Development. In J.H. Block (Ed.), Mastery Learning: Theory and Practice (pp. 47-63). Rinehart & Winston.

Boarelli, M. (2019). Contro il merito. Laterza.

Booth, A., Sutton, A., & Papaioannou, D. (2016). Systematic approaches to a successful literature review (2nd ed.). SAGE.

Bottani, N. (1986). La più bella del reame. Argomenti per una valutazione della scuola. Torino: SEI..

Bottani, N., & Benadusi, L. (a cura di). (2006). Uguaglianza e equità nella scuola. Erickson.

Bottani, N., & Checchi, D. (a cura di). (2012). La sfida della valutazione. Il Mulino.

Bottani, N., & Tuijnman, A. (1994). International education indicators: Framework, development and interpretation. In A. Tuijnman & N. Bottani (Eds.), Making education count: Developing and using international indicators. OECD Publishing.

Bourdieu, P., & Wacquant, L. J. D. (1992). An invitation to reflexive sociology. University of Chicago Press.

Bowen, G. A. (2009). Document analysis as a qualitative research method. Qualitative Research Journal, 9(2), 27-40. https://doi.org/10.3316/QRJ0902027

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

Braun, V., & Clarke, V. (2013). Successful qualitative research: A practical guide for beginners. Sage.

Braun, V., & Clarke, V. (2016). (Mis)conceptualising themes, thematic analysis, and other problems with Fugard and Potts' (2015) sample-size tool for thematic analysis. International Journal of Social Research Methodology, 19(6), 739–743

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health, 11(4), 589–597.

Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), Handbook of research methods in health social sciences. Springer.

Breakspear, S. (2012). The policy impact of PISA: An exploration of the normative effects of international benchmarking in school system performance. OECD Education Working Papers (No. 71). OECD Publishing.

British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.). BERA.

Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard University Press.

Brooks, C., Burton, R., & Sefton-Green, J. (2019). Feedback for learning. Routledge.

Brown, G. T. L. (2022). The past, present and future of educational assessment: A transdisciplinary perspective. Frontiers in Education, 7. DOI: https://doi.org/10.3389/feduc.2022.1060633

Bruner, J. S. (1996). The culture of education. Harvard University Press.

Bunda, M. A. (1979). Accountability and evaluation. Theory into Practice, 18(5), 357–362.

Burrell, G., & Morgan, G. (1979). Sociological paradigms and organisational analysis: Elements of the sociology of corporate life. Heinemann.

Cain, L. K. (2023). Establishing quality in qualitative research: trustworthiness, validity, and a lack of consensus. In International Encyclopedia of Education. Elsevier.

Calenda, M., & Milito, F. (2020). L'attualità degli studi docimologici. Italian Journal of Educational Research, 24, 102-119.

Caillaud, S., & Flick, U. (2021). *Triangulation. In Research methods in the social sciences: An A-Z of key concepts* (pp. 87–89).

Calvani, A. (2009). Teorie dell'istruzione e carico cognitivo. Erickson.

Carbrey, E. (2024). Scrutinizing the standards: A literature review of the advantages and disadvantages of standardized testing. Journal of Student Research at Indiana University East, 6(1).

Carless, D. (2015). Excellence in university assessment: Learning from award-winning practice. Routledge.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. Assessment & Evaluation in Higher Education, 43(8), 1315–1325.

Carretta, A., & Fontana, S. (2018). Le prove INVALSI: Strumenti per la valutazione degli apprendimenti. Il Mulino.

Castellana, M., & Corsini, C. (2018). La valutazione tra accountability e sviluppo professionale. In G. Asquini (a cura di), La valutazione degli apprendimenti: Strumenti e metodi per la ricerca educativa (pp. 45–60). Carocci.

Castoldi, M. (2009). Valutare gli apprendimenti. Un approccio metacognitivo. Carocci.

Castoldi, M. (2012). Valutare a scuola. Dagli apprendimenti alla valutazione di sistema. Carocci

Castoldi, M. (2013). Valutare a scuola: dagli apprendimenti alla valutazione delle competenze. Carocci.

Castoldi, M. (2015). Valutare e certificare le competenze. Carocci.

Castoldi, M. (2021). Valutare a scuola: Dagli apprendimenti alla valutazione per competenze. Carocci.

Cerini, G. (2015). Prove INVALSI: istruzioni per l'uso. La Scuola.

Checchi, D. (2006). Il valore della scuola. Istruzione e mercato del lavoro in Italia. Il Mulino.

Chelimsky, E. (2012). Valuing, evaluation methods, and the politicization of the evaluation process. New Directions for Evaluation, 2012(133), 77–83. https://doi.org/10.1002/ev.20045

Cheng, L., DeLuca, C., Braund, H., Yan, W., & Rasooli, A. (2020). Teachers' grading decisions and practices across cultures: Exploring the value, consistency, and construction of grades across Canadian and Chinese secondary schools. Studies in Educational Evaluation, 67, 100928. https://doi.org/10.1016/j.stueduc.2020.100928

Chiorri, C. (2020). Fondamenti di psicometria (3ª ed.). McGraw-Hill.

Chu, M.-W., Babenko, O., Cui, Y., & Leighton, J. P. (2014). Using HLM to explore the effects of perceptions of learning environments and assessments on students' test performance. International Journal of Testing, 14(2), 95-121. https://doi.org/10.1080/15305058.2013.841702

Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. The Psychologist, 26(2), 120–123.

Clarke, V., & Braun, V. (2016). Thematic Analysis. In E. Lyons & A. Coyle (Eds.), Analysing qualitative data in psychology (pp. 84-103), (2a ed.). Sage.

Coggi, C., & Notti, A. M. (2002). Docimologia: Teoria e pratica della valutazione. Carocci.

Corbetta, P. (2014). Metodologia e tecniche della ricerca sociale (4ª ed.). Il Mulino.

Corsini, C., & Losito, B. (2013). Le rilevazioni INVALSI: a che cosa servono? Cadmo, XXI(2), 55–76.

Corsini, C., & Nigris, E. (2025). La valutazione alla scuola primaria. Dalla vecchia alla nuova ordinanza: manuale d'uso della valutazione formativa. Sanoma.

Corsini, C. (2013). La valutazione a scuola: tra apprendimento e rendicontazione. FrancoAngeli.

Corsini, C. (2014). Flessibilità e ricerca educativa. La lezione dimenticata: l'INVALSI e la valutazione di scuole e insegnanti. In I. Giunta (a cura di), Flessibilmente. Un modello sistemico di approccio al tema della flessibilità (pp. 175–203). Pensa Multimedia.

Corsini, C. (2018). Scuola, prove Invalsi e culture della valutazione. In Le emergenze educative della società contemporanea (Vol. 2, pp. 937–942). Pensa MultiMedia.

Corsini, C. (2023). La valutazione che educa. Liberare insegnamento e apprendimento dalla tirannia del voto. FrancoAngeli.

Corsini, C., Scierri, I. D. M., & Scionti, R. (2017). INVALSI reading comprehension tests: An analysis of content validity (2010–2016). Italian Journal of Educational Research, 19, 85–104.

Corsini, C., Scierri, I. D. M., & Scionti, A. (2018). La validità delle prove INVALSI di comprensione della lettura (2010–2016). In A. M. Notti (Ed.), La funzione educativa della valutazione. Teorie e pratiche della valutazione educativa (pp. 335–351). Pensa MultiMedia.

Council of the European Union. (2018). Council recommendation of 22 May 2018 on key competences for lifelong learning (2018/C 189/01). Official Journal of the European Union, C 189, 1–13. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018H0604(01)

Cowie, B., & Bell, B. (1999). A Model of Formative Assessment in Science Education. Assessment in Education: Principles, Policy & Practice, 6(1), 101-116.

Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). Sage.

Creswell, J. W., & Poth, C. N. (2024). Qualitative inquiry and research design: Choosing among five approaches (5th ed.). Sage.

Crotty, M. (1998). The foundations of social research: Meaning and perspective in the research process. Sage.

Dahlberg, K., Dahlberg, H., & Nyström, M. (2008). Reflective lifeworld

research. Studentlitteratur.

Damiani, V., Losito, B., & Sanzo, A. (2018). Large-scale assessment on civic and citizenship education: historical perspectives and future research developments. Rivista di storia dell'educazione, 5(2), 185- 203. DOI: https://doi.org/10.4454/rse.v5i2.167

Damiano, E. (2013). La mediazione didattica: Per una teoria dell'insegnamento. FrancoAngeli.

Daniele, K. (2025). Oltre gli standard. Il rigore nella ricerca qualitativa educativa come pratica riflessiva e responsabile. Intervento presentato al Convegno SIPed, Università di Parma.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching. ASCD.

Darling-Hammond, L. (2004). Standards, Accountability, and School Reform. Teachers College Record, 106(6), 1047-1095.

Datnow, A., & Park, V. (2014). Data-driven leadership. Jossey-Bass.

De Angelis, M., Bruni, F., & Petti, L. (2023). Developing feedback literacy in initial teacher education: Results from an investigation. Education Sciences & Society. https://doi.org/10.3280/ess1-2023oa15189

de Landsheere, G. (1973). Elementi di docimologia. Valutazione continua ed esami. Firenze: La Nuova Italia.

de Landsheere, G. (1988). Storia della pedagogia sperimentale. Cento anni di ricerca educativa nel mondo. Roma: Armando.

De Luca, A. M., Lucisano, P., & Zanazzi, S. (2019). [Contributi su valutazione, competenze e transizione scuola-lavoro]. In A. Notti (Ed.), La funzione educativa della valutazione (pp. 647–664). Pensa MultiMedia.

De Mauro, T. (1980). Guida all'uso delle parole. Editori Riuniti.

De Mauro, T. (2016). Il Nuovo vocabolario di base della lingua italiana

De Santis, C. (2022). Competenze e valutazione nel modello DADA. Didattiche per ambienti di apprendimento: pratiche educative e punto di vista genitoriale. Uno studio di caso. Stamen.

De Santis, C., & Asquini, G. (2020). L'autovalutazione delle Scuole DADA: i RAV tra competenze chiave e ambiente di apprendimento RicercAzione, 12(2), 207-223. DOI: http://dx.doi.org/10.32076/RA12205

De Santis, C., & Mastrogiovanni, A. (2024). Comunità di Pratica: il gruppo di lavoro di italiano dell'INVALSI per la sperimentazione CBT di quinta primaria. Lifelong Lifewide Learning (LLL), 22(45), 461 – 469. https://doi.org/10.19241/lll.v22i45.891

De Santis, C., Mastrogiovanni, A., Resio, F., & Vendramin, A. (2024). Le prove dell'INVALSI di italiano per la scuola primaria verso il CBT. Il quadro teorico di riferimento e i costrutti operativi: metodi e strumenti. Working Papers INVALSI, 67/2024. https://www.invalsi.it/download2/wp/wp67_De%20Santis_Mastrogiovanni_Resio_Venframin.doc.pdf

Denzin, N. K., & Lincoln, Y. S. (Eds.). (1994). Handbook of qualitative research. Sage.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). McGraw-Hill.

Desimoni, M. (2018). I livelli descrittivi INVALSI: le basi metodologiche. INVALSI.

Desimoni, M., Cavicchiolo, E., Costanzo, A., & Di Chiacchio, C. (2020). Metodi di valutazione dell'item parameter drift nelle indagini su larga scala: Uno studio empirico sulla prova àncora INVALSI di italiano. In P. Falzetti (Ed.), Implementazione e miglioramento del dato: II Seminario "I dati INVALSI: uno strumento per la ricerca". FrancoAngeli.

Dewey, J. (1938). Experience and education. Macmillan.

Dewey, J. (1960). Teoria della valutazione. La Nuova Italia (ed. orig. 1939).

Dewey, J. (1961). Come pensiamo. La Nuova Italia.

Di Rienzo, P. (2017). Il riconoscimento, la validazione e la certificazione delle competenze nel sistema italiano. In C. Biasin (Ed.), Formazione Continua e Educazione degli Adulti: le dimensioni

formale, non formale e informale nel confronto fra Italia e Francia (pp. 33-41). Li-felong Lifewide Learning, 13 (30).

Domenici, G. (1993). Manuale della valutazione scolastica. Laterza.

Domenici, G. (2007). Manuale della valutazione scolastica. Roma-Bari: Laterza

Domínguez, A. (2024). Teaching dynamics to enhance critical thinking and knowledge socialization in the mathematics classroom. Frontiers in Education, 9, 1388720. https://doi.org/10.3389/feduc.2024.1388720

Dominici, P. (2011). La comunicazione nella società ipercomplessa: Condividere la conoscenza per governare il mutamento. FrancoAngeli.

Dwyer, S. C., & Buckle, J. L. (2009). The space between: On being an insider-outsider in qualitative research. International Journal of Qualitative Methods, 8(1), 54-63. https://doi.org/10.1177/160940690900800105

Earl, L. M. (2003). Assessment as Learning: Using Classroom Assessment to Maximize Student Learning. Corwin Press.

Earl, L. M. (2013). Assessment as learning: Using classroom assessment to maximize student learning (2nd ed.). Corwin Press

Elwood, S. A., & Martin, D. G. (2000).“Placing” interviews: Location and scales of power in qualitative research. The Professional Geographer, 52(4), 649–657. https://doi.org/10.1111/0033-0124.00253

European Commission (2025). Education and Training Monitor 2025: Comparative report. Luxembourg: Publications Office of the European Union

European Commission/EACEA/Eurydice (2020), Equity in school education in Europe: Structures, policies and student performance, Luxembourg, Publications Office of the European Union.

Falzetti, P. (Ed.). (2019). Implementazione e miglioramento del dato. II Seminario “I dati INVALSI: uno strumento per la ricerca” (pp. 9–33). Franco Angeli.

Ferraboschi, A., & Meini, C. (1993). La lettura: Processi cognitivi e comprensione del testo. Il Mulino.

Finlay, L. (2002). “Outing” the researcher: The provenance, process, and practice of reflexivity. Qualitative Health Research, 12(4), 531–545. https://doi.org/10.1177/104973202129120052

Fiorucci, M., & Zizioli, E. (2022). La formazione degli insegnanti: problemi, prospettive e proposte per una scuola di qualità e aperta a tutti e tutte. Pensa Multimedia.

Foy, P., Fishbein, B., Martin, M. O., & Mullis, I. V. S. (2020). TIMSS 2019 user guide for the international database. Boston College, TIMSS & PIRLS International Study Center.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 international report. Springer.

Freddano, M., & Pastore, S. (2020). Il sistema nazionale di valutazione e l’uso dei risultati delle rilevazioni nazionali INVALSI nelle scuole: quali cambiamenti? In P. Falzetti (a cura di), Il dato e il miglioramento scolastico. FrancoAngeli.

Freire, P. (2014). Pedagogia dell’autonomia: saperi necessari per la pratica educativa. Edizioni Gruppo Abele.

Fulcher, G. (2012). Assessment literacy for the language classroom. Language Assessment Quarterly, 9(2), 113–132. https://doi.org/10.1080/15434303.2011.642041

Fullan, M. (2006). The new meaning of educational change (4th ed.). Teachers College Press.

Fullan, M. (2011). Choosing the Wrong Drivers for Whole System Reform. Melbourne: Centre for Strategic Education.

Gamlem, S. M., Segaran, M., & Moltudal, S. (2024). Lower secondary school teachers' arguments on the use of a 26-point grading scale and gender differences in use and perceptions. Assessment in Education: Principles, Policy & Practice, 31(1), 56–74. https://doi.org/10.1080/0969594X.2024.2325365

Gambacorti-Passerini, M. B., & Palmieri, C. (2021). La postura riflessiva nella ricerca educativa. FrancoAngeli.

Gardner, J. (Ed.). (2006). Assessment and Learning. SAGE Publications.

Gentile, M. (2018). Valutare per apprendere: Strategie e strumenti per una valutazione formativa. Carocci.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). Impact Evaluation in Practice. World Bank Publications.

Gilbert, L. S., Jackson, K., & di Gregorio, S. (2014). Tools for analyzing qualitative data. The history and relevance of qualitative data analysis software. In M.J. Spector, M.D. Merrill, J. Elen & M.J. Bishop (Eds.), Handbook of research on educational communications and technology (pp. 221-236). Springer.

Gotch, C. M., & French, B. F. (2019). A systematic review of assessment literacy measures. Educational Measurement: Issues and Practice, 38(2), 14–25. https://doi.org/10.1111/emip.12230

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2017). An introduction to systematic reviews (2nd ed.). SAGE.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. Health Information and Libraries Journal, 26(2), 91-108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. Journal of Education Policy, 24(1), 23–37. https://doi.org/10.1080/02680930802412669

Grimaldi, E., & Pizzi, E. (2015). Le prove INVALSI: Problemi aperti e prospettive future. Giornale Italiano di Pedagogia, 68(2), 223-242.

Grimaldi, E., & Serpieri, R. (2014). Valutazione e governance della scuola: il caso italiano nel contesto europeo. Franc Angeli.

Grimaldi, E., & Serpieri, R. (2016). Neoliberalism, accountability and school development: Policy and practice in the Italian context. Educational Policy, 30(1), 158–180. https://doi.org/10.1177/0895904815614910

Gazzetta Ufficiale della Repubblica Italiana (1977). Legge 4 agosto 1977, n. 517. Norme sulla valutazione degli alunni e sull'abolizione degli esami di riparazione nonché altre norme di modifica dell'ordinamento scolastico. https://www.gazzetta-ufficiale.it/eli/id/1977/-08/18/077U0517/sg.

Gazzetta Ufficiale della Repubblica Italiana (1999). Decreto del Presidente della Repubblica 8 marzo 1999, n. 275. Regolamento recante norme in materia di autonomia delle istituzioni scolastiche, ai sensi dell'art. 21 della legge 15 marzo 1997, n. 59. - https://www.gazzetta-ufficiale.it/eli/id/1999/08/10/099G0339/sg.

Gazzetta Ufficiale della Repubblica Italiana (2017). Decreto Legislativo 13 aprile 2017, n.62. Norme in materia di valutazione e certificazione delle competenze nel primo ciclo ed esami di Stato, a norma

dell'articolo 1, commi 180 e 181, lettera i), della legge 13 luglio 2015, n. 107. (17G00070) (GU n.112 del 16-5-2017 - Suppl. Ordinario n.23).
https://www.gazzettaufficiale.it/eli/id/2017/05/16/17G00070/sg.

Gazzetta Ufficiale della Repubblica Italiana (2013). Decreto del Presidente della Repubblica 28 marzo 2013, n. 80. Regolamento sul sistema nazionale di valutazione in materia di istruzione e formazione. (13G00121). Serie Generale n.155 del 04-07-2013. https://www.gazzettaufficiale.it/eli/id/2013/07/04/13G00121/sg.

Geertz, C. (1973). *The interpretation of cultures*. Basic Books

Guba, E. G. (1990). *The paradigm dialog*. Sage.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth-generation evaluation*. Sage.

Gubrium, J. F., & Holstein, J. A. (2002). *Handbook of interview research: Context and method*. Sage.

Gubrium, J. F., Holstein, J. A., Marvasti, A. B., & McKinney, K. D. (Eds.). (2012). The SAGE Handbook of Interview Research: The Complexity of the Craft. Thousand Oaks, CA: Sage Publications

Guskey, T. R., Roy, P., & Von Frank, V. (2014). Reach the highest standard in professional learning: Data. Corwin Press.

Haladyna, T. M., Downing, S. M., & Rodríguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15(3), 309–334.

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2019). The Role of Standardized Tests in School Reform: A Historical Perspective. Routledge.

Happ, R., & Förster, M. (2018). The correlation between vocational school students' test motivation and the performance in a standardized test of economic knowledge: Using direct and indirect indicators of test motivation. Empirical Research in Vocational Education and Training, 10, 10. https://doi.org/10.1186/s40461-018-0071-x

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. Feminist Studies, 14(3), 575-599. https://doi.org/10.2307/3178066

Harks, B., Rakoczy, K., Hattie, J., Besser, M., Klieme, E., & Blum, W. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. Educational Psychology, 34(3), 269–290. https://doi.org/10.1080/01443410.2013.785384

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. Assessment in Education: Principles, Policy & Practice, 10(2), 169–207. https://doi.org/10.1080/0969594032000121270

Harlen, W., & Deakin Crick, R. (2021). Assessment for learning and building students' capacity for assessment. SAGE Publications.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. Review of Educational Research, 77(1), 81-112.

Hattie, J. (2009). Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge.

Hattie, J., & Clarke, S. (2018). Visible learning: Feedback. Routledge. https://doi.org/10.4324/9780429485480

Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. Educational Research Review, 17, 50–62. https://doi.org/10.1016/j.edurev.2015.12.002

Heritage, M. (2018). Formative Assessment: Making it Happen in the Classroom. Corwin Press.

Hill, H. C., Kapitula, L. R., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. American Educational Research Journal, 48(3), 794–831.

Hitchcock, G., & Hughes, D. (1995). Research and the teacher: A qualitative introduction to school-based research (2nd ed.). Routledge.

Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. Journal of Curriculum Studies, 40(4), 417–456. https://doi.org/10.1080/00220270801989818

Hughes, J. (2010). Philosophy of social research (3rd ed.). Pearson.

IEA (International Association for the Evaluation of Educational Achievement). (2019). IEA's vision and mission. IEA.

IEA (International Association for the Evaluation of Educational Achievement). (2021). Responses to Educational Disruption Survey (REDS): Technical report. IEA.

INVALSI. (2018). Rapporto nazionale INVALSI 2018. Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione.

INVALSI. (2019). Rapporto nazionale INVALSI 2019. Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione.

INVALSI. (2022). Rapporto sulle prove INVALSI. Roma: INVALSI. Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione.

INVALSI. (2024a). Rapporto prove INVALSI 2024. Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione.

INVALSI. (2024b). Tutorial alla lettura dei risultati delle prove INVALSI. Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione.

Italia. (1859) Regio decreto 13 novembre 1859, n. 3725. Riordinamento dell'istruzione pubblica. Gazzetta Piemontese, n. 285, 18 novembre 1859.

Italia. (1877). Legge 15 luglio 1877, n. 3961. Obbligo dell'istruzione elementare. Gazzetta Ufficiale del Regno d'Italia, n. 177, 30 luglio 1877.

Italia. (1923). Regio Decreto 6 maggio 1923, n. 1054. Riordinamento della istruzione media. Gazzetta Ufficiale del Regno d'Italia.

Italia. (1997). Legge 10 dicembre 1997, n. 425, recante disposizioni per la riforma degli esami di Stato conclusivi dei corsi di studio di istruzione secondaria superiore.

Italia. (1999) Decreto del Presidente della Repubblica 8 marzo 1999, n. 275. Regolamento recante norme in materia di autonomia delle istituzioni scolastiche, ai sensi dell'articolo 21 della legge 15 marzo 1997, n. 59. Gazzetta Ufficiale della Repubblica Italiana, n. 186, Supplemento ordinario n. 152, 10 agosto 1999.

Italia. (2004). Decreto legislativo 19 novembre 2004, n. 286, recante istituzione del Servizio nazionale di valutazione del sistema educativo di istruzione e formazione.

Italia. (2024). Legge 1° ottobre 2024, n. 150.

Italia. Decreto-Legge 9 settembre 2025, n. 127. Misure urgenti in materia di istruzione, università, ricerca e alta formazione artistica, musicale e coreutica. (Gazzetta Ufficiale n. 210 del 9 settembre 2025)

Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. The Qualitative Report, 17(42), 1–10. https://doi.org/10.46743/2160-3715/2012.1718

Joanna Briggs Institute. (2024). 10.3.6 Inclusion criteria. JBI Manual for Evidence Synthesis. https://jbiglobal.atlassian.net/wiki/spaces/MANUAL/pages/355862985/10.3.6+Inclusion+Criteria

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 17–64). American Council on Education/Praeger.

Karimov, J., Santos, O. C., & Boticario, J. G. (2024). Learning analytics: A comprehensive review of trends and challenges. Computers & Education: Artificial Intelligence, 5, 100160. https://doi.org/10.1016/j.caeai.2024.100160

Kefallinou, A., & Donnelly, V. (2016). Inclusive assessment: Issues and challenges. European Journal of Special Needs Education, 31(2), 206–220. https://doi.org/10.1080/08856257.2016.1141517

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge University Press.

Klenowski, V. (2012). Assessment for learning in the accountability era: Queensland, Australia. Studies in Educational Evaluation, 38(3–4), 78–83.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological Bulletin, 119(2), 254–284. https://doi.org/10.1037/0033-2909.119.2.254

Koffel, J. B. (2015). Use of recommended search strategies in systematic reviews and the impact of librarian involvement: A cross-sectional survey of recent authors. PLOS ONE, 10(5), e0125931. https://doi.org/10.1371/journal.pone.0125931

Kogan, M. (2022). Education accountability: An analytic overview. Routledge.

Koretz, D. (2008). Measuring up: What educational testing really tells us. Harvard University Press.

Kuckartz, U., & Rädiker, S. (2019). Analyzing qualitative data with MAXQDA: Text, audio, and video. Springer. https://doi.org/10.1007/978-3-030-15671-8

Kvale, S., & Brinkmann, S. (2009). InterViews: Learning the craft of qualitative research interviewing (2nd ed.). Sage.

Kamal, S. S. L. B. A. (2019). Research paradigm and the philosophical foundations of a qualitative study. PEOPLE: International Journal of Social Sciences, 4(3), 1386–1394. https://doi.org/10.20319/pijss.2019.43.13861394

Lamanauskas, V. (2024). Paradigmatic shifts in multiparadigmatic education reality. Problems of Education in the 21st Century, 82(4), 430–433.

Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. Journal of Applied Linguistics, 1(1), 49–72. https://doi.org/10.1558/japl.1.1.49.55872

Lave, J., & Wenger, E. (2012). Situated learning: Legitimate peripheral participation. Cambridge University Press.

Lavinio, C. (2000). Tipi testuali e processi cognitivi. In F. Camponovo & A. Moretti (a cura di), Didattica ed educazione linguistica. La Nuova Italia.

Lave, J., & Wenger, E. (1991/2012). Situated learning: Legitimate peripheral participation. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Le Rose, G., & Ricciardi, V. (2021). Prova strutturata vs prova con svolgimento aperto: gli effetti sul rendimento degli studenti in Italiano. Italiano a Scuola, 3(1), 83-98. https://doi.org/10.6092/issn.2704-8128/12937

Levatino, A., Verger, A., Camphuijsen, M., Termes, A., & Parcerisa, L. (2024). School governance through performance-based accountability: A comparative analysis of its side effects across different regulatory regimes. Review of Research in Education, 48(1), 248–286. https://doi.org/10.3102/0091732X241270672

Levy-Feldman, I. (2025). The role of assessment in improving education and promoting educational equity. Education Sciences, 15(2), 224. DOI: https://doi.org/10.3390/educsci15020224

Lichtner, M. (2004). La qualità delle azioni formative. FrancoAngeli.

Lincoln, Y.S. & Guba, E. G. (1985). *Naturalistic inquiry*. Vol. 75. Sage.

Lincoln, Y. S., & Guba, E. G. (2013). The constructivist credo. Left Coast Press.

Lingard, B., & Sellar, S. (2013). 'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling. Journal of Education Policy, 28(5), 634–656. https://doi.org/10.1080/02680939.2012.758815

Linn, R. L. (2000). Assessments and accountability. Educational Researcher, 29(2), 4–16.

Loiodice, I. (2016). Valutazione della scuola e formazione degli insegnanti: Un'analisi critica. Pedagogia Oggi, 3(2), 127-138.

Looney, J. W. (2011), "Integrating Formative and Summative Assessment: Progress Toward a Seamless System?" OECD Education Working Papers, No. 58, OECD Publishing. http://dx.doi.org/10.1787/5kghx3kbl734

Lopera-Oquendo, C., Lipnevich, A. A., & Máñez, I. (2024). Rating writing: Comparison of holistic and analytic grading approaches in pre-service teachers. Learning and Instruction, 94, Article 101992. https://doi.org/10.1016/j.learninstruc.2024.101992

Lucidi, F. (2013). Problemi di Attendibilità e Validità nelle Prove Standardizzate per la Valuta-zione del Sistema Educativo. https://www.fondazioneagnelli.it/2014/05/13/problemi-attendibilita-validita-nelle-prove-standardizzate-la-valutazione-del-sistema-educativo-fabio-lucidi

Lucidi, F. (2014). Psicometria: Teoria e metodi per la costruzione e la validazione dei test. McGraw-Hill Education.

Lucidi, F. (2016). Attendibilità e validità: le gambe con le quali intraprendere il difficile cammino verso una cultura della valutazione. Giornale italiano di psicologia, 43(3), 489-494.

Lucisano, P. (2008). Misurare la scuola: Conoscere gli apprendimenti e valutare il sistema educativo. Carocci.

Lucisano, P. (2016). La valutazione nella scuola italiana: problemi e prospettive. Carocci.

Lucisano, P., & Corsini, C. (2015). Docenti e valutazione di scuole e insegnanti. Giornale Italiano della Ricerca Educativa, VIII(15), 97–109.

Luhmann, N. (1995). Social systems. Stanford University Press.

Lumbelli, L. (1996). La comprensione come problema: Il punto di vista cognitivo. La Nuova Italia.

Lumbelli, L. (2009). La comprensione come problema: Il punto di vista cognitivo. Laterza.

Lundahl, C., & Waldow, F. (2009). Standardisation and "quick languages": The shape-shifting of standardised measurement of pupil achievement in Sweden and Germany. Comparative Education, 45(3), 365–385. https://doi.org/10.1080/03050060903184969

Lynch, K., & O'Riordan, C. (2020). Equality and Diversity in Education: An Introduction. Policy Press.

Mager, R. F. (1962). Preparing instructional objectives. Fearon Publishers.

Mager, R.F. (1987). Gli obiettivi didattici (G. Di Paolo, Trad.). Giunti & Lisciani. (Opera origi-nale pubblicata nel 1962).

Marion, S. F., Thompson, J., Evans, C., Martineau, J., & Dadey, N. (2019). The challenges and opportunities of balanced systems of assessment: A policy brief. National Center for the Improvement of Educational Assessment.

Marion, S. F., Pellegrino, J. W., & Berman, A. I. (Eds.). (2024). Reimagining balanced assessment systems. National Academy of Education.

Martin, M. O., & Mullis, I. V. S. (2012). Methods and procedures in TIMSS and PIRLS. Boston College.

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). TIMSS 2015 international results in science. Boston College.

Martini, A. (2018). L'effetto scuola (valore aggiunto) nelle prove INVALSI 2018. INVALSI

Mathison, S. (2005). The socio-politics of educational assessment. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), International handbook of educational policy (pp. 133–152). Springer.

Mattei, A., Mastrogiovanni, A., Garuti, R., & Pozio, S. (2016). I quadri di riferimento delle prove INVALSI. In B. Fiore & T. Pedrizzi (a cura di), Valutare per migliorare le scuole. Mondadori Università.

Mattei, F., & Vertecchi, B. (2016). Competenze senza cultura. Educazione. Giornale di pedagogia critica, 5(2), 1–6. https://doi.org/10.14668/Educaz_5201

McCaffrey, D. F., Casabianca, J. M., Ricker-Pedley, K., Lawless, R., & Wendler, C. (2021). *Best Practices for Constructed-Response Scoring*. ETS.Messick, 1989

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). Macmillan.

Mifsud, D., & Wilkins, A. W. (2025). A systematic review of school governance literature between 2000 and 2023. Leadership and Policy in Schools. https://doi.org/10.1080/15700763.2025.2539791

Ministero dell'Istruzione. (2020). Ordinanza ministeriale 4 dicembre 2020, n. 172, recante "Valutazione periodica e finale degli apprendimenti delle alunne e degli alunni delle classi della scuola primaria". https://www.miur.gov.it

Ministero dell'Istruzione e del Merito. (2025). Ordinanza ministeriale 9 gennaio 2025, n. 3, recante disposizioni per la valutazione degli apprendimenti e lo svolgimento degli esami di Stato. https://www.miur.gov.it

Ministero dell'Istruzione e del Merito. (2025). Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione (D.M. 9 dicembre 2025, n. 221).

Ministero dell'Istruzione e del Merito. (2026). Decreto ministeriale 9 gennaio 2026, n. 2 – Disposizioni tecniche e funzionali del Curriculum dello studente nella Piattaforma Unica. Gazzetta Ufficiale della Repubblica Italiana.

Ministero dell'Istruzione, dell'Università e della Ricerca. (2012). Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. (D.M. 254/2012). https://www.mim.gov.it/documents/20182/51310/DM+254_2012.pdf

Ministero dell'Istruzione, dell'Università e della Ricerca. (2018). Indicazioni nazionali e nuovi scenari.

MIUR (2012). Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. http://www.indicazioninazionali.it/wp-content/uploads/2018/08/Indicazioni_Annali_Definitivo.pdf.

Miyamoto, A., Murayama, K., & Lechner, C. M. (2020). The developmental trajectory of intrinsic reading motivation: Measurement invariance, group variations, and implications for reading proficiency. Contemporary Educational Psychology, 63, 101921. https://doi.org/10.1016/j.cedpsych.2020.101921

Montani, P. (2020). Emozioni dell'intelligenza. Un percorso nel sensorio digitale. Meltemi.

Morin, E. (1993). Introduzione al pensiero complesso. Gli strumenti per affrontare la sfida della complessità. Sperling & Kupfer.

Morin, E. (2001). La testa ben fatta: riforma dell'insegnamento e riforma del pensiero. Raffaello Cortina Editore.

Mortari, L. (2003) Apprendere dall'esperienza: Il pensare riflessivo nella formazione. Carocci.

Mortari, L. (2007). Cultura della ricerca e pedagogia: Prospettive epistemologiche. Carocci.

Moss, C. M., & Brookhart, S. M. (2019). Advancing formative assessment in every classroom: A guide for instructional leaders (2nd ed.). ASCD.

Moustakas, C. (1990). Heuristic research: Design, methodology, and applications. Sage Publications.

Mullis, I. V. S., & Martin, M. O. (2012). Using TIMSS and PIRLS to improve teaching and learning. Recherches en éducation, 14. https://doi.org/10.4000/ree.5835

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). TIMSS 2019 assessment frameworks. TIMSS & PIRLS International Study Center, Boston College; International Association for the Evaluation of Educational Achievement.

Mullis, I. V. S., & Martin, M. O. (2017). PIRLS 2016 assessment framework. TIMSS & PIRLS International Study Center.

Mullis, I. V. S., & Martin, M. O. (2019). PIRLS 2021 assessment framework. Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 international results in reading. Boston College.

Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). PIRLS 2021 international results in reading. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb5342

Mura, A., Carta, F., & Ledda, C. (2021). L'uso dei dati INVALSI nella pratica didattica: Percezioni e sfide degli insegnanti. Formazione & Isegnamento, 19(2), 143-157.

Murray, T. S., Kellaghan, T., & Greaney, V. (2009). Using the results of a national assessment of educational achievement. World Bank. https://doi.org/10.1596/978-0-8213-7929-5

NAEP. (2024). Scale Scores and Achievement Levels – Understanding Results. National Center for Education Statistics. Retrieved from https://nces.ed.gov/nationsreportcard/guides/scores_achv.aspx

Natali, C. (2024). L'intervista semi strutturata come strumento di ricerca. In La pratica della ricerca antropologica. Strumenti e metodologie (pp. 19–44). Carocci.

National Assessment Governing Board (NAGB). (2022). NAEP Achievement Levels: Understanding What Students Can Do. Retrieved from https://www.nagb.gov/content/dam/nagb/en/documents/naep/achievement-levels-descriptions.pdf

National Center for Education Statistics. (2022). The condition of education 2022 (NCES 2022-144). U.S. Department of Education.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of General Psychology, 2(2), 175-220. https://doi.org/10.1037/1089-2680.2.2.175

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199–218. DOI: https://doi.org/10.1080/03075070600572090

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. Assessment & Evaluation in Higher Education, 35(5), 501–517.

Nieminen, J. H., & Carless, D. (2023). Feedback literacy: A critical review of contemporary perspectives. Assessment & Evaluation in Higher Education.

Noble, H., & Smith, J. (2025). Ensuring validity and reliability in qualitative research. Evidence-Based Nursing. DOI: https://doi.org/10.1136/ebnurs-2024-104232

Notti, A. M. (2010). La funzione educativa della valutazione. Pensa MultiMedia

Notti, A. M. (2014). A scuola di valutazione. Pensa MultiMedia.

Notti, A. M. (2017). Valutazione educativa e miglioramento. Pensa MultiMedia.

OECD (2005). The definition and selection of key competencies: Executive summary. OECD.

OECD (2013). Synergies for better learning: An international perspective on evaluation and assessment. OECD Publishing. https://doi.org/10.1787/9789264190658-en

OECD (2019). PISA 2018 Assessment and Analytical Framework. OECD Publishing. DOI: https://doi.org/10.1787/b25efab8-en

OECD (2019). PISA 2018 Results (Volume I): What Students Know and Can Do. OECD Publishing

OECD (2023a). Education at a Glance 2023: OECD indicators. OECD Publishing. https://doi.org/10.1787/e13bef63-en

OECD (2023b). PISA 2022 results (Volume I): The state of learning and equity in education. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD (2024). Education at a glance 2024: OECD indicators.

OECD Publishing. https://doi.org/10.1787/c00cad36-en

OECD (2025). Education at a Glance 2025: OECD Indicators. OECD Publishing.

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. Systematic Reviews, 5, 210. https://doi.org/10.1186/s13643-016-0384-4

Ozga, J. (2016). Trust in numbers? Digital education governance and the inspection process. European Educational Research Journal, 15(1), 69–81. https://doi.org/10.1177/1474904115616636

Pagani, V. (2020). Dare voce ai dati. L'analisi dei dati testuali nella ricerca educativa. Junior.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. Frontiers in Psychology, 8, 422. https://doi.org/10.3389/fpsyg.2017.00422

Panciroli, C., & Rivoltella, P.C. (2023). Pedagogia algoritmica. Per una riflessione educativa sull'Intelligenza Artificiale. Scholé

Patton, M. Q. (2015). Qualitative research & evaluation methods (4th ed.). Sage Publications.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), Measures of personality and social psychological attitudes (pp. 17–59). Academic Press.

Pawson, R., & Tilley, N. (1997). Realistic Evaluation. SAGE Publications.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. Educational Psychology Review, 18(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). Knowing what students know: The science and design of educational assessment. National Academy Press.

Pellerey, M. (2004). Le competenze individuali e il portfolio. La Nuova Italia.

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes: Towards a wider conceptual field. Assessment in Education: Principles, Policy & Practice, 5(1), 85–102. https://doi.org/10.1080/0969595980050105

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Blackwell.

Poehner, M. E. (2008). Dynamic assessment: A Vygotskian approach to understanding and promoting L2 development. Springer.

Polikoff, M. S., & Hutt, E. L. (2020). Reforming testing: Lessons from the past for contemporary accountability policy. Harvard Education Press.

Pope, C., & Mays, N. (2006). Qualitative research in health care (3rd ed.). Blackwell Publishing.

Pratesi, D. (2000). Didattica della testualità: Teoria e metodologia della competenza testuale. Armando.

Previtali, D. (2023). Il Sistema nazionale di valutazione: Gli strumenti strategici per il miglioramento [Presentazione]. INVALSI

Purves, A. C. (1987). The evolution of the IEA: A memoir. IEA.

RAND Reading Study Group. (2002). Reading for understanding: Toward an R&D program in reading comprehension. RAND.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. University of Chicago Press.

Ravitch, D. (2010). The death and life of the great American school system: How testing and choice are undermining education. Basic Books.

Rethlefsen, M. L., Farrell, A. M., Osterhaus Trzasko, L. C., & Brigham, T. J. (2015). Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. Journal of Clinical Epidemiology, 68(6), 617–626. https://doi.org/10.1016/j.jclinepi.2014.11.025

Rethlefsen, M. L., Murad, M. H., & Livingston, E. H. (2014). Engaging medical librarians to improve the quality of review articles. JAMA, 312(10), 999–1000. https://doi.org/10.1001/jama.2014.9263

Ricci, R. (2008). La misurazione del valore aggiunto nella scuola. Fondazione Giovanni Agnelli, Programma Education, Working Paper n. 9.

Ricci, R. (2020). Il sistema delle prove INVALSI per il miglioramento. IPRASE.

Rice, M. L. (1986). The analysis of child discourse. In P. Fletcher & M. Garman (Eds.), Language acquisition: Studies in first language development (2nd ed., pp. 244–274). Cambridge University Press.

Richmond, B. (1994). Systems thinking/system dynamics: Let's just get on with it. System Dynamics Review, 10(2–3), 135–157.

Rossi, G., & Giannandrea, L. (2019). Le prove INVALSI come strumento per la valutazione formativa: Il punto di vista degli insegnanti. In Cultura e Scuola: Rivista di scienze dell'educazione (Vol. 1, pp. 11-26). Franco Angeli.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). Evaluation: A Systematic Approach (7th ed.). SAGE Publications.

Rumelhart, D. E. (1980). *Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension* (pp. 33–58). Lawrence Erlbaum.

Rutkowski, L., von Davier, M., & Rutkowski, D. (Eds.). (2014). *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. CRC Press.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

Sadler, D. R. (1989). *Formative assessment and the design of instructional systems*. Instructional Science, 18(2), 119-144.

Sadler, D. R. (2010). *Beyond feedback: Developing student capability in complex appraisal*. Assessment & Evaluation in Higher Education, 35(5), 535–550.

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). SAGE.

Samuk Carignani, Ș., & Burchi, S. (2022). *Preparing for online interviews during Covid-19: The intricacies of technology and online human interaction*. SN Social Sciences, 2, 210. https://doi.org/10.1007/s43545-022-00498-2

Sarracino, G. (2022). *La valutazione come leva per il miglioramento: Il caso delle prove INVALSI nel sistema educativo italiano*. Italian Journal of Educational Research, 15(30), 205-220. (Esempio di articolo italiano recente).

Schizzerotto, A., & Barone, C. (2021). *Sociologia dell'istruzione*. Il Mulino.

Schleicher, A. (2008). PIAAC: *A new strategy for assessing adult competencies. International Review of Education*, 54(5–6), 627–650.

Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. OECD Publishing.

Schön, D. A. (1983). The reflective practitioner: How professionals think in action. Basic Books.

Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). ICCS 2016 assessment framework. Springer.

Scierri, I., & Capperucci, D. (2024). Implementare strategie di autoregolazione dell'apprendimento in classe: una ricerca sulle percezioni di autoefficacia dei docenti e sul ruolo della dimensione valutativa. In R. Viganò & C. Lisimberti (Eds.), A cosa serve la ricerca educativa? Il dato e il suo valore sociale (pp. 819-829). Pensa Multimedia

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), Perspectives of curriculum evaluation (AERA Monograph Series on Curriculum Evaluation, No. 1) (pp. 39–83). Rand McNally.

Scriven, M. (1991). Evaluation thesaurus (4th ed.). SAGE.

Semeraro, R. (2011). *L'analisi qualitativa dei dati di ricerca in educazione*. Italian Journal of Educational Research, (7), 97-106.

Senge, P. M. (2006). *The Fifth Discipline: The Art and Practice of the Learning Organization* (Rev. ed.). New York: Doubleday.

Shepard, L. A. (2000). The role of assessment in a learning culture. Educational Researcher, 29(7), 4–14. https://doi.org/10.3102/0013189X029007004

Silver, C., & Lewins, A. (2014). *Using software in qualitative research: A step-by-step guide* (2nd ed.). SAGE.

Siniscalco, M. T., Bolletta, R., Mayer, M., & Pozio, S. (2007). *Le valutazioni internazionali e la scuola italiana.* Zanichelli.

Sjøberg, S. (2015). *PISA and global educational governance: A critique of the project, its uses and implications.* Eurasia Journal of Mathematics, Science and Technology Education, 11(1), 111–127. https://doi.org/10.12973/eurasia.2015.1310a

Skedsmo, G., & Huber, S. G. (2024). *Navigating educational challenges: Insights into accountability, support and teacher professionalism in diverse contexts. Educational Assessment, Evaluation and Accountability,* 36(1), 1–5. https://doi.org/10.1007/s11092-024-09430-z

Snow, C. E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* RAND.

Soto, C., Gutiérrez de Blume, A. P., Jacovina, M., McNamara, D., Benson, N., & Riffo, B. (2019). *Reading comprehension and metacognition: The importance of inferential skills. Cogent Education*, 6(1), Article 1565067. https://doi.org/10.1080/2331186X.2019.1565067

Stake, R. E. (1995). *The Art of Case Study Research*. SAGE Publications.

Stake, R. (2005). *Qualitative case studies*. In N. K. Denzin & Y. S. Lincoln (Eds.), The Sage handbook of qualitative research (3a ed., pp. 433-466). Sage.

Stame, N. (a cura di). (2001). *Valutazione 2001: Lo sviluppo della valutazione in Italia*. FrancoAngeli.

Stame, N. (2004). *Theory-based evaluation and types of complexity*. Evaluation, 10(1), 58–76. DOI: https://doi.org/10.1177/1356389004042329

Stame, N. (2016). *Valutazione pluralista*. FrancoAngeli. ISBN: 9788891728715.

Stiggins, R. (2005). *From formative assessment to assessment for learning: A path to success in standards-based schools*. Phi Delta Kappan, 87(4), 324–328.

Stiggins, R. J. (2007). *Conceiving and Communication Achievement Targets*. Pearson Education.

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.

Stockmann, R. (2014). *Evaluation and its social and political context. In R. Stockmann (Ed.), Evaluation: A systematic approach* (pp. 61–84). Springer.

Tariq, M. U. (2025). *Ensuring Trustworthiness and Rigor in Qualitative Research. In Qualitative Inquiry in Doctoral Research: Pathways to Effective Design and Implementation* (pp. 365–392). IGI Global. DOI: https://doi.org/10.4018/979-8-3693-7200-5.ch012

Tatto, M. T., Schwille, J., Senk, S., Rodriguez, M., Peck, R., Ingvarson, L., et al. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics: Findings from the IEA Teacher Education and Development Study in Mathematics* (TEDS-M). Springer.

Teltemann, J., & Jude, N. (2019). *Assessments and accountability in secondary education: International trends. Research in Comparative and International Education*, 14(2), 249–271.

Trinchero, R. (2014). Valutazione educativa: Principi, metodi e strumenti. FrancoAngeli

Trinchero, R. (2024). Valutare per apprendere. FrancoAngeli.

Tyler, R. W. (1949). Basic principles of curriculum and instruction. University of Chicago Press.

UNESCO (2006). Education for All Global Monitoring Report: Literacy for Life. Paris: UNESCO

UNESCO. (2020). Global Education Monitoring Report 2020: Inclusion and education: All means all. Paris: UNESCO.

Vaessen, J., & Stame, N. (Eds.) (2024). Theories of Change in Reality: Strengths, Limitations and Future Directions. Routledge

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. Academic Press.

van Manen, M. (1990). Researching lived experience: Human science for an action sensitive pedagogy. State University of New York Press.

van Manen, M. (2016). Researching lived experience: Human science for an action sensitive pedagogy (2nd ed.). Routledge. https://doi.org/10.4324/9781315421056

van Rijn, P., Por, H.-H., McCaffrey, D. F., Bhaduri, I., & Bertling, J. (2024). A framework for comparing large-scale survey assessments: Contrasting India's NAS, United States' NAEP, and OECD's PISA. Frontiers in Education, 9, 1422030. https://doi.org/10.3389/feduc.2024.1422030

Vena, D. (2022). Leggibilità e comprensibilità: un binomio necessario. Italiano LinguaDue, 14(1). https://doi.org/10.54103/2037-3597/18298

Vertecchi, B. (1999). L'archivio docimologico per l'autovalutazione delle scuole. FrancoAngeli.

Vertecchi, B. (2003). Manuale della valutazione. Analisi degli apprendimenti e dei contesti. FrancoAngeli.

Vertecchi, B. (1991). Origini e sviluppi della docimologia. Giunti e Lisciani.

Vertecchi, B. (2010). Decisione didattica e valutazione. La Nuova Italia.

Vertecchi, B. (2011). Valutare la scuola. Laterza.

Visalberghi, A. (1955). Misurazione e valutazione nel processo formativo. Edizioni di Comunità.

Visalberghi, A. (1984). Misurazione e valutazione nel processo educativo. La Nuova Italia.

von Davier, M., & Kennedy, A. M. (Eds.). (2024). PIRLS 2026 assessment frameworks. Boston College, TIMSS & PIRLS International Study Center.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

Vygotskij, L. S. (1990). Pensiero e linguaggio: Ricerche psicologiche (L. Mecacci, a cura di e trad.). Laterza.

Weeden, P., Winter, J., & Broadfoot, P. (2002). Assessment: What's in it for schools? Routledge. https://doi.org/10.4324/9780203468920White, H. (2009). *Theory-Based Impact Evaluation: Principles and Practice. Journal of Development Effectiveness,* 1(2), 85-113.

Wiliam, D. (2010). *Standardized testing and school accountability*. Educational Psychologist, 45(2), 107-122.

Wiliam, D. (2011). *What is assessment for learning? Studies in Educational Evaluation*, 37(1), 3-14. https://doi.org

Wiliam, D. (2018). *Embedded Formative Assessment*. Solution Tree Press.

Williams, A. (2024). Delivering effective student feedback in higher education: An evaluation of the challenges and best practice. International Journal of Research in Education and Science, 10(2), 473-501. DOI: https://doi.org/10.46328/ijres.3404

Wilson, M. (2004). *Assessment, accountability and the classroom: A community of judgment*. Teachers College Record, 106(14), 1–19.

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10, 3087. https://doi.org/10.3389/fpsyg.2019.03087

Yamamoto, K., He, Q., Shin, H. J., & von Davier, M. (2017). Developing a machine-supported coding system for constructed-response items in PISA (ETS Research Report No. RR-17-47). ETS. https://doi.org/10.1002/ets2.12169

Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and Stake. The Qualitative Report, 20(2), 134-152. https://doi.org/10.46743/2160-3715/2015.2102

Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). Sage.

APPENDICE 1 Tabella di sintesi degli studi inclusi nella revisione sistematica

Studio	Paese	Obiettivo dello studio	Popolazione	Campione	Grado scolastico	Tipo di *assessment*	Tipo di feedback	Disegno dello studio	Strumenti di raccolta dati	Outcome principali	Rilevanza per la revisione	Note
Barana et al. (2020)	Italia	Esaminare l'uso di un sistema di valutazione automatizzata per trasformare le prove standardizzate in attività di valutazione formativa e supportare l'insegnamento adattivo.	Insegnanti di scuola secondaria	17 insegnanti; 51 attività/quesiti prodotti dai docenti	Scuola secondaria di primo e secondo grado	*Assessment* automatizzato; trasformazione di item standardizzati in attività formative	Feedback automatico, immediato, interattivo; orientato al processo e alla correzione dell'errore	Studio empirico con intervento/formazione docenti	Questionari ai docenti; analisi delle attività prodotte; riflessioni post-sperimentazione	Percezioni dei docenti; uso del feedback; adattamento didattico; difficoltà degli studenti riportate dai docenti	Rilevante per il ruolo del feedback formativo e dell'*assessment* automatizzato nei processi di insegnamento-apprendimento.	Dati sugli studenti, prevalentemente indiretti e mediati dalle percezioni degli insegnanti.
Chu et al. (2014)	Canada e Stati Uniti	Analizzare le percezioni degli studenti e dei dirigenti su apprendimento e *assessment* e se siano associate alla performance nei test standardizzati.	Studenti quindicenni; dirigenti scolastici	Campione PISA su larga scala: studenti di Canada e Stati Uniti; scuole partecipanti alla rilevazione	Scuola secondaria	Valutazione standardizzata internazionale su larga scala (PISA)	Non centrale; assessment standardizzato e percezioni delle pratiche valutative	Studio quantitativo correlazionale	Dati PISA; questionari studenti; questionari dirigenti; punteggi di reading performance	Performance nei test; percezioni dell'ambiente di apprendimento; percezioni delle attività di *assessment*	Rilevante per il legame tra le percezioni degli studenti, il contesto valutativo e i risultati ai test standardizzati.	Studio che non consente inferenze causali.

Studio	Paese	Obiettivo dello studio	Popolazione	Campione	Grado scolastico	Tipo di *assessment*	Tipo di feedback	Disegno dello studio	Strumenti di raccolta dati	Outcome principali	Rilevanza per la revisione	Note
Arner et al. (2021)	Stati Uniti	Valutare un ambiente ITS gamificato per la pratica di strategie di comprensione della lettura e preparazione a prove standardizzate.	Studenti universitari	Circa 51 partecipanti nelle analisi finali	Target scolastico studenti di scuola secondaria;	Test di comprensione della lettura; pratica in ambiente ITS con compiti simili a prove standardizzate	Feedback automatico;	Studio sperimentale con un disegno sperimentale di controllo	Test di comprensione; questionari di percezione; dati di log del sistema	Percezioni degli studenti; progressione nel sistema; performance di comprensione della lettura	Rilevante per esplorare il potenziale delle tecnologie e del feedback automatico nella preparazione a prove di comprensione.	Studio condotto con studenti universitari ma relativo a un contesto scolastico di scuola secondaria)
Happ & Förster (2018)	Germania	Esaminare la relazione tra motivazione al test, interesse a ricevere feedback e performance in un test standardizzato di conoscenze economiche.	Studenti di scuole professionali	1018 studenti di scuole professionali tedesche	Secondaria superiore / vocational education	Test standardizzato di conoscenze economiche (TEL4-G)	Interesse a ricevere feedback	Studio quantitativo non sperimentale	Test standardizzato; scala di motivazione; indicatore di interesse al feedback;	Performance al test; motivazione; interesse al feedback; numero di risposte mancanti	Rilevante per il ruolo della motivazione nella validità e nell'interpretazione dei risultati delle prove standardizzate.	Feedback trattato come variabile motivazionale indiretta, non come feedback descrittivo individuale.

Studio	Paese	Obiettivo dello studio	Popolazione	Campione	Grado scolastico	Tipo di *assessment*	Tipo di feedback	Disegno dello studio	Strumenti di raccolta dati	Outcome principali	Rilevanza per la revisione	Note
Ahmed et al. (2026)	Arabia Saudita	Indagare gli atteggiamenti, le esperienze, le criticità e i suggerimenti degli studenti in relazione al test nazionale NAFS e alla sua ricaduta percepita sull'apprendimento.	Studenti del primo anno di scuola secondaria che avevano sostenuto il NAFS nell'ultimo anno della scuola secondaria di primo grado	533 studenti nella raccolta quantitativa; 15 studenti nella raccolta qualitativa	Scuola secondaria; esperienza riferita al passaggio dalla scuola secondaria di primo grado alla secondaria di secondo	Valutazione nazionale standardizzata NAFS	Assente	Mixed methods	Questionario su dimensioni cognitive, affettive e comportamentali; interviste semi-strutturate	Atteggiamenti verso il test; utilità percepita; motivazione; difficoltà organizzative; suggerimenti di miglioramento	Molto rilevanti per il valore percepito dell'*assessment*, la necessità di feedback individualizzato e la ricaduta sulla motivazione.	Studio centrato su una valutazione nazionale; evidenzia criticità legate a mancanza di feedback e basso valore percepito della prova.

APPENDICE 2 – Moduli consenso informato

CONSENSO INFORMATO - Modulo informativo (MAGGIORENNI)

Gentile partecipante,

lo studio per cui è richiesta la sua partecipazione si propone di indagare *l'opinione degli studenti e dei docenti in merito ai risultati delle prove INVALSI di Italiano*. La responsabile del progetto è **Antonella Mastrogiovanni**, dottoranda presso il Dipartimento di Scienze Umane dell'Università di Verona e Prima ricercatrice (Responsabile delle prove di Italiano) presso INVALSI, che si occuperà direttamente della raccolta dei dati.

Prima di decidere se dare o meno il suo consenso è importante che lei legga attentamente le seguenti informazioni relative agli obiettivi della ricerca e alle modalità con cui è condotta.

La preghiamo di prendere il tempo necessario per leggere le seguenti informazioni e di non esitare a chiedere chiarimenti o approfondimenti.

La ricerca ha l'obiettivo di comprendere **l'opinione degli studenti e dei docenti** riguardo i **risultati alle prove INVALSI** di Italiano; particolare interesse riguarderà l'opinione relativa alla certificazione rilasciata e alle descrizioni dei livelli che sono riportate sinteticamente nel certificato stesso.

Per realizzare questo studio stiamo chiedendo la partecipazione di **XX docenti** dell'ultimo anno della scuola secondaria di primo e secondo grado e di **XX studenti** dei medesimi gradi scolastici

La partecipazione a questo studio è volontaria, pertanto lei può rifiutarsi di dare il suo consenso. Se decide di accettare, le sarà chiesto di firmare il consenso alla partecipazione alla ricerca e il consenso al trattamento dei dati raccolti attraverso la sua partecipazione. Il consenso può essere ritirato in ogni momento, senza che ciò comporti alcuna conseguenza negativa, e senza la necessità di doverne specificare il motivo.

Se deciderà di partecipare, i dati per lo sviluppo della ricerca verranno raccolti attraverso interviste semi strutturate individuali, che saranno svolte in modalità online (tramite zoom) e avranno una durata di circa 40 minuti; tali interviste verranno registrate per consentire la successiva analisi dei contenuti (esclusivamente la parte audio).

Se decide di accettare, le sarà chiesto di firmare il consenso alla audio-videoregistrazione della intervista.

I dati saranno trattati ai sensi dell'articolo 13 del Regolamento (UE) n. 679/2016 e dal D.L. 196/2003, così come adeguato al D.L. 101/2018 in tema di protezione dei dati personali. Titolare del trattamento

è l'Università degli Studi di Verona, con sede in Via dell'Artigliere n. 8, IT37129, Verona (e-mail: privacy@ateneo.univr.it, PEC: ufficio.protocollo@pec.univr.it, tel. +39 045.8028777). Informazioni ulteriori sul trattamento e la protezione dei dati personali nel caso di ricerche svolte presso l'Università degli Studi di Verona sono reperibili sul sito https://www.univr.it/it/privacy, in particolare in riferimento alle parti che riguardano il trattamento e la protezione dei dati per le persone coinvolte in attività di ricerca. I dati raccolti saranno utilizzati esclusivamente per scopi di ricerca scientifica. Tutte le informazioni raccolte saranno conservate in modo sicuro e ne sarà impedita la visione da parte di estranei. **Qualsiasi informazione che possa identificare i partecipanti sarà rimossa per garantirne l'anonimato. Il materiale sarà custodito dalla responsabile dello studio.**

Le informazioni che verranno raccolte aiuteranno a conoscere in maniera più approfondita il possibile impatto nella scuola dei modelli di certificazioni rilasciate in esito allo svolgimento delle prove INVALSI, in particolare riguardo alle descrizioni che sono riportate come giudizio esplicativo rispetto al livello raggiunto. I risultati della ricerca saranno oggetto della tesi di dottorato di Antonella Mastrogiovanni e saranno divulgati attraverso i consueti canali scientifici: convegni, seminari, pubblicazioni.

In ogni caso questi risultati saranno divulgati in forma aggregata e in nessun caso in forma individuale.

Per qualsiasi domanda o chiarimento è possibile contattare la responsabile dello studio ai seguenti indirizzi mail: antonella.mastrogiovanni@univr.it; a.mastrogiovanni@invalsi.it

Qualora decidesse di ritirare il suo consenso ed interrompere la sua partecipazione alla ricerca, non rinuncerà ad alcun diritto legale acquisito attraverso la partecipazione alla ricerca. Se avesse delle domande in merito ai suoi diritti durante la partecipazione a questa ricerca, o avesse dei dubbi, dei suggerimenti o volesse parlare della ricerca con altri che non siano i ricercatori coinvolti, può contattare Margherita Pasini, presidente del Comitato Etico presso il Dipartimento di Scienze Umane dell'Università di Verona, al numero 0458208558 o scrivere all'indirizzo email: margherita.pasini@univr.it

La ringraziamo per la collaborazione.

Consenso alla partecipazione allo studio

Ho letto (o qualcuno ha letto a me) questo modulo e sono consapevole che mi è stato chiesto di partecipare ad una ricerca. Ho avuto la possibilità di fare domande e di avere risposte soddisfacenti. Io volontariamente sono d'accordo di partecipare a questo studio.

Non rinuncio ad alcun diritto legale firmando questo modulo. Riceverò una copia di questo modulo.

COGNOME E NOME IN STAMPATELLO DEL/LA PARTECIPANTE	**FIRMA DEL/LA PARTECIPANTE**
	Data e ora

Consenso al trattamento dei dati

Acconsento al trattamento dei dati derivanti dalla mia partecipazione.

COGNOME E NOME IN STAMPATELLO DEL/LA PARTECIPANTE	**FIRMA DEL/LA PARTECIPANTE**
	Data e ora

Consenso alla audio-videoregistrazione dell'intervista

Acconsento alla audio-videoregistrazione dell'intervista.

COGNOME E NOME IN STAMPATELLO DEL/LA PARTECIPANTE	**FIRMA DEL/LA PARTECIPANTE**
	Data e ora

Ho spiegato la ricerca al/la partecipante o al/la suo/a rappresentante prima di richiedere la firma (le firme) di cui sopra. Non ci sono parti non compilate in questo documento. Una copia di questo modulo è stata consegnata al/la partecipante o al/la suo/a rappresentante.

Cognome e nome in stampatello della persona che ha ottenuto il consenso	**Firma della persona che ha ottenuto il consenso**
	Data e ora

Nota: una fotocopia di tale modulo firmato va consegnato al/la partecipante.

CONSENSO INFORMATO - Modulo informativo

Gentile genitore/legale rappresentante,

lo studio per cui viene richiesta la partecipazione di sua/o figlia/o si propone di indagare *l'opinione degli studenti e dei docenti in merito ai risultati delle prove INVALSI di Italiano*. La responsabile del progetto è **Antonella Mastrogiovanni**, dottoranda presso il Dipartimento di Scienze Umane dell'Università di Verona e Prima ricercatrice (Responsabile delle prove di Italiano) presso INVALSI, che si occuperà direttamente della raccolta dei dati.

Prima di decidere se dare o meno il suo consenso alla partecipazione di sua/o figlia/o è importante che lei legga attentamente le seguenti informazioni relative agli obiettivi della ricerca e alle modalità con cui sarà condotta.

La preghiamo di prendere il tempo necessario per leggere le seguenti informazioni e di non esitare a chiedere chiarimenti o approfondimenti ai recapiti forniti di seguito. Le sottolineiamo che da ora in avanti useremo il maschile intendendo in realtà sia bambini di sesso maschile che bambine di sesso femminile, per non appesantire il testo.

La ricerca ha l'obiettivo di comprendere **l'opinione degli studenti** e dei docenti riguardo i **risultati alle prove INVALSI** di Italiano; particolare interesse riguarderà l'opinione relativa alla certificazione rilasciata e alle descrizioni dei livelli che sono riportate sinteticamente nel certificato stesso.

Per realizzare questo studio stiamo chiedendo la partecipazione di **XX studenti** dell'ultimo anno della scuola secondaria di primo e secondo grado e di **XX docenti** dei medesimi gradi scolastici.

La partecipazione a questo studio è volontaria, pertanto lei può rifiutare di dare il suo consenso. Se decide di accettare, le sarà chiesto di firmare il consenso alla partecipazione alla ricerca e il consenso all'uso dei dati raccolti. Il consenso può essere ritirato in ogni momento, senza che ciò comporti alcuna conseguenza negativa, e senza la necessità di doverne specificare il motivo.

I dati per lo sviluppo di tale ricerca verranno raccolti attraverso interviste semi strutturate individuali, che si svolgeranno in modalità online (tramite zoom) e che verranno registrate per consentire la successiva analisi dei contenuti (esclusivamente la parte audio).

Ai partecipanti verrà chiesto di rispondere ad alcune domande anche attraverso la lettura condivisa del modello di certificato che riceveranno come esito finale allo svolgimento delle prove INVALSI di Italiano.

I dati saranno trattati ai sensi dell'articolo 13 del Regolamento (UE) n. 679/2016 e dal D.L. 196/2003, così come adeguato al D.L. 101/2018 in tema di protezione dei dati personali. Titolare del trattamento è l'Università degli Studi di Verona, con sede in Via dell'Artigliere n. 8, IT37129, Verona (e-mail: privacy@ateneo.univr.it, PEC: ufficio.protocollo@pec.univr.it, tel. +39 045.8028777). Informazioni ulteriori sul trattamento e la protezione dei dati personali nel caso di ricerche svolte presso l'Università degli Studi di Verona sono reperibili sul sito https://www.univr.it/it/privacy, in particolare in

riferimento le parti che riguardano il trattamento e la protezione dei dati per le persone coinvolte in attività di ricerca.

I dati raccolti saranno utilizzati esclusivamente per scopi di ricerca scientifica. Tutte le informazioni raccolte saranno conservate in modo sicuro e ne sarà impedita la visione da parte di estranei. Qualsiasi informazione che possa identificare i partecipanti sarà rimossa per garantirne l'anonimato. Il materiale sarà custodito dalla responsabile dello studio.

Le informazioni che verranno raccolte, che derivano dalla partecipazione di sua/o figlia/o aiuteranno a conoscere in maniera più approfondita il possibile impatto nella scuola dei modelli di certificazioni rilasciate in esito allo svolgimento delle prove INVALSI, in particolare riguardo alle descrizioni che sono riportate come giudizio esplicativo rispetto al livello raggiunto.

I risultati della ricerca saranno oggetto della tesi di dottorato di Antonella Mastrogiovanni e saranno divulgati attraverso i consueti canali scientifici: convegni, seminari, pubblicazioni.

In ogni caso questi risultati saranno divulgati in forma aggregata e in nessun caso in forma individuale.

Per qualsiasi domanda o chiarimento è possibile contattare la responsabile dello studio ai seguenti indirizzi mail: antonella.mastrogiovanni@univr.it; a.mastrogiovanni@invalsi.it

Qualora decidesse di ritirare il suo consenso ed interrompere la sua partecipazione alla ricerca, non rinuncerà ad alcun diritto legale acquisito attraverso la partecipazione alla ricerca.

Se avesse delle domande in merito ai suoi diritti durante la partecipazione a questa ricerca, o avesse dei dubbi, dei suggerimenti o volesse parlare della ricerca con altri che non siano i ricercatori coinvolti, può contattare Margherita Pasini, presidente del Comitato Etico presso il Dipartimento di Scienze Umane dell'Università di Verona, al numero 0458208558 o scrivere all'indirizzo email: margherita.pasini@univr.it

La ringraziamo per la collaborazione.

Antonella Mastrogiovanni

Data e luogo

Consenso alla partecipazione allo studio del proprio figlio/a

(IL MODULO DEVE ESSERE FIRMATO DA ENTRAMBI I GENITORI)

Ho letto (o qualcuno ha letto a me) questo modulo e sono consapevole che mi è stato chiesto di dare l'autorizzazione affinché mio/a figlio/a partecipi ad una ricerca. Ho avuto la possibilità di fare domande e di avere risposte soddisfacenti. Io volontariamente sono d'accordo affinché mia/o/figlia/o partecipi a questo studio.

Non rinuncio ad alcun diritto legale firmando questo modulo. Riceverò una copia di questo modulo.

COGNOME E NOME IN STAMPATELLO

DEL/LA PARTECIPANTE

COGNOME E NOME IN STAMPATELLO

DEL GENITORE

FIRMA DEL GENITORE

COGNOME E NOME IN STAMPATELLO

DEL GENITORE

FIRMA DEL GENITORE

O COGNOME E NOME IN STAMPATELLO

DEL LEGALE RAPPRESENTANTE

FIRMA DEL LEGALE RAPPRESENTANTE

Data e ora

Consenso al trattamento dei dati

Acconsento al trattamento dei dati derivanti dalla partecipazione di mia/o/figlia/o

COGNOME E NOME
DEL/LA PARTECIPANTE IN STAMPATELLO

COGNOME E NOME DEL GENITORE IN STAMPATELLO	**FIRMA DEL GENITORE**
COGNOME E NOME DEL GENITORE IN STAMPATELLO	**FIRMA DEL GENITORE**
COGNOME E NOME DEL LEGALE RAPPRESENTANTE IN STAMPATELLO	**FIRMA DEL LEGALE RAPPRESENTANTE**
	Data e ora

Ho spiegato la ricerca al/la partecipante o al/la suo/a rappresentante prima di richiedere la firma (le firme) di cui sopra. Non ci sono parti non compilate in questo documento. Una copia di questo modulo è stata consegnata al/la partecipante o al/la suo/a rappresentante.

Nome in stampatello della persona che ha ottenuto il consenso	**Firma della persona che ha ottenuto il consenso**
	Data e ora

Nota: una fotocopia di tale modulo firmato va consegnato al/la partecipante.