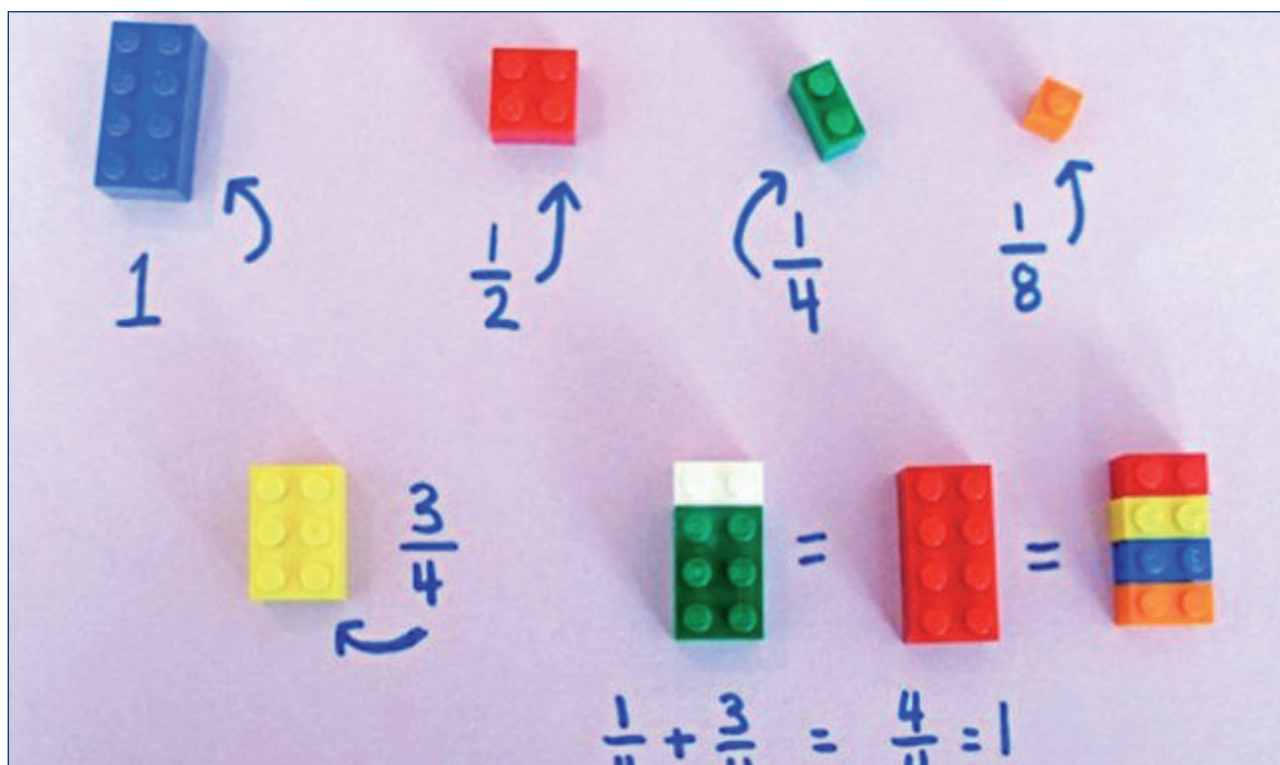




I dati INVALSI: uno strumento per la ricerca

A cura di Patrizia Falzetti





Il presente volume è pubblicato in open access, ossia il file dell'intero lavoro è liberamente scaricabile dalla piattaforma **FrancoAngeli Open Access** (<http://bit.ly/francoangeli-oa>).

FrancoAngeli Open Access è la piattaforma per pubblicare articoli e monografie, rispettando gli standard etici e qualitativi e la messa a disposizione dei contenuti ad accesso aperto. Oltre a garantire il deposito nei maggiori archivi e repository internazionali OA, la sua integrazione con tutto il ricco catalogo di riviste e collane FrancoAngeli ne massimizza la visibilità e favorisce la facilità di ricerca per l'utente e la possibilità di impatto per l'autore.

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio "Informatemi" per ricevere via e.mail le segnalazioni delle novità.



I dati INVALSI: uno strumento per la ricerca

A cura di Patrizia Falzetti

FrancoAngeli

Le opinioni espresse in questo lavoro sono attribuibili esclusivamente agli autori e non impegnano in alcun modo la responsabilità dell'Istituto. Nel citarlo, non è, pertanto, corretto attribuire le argomentazioni ivi espresse all'INVALSI.

Copyright © 2017 by FrancoAngeli s.r.l., Milano, Italy &
INVALSI – Istituto Nazionale per la Valutazione del Sistema educativo di Istruzione e di formazione.

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore ed è pubblicata in versione digitale con licenza
Creative Commons Attribuzione-Non Commerciale-Non opere derivate 3.0 Italia (CC-BY-NC-ND 3.0 IT)

*L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni
della licenza d'uso dell'opera previste e comunicate sul sito
<http://creativecommons.org/licenses/by-nc-nd/3.0/it/legalcode>*

Introduzione

di Anna Maria Ajello

pag. 9

Parte prima I dati INVALSI per la didattica

1. **I dati INVALSI come strumento per la formazione dei docenti**
INVALSI assessment data as a tool for teachers' training
di Roberto Capone, Cristina Coppola » 13
2. **Dalle prove INVALSI alle pratiche didattiche**
From INVALSI tests to the teaching practices
di Mario Castoldi » 19
3. **I legami tra i risultati delle prove di pre-lettura, difficoltà di lettura e reading literacy**
Relations between performance on pre-reading skills, reading difficulty and reading literacy
di Graziella Marrone, Marianna Rasetta, Giselda Di Cesare, Sergio Di Sano » 33
4. **Il database GESTINV delle prove standardizzate INVALSI: uno strumento per la ricerca.**
Alcuni esempi di utilizzo nell'ambito della matematica
The GESTINV database of INVALSI standardized tests: a research tool.
Some examples of utilization in Mathematics
di Giorgio Bolondi, Federica Ferretti, Alessandro Gambini » 43
5. **Un'analisi qualitativa delle prove di matematica**
A qualitative analysis of Mathematics tests
di Federica Ferretti, Alice Lemmo, Andrea Maffia » 49
6. **Le prove INVALSI quale strumento di miglioramento**
INVALSI tests as a tool for improvement
di Maria Brutto » 57
7. **Somministrazione delle prove INVALSI dal 2009 al 2015:
un patrimonio d'informazioni tra evidenze psicometriche e didattiche**
INVALSI tests' administration from 2009 to 2015.
Big data assets between Psychometrics and Didactics
di Giorgio Bolondi, Clelia Cascella » 75

-
8. **“Questione di feedback”:** dati INVALSI e pratiche di valutazione in classe
“A matter of feedback”: INVALSI data and classroom assessment practices
 di Serafina Pastore, Michela Freddano pag. 89
9. **Uno strumento per analizzare l’impatto di una variazione nella formulazione di un quesito INVALSI di matematica**
A tool for analyzing the impact of a variation in the formulation of an INVALSI question in Mathematics
 di Rebecca Boninsegna, Giorgio Bolondi, Laura Branchetti, Chiara Giberti, Alice Lemmo » 101
10. **LIM e rendimenti scolastici degli studenti italiani: un’analisi di impatto su larga scala**
The effects of interactive whiteboards on students’ academic performances: a large scale impact analysis
 di Gianluca Argentin, Tiziano Gerosa » 111

Parte seconda
I dati INVALSI per il governo del sistema scolastico

11. **Il tempo pieno e la dispersione dei voti**
Full-time school scheme and inequality in students’ learning outcomes
 di Giulia Bovini, Marta De Philippis, Paolo Sestito » 125
12. **Predittori della comprensione del testo nei primi anni di scuola primaria: un’analisi multilivello**
Predictors of reading comprehension in early primary school grades: a multilevel analysis
 di Marta Desimoni, Antonella Mastrogiovanni, Alessia Mattei » 137
13. **La Prova nazionale INVALSI e l’esame conclusivo del primo ciclo d’istruzione**
The role of INVALSI national test in the leaving exam of lower secondary education
 di Angela Martini » 153
14. **Stima pesata delle abilità degli studenti nei test standardizzati di profitto attraverso modelli IRT multidimensionali**
Weighted estimate of students’ abilities in standardized assessment tests through multidimensional IRT models
 di Simone Del Sarto, Michela Gnaldi » 165
15. **Tecniche statistiche avanzate per la previsione della performance scolastica**
Advanced statistical models for predicting student performances
 di Anna Siri, Nicola Luigi Bragazzi, Luca Oneto » 175
16. **Oltre l’effetto “in media”: uno studio sulle prestazioni degli studenti nei test INVALSI utilizzando l’approccio quantile**
Thinking beyond the “average case”: exploring students’ performance in INVALSI test through a quantile regression perspective
 di Antonella Costanzo, Marta Desimoni » 185
17. **Non proprio la stessa scuola. Segregazione degli insegnanti tra scuole e abbinamento insegnanti-studenti come meccanismi nascosti di disuguaglianza nel sistema scolastico italiano**
Not really the same school. Teachers inter-school segregation and teachers-students matching as hidden inequality mechanisms in the Italian school system
 di Gianluca Argentin, Giovanni Abbiati, Tiziano Gerosa » 199
-

**18. Differenze di genere e di status socio-economico nel rendimento scolastico:
evidenze empiriche nella scuola primaria**
*Gender and socio-economic gaps in educational achievement:
empirical evidence from primary school*
di Clelia Cascella, Elisa Cavicchiolo

pag. 211

Gli autori

» 223

Introduzione

di Anna Maria Ajello

A che serve la storia? Così incomincia un libro famoso di Marc Bloch intitolato *Apologia della storia o Mestiere di storico*, in cui l'autore immagina di rispondere a questa impegnativa domanda rivolta a lui dal figlio.

Analogamente si potrebbe pensare a questo volume come idealmente rispondente a una domanda simile: a che servono i dati INVALSI?

In diciotto capitoli sono presentati approfondimenti diversi, muovendo da quei dati e sviluppando specifici temi che hanno a che fare direttamente con la didattica, con la formazione dei docenti, con la ricerca e con le scelte politiche conseguenti.

Si tratta di una mole imponente di dati che si presentano in primo luogo come un servizio al Paese, poiché si mette a disposizione della comunità nazionale, siano essi docenti, ricercatori e autorità politiche, un insieme di informazioni che possono dar luogo a scelte politiche fondate su dati attendibili, a comportamenti professionali riflessivi e a sviluppi ulteriori della ricerca.

La varietà e la ricchezza delle ricerche presentate rendono impossibile una loro presentazione in questa sede, per quanto sintetica; mi limiterò a richiamarle inquadrando nei tre ambiti già indicati, quello della formazione docenti, della didattica e della ricerca evidenziando la funzione di fondamento che tali ricerche possono avere per assumere decisioni politiche sulla base di dati attendibili.

Per quanto riguarda la formazione dei docenti, il tema è focalizzato riferendolo ad attività formative che hanno fatto perno sulle prove INVALSI per progettare curricoli sulle competenze, per progettare attività atte a promuovere il miglioramento degli esiti degli alunni, per individuare pratiche didattiche conseguenti all'impostazione riconosciuta a fondamento delle prove che è quella rintracciabile nelle Indicazioni nazionali e nei Quadri di riferimento. Vi è poi uno specifico studio che analizza il feedback fornito ai docenti mediante la restituzione dei dati; mediante l'analisi delle risposte al Questionario infatti, si mette in luce la forte discrasia esistente tra valutazione didattica e valutazione su larga scala; di quest'ultima si sottolinea l'estraneità percepita da parte dei docenti che avvertono la presenza di queste prove come un intervento estemporaneo e non utilizzabile a fini di sviluppo ulteriore della propria professionalità e dell'apprendimento degli studenti.

Un'altra parte piuttosto corposa del volume riporta l'uso dei dati INVALSI per fini di ricerca, come base dati da cui muovere con specifiche ipotesi di lavoro e metodologie conseguenti.

Così vengono espone considerazioni e analisi relative all'uso di database delle prove standardizzate di Matematica, all'impatto su larga scala delle LIM rispetto al rendimento scolastico, all'uso di modelli IRT multidimensionali per la stima pesata delle abilità degli studenti nei test standardizzati.

Questi sono soltanto alcuni dei temi presentati che riguardano un approfondimento specialistico nei metodi e nelle tecniche, a partire dai dati INVALSI; vi sono altri studi che approfondiscono aspetti che hanno una più evidente dimensione politica e informano per consentire quindi decisioni fondate su dati attendibili.

Si tratta di ricerche che riflettono sull'effetto del tempo pieno sulla dispersione dei voti; sulla prova nazionale INVALSI e l'esame conclusivo del primo ciclo di istruzione; sulla segregazione di insegnanti e studenti nella formazione delle classi come meccanismi nascosti di diseguaglianza nel sistema scolastico italiano; sulle differenze di genere e di status socio-economico rispetto al rendimento scolastico.

Come si può vedere tutto il volume si presenta come un panorama ricco adatto a visitatori che hanno interessi diversi, ciascuno dei quali può fare anche incursioni in ambiti meno familiari, sollecitati dalla curiosità intellettuale e/o

dalla rilevanza del tema. A tutti costoro sono rivolti questi studi nella convinzione che già soltanto la presa d'atto della molteplicità delle prospettive orienta l'attenzione verso un atteggiamento attento e sorvegliato quando si parla di scuola perché non basta essere stati studenti per conoscere le caratteristiche articolate del nostro sistema scolastico ed essere abilitati a giudizi sensati.

Alla costruzione di una simile convinzione sono rivolte implicitamente tutte le attività che l'INVALSI svolge, anche se la valutazione mediante prove standardizzate e, più di recente, l'autovalutazione delle scuole mediante il Rapporto di autovalutazione sono quelle più note.

Parte prima
I dati INVALSI per la didattica

1. I dati INVALSI come strumento per la formazione dei docenti

INVALSI assessment data as a tool for teachers' training

di Roberto Capone, Cristina Coppola

In questo lavoro si parla di un progetto di formazione rivolto ai docenti della scuola primaria e secondaria di I grado, voluto dall'USR della Regione Campania. Le domande che hanno condotto alla sua nascita erano "Perché valutare? Che cosa valutare? Come valutare?". Al progetto ha fatto seguito la richiesta da parte di molti dirigenti scolastici di formare i docenti per avvicinarli a una didattica innovativa e accattivante, che sfruttasse le nuove metodologie. Il progetto, già avviato in quattro scuole primarie della Campania, parte dall'analisi dei dati INVALSI per costruire, da questa analisi, l'impalcatura di una progettazione didattica per competenze. Scopo degli incontri di formazione è guidare gli utenti al superamento della logica della didattica trasmissiva e della valutazione di sole conoscenze e abilità, pervenendo a un sistema di valutazione delle competenze e a un sistema di progettazione coerente delle attività didattiche. La direzione è orientata verso l'apprendimento per competenze, nell'ottica della progettazione basata sul lavoro sinergico dei dipartimenti, dei gruppi di classi parallele, delle commissioni, dei consigli di classe, fino all'ambizione più ampia e complessa delle reti di scuole sul territorio. Sono illustrate le diverse tipologie di valutazione, le strategie per la realizzazione di prove orientate alla valutazione di competenze, con riferimento anche ai casi INVALSI e OCSE-PISA. Nel contributo sarà messo in evidenza come la raccolta e l'analisi dei dati è di notevole importanza per strutturare un'azione di formazione dei docenti efficace.

In this work we present an educational project, addressed to primary and secondary school teachers. The project was solicited by the USR of the Campania region. It arose from questions such as "Why assessing? What assessing? How assessing?". It received great interest from the school heads as an activity useful to let the teachers approach some innovative and appealing educational methodologies. The project, which had been already carried out in four primary schools in Campania, begins from the analysis of the INVALSI data to build the foundations of an education by competences. The meetings aimed at driving the participants toward the overcome of the logic of transmissive teaching and of the evaluation based only on knowledge and abilities. The goal was to reach a system of assessment coherent with the educational activities. The direction is that of learning by competences, but with the perspective of a project based on the synergic work among departments, groups of parallel classes, commissions, school boards and the complex network of the schools on the territory. In this work, we describe different kinds of evaluation and strategies adopted to design tests oriented to the evaluation by competences, with a reference to the INVALSI and OCSE-PISA examples. It will also be shown how data collection and analysis are important to organize an effective teachers' education activity.

Nonostante i più recenti dati statistici rilevino miglioramenti dei livelli di competenza in Matematica, specialmente per gli studenti della scuola primaria, esiste ancora un divario tra alcuni Paesi dell'Unione Europea e l'Italia (OECD, 2013). Se poi entriamo nel dettaglio dei risultati italiani, esistono marcate differenze territoriali che distinguono le diverse zone del Paese e i diversi canali della scuola secondaria di II grado, e il divario si accentua con il progredire dei livelli scolari: nella scuola primaria le differenze territoriali sono più ridotte. Con particolare riferimento alla matematica, le rilevazioni nazionali degli apprendimenti 2014-15 e 2015-16 evidenziano risultati eccellenti nel Nord-Est e Lombardia, mentre per il Sud si distingue solo la Puglia. L'attenzione di molti dirigenti scolastici della regione Campania, in linea con l'Ufficio scolastico regionale, che ha attuato un Piano per lo sviluppo del Sistema nazionale di valutazione attra-

verso seminari formativi sulla didattica per competenze della matematica, si sta dirigendo sempre più sulla formazione degli insegnanti in servizio.

Gli insegnanti non hanno solamente il compito e la responsabilità di far imparare nuove conoscenze e abilità ai loro studenti ma sentono essi stessi il bisogno e la responsabilità di imparare e di rinnovare costantemente le loro competenze professionali. Infatti, i docenti che avvertono l'esigenza di ricostruire continuamente le proprie competenze attraverso la formazione permanente, che consente loro di maturare una didattica più flessibile rispetto ai mutati contesti socio-culturali, sono quelli che maggiormente riescono a proporsi come facilitatori dell'apprendimento dei loro studenti. Si è pensato così di strutturare delle attività formative rivolte ai docenti in maniera tale che questi ultimi possano costruire e ricostruire attivamente nuovi modi di pensare, di realizzare e di gestire i propri interventi educativi e possano essere incoraggiati a uscire dalle pratiche di routine e orientarsi verso la rielaborazione e ristrutturazione del proprio modo di fare didattica (Brophy, 2006). Questo processo di rinnovamento non deve essere del singolo docente ma del docente in quanto soggetto inserito in un preciso contesto socio-culturale, che interagisce con la comunità di appartenenza ed è inserito in un processo ciclico di arricchimento della comunità stessa mettendo a disposizione il proprio *know-how* culturale e metodologico. Egli si arricchisce attraverso un processo iterativo di input-output gnostico calandosi in esperienze formative significative. Un'attenta analisi dei risultati delle prove INVALSI può essere un buon punto di partenza per mettere a punto e delineare percorsi formativi per i docenti. Solitamente, di fronte all'evidente risultato negativo alle prove INVALSI, si decide di mettere in atto strategie didattiche direttamente rivolte al gruppo classe, a volte ricorrendo a risorse esterne alla scuola, a volte intensificando le ore di insegnamento disciplinare attraverso attività pomeridiane. Un'altra direzione da seguire, invece, potrebbe essere quella di pensare a percorsi diretti ai docenti, sempre a partire da un'analisi dei risultati delle prove.

In quattro scuole della regione Campania – l'IC “Mercogliano-Guadagni” di Cimitile (NA), l'IC “Picentia” di Pontecagnano (SA), l'IC “Stroffolini” di Casapulla (CE), il Circolo didattico di Baronissi (SA) – il gruppo di ricerca di Didattica della Matematica dell'Università di Salerno sta attuando una sperimentazione didattica per il recupero delle competenze in Matematica, puntando l'attenzione sulla formazione docenti. L'azione di formazione parte dall'analisi dei bisogni formativi dei docenti stessi analizzati attraverso un questionario anonimo elaborato con “Google moduli” e attraverso l'analisi dei risultati delle prove INVALSI. La scelta di consentire ai docenti di rispondere in modo anonimo ha lo scopo di renderli liberi di esprimere i propri bisogni e pensieri.

Entrando nel dettaglio della prova di matematica relativa all'anno scolastico 2014-15 (sono stati analizzati i protocolli delle classi seconde e quinte della scuola primaria e delle classi terze della scuola secondaria di I grado), nelle quattro scuole a cui facciamo riferimento abbiamo potuto osservare forti criticità in tutte le aree; le differenze più significative riguardano gli ambiti Dati e previsioni e Spazio e figure (circa -5% rispetto alla media nazionale). Si ricorda che la prova di matematica è costituita da una serie di item su quattro ambiti (Numeri, Spazio e figure, Dati e previsioni, Relazioni e funzioni). Inoltre, ogni domanda viene collegata a un traguardo delle Indicazioni nazionali, e i traguardi sono a loro volta accorpati in dimensioni (Conoscere, Risolvere problemi, Argomentare). Per quanto riguarda le dimensioni, gli alunni hanno incontrato minori difficoltà nella prima parte (Conoscere), mentre si sono riscontrate maggiori difficoltà nella risoluzione dei problemi e nell'argomentazione.

Dai test somministrati ai docenti si evince che gli ambiti sopracitati sono stati affrontati di meno nella pratica didattica quotidiana. Inoltre, i docenti stessi manifestano un'esigenza formativa in tali ambiti e non solo dal punto di vista dei contenuti quanto piuttosto nei metodi adoperati per proporre agli studenti alcuni argomenti. Si è così pensato di fornire loro proposte didattiche, mostrare attività laboratoriali, coinvolgerli in sperimentazioni. Lo scopo è stato quello di promuovere lo sviluppo di buone/efficaci competenze dei docenti nell'ambito del processo di insegnamento-apprendimento della matematica, tenendo conto sia di saperi didattici generali (connessi all'ambito della progettazione e della valutazione), sia di saperi specifici di Didattica della matematica, e di rafforzare e potenziare le competenze dei docenti nell'ambito della progettualità didattica e della valutazione (*Formative and summative assessment; Assessment for learning*). Si tratta infatti di competenze da considerare in qualche modo trasversali in quanto di fondamentale supporto a quelle più specificamente legate alla didattica disciplinare e che rientrano nell'area di quella che viene definita “professionalità insegnante” (Ferretti e Lovece, 2015). Queste competenze sono di supporto agli insegnanti per individuare contenuti, tecniche, metodologie, strumenti per promuovere e valutare gli apprendimenti in termini di prodotti e di processi (Betti *et al.*, 2014).

Il progetto didattico, a carattere strettamente laboratoriale e pratico, è stato articolato in tre fasi. Una fase preliminare, una fase seminariale, una fase laboratoriale. La fase preliminare del progetto può essere così riassunta:

- il gruppo di ricerca ha analizzato le prove INVALSI relative all'anno scolastico 2014-15 cercando la correlazione tra la valutazione del gruppo classe effettuata dal docente e i risultati ottenuti;
- il gruppo di ricerca ha analizzato i principali bisogni formativi degli studenti in relazione ai risultati delle prove, tenendo conto delle Indicazioni nazionali ma anche di contesti particolari;
- il gruppo di ricerca ha somministrato ai docenti un *test affect* (Zan e Di Martino, 2004; Coppola *et al.*, 2012) da cui si potesse evincere la predisposizione del docente stesso all'insegnamento della Matematica piuttosto che di altre discipline; i test sono stati successivamente analizzati;
- il gruppo di ricerca ha effettuato un'analisi correlata dei bisogni formativi degli studenti a partire dai bisogni formativi dei docenti;
- il gruppo di ricerca ha elaborato un modello di formazione per docenti che ha previsto incontri seminariali sulle moderne metodologie (*role playing*, ricerca-azione, *digital storytelling*) e sull'uso degli strumenti informatici.

Il percorso di formazione attivato parte, dunque, da un'analisi delle prove INVALSI, da un'attenta lettura dei verbali dei nuclei di valutazione di ciascuna scuola e dalle Indicazioni nazionali, e mira a individuare e attivare metodologie didattiche tenendo presenti i traguardi formativi in campo logico-matematico.

Il progetto parte dalla rilevazione delle credenze e delle pratiche didattiche quotidiane, in base alle quali poter individuare i bisogni formativi cui rispondere attraverso la creazione di modelli e strumenti di formazione in grado di promuovere una riflessione critica e la conseguente adozione di metodologie didattiche più efficaci (Ferretti e Lovece, 2015).

La realizzazione dell'indagine correlazionale, effettuata attraverso metodi specifici qualitativi e quantitativi della ricerca (osservazioni, interviste, questionari ecc.), ha messo in rilievo come il successo formativo dello studente sia spesso strettamente legato all'atteggiamento del docente nei confronti dell'insegnamento della matematica. Una delle domande di ricerca emerge dal dubbio ricorrente evidenziato dall'analisi dei *test affects* (Coppola *et al.*, 2012) somministrati ai docenti, ed è come conciliare una didattica per competenze della matematica con le richieste delle prove INVALSI. Queste, infatti, vengono spesso vissute dagli insegnanti come una frattura rispetto al percorso didattico abituale e, per superare l'ostacolo, si ricorre a un mero "addestramento". Per smontare questa falsa credenza, si è partiti dal chiarire il concetto di didattica per competenza. La fase iniziale di formazione, gestita attraverso seminari, ha mirato, infatti, a far entrare il docente nell'ottica polisopica della didattica per competenze. Riferendoci al documento OECD (2013), per competenza matematica si intende "la capacità di una persona di *formulare, utilizzare e interpretare* la matematica in svariati contesti. Tale competenza comprende la capacità di ragionare in modo matematico e di utilizzare concetti, procedure, dati e strumenti di carattere matematico per descrivere, spiegare e prevedere fenomeni. Aiuta gli individui a riconoscere il ruolo che la matematica gioca nel mondo, a operare valutazioni e a prendere decisioni fondate che consentano loro di essere cittadini impegnati, riflessivi e con un ruolo costruttivo" (OECD, 2013, p. 25).

Il costrutto della competenza matematica, così come definita da PISA, pone fortemente l'accento sulla necessità di sviluppare le capacità degli studenti di utilizzare la matematica in un contesto di vita reale. Per poter raggiungere tale capacità, è importante che gli studenti abbiano delle esperienze significative durante le lezioni di matematica in classe. Nella definizione di competenza matematica si sottolinea l'importanza del coinvolgimento attivo in matematica e tale coinvolgimento deve comprendere il ragionamento matematico e l'uso di concetti, procedure, fatti e strumenti matematici per descrivere, spiegare e prevedere fenomeni. In particolare, i verbi "formulare", "utilizzare" ed "interpretare" indicano i tre processi nei quali gli studenti sono coinvolti nel momento in cui risolvono problemi in modo attivo (OECD, 2013). Accanto a questi fattori, riteniamo importante fare riferimento alle componenti affettive e motivazionali necessarie per una buona didattica per competenze della matematica, così come messo in evidenza da Pellerey (2003): la competenza può essere definita come "un sistema coordinato di conoscenze e abilità che sono mobilitate dal soggetto in relazione a uno scopo (un compito, un insieme di compiti o un'azione) che lo interessano e che favoriscono buone disposizioni interne motivazionali e affettive" (Sbaragli, 2011, p. 143). La competenza, inoltre, è la capacità di affrontare un compito o una situazione problematica, riuscendo a mettere in moto e a saper gestire le proprie risorse interne, cognitive, affettive e volitive e a utilizzare opportunamente le risorse esterne disponibili (Pellerey, 2004).

Entrando nello specifico della didattica della matematica, facciamo riferimento agli studi di Zan e Di Martino (2004). Per fare entrare i docenti nell'ottica delle prove INVALSI, si è posto l'accento sui seguenti aspetti, tra loro interconnessi, necessari per la "costruzione" delle prove stesse:

- i processi matematici necessari per collegare il contesto del problema alla matematica e quindi risolvere il problema, e le sette capacità che sottostanno a questi processi: comunicazione, matematizzazione, rappresentazione, ragionamento e argomentazione, escogitare strategie per risolvere problemi, usare un linguaggio simbolico formale e tecnico, usare strumenti matematici;
- i contenuti matematici: Numeri, Spazi e figure, Relazioni e funzioni, Dati e previsioni;
- i contesti nei quali sono ambientati i quesiti: un aspetto importante della competenza matematica è il fatto che la matematica deve essere appresa in situazione e i problemi ambientati in un determinato contesto (Lave e Wenger, 1991). La fase di formazione teorica ha tenuto conto del prospetto sinottico presentato in tab. 1.

Tab. 1 – Fase di formazione teorica: quadro sinottico

Contenuti	Attività	Tempi
Progettazione e valutazione per competenze: il substrato teorico e culturale	Incontro formalizzato attraverso una didattica interattiva. Sono state ripercorse le tappe culturali che hanno condotto alla teorizzazione della didattica per competenze e le tappe normative. Sono state illustrate le diverse tipologie di valutazioni, le strategie per la realizzazione di prove orientate alla valutazione di competenze, con particolare riferimento ai casi INVALSI e OCSE-PISA	2 ore
Progettazione e valutazione per competenze: metodologie didattiche e esempi di attività già realizzate	Incontro formalizzato attraverso attività laboratoriali e didattica <i>peer to peer</i> : sono stati mostrati esempi di progettazioni didattiche per competenze e attività già realizzate	2 ore
Esempi di realizzazione di prove autentiche e delle corrispondenti rubriche valutative in ambito scientifico-matematico	Sono state analizzate alcune prove INVALSI e OCSE-PISA come esempi di prove autentiche. Gli utenti hanno prodotto prove autentiche da sperimentare in aula	2 ore

Nella seconda fase, a carattere laboratoriale, i docenti hanno progettato un'unità di competenza da sperimentare in classe attraverso l'utilizzo di una delle nuove metodologie apprese. Durante la fase della progettazione, i docenti sono stati invitati a lavorare seguendo un'impostazione *Scrum*, non fornendo una struttura rigida di attività che devono essere eseguite pedissequamente, ma offrendo una serie di strumenti (*best practices*) tra cui scegliere quelli che meglio possono essere asserviti al raggiungimento di obiettivi specifici (Mahnic, 2010). La metodologia *Scrum* (termine preso in prestito dal mondo del rugby) è molto utilizzata nella pratica aziendale ed è estendibile anche al mondo della scuola. Essa mira a sfruttare le potenzialità di ciascun individuo nelle dinamiche di gruppo, cercando di creare una pianificazione sinergica delle attività da parte dei partecipanti.

Gli insegnanti hanno lavorato, in un primo momento, in gruppi omogenei, per classi parallele, ponendosi obiettivi unitari. Successivamente, gli insegnanti dei due segmenti dell'istruzione primaria e secondaria di I grado hanno lavorato anche in gruppi eterogenei per provare a costruire percorsi verticali. L'organizzazione del curriculum verticale ha stimolato innovazioni sia sul piano metodologico e dell'organizzazione formativa delle discipline, sia per facilitare connessioni, rapporti, consapevolezza. La convinzione è che se gli insegnanti sono stati aiutati a fare per primi l'esperienza dell'apprendimento cooperativo, possono conoscere meglio la validità di tale approccio, individuarne le risorse e proporlo con entusiasmo e convinzione agli studenti (Polito, 2003).

La cooperazione tra insegnanti è richiesta dall'estrema complessità della realtà formativa, così il ruolo dell'insegnante si dilata ed egli viene a svolgere nuove funzioni educative e didattiche. Successivamente alle attività già realizzate, indichiamo di seguito le attività *work in progress*. In classe, il docente sperimenterà una o più metodologie illustrate e realizzerà un prodotto di intervento da condividere con gli altri docenti attraverso delle giornate di studio e attraverso l'implementazione di un *repository web* per la formazione degli insegnanti in situazioni di insegnamento-apprendimento della matematica. Questo ambiente di apprendimento fungerà da supporto per la formazione degli insegnanti di matematica in quanto vi saranno raccolte diverse tipologie di materiali didattici (esempi di contesti di apprendimento, video di situazioni di insegnamento della matematica, strumenti di valutazione, percorsi di formazione ecc.), da usare per promuovere un corretto utilizzo della valutazione formativa in situazioni di insegnamento-apprendimento (Ferretti e Lovece, 2015).

Ci saranno poi due ulteriori incontri dedicati al confronto e alla condivisione. I docenti mostreranno il prodotto di intervento relativo all'unità di competenza progettata, renderanno partecipi i loro colleghi dell'attività svolta, della metodologia utilizzata, dei risultati ottenuti, delle difficoltà riscontrate. Al gruppo classe, nel frattempo, sarà somministrata

una prova INVALSI per valutare l'efficacia dell'azione formativa e i risultati, debitamente analizzati, saranno mostrati al collegio. La tab. 2 illustra il modo in cui gli incontri sono stati progettati.

Tab. 2 – Progettazione incontri

Contenuti	Attività	Tempi
Laboratorio di rinforzo e condivisione di esperienze in ambito scientifico -matematico	Incontro destinato all'analisi delle attività proposte agli utenti attraverso il confronto condiviso	2 ore
Valutazione a medio termine delle attività d'aula e conclusione plenaria del percorso	Incontro destinato all'analisi delle esperienze realizzate durante il percorso attraverso il confronto condiviso. Sintesi conclusiva	2 ore

Siamo fermamente convinti che “L'insegnante ben preparato appare chiave di volta di tutte le innovazioni educativo-didattiche, quindi fattore determinante per la qualità della scuola. Ogni importante innovazione educativa, infatti, richiede sì una spinta esterna, ma l'attuazione, sia pur lenta, esige l'intelligenza, la preparazione e l'attiva volontà dei docenti” (Chang e Astin, 1997, p. 612).

Pertanto, insistere sulla formazione del docente e invitarlo a sperimentare e innovare pensiamo possa migliorare il rendimento degli alunni percorrendo nuove strade di insegnamento. Inoltre la divulgazione della sua ricerca, favorisce la formazione dei suoi colleghi e la crescita dell'intera comunità scolastica. Sarebbe inoltre interessante analizzare la correlazione tra i risultati delle prove INVALSI e le predisposizioni emotive degli insegnanti verso l'insegnamento di alcuni ambiti della matematica.

Riferimenti bibliografici

- Betti M., Ciani A., Lovece S., Tartufoli L. (2014), “Costruire competenze progettuali e valutative attraverso la didattica laboratoriale. Una ricerca esplorativo-qualitativa nel corso di laurea magistrale in Scienze della formazione primaria dell'Università di Bologna”, *Giornale italiano della ricerca educativa – Italian Journal of Educational Research*, 13: 29-48.
- Brophy J. (2006), *Grade Repetition*, International Institute for Educational Planning, Paris.
- Chang M.J., Astin A.W. (1997), “Who benefits from racial diversity in higher education”, *Diversity Digest*, 1, 2: 13-16.
- Coppola C., Di Martino P., Pacelli T., Sabena C. (2012), “Primary teachers' affect: a crucial variable in the teaching of mathematics”, *Nordic Studies in Mathematics Education*, 17, 3-4: 107-123.
- Di Martino P., Zan R. (2011), “Attitude towards mathematics: a bridge between beliefs and emotions”, *ZDM: The International Journal on Mathematics Education*, 43, 4: 471-482.
- Ferretti F., Lovece S. (2015), “La valutazione formativa per la didattica della matematica nell'ambito del progetto FAMT&L. Le concezioni degli studenti di scuola media nei confronti degli strumenti di verifica utilizzati in classe”, *Ricerche di pedagogia e didattica – Journal of Theories and Research in Education*, 10, 2: 39-68.
- Lave J., Wenger E. (1991), *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press, Cambridge.
- Mahnic V. (2010), “Teaching Scrum through team-project work: students' perceptions and teacher's observations”, *International Journal of Engineering Education*, 26, 1: 96-110.
- OECD (2013), *PISA 2012. Quadro di riferimento analitico per la matematica, la lettura, le scienze, il problem solving e la financial literacy*, testo disponibile al sito: <http://www.invalsi.it/invalsi/ri/pisa2012/documenti/Matematica.pdf>, data di consultazione: 5 maggio 2017.
- Pellerey M. (2003), “Metacognizione e processi affettivi, motivazionali e volitivi”, in O. Albanese (a cura di), *Percorsi metacognitivi. Esperienze e riflessioni*, FrancoAngeli, Milano: 57-74.
- Pellerey M. (2004), *Le competenze individuali e il portfolio*, La Nuova Italia, Firenze.
- Polito M. (2003), *Comunicazione positiva e apprendimento cooperativo: strategie per intrecciare benessere in classe e successo formativo*, Erickson, Gardolo (TN).
- Sbaragli S. (2011), “Le competenze nell'ambito della matematica”, *Difficoltà in matematica*, 7, 2: 143-156.
- Zan R., Di Martino P. (2004), “The role of affect in the research on affect: the case of attitude”, in M.A. Mariotti (ed.), *CERME 3: Proceedings of the Third Conference of the European Society for Research in Mathematics Education*, PLUS, Pisa.

2. Dalle prove INVALSI alle pratiche didattiche

From INVALSI tests to the teaching practices

di Mario Castoldi

L'obiettivo del lavoro di ricerca riguardava il verificare in quale misura i risultati delle prove INVALSI restituiti alle scuole potessero essere utili per avviare azioni di miglioramento della didattica, in particolare dell'italiano e della matematica. La ricerca è stata realizzata in un istituto comprensivo della provincia di Biella e si è caratterizzata come ricerca-intervento: il ricercatore ha formulato una proposta di lavoro al gruppo dei docenti in modo da definire come adattarla allo specifico contesto e svilupparla in termini operativi. La modalità di lavoro concordata con la dirigente e i docenti ha previsto la formazione di sei gruppi di insegnanti suddivisi per gli ambiti (matematica e italiano) e le classi (II e V primaria, e III secondaria di I grado) interessate alle prove. Il percorso di accompagnamento ha previsto cinque incontri: uno iniziale e uno finale svolti in plenaria con tutti i docenti che partecipavano alla formazione e tre limitati ai soli coordinatori. Ciascuno dei sei gruppi di docenti ha svolto tre incontri di lavoro, secondo la seguente scansione:

- *prima fase: dai risultati alle prove.* La prima fase del lavoro consisteva nel partire dai risultati restituiti da INVALSI sulle prove del 2014 per focalizzare l'attenzione sulle competenze richieste dalle prove. La scelta è stata di approfondire nei gruppi i punteggi globali delle classi ritenendoli sia un utile strumento di diagnosi per migliorare l'offerta formativa all'interno della scuola, sia un mezzo per individuare aree di eccellenza e di criticità al fine di potenziare e migliorare l'azione didattica. Completata la fase di analisi generale dei dati, si è proceduto a prendere in esame le tabelle riportanti il dettaglio delle risposte per gli item di matematica e di italiano;
- *seconda fase: dalle prove all'idea di apprendimento.* Durante il secondo incontro, partendo dai dati ottenuti nella prima fase del lavoro, si è spostata l'attenzione sulla quotidianità della didattica d'aula. L'obiettivo era quello di riflettere in gruppo in merito all'idea di apprendimento prevalente tra i docenti, alle prassi didattiche e alle prassi valutative;
- *terza fase: dalle pratiche didattiche alle azioni di miglioramento.* Durante il terzo e ultimo incontro, partendo da alcune domande guida e dal quadro d'insieme, è stato richiesto ai gruppi di predisporre proposte concrete di attività didattiche per poter portare in aula un nuovo modo di fare scuola alla luce dei punti di forza e di debolezza evidenziati nelle precedenti fasi di lavoro.

Oltre alla ricostruzione della procedura nel capitolo vengono richiamati alcuni esempi di titoli delle attività ipotizzate per quanto riguarda il gruppo degli insegnanti delle classi II di matematica, come esempio dei risultati conseguiti.

The goal of the research work concerned the check to what extent the results of the tests INVALSI returned to schools could be helpful to improve the teaching, especially of Italian and Mathematics. The research was carried out in a comprehensive school in the province of Biella and is characterized as action research: the researcher has made a proposal to the teachers' group in order to define how to adapt it to the specific context and develop it in operational terms. The working arrangements agreed with the head of school and the teachers provided six groups of teachers divided by areas (Mathematics and Italian) and classes (II, V and VIII year of school). The accompanying path is structured in five meetings: an initial and a final held in plenary and addressed to all teachers and three limited to only coordinators. Each of the six groups of teachers conducted three working meetings, according to the followings can:

- *first phase: from the tests to the results.* The first phase of the work consisted in starting from the results returned by INVALSI on the 2014 tests to focus on the learning required. The choice has been to deepen the overall scores of the school taking them as a useful diagnostic tool to improve the provision of training within the school and as a mean of identifying areas of strength and weakness in order to enhance and im-

prove the didactic action. After the phase of general data analysis the groups examined the tables bearing the details of the answers to the items of Mathematics and Italian;

- *second phase: from evidence to the idea of learning.* During the second meeting the groups has shifted attention from data analysis to the everyday classroom teaching. The goal was to reflect in group about the idea of learning prevalent among teachers, teaching practices and evaluation practices;
- *third phase: from teaching practices to the improvement actions.* During the third and final meeting, starting from some guiding questions and the big picture, the groups prepared concrete proposals for educational activities in the classroom in order to bring a new kind of school in the light of the strengths and weakness highlighted in previous work steps.

In addition to the reconstruction of the steps in the article are recalled some examples of titles regarding the group of teachers in classes II of Mathematics, as an example of the results achieved.

1. Premessa

I dati delle prove INVALSI rappresentano un’opportunità per rivedere le proprie pratiche didattiche e valutative in chiave migliorativa. Ovviamente non sono l’unico riferimento possibile con cui intraprendere una revisione critica della propria azione professionale; accanto e in alternativa a essi vi sono altre potenziali chiavi di accesso: i dati delle prove di verifica impiegate nell’azione didattica ordinaria, i giudizi valutativi complessivi forniti dalla scuola, le opinioni di studenti e genitori in rapporto al servizio formativo, una valutazione critica del processi di insegnamento “tra pari” ecc. I risultati delle prove INVALSI, peraltro, presentano molteplici motivi di interesse, tali da giustificarne un impiego sistematico e intenzionale come spunto di riflessione in chiave didattico-educativa, da soli o in connessione con altri tipi di dati tra quelli elencati:

- possono essere comparati non solo tra le diverse classi della scuola (come nel caso di prove di verifica di istituto), ma anche con altri istituti scolastici ai diversi livelli territoriali e in relazione ai background delle famiglie degli allievi;
- si fondano su Quadri di riferimento espliciti in merito all’apprendimento che intendono verificare, facilitando quindi una riflessione in chiave professionale;
- si ispirano alle prove di verifica degli apprendimenti adottate a livello internazionale (OCSE-PISA, IEA-TIMMS, IEA-PIRLS), rappresentando quindi un riferimento prospettico per ragionare su quanto l’offerta formativa della scuola risulti allineata con ciò che viene richiesto dal confronto europeo e internazionale;
- prevedono una somministrazione censuaria rivolta a tutte le classi del sistema pubblico di istruzione, costituendo quindi una piattaforma comune tra i diversi istituti scolastici e tra i differenti ordini di scuola su come concettualizzare e come verificare alcuni apprendimenti di base linguistici e matematici;
- richiamano esplicitamente le Indicazioni nazionali e gli altri documenti programmatici vigenti a livello di sistema scolastico nazionale, rappresentando un tentativo di andare oltre a traguardi formativi di tipo generale e di declinare tali riferimenti programmatici comuni in standard di apprendimento a livello operativo;
- rappresentano il tentativo, più o meno riuscito, del sistema scolastico italiano di avvicinarsi a forme di *testing* sugli apprendimenti standardizzate e di sistema presenti in tutti gli altri sistemi scolastici dei Paesi economicamente avanzati.

In questo contributo si intende proporre un percorso operativo per la lettura dei risultati delle prove INVALSI a livello di aula: dopo aver richiamato alcuni principi guida, si richiameranno i passaggi chiave del percorso e si presenteranno alcuni esempi tratti dalla sperimentazione del percorso stesso compiuta con scuole e gruppi di scuole.

2. Principi-guida

Le diffuse forme di diffidenza e resistenza manifestate nel corpo docente anche negli ultimi anni sono indizio di una scarsa comprensione del significato potenziale di queste prove in chiave professionale; senza addentrarci troppo su questi aspetti intendiamo solo segnalare la difficoltà di utilizzare un linguaggio professionale nella comunicazione con i docenti, per il prevalere di altri linguaggi più rumorosi e consueti: linguaggi politico-sindacali, linguaggi burocratico-

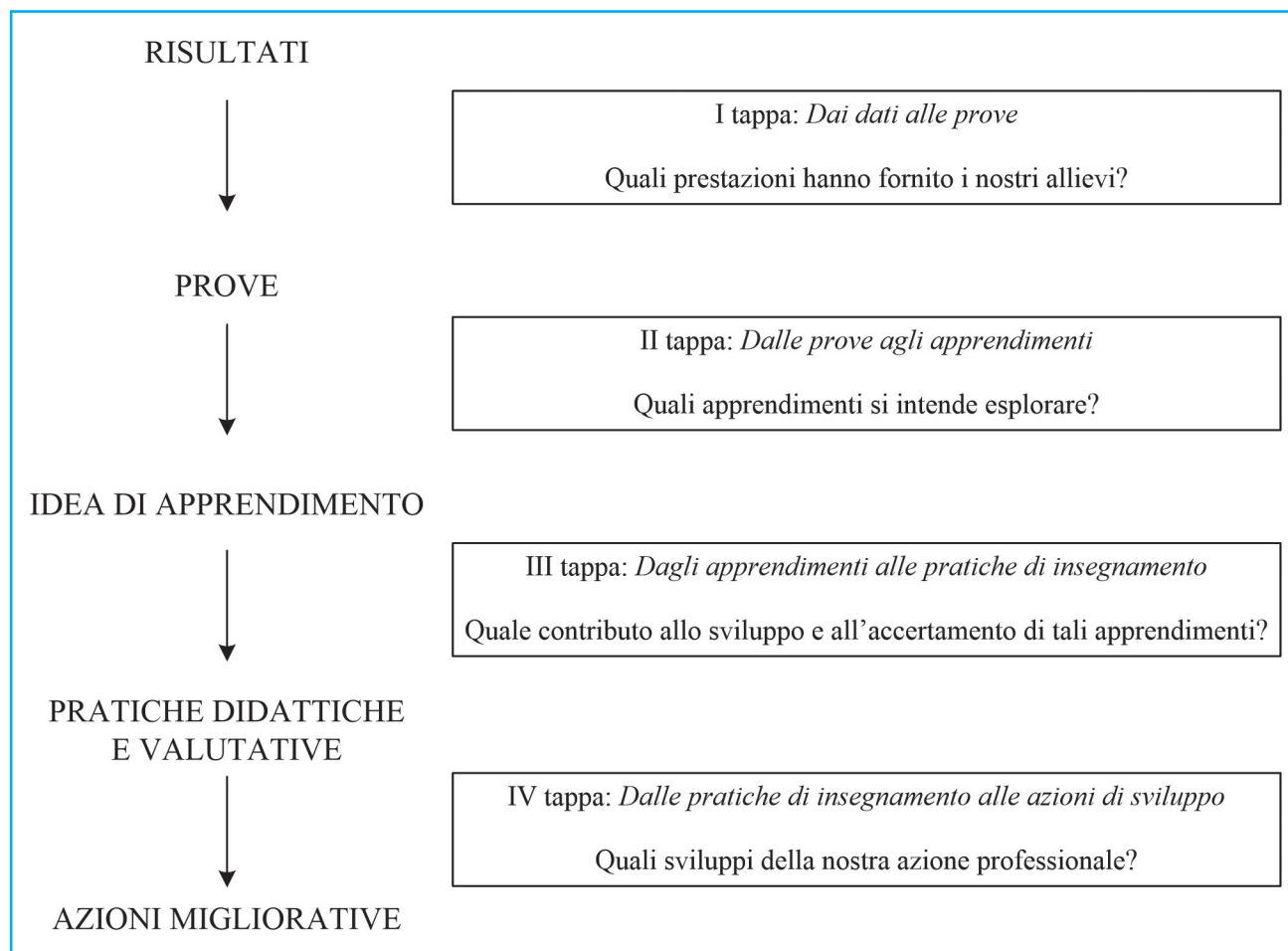
impiegatizi, linguaggi giuridico-normativi, linguaggi vocazional-moraleggianti. Entrare nel merito delle prove INVALSI, dei loro pregi e dei loro difetti, significa innanzitutto assumerle come un oggetto professionale con cui confrontarsi, in modo laico e disincantato.

In tale prospettiva i dati INVALSI consentono di sollecitare un processo a ritroso in chiave professionale che parta dai risultati per interrogarsi sulla natura delle prove di verifica impiegate, sull'idea di apprendimento a cui tali prove fanno riferimento, sulle proprie pratiche didattiche e valutative, sulla definizione di possibili azioni migliorative (fig. 1). L'espressione "a ritroso" richiama un approccio progettuale, denominato appunto "progettazione a ritroso", che si fonda su una sorta di ribaltamento tra momento progettuale e momento valutativo (Wiggins e McTighe, 2004) e propone di avviare lo sviluppo di un percorso progettuale a partire da due interrogativi tipicamente valutativi:

- qual è il profilo di competenza che voglio contribuire a sviluppare con il mio percorso?
- in termini operativi, quale prova di competenza mi aspetto che i miei allievi possano affrontare a conclusione del percorso?

Si tratta, come si vede, di anteporre alcune questioni tipicamente valutative alla strutturazione del percorso progettuale, allo scopo di poterlo trarre in relazione a una idea di competenza definita e articolata. Ciò implica l'esigenza di definire i propri traguardi di competenza e di analizzarli in profondità per individuare le dimensioni che concorrono alla loro manifestazione; il riferimento al profilo di competenza, infatti, richiama uno strumento come la rubrica valutativa orientata a descrivere operativamente tale profilo (Castoldi, 2016).

Fig. 1 – Un percorso a ritroso



Come si vede, l'espressione "a ritroso" richiama l'andamento del percorso progettuale proposto: si parte da alcune domande valutative, che sollecitano ad analizzare gli apprendimenti che si intende promuovere, per poi andare a strutturare il percorso formativo, definendo i contenuti di sapere, le metodologie didattiche, le scansioni operative che si

intendono utilizzare (Castoldi, 2017). Il lavoro di analisi delle prove INVALSI riproduce un processo analogo, sebbene in una prospettiva che non è strettamente progettuale bensì riflessiva in rapporto alla propria azione professionale. Anche in questo caso si tratta di partire da alcuni dati valutativi per interrogarsi sul loro significato e metterli in relazione alle proprie pratiche professionali. Come vedremo meglio nei prossimi paragrafi le fonti su cui sviluppare tale percorso a ritroso inizialmente sono i dati di risultato delle prove (quali prestazioni hanno fornito i nostri allievi?), nel secondo passaggio divengono i Quadri di riferimento e i materiali forniti dall'INVALSI per esplorare i passaggi a monte delle prove (quali apprendimenti si intende esplorare?), nel terzo passaggio divengono le pratiche didattiche e valutative degli insegnanti coinvolti nel processo riflessivo (quale contributo riteniamo di fornire allo sviluppo e all'accertamento di tali apprendimenti?), nel quarto passaggio gli orientamenti progettuali degli stessi insegnanti (quali sviluppi della nostra azione professionale?).

Tutto ciò, ovviamente, non presuppone un'accettazione acritica e incondizionata della proposta valutativa realizzata dall'INVALSI; semplicemente quest'ultima viene utilizzata come riferimento con cui mettersi in relazione e confrontarsi. In rapporto, in particolare, agli ultimi due passaggi indicati – il contributo didattico fornito dagli insegnanti e le azioni future di miglioramento della propria didattica – la riflessione e il confronto tra gli insegnanti è sollecitata ad andare “oltre” le stesse prove INVALSI e a prefigurare linee d'azione in una prospettiva più ampia e comprensiva. Il punto è, come già anticipato, desacralizzare il “feticcio” delle prove INVALSI e assumerlo come oggetto professionale da smontare e ricostruire in rapporto alle proprie sensibilità ed esperienze professionali; le prove INVALSI in tale prospettiva divengono il punto di partenza di un percorso critico-riflessivo tra gruppi di insegnanti (Schön, 1983).

Si tratta, come si può facilmente intuire, di una prospettiva di sviluppo professionale, per la quale i dati INVALSI divengono un pretesto per percorsi formativi e di confronto professionale tra gli insegnanti; percorsi che muovono da alcuni dati concreti, le prove e i loro risultati, per svilupparsi in una riflessione collegiale più ampia in merito alla propria azione didattica e valutativa e alle sue linee di miglioramento. La valenza formativa della proposta va sottolineata in quanto contribuisce ad allontanarsi da letture inquisitorie e fantasmatiche di controllo spesso prevalenti nella percezione degli insegnanti; il senso del lavoro non è quello di interrogarsi su “quanto raggiungiamo gli obiettivi richiesti dall'INVALSI?”, magari per elaborare graduatorie tra gli insegnanti o individuare chi è più “filo-INVALSI”, quanto di riflettere sulle proprie pratiche didattiche a partire dal pretesto offerto dalle prove nazionali.

Da questo punto di vista è curioso che, in un periodo in cui le risorse dedicate alla formazione sono ridotte al lumicino e gli insegnanti più avvertiti si lamentano sulla carenza di opportunità formative per la propria riqualificazione professionale, siano una quota irrisoria gli istituti scolastici che colgono l'opportunità offerta dalle prove INVALSI per strutturare percorsi di auto-aggiornamento. Percorsi che potrebbero essere realizzati senza costi aggiuntivi, solo mettendo in circolazione e valorizzando risorse disponibili in rete, a partire dai dati stessi sui risultati delle prove. Un paradosso che la dice lunga sul faticoso cammino che le scuole italiane sono chiamate a intraprendere in direzione di comportamenti professionali autonomi e responsabili (è molto più facile lamentarsi sulla carenza di risorse e sulle disquisizioni sindacali in merito a chi spetta e fino a quale soglia la somministrazione e la correzione delle prove INVALSI).

Come abbiamo anticipato i materiali di riferimento con cui sviluppare tali percorsi di analisi e riflessione sui risultati delle prove INVALSI sono rappresentati, in primo luogo dalla documentazione fornita dall'Istituto di valutazione. Attualmente si possono identificare i seguenti materiali¹:

- dati relativi ai risultati delle prove da parte dei propri allievi;
- testi delle prove impiegate e relative chiavi di correzione;
- Quadri di riferimento sulla cui base sono state elaborate le prove di verifica;
- guide alla lettura delle singole prove, che propongono un'analisi particolareggiata delle singole prove item per item;
- quaderni del Servizio nazionale di valutazione che propongono percorsi di riflessione su aspetti positivi e criticità che emergono dalle risposte date dagli studenti alle prove.

Si tratta di un materiale molto articolato e ampiamente sconosciuto tra i docenti, a conferma delle considerazioni svolte in precedenza; la maggior parte di essi, infatti, ha preso visione solo dei testi delle prove e delle relative chiavi di correzione (spesso perché obbligata a farlo...), solo una piccola quota conosce i dati relativi ai risultati o i Quadri di riferimento, quasi nessuno sa dell'esistenza delle guide alla lettura e dei quaderni SNV.

¹ Tutti i materiali indicati sono scaricabili dal sito www.invalsi.it; tranne i dati relativi ai risultati delle prove, il cui accesso è riservato, sono tutti materiali pubblici.

Allo scopo di orientarsi nell'analisi e nell'impiego dei materiali proposti vorremmo presentare uno strumento che fornisce le chiavi di lettura essenziali per gestire le operazioni di smontaggio e rimontaggio che abbiamo richiamato: la matrice processi-contenuti. Sia per le prove di italiano sia per quelle di matematica, infatti, i Quadri di riferimento propongono alcuni ambiti di processo e di contenuto intorno a cui sono organizzate le prove stesse; l'incrocio tra i due parametri consente di rappresentare il territorio esplorato dalle prove, ovvero la natura dell'apprendimento che è oggetto dell'indagine. Le tabb. 1 e 2 riportano le due matrici², qui ci interessa solo richiamare il loro valore strategico per la proposta di analisi e riflessione critica sui dati INVALSI, in quanto ci forniscono i parametri e le categorie di riferimento per sviluppare i diversi passaggi.

Tab. 1 – Matrice dei processi/contenuti di italiano

	Testo narrativo	Testo espositivo	Testo non continuo	Grammatica
Individuare informazioni				
Ricostruire il significato del testo				
Interpretare e valutare				
Riflessione sulla lingua				

Nota: Le domande di grammatica sono tutte riconducibili al processo "Riflessione sulla lingua"; la tipologia testuale "Testo non continuo" è presente solo nelle prove di seconda superiore.

Tab. 2 – Matrice dei processi/contenuti di matematica

	Numeri	Spazio e figure	Dati e previsioni	Relazioni e funzioni
Formulare				
Utilizzare				
Interpretare				

3. Livelli di analisi e domande-guida

Sulla base del percorso a ritroso rappresentato nella fig. 1 proveremo ad articolare un insieme di domande guida che possano aiutare a strutturare il processo di lettura dei risultati proposto. Le quattro tappe indicate rappresentano altrettanti livelli di analisi, che muovono da una lettura analitica dei risultati forniti dalle prove INVALSI, si allargano sull'idea di apprendimento sottesa alle prove, interrogano le proprie pratiche didattiche e valutative, si orientano verso la definizione di azioni di miglioramento del lavoro d'aula. Ovviamente si tratta di un'analisi critica, per la quale ogni passaggio implica non solo una ricognizione degli elementi emergenti, ma anche una loro problematizzazione aperta a molteplici prospettive di lettura.

Per i diversi livelli proposti proveremo a indicare i materiali che possono essere utilizzati e una traccia di interrogativi su cui fondare l'analisi dei materiali da parte di gruppi operativi di docenti (consigli di classe e/o dipartimenti disciplinari).

3.1. Primo step – Dai dati alle prove

La tab. 3 si riferisce al primo passaggio, finalizzato a una lettura analitica dei risultati forniti dalle prove INVALSI; particolarmente utili per questo passaggio risultano le guide alla lettura fornite dall'INVALSI per ciascuna prova impiegata, un documento nel quale item per item si forniscono un insieme di informazioni relative sia alle caratteristiche dell'item, sia ai risultati ottenuti dal campione nazionale, sia agli elementi di interesse e alle potenziali difficoltà che l'item può rilevare sul piano didattico.

² Le categorie utilizzate nelle due matrici fanno riferimento a quelle impiegate dall'INVALSI nelle edizioni 2013-14 e 2014-15.

Tab. 3 – Dai dati alle prove – Domande-guida

Materiali di riferimento	Domande-guida
– Grafici che, separatamente per ciascuna classe, mostrano le differenze, item per item, dei punteggi medi degli alunni della classe rispetto alla media nazionale	– Su quali item si concentrano i risultati migliori e quelli peggiori?
– Dettagli della prova di italiano:	– In rapporto alla matrice processi/contenuti dove si collocano gli item che presentano punti di flesso?
• parti del testo	– Su quali ambiti di contenuto i risultati sono più critici?
• processi	– Su quali processi i risultati sono più critici?
– Dettagli della prova di matematica:	– Utilizzando la matrice processi/contenuti come possiamo rappresentarci il quadro emergente dalle prove?
• ambiti	– Analizzando gli item più critici qual è la percentuale di mancate risposte?
• processi	– Quali ragioni si possono ipotizzare rispetto alle mancate risposte?
– Dettaglio risposte per item – italiano e matematica (valori percentuali)	– Analizzando gli item a risposta multipla più critici su quali distrattori si concentrano le risposte errate?
– Guide alla lettura per le singole prove	– Quali ragioni si possono ipotizzare rispetto alla concentrazione di risposte errate?
– Quaderni SNV per eventuali approfondimenti	– Le ragioni ipotizzate trovano conferma nell’analisi degli item proposta dai materiali INVALSI?

Le informazioni fornite dalla guida sono richiamate sia per collocare le caratteristiche degli item su cui si concentrano i risultati migliori e peggiori, sia per analizzare gli apprendimenti che ciascun item intende rilevare e le sue potenziali difficoltà. Per la prima operazione indicata si propone la matrice processi/contenuti come strumento di rappresentazione degli item migliori e peggiori (per un esempio si veda la tab. 4); la matrice può essere indicata anche per portare a sintesi la lettura dei risultati della prova, attraverso una rappresentazione grafica che aiuta a spostare l’attenzione sui livelli di analisi successivi (per un esempio si veda la tab. 5).




Tab. 4 – Rappresentazione degli item migliori e peggiori della prova di italiano attraverso la matrice processi/contenuti

	Testo narrativo	Testo espositivo	Testo non continuo	Grammatica
Individuare informazioni	<i>a1 a3</i>	<i>b2</i>	<i>c2</i>	
Ricostruire il significato del testo	a5	<i>b6</i>	c4	
Interpretare e valutare	a8		c6	
Riflessione sulla lingua				d4

Nota: In corsivo gli item migliori; in grassetto gli item peggiori.

Tab. 5 – Rappresentazione sintetica dei risultati della prova di matematica tramite la matrice processi/contenuti

	Numeri	Spazio e figure	Dati e previsioni	Relazioni e funzioni
Formulare				
Utilizzare				
Interpretare				

-  Risultati superiori alla media di riferimento
-  Risultati intermedi rispetto alla media di riferimento
-  Risultati inferiori alla media di riferimento

3.2. Secondo step – Dalle prove agli apprendimenti

La tab. 6 si riferisce al secondo passaggio, finalizzato a un’analisi critica in merito all’idea di apprendimento sottesa alle prove; il riferimento primario per questo passaggio sono i Quadri di riferimento proposti dall’INVALSI, che mirano a esplicitare l’idea di apprendimento su cui sono costruite le prove. In tal modo ci si allontana dal dato empirico connesso ai risultati delle prove e si sposta l’attenzione più a monte sull’idea di apprendimento che esse riflettono; si tratta di un distanziamento importante da operare, in modo da non rimanere eccessivamente ancorati al dato particolare e assumerlo come opportunità per una riflessione più globale sull’idea di apprendimento che veicola. Ciò consente di superare una logica di analisi delle prove troppo centrata sulla prestazione, che rischia di veicolare un messaggio di addestramento

alle prove nella prospettiva del *teaching to the test*, e di spostare l'attenzione sui processi di apprendimento connessi alle prove stesse e su una riflessione più ampia sui compiti formativi affidati alla scuola.

Si suggerisce di mettere a confronto l'idea di apprendimento proposta dai Quadri di riferimento sia con le prove INVALSI, in modo da mantenere la connessione tra dimensione teorica e operativa; sia con i documenti programmatici a livello nazionale, in modo da verificare la congruenza tra proposta INVALSI e riferimenti normativi; sia con le proprie idee di apprendimento, in modo da dialettizzare l'analisi della proposta INVALSI e connetterla più strettamente alla propria dimensione professionale. Il tipo di analisi proposto in questo passaggio rimane comunque centrato sull'idea di apprendimento più che sulla pratica didattica, la quale verrà sviluppata nel passaggio successivo; si suggerisce quindi di evitare di spostare l'attenzione su aspetti didattici e/o valutativi e di rimanere centrati sul processo di apprendimento. Il focus di questo passaggio è “che cosa significa apprendere in matematica/italiano?”, non “come insegnare matematica/italiano”, la prospettiva è centrata sul soggetto che apprende (l'allievo), non sul soggetto che insegna (l'insegnante).

Tab. 6 – Dalle prove agli apprendimenti – Domande-guida

Materiali di riferimento	Domande-guida
<ul style="list-style-type: none"> – Quadri di riferimento INVALSI – Quaderni SNV per eventuali approfondimenti – Indicazioni normative a livello nazionale (Indicazioni/Linee guida) 	<ul style="list-style-type: none"> – Quale idea di apprendimento in lettura-grammatica/matematica emerge dai Quadri di riferimento? – Quali sono gli aspetti che vengono maggiormente sottolineati? – Su quali ambiti di contenuto ci si concentra? – Su quali processi ci si concentra? – Come l'idea di apprendimento emergente dal QdR si riflette sulla prova che abbiamo esaminato? – In che misura l'idea di apprendimento emergente è congruente con le indicazioni programmatiche proposte a livello nazionale (Indicazioni nazionali/Linee guida)? – Quali sono gli aspetti maggiormente trascurati che meriterebbero più attenzione? – Quali sono i maggiori elementi di affinità con la nostra idea di apprendimento in lettura-grammatica/matematica? – Quali sono i maggiori elementi di differenza con la nostra idea di apprendimento in lettura-grammatica/matematica?

Ancora una volta la matrice processi/contenuti proposta sia per le prove di italiano sia per quelle di matematica può rappresentare un ottimo strumento di rappresentazione del territorio delle prove, da utilizzare come griglia di lettura dei documenti programmatici nazionali e delle idee prevalenti di apprendimento. La tab. 7 fornisce un esempio di uso della matrice matematica per sondare le idee di apprendimento degli insegnanti.

Tab. 7 – Autoriflessione sull'idea di apprendimento. Cosa è importante nell'apprendimento della matematica?

	Numeri	Spazio e figure	Dati e previsioni	Relazioni e funzioni
Formulare				
Utilizzare				
Interpretare				

Consegna: Tra le dodici celle proposte nella matrice annerisci completamente quelle che ritieni più importanti (almeno 4), annerisci parzialmente quelle intermedie (almeno 4), lascia in bianco quelle che ritieni meno importanti (almeno 4).

3.3. Terzo step – Dagli apprendimenti alle pratiche di insegnamento

La tab. 8 si riferisce al terzo passaggio, che sposta lo sguardo sulle prassi didattiche e valutative, ovvero a quanta attenzione viene posta agli apprendimenti oggetto delle prove INVALSI nel proprio lavoro didattico. Alcune evidenze, tra quelle restituite dall'INVALSI, che possono essere utilizzate in questo passaggio riguardano la distribuzione degli allievi della/e classe/i nei diversi livelli di apprendimento e la correlazione tra i risultati delle prove INVALSI e i voti assegnati dall'insegnante.

Tab. 8 – Dagli apprendimenti alle pratiche di insegnamento – Domande-guida

Materiali di riferimento	Domande-guida
<ul style="list-style-type: none"> – Distribuzione degli studenti per livelli di apprendimento: <ul style="list-style-type: none"> • italiano • matematica • italiano/matematica – Correlazioni: <ul style="list-style-type: none"> • correlazione tra risultati nelle prove INVALSI e voto di scuola – Grafici di confronto tra i punteggi medi nella prova INVALSI e il voto assegnato dal docente nell'ultimo scrutinio intermedio 	<ul style="list-style-type: none"> – Come si distribuiscono gli studenti nei diversi livelli di apprendimento? – Quale connessione si evidenzia tra risultati degli studenti in italiano e matematica? – Quale correlazione tra i risultati delle prove INVALSI e i voti scolastici? – In che misura le nostre pratiche didattiche sono congruenti con le richieste delle prove? – A quali contenuti/processi dedichiamo maggiore attenzione nel nostro insegnamento? – A quali contenuti/processi dedichiamo minore attenzione nel nostro insegnamento? – Quali “buone pratiche” didattiche ci vengono in mente nella prospettiva delle prove INVALSI? – Utilizzando la matrice processi/contenuti come possiamo rappresentare le nostre pratiche didattiche? – In che misura le nostre pratiche valutative sono congruenti con le richieste delle prove? – A quali contenuti/processi dedichiamo maggiore attenzione nella nostra valutazione? – A quali contenuti/processi dedichiamo minore attenzione nella nostra valutazione? – Quali “buone pratiche” didattiche ci vengono in mente nella prospettiva delle prove INVALSI? – utilizzando la matrice processi/contenuti come possiamo rappresentare le nostre pratiche valutative?

Per quanto riguarda la distribuzione degli allievi nei diversi livelli di apprendimento può essere interessante per capire se segue una distribuzione normale nelle varie classi o vi sono alcune anomalie, in termini di maggiore concentrazione nelle zone basse o nelle zone alte; è interessante anche l'incrocio tra risultati in italiano e in matematica, per riconoscere eventuali aree di criticità (per esempio gruppi numerosi di allievi che eccellono in matematica ma hanno risultati scadenti in italiano o viceversa). La relazione tra risultati delle prove e voti assegnati può essere invece un indizio interessante in rapporto alle modalità e ai criteri della valutazione, ponendo attenzione al grado di correlazione tra i due giudizi, in rapporto sia al dato medio di classe sia ai singoli allievi, e alla variabilità dei giudizi stessi dentro la classe e tra le classi (per esempio voti medi tra le classi poco differenziati a fronte di risultati nelle prove INVALSI fortemente differenziati).

Al di là di queste evidenze il lavoro più interessante in questa fase è quello auto-riflessivo, attraverso il confronto collettivo tra il *framework* INVALSI e le proprie prassi didattiche e valutative. Ancora una volta la matrice processi/contenuti può guidare la riflessione, sia attraverso l'individuazione di attività didattiche e valutative realizzate dai docenti nelle classi che si connettono alle diverse celle della matrice (si veda tab. 9), sia attraverso un'autovalutazione complessiva della propria azione didattica e valutativa in rapporto al territorio esplorato dalle prove INVALSI (si veda tab. 10 per un esempio).

Tab. 9 – Ricognizione sulle attività didattiche in rapporto alla matrice INVALSI. Quali attività didattiche facciamo in classe in rapporto alle varie celle della matrice?

	Testo narrativo	Testo espositivo	Testo non continuo
Individuare informazioni	Schede di comprensione sui racconti	Domande di analisi sul sussidiario	Individuazione informazioni contenute nelle tabelle
Ricostruire il significato del testo	Riassunti Dare il titolo a un racconto	Mappe concettuali di sintesi	
Interpretare e valutare	Discussione sulla relazione tra racconti e proprie esperienze	Confronto tra fonti differenti	

Tab. 10 – Auto-riflessione sulle pratiche didattiche – esempio. Su che cosa lavoro in matematica?

	Numeri	Spazio e figure	Dati e previsioni	Relazioni e funzioni
Formulare				
Utilizzare				
Interpretare				

Consegna: A che cosa dedico più attenzione nella mia didattica? Tra le dodici celle proposte nella matrice annerisci completamente quelle che a cui dedichi molta attenzione, annerisci parzialmente quelle a cui dedichi abbastanza attenzione, lascia in bianco quelle a cui dedichi poca attenzione.

3.4. Quarto step – Dalle pratiche di insegnamento alle azioni di sviluppo

La tab. 11 si riferisce al quarto passaggio, che sollecita la transizione dall'analisi all'azione migliorativa, dal lavoro di riflessione alle scelte individuali e collegiali. Il valore di questo passaggio sta proprio nel non esaurire l'analisi delle prove INVALSI a un mero lavoro retrospettivo, di riflessione a posteriori, bensì di orientarlo in chiave proattiva, come sollecitazione a ripensare le proprie pratiche didattiche e valutative. È importante connettere quest'ultimo passaggio ai precedenti, partendo proprio dal confronto tra i diversi livelli di analisi proposti per individuare alcune piste migliorative (si veda fig. 2); ancora una volta la matrice processi/contenuti può fornire lo strumento operativo per fare questa operazione, consentendo di disporre di un quadro di sintesi del percorso auto-riflessivo sui dati INVALSI (si veda la tab. 12). Ovviamente il confronto può essere fatto anche sulla base di dati non completi: più che uno schema operativo, si tratta di uno schema logico.

Tab. 11 – Dalle pratiche di insegnamento alle azioni di sviluppo – Domande-guida

Materiali di riferimento	Domande-guida
Analisi svolta nelle tra fasi precedenti, sia in rapporto ai risultati delle prove, sia all'idea di apprendimento, sia alle pratiche di insegnamento e valutazione (con particolare riferimento alla matrice processi/contenuti come strumento di confronto tra i tre passaggi)	– Quali traguardi di miglioramento possiamo porci in rapporto alle prove INVALSI?
	– Quali attenzioni nella didattica quotidiana?
	– Quali strumenti/materiali didattici introdurre?
	– Quali strumenti/materiali didattici eliminare?
	– Quali percorsi didattici aggiungere/potenziare nella proposta di lavoro alle classi?
	– Quali percorsi didattici eliminare/ridimensionare nella proposta di lavoro alle classi?
	– Quali prove valutative aggiungere/potenziare?
	– Quali prove valutative eliminare/ridimensionare?
	– I libri di testo che impieghiamo sono funzionali allo sviluppo degli apprendimenti richiesti?
	– Che cosa fare collegialmente (a livello di dipartimento o di consiglio di classe)?

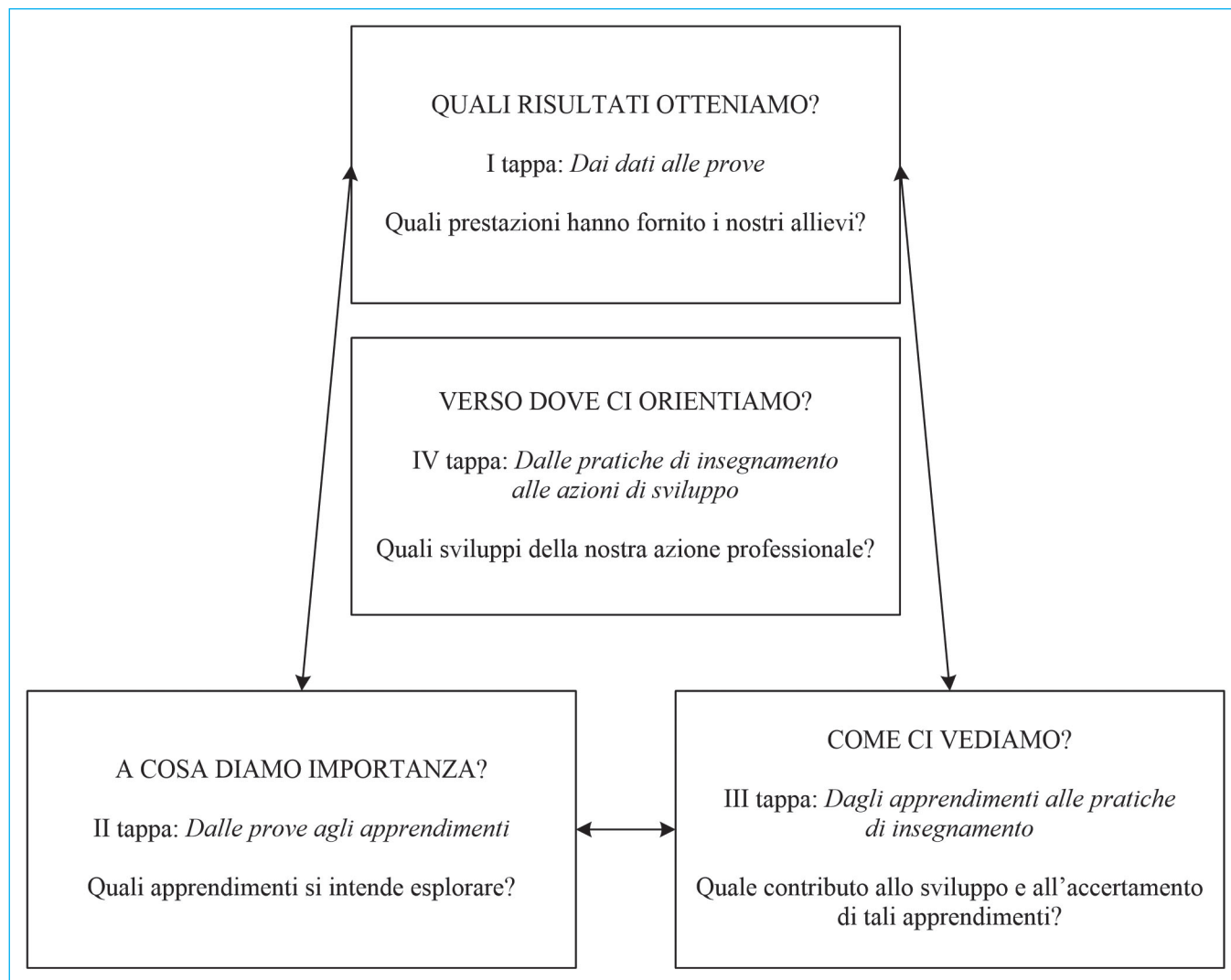
Ciò che rende interessante lo schema logico richiamato nella tab. 12 è proprio la chiave di lettura proposta per indagare i risultati delle prove INVALSI emersi come non pienamente soddisfacenti. Emerge infatti una distinzione tra gli aspetti del *framework* che risultano più scoperti in rapporto all'idea di matematica e alle prassi degli insegnanti (quelli che evidenziano giudizi medio-bassi sulle tre voci di importanza, didattica e valutazione); gli aspetti del *framework* che risultano insoddisfacenti nonostante l'attenzione posta dagli insegnanti (quelli che evidenziano giudizi medio-alti sulle tre voci di importanza, didattica e valutazione); gli aspetti del *framework* che risultano contraddittori nel lavoro degli insegnanti (quelli che evidenziano giudizi alti a fianco di giudizi bassi su almeno una delle tre voci). Evidentemente i tre aspetti si possono affrontare diversamente in chiave proattiva, nel momento in cui si ipotizzano azioni di miglioramento.

Tab. 12 – Matrice processi/contenuti – quadro di sintesi

	Numeri	Spazio e figure	Dati e previsioni	Relazioni e funzioni
Formulare			Importanza: alta Didattica: bassa Valutazione: media	
Utilizzare		Importanza: alta Didattica: alta Valutazione: media		Importanza: bassa Didattica: media Valutazione: alta
Interpretare	Importanza: alta Didattica: media Valutazione: alta		Importanza: media Didattica: bassa Valutazione: bassa	Importanza: bassa Didattica: bassa Valutazione: media

Nota: Le celle in grigio segnalano risultati al di sopra della media di riferimento.

Fig. 2 – Modello di relazione tra i quattro passaggi proposti



4. Un'esperienza concreta

Tra le esperienze di applicazione del percorso proposto si richiama quella svolta presso l'istituto comprensivo di Gaglianico (BI) nell'a.s. 2014-15³, in riferimento ai risultati delle prove INVALSI ottenuti dalle classi dell'istituto nel 2014. La modalità di lavoro ha previsto la formazione in sei gruppi di insegnanti suddivisi per ambiti (matematica e italiano) e classi (II e V primaria, e III secondaria di I grado) interessati dalle prove. La partecipazione è stata volontaria e per ogni gruppo era previsto un coordinatore che seguiva gli incontri di supervisione con il formatore (il sottoscritto) e poi organizzava e gestiva il lavoro del proprio gruppo. Il percorso si è distribuito in cinque incontri: uno iniziale e uno finale, svolti in plenaria con tutti i docenti che partecipavano alla formazione, e tre coi coordinatori che ne hanno svolti altrettanti con i gruppi di lavoro (tab. 13).

³ Il resoconto dell'esperienza riprende la documentazione dell'esperienza curata dall'insegnante Manuela Morello in collaborazione con la dirigente Claudia Valz, alla quale vanno i miei ringraziamenti.

Tab. 13 – Percorso di lavoro

Lunedì 2 febbraio 2015	Incontro con i coordinatori dei gruppi (2 h)	Impostazione del percorso di lavoro dei gruppi e definizione primo incontro
	Incontro con il gruppo docente (2 h)	Condivisione del senso del percorso
Febbraio 2015	I incontro dei gruppi di lavoro (2 h)	Analisi risultati prove INVALSI
Martedì 3 marzo 2015	Incontro con i coordinatori dei gruppi (3 h)	Verifica primo incontro e definizione secondo incontro
Marzo 2015	II incontro dei gruppi di lavoro (2 h)	Auto-riflessione sulle proprie pratiche professionali
Martedì 31 marzo 2015	Incontro con i coordinatori dei gruppi (3 h)	Verifica secondo incontro e definizione terzo incontro
Aprile 2015	III incontro dei gruppi di lavoro (2 h)	Individuazione proposte di miglioramento sul piano didattico e valutativo
Martedì 12 maggio 2015	Incontro con i coordinatori dei gruppi (2 h)	Verifica percorso complessivo e indicazioni di sviluppo
	Incontro con il gruppo docente (2 h)	Condivisione risultati dei gruppi e linee di sviluppo

La prima fase del lavoro prevedeva di partire dai risultati restituiti da INVALSI per focalizzare successivamente l'attenzione sugli apprendimenti richiesti dalle prove. Si è partiti dalle tavole restituite dall'INVALSI alle singole scuole che riportano i punteggi generali ottenuti dalle classi e dall'istituto, ma per attenuare il riferimento ai singoli docenti i valori delle singole classi dei plessi (5 classi e 3 plessi per la primaria; 4 classi e 2 plessi per la secondaria) sono stati aggregati e poi analizzati come dati di istituto. Questi ultimi, pur non essendo estremamente bassi, risultavano inferiori rispetto alle medie di riferimento regionali e nazionali e a quelle del campione di scuole avente un background socio-culturale simile.

Si è poi passati ad analizzare le tabelle restituite dall'INVALSI sui dati disaggregati per ambiti e per processi. Completata la fase di analisi generale dei dati, si è proceduto a prendere in esame le tabelle riportanti il dettaglio delle risposte per i singoli item, rielaborate dal gruppo dei coordinatori a livello di dati globali di istituto. In particolare si è centrata l'attenzione sugli item in cui più del 50% degli alunni aveva dato una risposta non corretta. In questa fase del lavoro è stato indispensabile l'utilizzo della guida alla lettura elaborata dall'INVALSI per ciascuna prova, da cui è stato possibile reperire per ciascun item tutte le informazioni riconducibili a processi e ambiti.

In conclusione è stata predisposta da ogni gruppo una tabella di sintesi basata sulle differenze tra la percentuale del punteggio medio dell'istituto e il punteggio del campione nazionale per ciascun ambito e processo. Dopo aver inserito tutti i valori, se l'intersezione tra ambito e processo corrispondeva a due valori positivi, cioè entrambi sopra la media, la cella è stata colorata di verde in quanto "risultato soddisfacente"; se l'intersezione riguardava un valore positivo e uno negativo si è colorata di giallo in quanto risultato "abbastanza soddisfacente"; se l'intersezione riguardava due valori negativi, cioè entrambi sotto la media, si è colorata di rosso in quanto "risultato insoddisfacente" (tab. 5; i tre colori indicati sono stati resi con diverse tonalità di grigio).

Durante il secondo incontro, partendo dai dati ottenuti nella prima fase del lavoro, si è spostata l'attenzione sulla quotidianità della didattica d'aula. L'obiettivo era quello di operare una riflessione collegiale sugli aspetti dell'apprendimento richiamati nella matrice ambiti/processi a cui si attribuiva importanza e a cui si prestava attenzione nelle pratiche didattiche e valutative (tab. 7 e tab. 10). Per arrivare a formulare queste valutazioni è stato indispensabile approfondire i Quadri di riferimento delle prove forniti da INVALSI, recuperare i nostri documenti di programmazione e rievocare le più frequenti attività didattiche e valutative svolte dai docenti in aula.

Durante il terzo e ultimo incontro, partendo dalle domande guida proposte (tab. 11) e dal quadro d'insieme dei passaggi precedenti (tab. 12), è stato richiesto ai gruppi di predisporre proposte concrete di attività didattiche per poter portare in aula un nuovo modo di fare scuola alla luce dei punti di forza e di debolezza evidenziati nelle precedenti fasi di lavoro.

Di seguito alcuni esempi delle attività proposte dal gruppo di matematica alle classi seconde dell'istituto comprensivo di Gaglianico (BI); per ciascuno di essi il gruppo di docenti ha provato a evidenziare su quali aspetti della matrice ambiti/processi si concentrava l'attenzione dell'attività proposta.

Bisogna organizzare una visita al Museo del territorio di Biella, come possiamo fare?

Lo scopo di questa attività è far sì che ogni gruppo in cui è stata suddivisa la classe porti a termine una tappa del lavoro:

- saper leggere la tabella orario del pullman di linea e scegliere le due corse più idonee per il viaggio di andata e ritorno, calcolare quante ore durerà l'uscita;
- organizzare i momenti fondamentali dell'uscita, suddividendoli nelle diverse parti del giorno, e preparare una piccola brochure da consegnare ai genitori;
- recarsi in tabaccheria con l'insegnante per acquistare i biglietti del pullman e controllare il resto della quota totale;
- fare un'ipotesi del tempo atmosferico che ci sarà il giorno dell'uscita e costruire un piccolo grafico che sarà confrontato a uscita avvenuta.



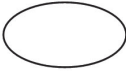
Matrice processi/contenuti – Visita al Museo del territorio di Biella

	<i>Numeri</i>	<i>Spazio e figure</i>	<i>Dati e previsioni</i>
Formulare	x		x
Utilizzare	x		
Interpretare	x		x

Attività presso la pinacoteca del Museo del territorio

Quali sono le forme geometriche che conoscete? Gli alunni faranno un elenco di figure geometriche solide e piane. Osserviamo i quadri e le loro cornici. Di che forma sono le cornici dei quadri? In seguito gli alunni completeranno una tabella a doppia entrata.

Tabella a doppia entrata per l'attività presso la pinacoteca del Museo del territorio

Cornice	Quadro grande	Quadro piccolo
		
		
		

L'attività sarà l'occasione di completare una semplice rete semantica sulle figure geometriche.

L'insegnante chiederà agli allievi di darsi la mano e formare cerchi, rettangoli, quadrati, il docente fotograferà dall'alto la "figura" formata dagli alunni.

Matrice processi/contenuti – Attività presso la pinacoteca del Museo del territorio

	<i>Numeri</i>	<i>Spazio e figure</i>	<i>Dati e previsioni</i>
Formulare		x	
Utilizzare		x	
Interpretare		x	

Orto in condotta

La scuola primaria di Gaglianico ha aderito al progetto “Orto in condotta” che ha come fine la realizzazione di orti per promuovere e sviluppare l’educazione alimentare e ambientale nelle scuole.

Il gruppo di lavoro ha preso spunto dal progetto per proporre un’attività da svolgere con gli alunni durante il trapianto delle piantine seminate nei mesi precedenti.

Avendo a disposizione 48 piantine di piselli, 12 di pomodori e dovendo trapiantarli in file con un uguale numero di piantine, come si può procedere per trovare la soluzione?

Gli alunni saranno invitati a esprimere una soluzione al problema posto e la strategia adottata.

In seguito al trapianto, si chiederà agli alunni di ipotizzare alcune previsioni sulla quantità in ogni pianta di baccelli e pomodori che saranno presenti a maturazione avvenuta. Sarà compito del docente fotografare le piantine con i frutti per poter svolgere in seguito un confronto tra i dati ipotizzati e quelli reali.

Matrice processi/contenuti – Orto in condotta

	<i>Numeri</i>	<i>Spazio e figure</i>	<i>Dati e previsioni</i>
Formulare	x		
Utilizzare	x		
Interpretare	x		x

Sudokino e sudoku

Al fine di far “giocare” gli alunni con i numeri, si chiede loro di risolvere, suddivisi in piccoli gruppi, prima un sudokino e in seguito un sudoku.

Ogni colonna, ogni riga e ogni regione devono contenere una sola volta i numeri da 1 a 4.

Sudokino

2			
4	1		
	4		
			3

Matrice processi/contenuti – Sudokino e sudoku

	<i>Numeri</i>	<i>Spazio e figure</i>	<i>Dati e previsioni</i>
Formulare	x		
Utilizzare	x		
Interpretare	x		

5. Conclusioni

Le prove INVALSI rappresentano a tutt’oggi un’occasione mancata per la scuola italiana. Al di là delle ragioni di questo difficile incontro tra la scuola reale e il Sistema nazionale di valutazione così come si è andato configurando in questi anni, che potrebbero sollecitare svariate riflessioni, in questo contributo abbiamo voluto provare a esemplificare in che senso si tratta di un’occasione mancata. Lo abbiamo fatto prefigurando una semplice proposta auto-formativa di analisi e riflessione collegiale sui risultati delle prove INVALSI, tra le tante possibili, che ben evidenzia le potenzialità di questa operazione per lo sviluppo professionale degli insegnanti italiani.

Ovviamente ci sono altri possibili piani di analisi e impiego delle prove INVALSI, a partire da un utilizzo per una valutazione complessiva dell'istituto scolastico connessa al Rapporto di autovalutazione e all'elaborazione del Piano di miglioramento (si veda, tra gli altri, Fiore e Pedrizzi, 2016). A noi premeva rimanere centrati sul livello dell'aula, per cercare di argomentare come anche su un piano strettamente didattico le prove INVALSI presentino molteplici spunti di interesse e siano un'opportunità di crescita professionale. In un momento in cui con il Piano formazione docenti 2016-19 si intendono porre le basi per la strutturazione di un sistema di formazione in servizio sistematico e organico la messa a tema e la riscoperta di queste opportunità risulta ancor più significativa e attuale.

In altre parole si tratta di un semplice tentativo per prendere sul serio le prove INVALSI, nel contesto di una scuola italiana in cui tendono a prevalere comportamenti volti a neutralizzare, depotenziare, banalizzare il loro significato (si veda l'incredibile successo editoriale delle guide di preparazione alle prove INVALSI). Per dirla con Watzlawick potremmo dire che la proposta di smontaggio delle prove INVALSI richiamata in queste pagine può essere vista come un modesto esempio per passare da un cambiamento di tipo 1, inteso come un tipo di cambiamento che rimane all'interno di un dato sistema lasciando immutate le sue caratteristiche costitutive, a un cambiamento di tipo 2, inteso come un tipo di cambiamento che punta a modificare le premesse del sistema stesso (Watzlawick, Weakland e Fisch, 1973).

Riferimenti bibliografici

- Castoldi M. (2014), *Capire le prove INVALSI*, Carocci, Roma.
- Castoldi M. (2016), *Valutare e certificare le competenze*, Carocci, Roma.
- Castoldi M. (2017, in corso di pubblicazione), *Progettazione a ritroso. Costruire unità di apprendimento per competenze*, Carocci, Roma.
- Dumont A., Istance D., Benavides F. (eds.) (2010), *The Nature of Learning*, OECD, Paris.
- Fiore B., Pedrizzi T. (a cura di) (2016), *Valutare per migliorare le scuole*, Mondadori Education, Firenze.
- Schön D. (1983), *The Reflective Practitioner. How Professionals think in Action*, Temple Smith, London; trad. it. *Il professionista riflessivo. Per una nuova epistemologia della pratica professionale*, Dedalo, Bari, 1993.
- Watzlawick P., Weakland J.H., Fisch R. (1974), *Change: The Principles of Problem Formation and Problem Resolution*, Norton, New York; trad. it. *Change: la formazione e la soluzione dei problemi*, Astrolabio, Roma, 1974.
- Wiggins, G., McTighe J. (2004), *Understanding by Design*, Association for Supervision and Curriculum Development. Alexandria (VA); trad. it. *Fare progettazione*, LAS, Roma, 2004.

3. I legami tra i risultati delle prove di pre-lettura, difficoltà di lettura e reading literacy

Relations between performance on pre-reading skills, reading difficulty and reading literacy

di Graziella Marrone, Marianna Rasetta, Giselda Di Cesare, Sergio Di Sano

L'istituto comprensivo di Loreto Aprutino (PE) dal 2013 ha scelto, come campo di studio e di ricerca, l'apprendimento della lettoscrittura nelle classi prima e seconda della scuola primaria secondo l'approccio *Response to Intervention* (RTI). Si tratta di un approccio multilivello, che si è ampiamente affermato negli Stati Uniti, basato sul principio di non attendere che il bambino presenti delle difficoltà di apprendimento importanti ma intervenire preventivamente in maniera ecologica e integrata nella didattica.

Il progetto RTI, frutto di una collaborazione tra Università di Chieti, USR Abruzzo, ASL di Pescara e una rete di scuole coordinata dall'istituto comprensivo di Loreto Aprutino, capofila, ha avuto come obiettivo principale quello di garantire un insegnamento efficace in ambiti importanti sul piano dell'apprendimento (lettura, scrittura) al fine di evitare l'invio di alunni per una diagnosi di apprendimento laddove le difficoltà scolastiche dipendano invece dall'istruzione inadeguata.

Il progetto si è basato sullo studio e sul confronto tra i diversi documenti (le Indicazioni nazionali, la normativa recente in materia di DSA, i dati OCSE-PISA e le ultime ricerche sull'apprendimento della letto-scrittura) che hanno determinato la progettazione di un curriculum di ricerca-azione che ha condotto a significativi risultati nella prova preliminare di lettura delle prove INVALSI nella classe seconda della scuola primaria nel maggio 2015.

Since 2013, the Comprehensive School of Loreto Aprutino (Pescara) has chosen literacy learning in 1st and 2nd grade of Primary School, according to the *Response to Intervention* approach (RTI), as a field of study and research. This is a multilevel approach well established in the United States, which privileges a preventive, ecological and integrated action in terms of teaching practices instead of waiting for the emergence of students' learning problems.

The aim of the RTI approach, supported by scientific research on reading, is the search of effective methods to teach reading. In this way, it highlights five components for reading instruction: phonemic awareness, phonetics, fluency, vocabulary, reading comprehension. It is considered the essence of reading.

The RTI Project, arisen from the collaboration between University of Chieti, USR Abruzzo, ASL of Pescara, and the network of schools coordinated by the Comprehensive School of Loreto Aprutino as leader, has the following goal: to ensure effective teaching in important areas of school learning (reading, writing) in order to avoid requests for diagnosis of pupils' learning problems when difficulties in schooling depend instead on inadequate teaching.

The project has begun with the study and comparison of various documents (the National Guidelines on LD, OECD-PISA data, latest research on reading and writing) that led us to design a curriculum for action research which has permitted to obtain better results in INVALSI pre-reading test in May 2015 (INVALSI data for second grade preliminary reading test).

1. Introduzione

In Italia la lettura ha sempre avuto un ruolo centrale nel percorso educativo della scuola primaria. Essa è una delle abilità più importanti che gli studenti acquisiscono nei primi anni scolastici ed è indispensabile per l'apprendimento

delle altre discipline. All'interno delle Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione del 2012 si dice che:

La pratica della lettura, centrale in tutto il primo ciclo di istruzione, è proposta come momento di socializzazione e di discussione dell'apprendimento di contenuti, ma anche come momento di ricerca autonoma e individuale, in grado di sviluppare la capacità di concentrazione e di riflessione critica, quindi come attività particolarmente utile per favorire il processo di maturazione dell'allievo. Per lo sviluppo di una sicura competenza di lettura è necessaria l'acquisizione di opportune strategie e tecniche, compresa la lettura a voce alta, la cura dell'espressione e la costante messa in atto di operazioni cognitive per la comprensione del testo. Saper leggere è essenziale per il reperimento delle informazioni, per ampliare le proprie conoscenze, per ottenere risposte significative. [...] A scuola si apprende la strumentalità del leggere e si attivano i numerosi processi cognitivi necessari alla comprensione (MIUR, 2012, pp. 28-29).

La legge n. 170 dell'8 ottobre 2010 (Nuove norme in materia di disturbi specifici di apprendimento in ambito scolastico) riconosce e definisce i Disturbi specifici di apprendimento (in seguito DSA), intendendo per dislessia:

Un disturbo specifico che si manifesta con una difficoltà nell'imparare a leggere, in particolare nella decifrazione dei segni linguistici, ovvero nella correttezza e nella rapidità della lettura. [...] È compito delle scuole di ogni ordine e grado, comprese le scuole dell'infanzia, attivare, previa apposita comunicazione alle famiglie interessate, interventi tempestivi, idonei a individuare i casi sospetti di DSA degli studenti, sulla base dei protocolli regionali di cui all'articolo 7, comma 1. L'esito di tali attività non costituisce, comunque, una diagnosi di DSA (artt. 1 e 3, comma 3).

La scarsa conoscenza della normativa in materia di DSA e la mancanza di un linguaggio condiviso tra la scuola, la famiglia e i servizi specialistici, hanno portato a una individuazione tardiva dei DSA con conseguenze negative a livello psicologico per l'alunno. L'individuazione precoce dei bambini "a rischio" è importante per evitare possibili ripercussioni psicologiche, legate a situazioni di disagio, e per avviare e strutturare un percorso di potenziamento didattico rivolto alle abilità carenti. Come notano Stella e Grandi:

Negli ultimi anni si sono diffusi scale e strumenti per l'identificazione precoce di alunni "a rischio" di disturbi dell'apprendimento, alcuni di questi si basano su conoscenze relative alle tappe di acquisizione della letto-scrittura, sulla scrittura spontanea e sulla valutazione delle competenze meta-fonologiche dei bambini, altri strumenti si avvalgono della tecnologia per offrire all'insegnante un supporto nella valutazione e nell'identificazione precoce (Stella e Grandi, 2011, p. 57).

Le classi sono caratterizzate da molteplici diversità, legate alle differenze nei modi e nei livelli di apprendimento, alle specifiche inclinazioni, ai personali interessi e a particolari stati emotivi e affettivi.

Ancor prima di intervenire con un potenziamento didattico, le istituzioni scolastiche dovrebbero realizzare, già dai primi mesi di scuola, un curricolo per l'apprendimento della strumentalità del leggere attivando i numerosi processi cognitivi necessari alla comprensione, valorizzando le differenze individuali (MIUR, 2012).

Negli Stati Uniti, nell'ambito dello *screening* e dell'identificazione dei bambini con difficoltà di apprendimento, si è sviluppato l'approccio *Response to Intervention* (RTI) secondo il quale, prima di ipotizzare una diagnosi, si dovrebbe valutare l'adeguatezza dell'insegnamento, ponendo quindi al centro del processo il ruolo del docente (Fuchs e Fuchs, 2006).

Inoltre, la metodologia RTI prevede l'impiego dell'approccio *Curriculum Based Measurement* (CBM), basato su misure di fluency della lettura, che fornisce una stima globale del successo scolastico e dell'evoluzione nel tempo della prestazione scolastica del bambino (Deno *et al.*, 2001).

Queste misure vengono impiegate nel monitoraggio dei progressi, per decidere forme di insegnamento più intense per bambini che "non rispondono" a una didattica *evidence-based* rivolta all'intera classe. Le misure di fluency della lettura sono risultate particolarmente utili negli studi inter-culturali, in quanto sembrano fornire criteri generali validi per culture e lingue diverse (Abadzi, 2012).

Nel complesso, il tasso di lettura, misurato in termini di numero di parole pronunciate correttamente in un minuto, sembra un indicatore valido e affidabile del rendimento scolastico per molte lingue diverse, e l'intervallo di 45-60 parole risulta un valore minimo di riferimento. Se il testo viene letto troppo lentamente, per esempio 25 parole al minuto, poco di esso può essere compreso.

-
- Il progetto qui presentato, denominato *Response to Intervention Abruzzo*, si propone di perseguire i seguenti obiettivi:
- 1) identificare il modo migliore per aiutare i bambini che “fanno fatica” nella lettura;
 - 2) usare, come prova di *screening*, una prova di fluenza nella lettura di sillabe nella classe prima e una prova di fluenza nella lettura di brani (di tipo narrativo) nella classe seconda della scuola primaria;
 - 3) usare una prova di fluenza nella lettura di sillabe/parole (classi prime) e di brani narrativi (classi seconde) come prova di monitoraggio;
 - 4) indagare la capacità predittiva delle prove di fluenza e del tasso di crescita (misurato come incremento settimanale nel numero di unità, sillabe o parole, lette correttamente in un minuto).

2. Gli approcci alla dislessia

L’obiettivo principale del progetto è stato quello di identificare il modo migliore per aiutare i bambini che faticano nell’abilità di lettura e per i quali l’acquisizione di tale abilità costituisce una sfida significativa (Lyytinen *et al.*, 2008).

Rispetto a questo obiettivo sono stati individuati due approcci possibili, che fanno riferimento a due modelli distinti: *Wait to Fail* (WTF) e *Response to Intervention* (RTI). Nel primo modello, si aspetta di poter dire con certezza che il bambino è dislessico prima di avviare un intervento. Nel secondo modello, si interviene prima in ambito educativo sui bambini con difficoltà di lettura o a rischio e successivamente i bambini che “resistono” all’intervento educativo vengono inviati per una valutazione da parte di professionisti. Il secondo modello risulta più efficace del primo, in quanto consente di intervenire fin da subito sui bambini della scuola dell’infanzia, in caso di cadute sui predittori della lettura, e della scuola primaria, in caso di conclamate difficoltà di lettura. Questo tipo d’intervento porta beneficio sia ai bambini che recuperano, sia a quelli che risulteranno dislessici perché evita un aggravamento del disturbo minimizzando l’impatto emotivo.

Da quanto detto, risulta chiara l’importanza di intervenire non solo sul disturbo ma anche sulla difficoltà, graduando l’intervento in modo dinamico in base alle effettive esigenze degli alunni.

Un secondo aspetto da considerare è l’importanza di un approccio predittivo al problema della dislessia. In questo ambito, la ricerca internazionale sulla dislessia ha impiegato una procedura correlazionale valutando determinate abilità cognitive come potenziali predittori al tempo T1 (per esempio la competenza fonologica nell’ultimo anno dell’infanzia) e l’abilità criterio come la lettura al tempo T2 (per esempio, la correttezza e/o velocità nella lettura di un brano nella seconda classe della primaria).

In alcuni casi, la ricerca ha coinvolto la popolazione tipica, cioè normale, e in altri si è focalizzata su popolazioni a rischio per i disturbi di lettura, per esempio bambini con pregresso disturbo del linguaggio, o con familiarità per la dislessia. I risultati della ricerca sui predittori hanno evidenziato una serie di abilità cognitive come la consapevolezza fonologica, la conoscenza delle lettere, la memoria verbale, la velocità di denominazione, la velocità di articolazione, il cui mancato sviluppo nella scuola dell’infanzia consente di definire un certo grado di rischio per le difficoltà di lettura negli anni successivi della scuola primaria (Muter, 2006).

Un terzo aspetto da considerare riguarda la relazione tra lettura, intesa come decodifica o lettura ad alta voce, e comprensione del testo. L’abilità strumentale di lettura consente di decodificare la sequenza di lettere costitutive del testo con correttezza e velocità, mentre la comprensione del testo richiede di accedere al suo significato. È importante non sottovalutare la natura interattiva dei due processi: si possono attivare processi di anticipazione semantica che consentono di prevedere le parole successive prima di leggerle e, viceversa, le difficoltà di decifrazione possono compromettere la comprensione e lo sviluppo del vocabolario. Data la relazione tra la comprensione del testo e la decodifica delle lettere e la sua evoluzione nel corso del tempo, risulta importante valutare sempre entrambi gli aspetti. Occorre tenere presente che le conseguenze negative della dislessia dipendono soprattutto dallo sforzo che i bambini fanno nel tentativo di comprendere i testi che leggono: molte risorse cognitive sono impegnate nella decifrazione e ciò appesantisce il carico della memoria di lavoro.

La ricerca sui predittori si inserisce in questa visione: predittori diversi possono essere collegati ad aspetti diversi della lettura. Per esempio, la consapevolezza fonologica risulta un predittore importante per la decodifica, mentre la velocità di denominazione è un predittore della rapidità di lettura e il vocabolario lo è per la comprensione del testo.

In realtà, il quadro è più complesso in quanto entrano in gioco altri fattori, per esempio la memoria e l'attenzione, che interagiscono con gli altri predittori nel determinare la fenomenologia del disturbo.

Le difficoltà incontrate dai programmi di intervento, diretti a migliorare abilità come la fluency nella lettura, la conoscenza del vocabolario e la comprensione potrebbero essere il risultato di un effetto Matthew nella lettura in base al quale i poveri diventano sempre più poveri e i ricchi sempre più ricchi (Stanovich, 1986).

Il deficit di elaborazione fonologica impedisce lo sviluppo del riconoscimento delle parole, di conseguenza i bambini fanno meno pratica con la lettura e di fronte a materiali difficili da leggere evitano di leggerli. Al contrario, i bambini che non hanno problemi di lettura fanno più pratica, accrescono il vocabolario, comprendono frasi complesse, arricchiscono la propria base di conoscenza per la comprensione e hanno maggiori opportunità di migliorare la fluency (Tunmer e Greaney, 2008).

Un quarto aspetto da considerare è l'importanza che riveste la collaborazione tra insegnante, genitore e professionista per identificare le difficoltà di lettura del bambino e intervenire precocemente, tenendo conto di quanto avviene sia a scuola sia a casa. Il ruolo dei genitori risulta importante non solo per favorire le attività di sensibilizzazione verso la lettura che possono essere svolte a casa, ma anche per fornire informazioni sulla storia del bambino e sugli aspetti emotivi coinvolti nelle difficoltà di lettura.

3. La ricerca-azione RTI Abruzzo

Il progetto RTI Abruzzo nasce come ricerca-azione nell'anno scolastico 2013-14 da una collaborazione tra il Dipartimento di Neuroscienze e imaging dell'Università "G. d'Annunzio" di Chieti-Pescara, il Servizio di Neuropsichiatria infantile dell'ASL di Pescara, l'Ufficio scolastico regionale dell'Abruzzo e quattro istituti comprensivi delle province di Pescara, Chieti e Teramo. Il progetto, che si ispira all'approccio RTI, implica lo svolgimento di attività di prevenzione e identificazione di problemi di apprendimento, in base a quanto previsto dalla normativa ministeriale in materia di DSA.

Il progetto di ricerca-azione RTI Abruzzo prevede tre livelli di intervento:

- *livello I*: consiste in un intervento sulle abilità di lettura basato su attività curriculari, rivolto a tutta la classe e svolto almeno tre volte a settimana, con un monitoraggio continuo dei progressi, una valutazione iniziale nel mese di ottobre e una valutazione finale nel mese di maggio da parte dell'insegnante curriculare; la durata delle attività è di 90 minuti al giorno;
- *livello II*: consiste in un intervento aggiuntivo rivolto ai bambini a rischio, svolto in piccolo gruppo con un numero massimo di 5-6 alunni, per quattro volte a settimana, e prevede inoltre un monitoraggio continuo dei progressi; questo intervento è svolto da psicologhe tirocinanti, dopo un periodo di formazione sui disturbi specifici di apprendimento; la durata delle attività è di 30 minuti al giorno, per 4 giorni a settimana;
- *livello III*: consiste nella valutazione globale del profilo psicologico e dello stato degli apprendimenti per i bambini che non recuperano dopo l'intervento di I e II livello; la valutazione è svolta da professionisti dell'ASL di competenza o di un centro qualificato per un inquadramento del caso e l'eventuale successivo avvio di un intervento riabilitativo di natura clinica.

All'interno della ricerca-azione si effettua un confronto tra i progressi compiuti da tre gruppi di alunni:

- *gruppo di controllo*: fanno parte del gruppo gli alunni che non effettuano attività curriculari del I livello ma solo le valutazioni iniziali nel mese di ottobre e quelle finali nel mese di maggio, da parte di personale esterno (psicologhe tirocinanti);
- *gruppo curriculare*: fanno parte del gruppo gli alunni che effettuano l'attività didattica di I livello;
- *gruppo di potenziamento*: fanno parte del gruppo un esiguo numero di alunni "a rischio" per le abilità di lettoscrittura che usufruiscono di un intervento da parte di uno psicologo tirocinante, in aggiunta all'intervento di I livello di tipo curriculare. Questi bambini sono stati successivamente confrontati con un gruppo a rischio "paragonabile" di bambini che non partecipavano però all'intervento di II livello.

3.1. Articolazione del progetto

3.1.1. Partecipanti

Il progetto di ricerca-azione “RTI Abruzzo” ha coinvolto:

- gli alunni delle classi prime e seconde delle scuole primarie degli istituti comprensivi di Loreto Aprutino, Pianella e Collecervino in provincia di Pescara, dell’istituto comprensivo di Pineto in provincia di Teramo e dell’istituto comprensivo IV di Chieti;
- un gruppo di lavoro composto da uno staff di sei docenti facenti parte degli istituti comprensivi di Loreto Aprutino, Pianella, Pineto e Chieti, per la preparazione di attività curriculari e del relativo monitoraggio settimanale;
- un gruppo di lavoro costituito da psicologi e tirocinanti per la predisposizione di attività di potenziamento rivolte agli alunni individuati “a rischio”;
- un gruppo di lavoro, denominato “staff di progetto”, costituito da psicologi, tirocinanti e insegnanti per la fase progettuale.

I componenti dei gruppi di lavoro hanno effettuato un periodo di formazione di 20 ore, integrato con attività di supervisione. Inoltre, hanno effettuato riunioni con cadenza settimanale e/o quindicinale per la predisposizione del materiale, la discussione sull’implementazione del programma e la rilevazione dei punti di forza e di debolezza.

3.1.2. La ricerca-azione: intervento curriculare di I livello

Il percorso di ricerca-azione sulle abilità di lettura e comprensione si è svolto negli anni scolastici 2013-14 e 2014-15 mediante incontri di:

- formazione iniziale, nel mese di settembre, sugli strumenti per la valutazione delle abilità di letto-scrittura: *screening* iniziale e finale da somministrare agli alunni coinvolti;
- studio e lavoro di gruppo, a cadenza settimanale, tra gli insegnanti degli istituti comprensivi che hanno aderito al progetto e gli psicologi, per la preparazione delle attività del curricolo, del monitoraggio settimanale (I livello) per le classi prime e seconde della scuola primaria, dal mese di settembre a maggio;
- lavoro di gruppo, a cadenza quindicinale, per la preparazione delle attività di potenziamento per gli alunni a rischio, dal mese di ottobre a maggio;
- monitoraggio e supervisione sulle attività di intervento di I e II livello per gli insegnanti e gli psicologi;
- monitoraggio delle attività svolte e dell’intero progetto di ricerca-azione all’interno delle singole istituzioni scolastiche;
- restituzione dei dati sugli alunni, elaborati e analizzati dagli psicologi tirocinanti, agli insegnanti e ai genitori;
- monitoraggio e valutazione dell’intero impianto progettuale da parte dei gruppi di lavoro.

Le attività curriculari, preparate dagli insegnanti, hanno seguito un protocollo stabilito dal gruppo di studio e dagli psicologi da implementare per 5 giorni settimanali e per la durata di 90 minuti di insegnamento quotidiano di italiano. A titolo di esempio si indica una tipologia di unità didattica utilizzata per la classe prima, per una settimana.

Ogni giorno, prima dell’avvio dell’attività didattica, l’insegnante svolge l’“appello emozionale”, della durata di circa 30 minuti: si tratta di un momento di condivisione delle emozioni nel quale ogni bambino che risponde all’appello dice come si sente, se la mattina si sente felice lo esprime senza problemi, se è triste può trovare aiuto nei compagni e nelle maestre che capiscono il suo stato d’animo.

Il primo giorno di attività prevede la lettura di una storia, di un racconto o di filastrocche da parte dell’insegnante a cui fa seguito la ricerca orale delle parole che contengono il fonema o la sillaba “bersaglio” attraverso l’utilizzo di giochi fonologici.

Il giorno successivo l’attività prevalente è la comprensione mediante la drammatizzazione del racconto proposto il giorno precedente. In seguito, con l’aiuto dell’insegnante, gli alunni individuano le sequenze del racconto in successione temporale utilizzando le parole “bersaglio” conosciute. In tal modo gli alunni arricchiscono il loro vocabolario e sviluppano l’oralità.

Nei giorni successivi, l’insegnante attraverso giochi, dettato di parole, lettura di grafemi e parole, pone l’attenzione allo sviluppo di processi cognitivi di discriminazione visuo-spaziale e di discriminazione uditiva. L’ultimo giorno di attività settimanale è dedicato al consolidamento di quanto svolto e all’attività di monitoraggio.

Quest'ultima attività consiste in una prova di fluency della durata di un minuto: l'insegnante, in un ambiente fuori dall'aula, con il cronometro, sottopone l'alunno, individualmente, a una prova di lettura ad alta voce. In una griglia appositamente predisposta registra il tempo di lettura e annota gli errori e le omissioni effettuate dall'alunno durante la prova. Inizialmente gli alunni leggono prima i singoli grafemi, le sillabe, poi le parole. Attraverso le attività di monitoraggio settimanali gli insegnanti hanno l'opportunità di rilevare l'andamento dell'apprendimento della lettura degli alunni della classe e di avere un primo riscontro in merito all'efficacia della propria azione didattica.

Per il percorso di potenziamento sono state costruite delle attività per gruppi di alunni con profili simili e sono stati proposti materiali per gradi di difficoltà. Per la classe seconda il protocollo di implementazione del curricolo è simile a quello della classe prima, con una variazione nelle attività.

In data 6 maggio 2015 le classi seconde sono state sottoposte alle Prove nazionali di valutazione in italiano: la prova preliminare di lettura e la prova di comprensione.

4. I risultati INVALSI

Dall'anno scolastico 2013-14 l'istituto comprensivo di Loreto Aprutino ha avviato una riflessione, uno studio e un'analisi dei risultati delle prove INVALSI concentrando l'attenzione sulla prova preliminare di lettura per rilevare quanto l'approccio RTI abbia inciso sull'apprendimento della letto-scrittura negli alunni delle classi che hanno partecipato al progetto di ricerca-azione.

Lo scopo di tale prova è di verificare quale percentuale di alunni non abbia ancora raggiunto un sufficiente grado di automatismo nella decodifica di parole scritte, misurato dal numero di parole lette correttamente nel tempo assegnato.

La capacità strumentale di lettura costituisce un indispensabile pre-requisito per lo sviluppo della capacità di comprensione.

Analizzando la tabella sui punteggi generali della prova preliminare di lettura, nell'a.s. 2013-14 sono stati rilevati punti di criticità e caduta in tutte le parti della prova, in particolare nella decodifica di parole trisillabe, quadrisillabe e polisillabe; la maggior parte degli alunni non ha terminato la prova.

Questo dato ha confermato che nelle classi prime e seconde sarebbe stato necessario predisporre un intervento curricolare di I livello da parte degli insegnanti per garantire il successo formativo degli alunni, puntando su fluency e correttezza della lettura di parole per poi ottenere dei miglioramenti nella comprensione del testo.

Tab. 1 – Risultati della prova preliminare di lettura di II primaria – IC Loreto Aprutino

	<i>Punteggio</i>	<i>% bisillabe corrette</i>	<i>% trisillabe corrette</i>	<i>% quadrisillabe corrette</i>	<i>% polisillabe corrette</i>
Istituto comprensivo di Loreto Aprutino	76,1	94,3	81,0	51,7	30,5
Abruzzo	78,2	96,7	85,6	50,1	31,5
Differenze nei risultati rispetto a scuole con background familiare simile	-9,3	-2,4	-9,0	-16,8	-26,4
Sud	79,4	95,0	83,2	57,6	45,5
Italia	76,6	96,0	82,9	48,4	32,1

Fonte: elaborazione propria su dati INVALSI 2014.

A settembre 2015, i risultati delle prove somministrate nell'anno 2014 sono stati comparati con le prove somministrate a maggio 2015 nelle classi che hanno seguito l'approccio RTI.

Dalla fig. 1 possiamo ricavare due informazioni importanti: confrontando il punteggio ottenuto nelle due annualità (86,7% nell'a.s. 2014-15 e 76,1% nell'a.s. 2013-14) si osserva che i risultati delle prove preliminari di lettura sono migliorati di 10,6 punti percentuali. Inoltre rispetto al precedente anno scolastico, in cui solo il 30,5% degli studenti aveva terminato la prova (polisillabe), nel 2014-15 si è riscontrato un incremento di questa percentuale di circa il 7,8%. Il 38,3% degli studenti ha completato la prova nei tempi stabiliti.

Questi risultati ci permettono di constatare che l'intervento di prevenzione del progetto RTI, rivolto a tutte le classi prime e seconde della scuola primaria, ha migliorato la fluency nella lettura.

Fig. 1 – Confronto prova preliminare di lettura tra le due annualità (istituto comprensivo di Loreto Aprutino)

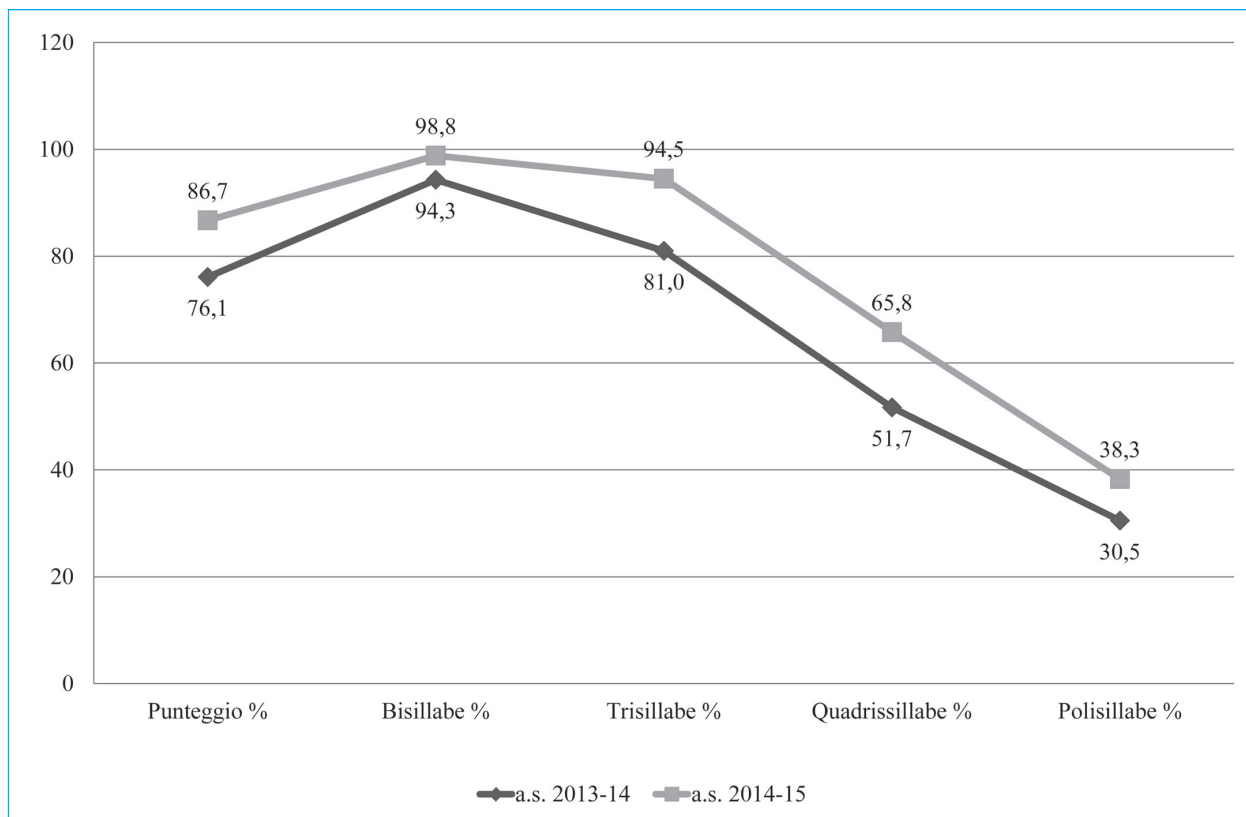
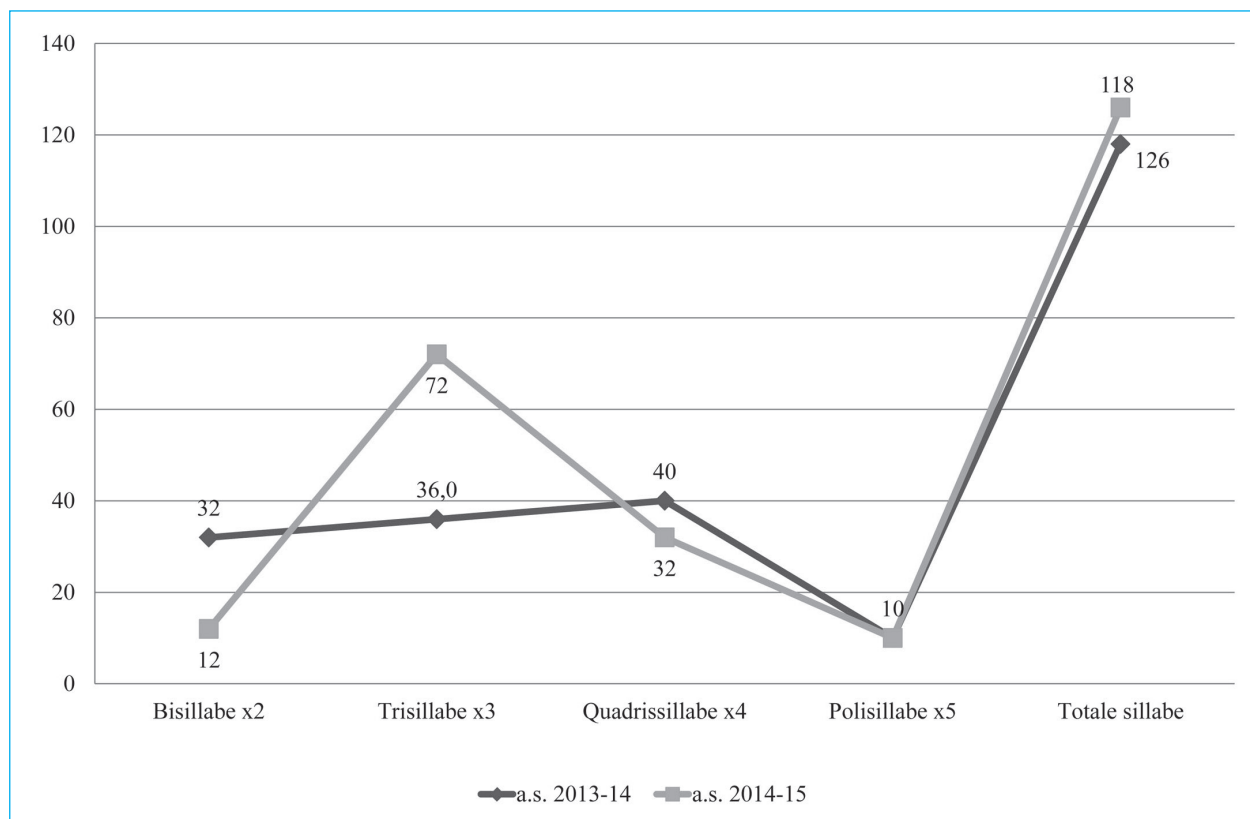


Fig. 2 – Prova preliminare di lettura – confronto numero sillabe (istituto comprensivo di Loreto Aprutino)



A livello d'istituto abbiamo inoltre analizzato e comparato la prova preliminare usata nelle due annualità di riferimento per osservare le analogie e le differenze nella strutturazione della prova e nel livello di complessità. Nella fig. 2 si possono ricavare alcune informazioni utili.

La prima informazione riguarda il numero di sillabe da leggere in 120 secondi: nell'a.s. 2013-14 la prova preliminare era costituita da un totale di 118 sillabe mentre nell'a.s. 2014-15 la prova somministrata era composta da 126 sillabe. Questo dato ci fa constatare che la prova oggetto di studio proposta nell'a.s. 2014-15 aveva 8 sillabe in più rispetto alla prova proposta nell'a.s. 2013-14.

La seconda informazione riguarda il confronto tra il numero delle parole bisillabe e trisillabe presenti nelle due prove: nella prova del 2013-14 erano presenti 16 parole bisillabe per un totale di 32 sillabe; nella prova del 2014-15 il numero di parole bisillabe da leggere era solo 6 per un totale di 12 sillabe mentre aumentava notevolmente il numero delle parole trisillabe.

Queste informazioni ci permettono di constatare che la prova preliminare di lettura somministrata nel 2014-15 risultava essere più complessa rispetto alla prova del 2013-14. Quindi l'approccio RTI e gli interventi a vari livelli hanno contribuito al miglioramento nella fluenza della lettura.

5. La ricerca RTI nelle azioni progettuali e didattiche

Il percorso di miglioramento dei livelli delle prove INVALSI e dei relativi processi cognitivi risponde al Rapporto di autovalutazione (RAV), al Piano di miglioramento (PDM), e agli obiettivi indicati dall'Ufficio scolastico regionale per le istituzioni scolastiche abruzzesi. Risponde anche alla necessità di rendere più trasparenti, condivise e consapevoli le scelte progettuali e didattiche relative all'insegnamento-apprendimento e alla valutazione della lettura.

Il protocollo di ricerca applicato alla didattica d'aula rende il docente riflessivo, lo aiuta a comprendere gli eventuali errori nei processi attivati, lo motiva nella produzione di testi e letture che rispondano ai processi cognitivi da innescare. Contestualmente, permette di mantenere la specificità della scuola primaria che fa leva sulle emozioni, sul pensiero narrativo degli alunni, per conservare negli anni i livelli di partecipazione, interesse, curiosità e lettura autonoma.

Il percorso strutturato di ricerca RTI aiuta a individuare un possibile "cattivo insegnamento" della lettura, cioè un insegnamento poco efficace, e facilita la comprensione del testo.

Il progetto curricolare, all'interno della ricerca-azione RTI, coinvolge gli alunni a partire dalle classi prime, sviluppando quei processi che gli studi recenti (Taylor, 2008) sui disturbi di apprendimento hanno dimostrato essere rilevanti, come la consapevolezza fonemica, la fonetica, la fluenza, l'ampliamento del vocabolario e la comprensione (Taylor, 2008; National Reading Panel, 2000). Le attività si declinano in una struttura flessibile ma chiara, in un cronoprogramma condiviso anche con le famiglie che sono guidate a comprendere sia processi "dal basso in alto" (analisi fonologica e visiva) sia processi "dall'alto in basso" (vocabolario e narrazione).

Le vocali, le consonanti, le sillabe, le parole e i testi, presentati secondo i principi di gradualità, prendono vita in un racconto autoprodotta insieme ai bambini. Inoltre i vari fonemi, scelti in base al protocollo, entrano in contatto fra loro in forma di personaggi nelle semplici storie che quotidianamente arricchiscono le letture dei bambini. Ogni racconto diviene filastrocca in rima, ma anche rielaborazione di conoscenze scientifiche. La produzione di storie, emotivamente vissute e coinvolgenti, rende la ricerca piacevole per ogni protagonista coinvolto, che diviene un lettore sempre più abile, fluido ed espressivo, migliorando la fluenza della lettura e la comprensione del testo.

In tal senso, l'intervento si qualifica all'interno della relazione docente di classe-studente che, attraverso la ricerca RTI, porta il docente a comprendere meglio gli aspetti fondamentali della competenza nella lettura e il senso della prova di pre-lettura, rintracciabili nel *Quadro di riferimento della lingua italiana* a cura dell'INVALSI, nella versione aggiornata al 28 febbraio 2011 (INVALSI, 2012). In questo modo la ricerca, avviata per riconoscere i segnali di probabili disturbi specifici di lettura, di "cattivo insegnamento", promuove anche lo sviluppo professionale degli insegnanti e l'attivazione di approcci didattici fondati su costrutti scientifici, condivisi all'interno di una comunità, basati su contenuti didattici sempre più coerenti con le Indicazioni nazionali.

6. Conclusioni: i risultati raggiunti con l'approccio RTI

Prima dell'avvio del progetto, nel percorso di formazione iniziato con il master di I livello in Didattica e psicopedagogia per i Disturbi specifici di apprendimento (DSA), Bando MIUR, i docenti coinvolti hanno condiviso le potenzialità dell'approccio RTI nel riuscire a soddisfare le esigenze formative e di apprendimento degli studenti. In particolare, gli scopi dichiarati riguardavano la necessità di restringere il divario esistente nei livelli degli apprendimenti della lettura degli studenti per promuovere progressi costanti, anche se gradualmente, e prevenire comportamenti indesiderati.

L'avvio della ricerca ha coinvolto la comunicazione degli obiettivi e dei risultati attesi per creare le basi per l'implementazione dell'innovazione. In seguito la ricerca RTI ha coinvolto tutti i docenti e gli alunni nei primi due anni di scuola primaria. La ricerca, stabilendo obiettivi chiari, ha reso il protocollo RTI un elemento fondamentale e necessario per soddisfare le esigenze dei bambini e dei docenti circa la scelta della strategia migliore per acquisire la lettura.

I cambiamenti, nei tre anni scolastici, a partire dal 2013-14, sono stati monitorati e i diversi professionisti coinvolti (docenti, gruppi di lavoro, staff di dipartimento) hanno concordato che il cambiamento ha avuto inizio nel momento in cui lo staff della scuola e quello dell'università hanno lavorato insieme, durante le riunioni di gruppo e le sessioni di lavoro per la predisposizione del percorso didattico e delle prove di monitoraggio. Anche i genitori hanno rilevato cambiamenti negli incontri periodici e in fase di restituzione dei dati. Negli allievi si è manifestata la sensazione che "ogni studente può imparare", nei docenti che "tutto il personale può insegnare agli studenti con la consapevolezza dei processi cognitivi coinvolti".

Per i docenti dell'istituto comprensivo di Loreto Aprutino, questo cambiamento ha rappresentato una sfida significativa e difficile da affrontare secondo le prassi abituali. I docenti, inoltre, hanno avuto modo di constatare che le attività di implementazione RTI richiedono attenzione, pianificazione e tempo per la necessaria modifica delle pratiche quotidiane. Allo stesso tempo, però, una volta che le pratiche di implementazione hanno cominciato a produrre risultati attribuibili all'approccio RTI, il personale docente ha compreso la trasformazione positiva che il percorso stava producendo negli studenti. La perseveranza nel tempo (dal 2013 al 2016) ha permesso agli insegnanti di maturare la capacità di essere flessibili e creativi durante il processo di implementazione, consentendo loro la libertà di assumere rischi, provare strategie nuove, imparare da ciò che non funziona e andare avanti con ciò che funziona.

La ricerca RTI ha consentito uno sviluppo professionale continuo, ha dato risposte a molte preoccupazioni, ha apportato modifiche strutturali che hanno favorito la collaborazione e il coinvolgimento pratico nel processo decisionale. Lo staff di ricercatori (psicologi), di docenti referenti e coordinatori della scuola, di insegnanti impegnati in prima linea nelle classi, ha assicurato che gli studenti coinvolti fossero reattivi agli interventi, assicurando la trasparenza del processo decisionale e delle procedure basate sui dati che informano le decisioni didattiche adottate. Negli incontri di restituzione dei dati al collegio dei docenti tutto il personale della scuola ha avuto l'opportunità di essere coinvolto nel processo decisionale, almeno una volta durante l'anno scolastico. Tutti hanno partecipato, collaborato e condiviso le conoscenze e i risultati positivi registrati con le classi seconde nella prova preliminare di lettura.

Riferimenti bibliografici

- Abadzi H. (2012), "Developing cross-language metrics for reading fluency measurement: some issues and options", *Global Partnership for Education (GPE) Working Paper Series on Learning*, 6, testo disponibile al sito: <https://openknowledge.worldbank.org/bitstream/handle/10986/26819/797740WP0wpm0e0Box0379789B00PUBLIC0.pdf?sequence=1&isAllowed=y>, data di consultazione: 19 giugno 2017.
- Deno S.L., Fuchs L.S., Marston D., Shin J. (2001), "Using curriculum based measurement to establish growth standards for students with learning disabilities", *School Psychology Review*, 30, 4: 507-524.
- Fuchs D., Fuchs L.S. (2006), "Introduction to response to Intervention: what, why, and how valid is it?", *Reading Research Quarterly*, 41, 1: 93-99.
- INVALSI (2012), *Quadro di riferimento della prova di italiano*, testo disponibile al sito: http://www.invalsi.it/snv2012/documenti/QDR/QdR_Italiano.pdf, data di consultazione: 20 marzo 2017.
- Lyytinen H., Erskine J., Ahonen T., Aro M., Eklund K., Guttorm T., Hintikka S., Hämäläinen J., Ketonen R., Laakso M.-L., Leppänen, P.H.T., Lyytinen P., Poikkeus A.-M., Puolakanaho A., Richardson U., Salmi P., Tolvanen A., Torppa M., Viholainen H. (2008), "Early identification and prevention of dyslexia: results from a prospective follow-up study of children at familial

-
- risk for dyslexia”, in G. Reid, A.J. Fawcett, F. Manis, L.S. Siegel (eds.), *The Sage Handbook of Dyslexia*, Sage, Thousand Oaks (CA): 121-146.
- MIUR (2012), *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione*, Roma, testo disponibile al sito: www.indicazioninazionali.it/documenti_Indicazioni_nazionali/indicazioni_nazionali_infanzia_primo_ciclo.pdf, data di consultazione: 20 giugno 2017.
- Muter V. (2006), “The Prediction and Screening of Children’s Reading Difficulties”, in M.J. Snowling, J. Stackhouse (eds.), *Dyslexia, Speech and Language: A Practitioner’s Handbook*, Whurr Publishers, London-Philadelphia, 2nd ed.: 54-72.
- National Reading Panel (2000), *Report of the National Reading Panel: Teaching Children To Read: An Evidence-Based Assessment Of The Scientific Research Literature On Reading And Its Implications For Reading Instruction: Reports Of The Subgroups*, National Institute of Child Health and Human Development, Washington (DC).
- Stanovich K.E. (1986), “Matthew effects in reading: some consequences of individual differences in the acquisition of literacy”, *Reading Research Quarterly*, 21, 4: 360-407.
- Stella G., Grandi L. (2011), *La dislessia e i DSA*, Giunti, Brescia.
- Taylor B.M. (2008), “Tier 1: Effective classroom reading instruction in the elementary grades”, in D. Fuchs, L.S. Fuchs, S. Vaughn (eds.), *Response to Intervention: A Framework for Reading Educators*, International Reading Association, Newark (DE): 5-25.
- Tunmer W.E., Greaney K.T. (2008), “Reading intervention research: an integrative framework”, in G. Reid, A.J. Fawcett, F. Manis, L.S. Siegel (eds.), *The Sage Handbook of Dyslexia*, Sage, Thousand Oaks (CA): 241-267.

4. Il database GESTINV delle prove standardizzate INVALSI: uno strumento per la ricerca.

Alcuni esempi di utilizzo nell'ambito della matematica

The GESTINV database of INVALSI standardized tests: a research tool.

Some examples of utilization in Mathematics

di Giorgio Bolondi, Federica Ferretti, Alessandro Gambini

Le valutazioni standardizzate degli apprendimenti sono progettate con il fine di avere un impatto a livello sistemico. Un problema al centro di un forte dibattito epistemologico, ideologico e didattico è come integrare i risultati, i metodi, i quadri teorici e in generale gli strumenti di queste valutazioni nelle azioni locali di insegnanti e scuole. Oggetto di questo capitolo è la descrizione del progetto di ricerca GESTINV, destinato a fornire strumenti e modelli di azione per affrontare questo problema. Il database GESTINV contiene più di 1.400 item delle rilevazioni standardizzate nazionali effettuate dal SNV per INVALSI ed è utilizzato in ricerche in didattica della matematica e in programmi di sviluppo professionale per i docenti in servizio e in formazione.

An issue at the core of an intense epistemological, didactical and, sometimes, ideological debate is how to integrate results, methods, theoretical frameworks and tools of standardised assessments –which are designed in order to impact at a systemic level – into the local actions of teachers and schools. This appears as an important research topic in teacher training research. We describe a project (the GESTINV project) intended to provide large-scale tools and models of action for addressing this issue. The GESTINV database contains more than 1.400 items administered in the Italian national standardised tests (SNV) for INVALSI and is used in professional development programs developed by schools and in many researches in math education.

1. Le valutazioni standardizzate in ottica formativa

A livello internazionale la valutazione degli studenti sta assumendo sempre più un ruolo centrale anche sul piano politico-istituzionale. Da diversi anni la maggior parte dei Paesi partecipanti alle rilevazioni OECD-PISA 2015 (Looney, 2011) ha decentrato i sistemi di istruzione per cercare di soddisfare sempre più le esigenze locali e ha sviluppato le valutazioni nazionali su larga scala per monitorare il sistema scolastico. Si sta sempre più andando verso la direzione di vedere la scuola in generale come quell'istituzione che accompagna, aiuta e sostiene gli studenti nei loro percorsi di apprendimento, e sta svanendo sempre più l'idea dell'insegnante valutatore che "etichetta" lo studente con un valore. Ed è proprio in questa ottica che si sta sviluppando l'idea di utilizzare i dati forniti dalle valutazioni standardizzate per identificare i punti di forza e di debolezza non solo del sistema scuola, ma anche delle prestazioni degli studenti. Come fare a integrare la valutazione standardizzata con la valutazione formativa? Quali strumenti possono essere utili a tal fine?

In letteratura esistono diverse definizioni di valutazione formativa; noi ci riferiamo a quella condivisa nata all'interno del progetto LLP-Comenius FAMT&L – *Formative Assessment for Teaching and Learning in Mathematics*¹ (fig. 1).

¹ Il progetto Comenius FAMT&L rientra nei programmi europei LLP e ha come focus principale quello di costruire un modello di formazione degli insegnanti volto a migliorare le loro competenze nell'ambito della valutazione formativa in matematica. I cinque membri partner sono tutte istituzioni universitarie di Paesi europei e hanno complessivamente competenze in pedagogia, didattica e in matematica. In particolare il partner capofila è l'Università di Bologna, con la prof.ssa Vannini (Dipartimento di Scienze dell'educazione) e il prof. Bolondi (Dipartimento di Matematica).

La valutazione formativa (VF) è connessa con un concetto di apprendimento secondo cui tutti gli studenti sono in grado di acquisire, a un livello adeguato, le competenze di base di una disciplina. L'apprendimento passa attraverso l'utilizzo di metodologie di insegnamento che possono rispondere efficacemente ai tempi di apprendimento diversi per ogni studente, ai loro diversi stili di apprendimento, alle loro zone di sviluppo prossimale.

Inoltre la VF:

- è parte del processo di insegnamento-apprendimento e lo regola;
- identifica, in modo analitico, i punti forti e quelli deboli dell'apprendimento dell'allievo, al fine di consentire agli insegnanti di riflettere sulle proprie pratiche didattiche e di modificarle;
- permette un feedback formativo al fine di stabilire un dialogo tra docente e studente e per programmare interventi didattici finalizzati al recupero;
- promuove e favorisce l'apprendimento di tutti gli studenti attraverso l'insegnamento differenziato che garantisce a ogni studente ritmi diversi e diverse strategie di insegnamento e apprendimento;
- coinvolge lo studente nell'analisi dei propri errori o debolezze e delle proprie capacità per promuovere sia l'autovalutazione sia la valutazione tra pari e la partecipazione attiva nel processo di insegnamento-apprendimento.

Fonte: Ferretti e Lovece (2015, p. 51).

Abbracciando la definizione precedente, ci troviamo all'interno di una visione di *valutazione per l'apprendimento* che interviene nel processo didattico in modo formativo sia per lo studente sia per l'insegnante. La principale funzione della valutazione diventa così quella regolativa, nel senso dell'adattamento e della rimodulazione continua dei percorsi didattici in base alle esigenze degli studenti. La valutazione diventa così anche uno strumento per garantire la qualità del livello di competenze raggiunto dagli studenti e il raggiungimento degli obiettivi curricolari (Vannini, 2009). Ricerche internazionali hanno infatti mostrato che il coinvolgimento attivo degli studenti nel processo valutativo, la restituzione di prove scritte e orali in modalità di feedback formativi e l'utilizzo della valutazione come strumento di incremento della motivazione e dell'autostima degli studenti sono tutti fattori determinanti per l'efficacia della funzione formativa della valutazione e quindi per una ricaduta positiva sul processo di apprendimento (Black e William, 1998).

In particolare, nel contesto italiano, il ruolo centrale che deve avere la valutazione all'interno dei processi di apprendimento/insegnamento è anche esplicitato nelle Indicazioni nazionali (MIUR, 2012), in cui vengono sottolineati la funzione formativa della valutazione e il fatto che essa viene affidata a insegnanti, istituzioni scolastiche e istituzioni ministeriali (come l'INVALSI). Infatti il Servizio nazionale di valutazione (SNV) effettua a livello di sistema valutazioni standardizzate in modo censuario dal 2008 in italiano e matematica su diversi livelli (con inizio in diversi anni) che ricoprono tutti i segmenti scolastici del sistema di istruzione italiano ed effettua una restituzione campionaria per ogni item delle rilevazioni svolte. Questa mole di dati fornisce informazioni puntuali sia a livello di sistema sia a livello di performance degli studenti mostrando fenomeni rilevanti su larga scala.

A livello nazionale sono sempre più numerose le ricerche che partono da evidenze di dati di queste valutazioni standardizzate, sia in direzione della formazione insegnanti (Martignone, 2016; Bolondi, Ferretti e Spagnuolo, 2016) sia in termini di impatto che questi fenomeni possono avere sulle pratiche didattiche (Ferretti, Lemmo e Maffia, 2015).

In questo contributo mostriamo uno strumento nato anche per mettere a disposizione degli insegnanti, in maniera ragionata e strutturata, tutti i materiali del sistema di prove INVALSI (Quadri di riferimento, prove rilasciate, risultati) di matematica, e una ricerca costruita per validare modelli di utilizzo di questo strumento.

2. Il database: uno strumento di ricerca in mano agli insegnanti

Diverse ricerche stanno mostrando come, integrando analisi quantitative e qualitative, si possano identificare nuove metodologie di ricerca (Ferretti, Lemmo e Maffia, 2016) volte a utilizzare in modo formativo i risultati delle valutazioni standardizzate. In linea con l'idea di poter costruire e utilizzare strumenti e metodologie, è nato il progetto di ricerca GESTINV, destinato a fornire strumenti e modelli per suggerire piste di azione in questa direzione. La nostra ipotesi è che i test standardizzati possano restituire anche all'insegnante sul campo una grande quantità di informazioni sui processi di apprendimento e fornire diversi feedback in termini di efficacia del proprio insegnamento. Le informazioni

sono contenute non solo nei punteggi globali (misurati dai modelli statistici), ma anche in fatti puntuali, osservabili nelle risposte date alle singole domande. In particolare, i risultati delle prove INVALSI mettono in evidenza e quantificano molti macro-fenomeni rilevanti che possono venire interpretati grazie a metodi e risultati della ricerca in didattica. La possibilità di utilizzo di queste informazioni da parte degli insegnanti è un tema di ricerca estremamente interessante, per le sue ricadute sia in termini operativi sulla didattica messa in campo, sia in termini di immagine e accettazione delle prove standardizzate.

Il cuore della ricerca è un database di semplice utilizzo contenente circa 1.400 item sviluppato da ForMATH Project (un gruppo di giovani ricercatori in Didattica della matematica), accessibile dal sito www.gestinv.it. In questo contributo analizzeremo, in particolare, il contenuto del database per quanto riguarda la matematica e alcune sue possibili applicazioni.

3. Il database delle prove di matematica

Il progetto è incentrato su un database online contenente i quesiti somministrati nelle valutazioni standardizzate INVALSI insieme a risultati, commenti, approfondimenti didattici, metadati, statistiche e analisi. Il database, a oggi, contiene 1.469 item di matematica² delle seguenti rilevazioni standardizzate nazionali effettuate dall'INVALSI:

- a.s. 2007-08: grado 8, Prova nazionale (scuola secondaria di I grado);
- a.s. 2008-09: grado 2 e grado 5 (scuola primaria) e grado 8, Prova nazionale (scuola secondaria di I grado);
- a.s. 2009-10: grado 2 e grado 5 (scuola primaria), grado 6 e grado 8, Prova nazionale (scuola secondaria di I grado);
- a.s. 2010-11: grado 2 e grado 5 (scuola primaria), grado 6 e grado 8, Prova nazionale (scuola secondaria di I grado) e grado 10 (scuola secondaria di II grado);
- a.s. 2011-12: grado 2 e grado 5 (scuola primaria), grado 6 e grado 8, Prova nazionale (scuola secondaria di I grado) e grado 10 (scuola secondaria di II grado);
- a.s. 2012-13: grado 2 e grado 5 (scuola primaria), grado 6 e grado 8, Prova nazionale (scuola secondaria di I grado) e grado 10 (scuola secondaria di II grado);
- a.s. 2013-14: grado 2 e grado 5 (scuola primaria), grado 8, Prova nazionale (scuola secondaria di I grado) e grado 10 (scuola secondaria di II grado);
- a.s. 2014-15: grado 2 e grado 5 (scuola primaria), grado 8, Prova nazionale (scuola secondaria di I grado) e grado 10 (scuola secondaria di II grado);
- a.s. 2015-16: grado 2 e grado 5 (scuola primaria), grado 8, Prova nazionale (scuola secondaria di I grado) e grado 10 (scuola secondaria di II grado).

All'interno del database, per ogni rilevazione c'è il PDF della prova completa e ogni item è accompagnato, in linea con la Guida alla lettura delle prove INVALSI, da risultati dettagliati, dati statistici e classificazioni in diverse categorie. In riferimento a ogni item c'è l'immagine della domanda, l'ambito di contenuto, il processo, il riferimento alle Indicazioni nazionali o Linee guida, alcune parole chiave caratterizzanti i contenuti in gioco, il testo della domanda in Microsoft Word, la risposta corretta o l'immagine della risposta corretta, le percentuali di risposta nazionali e altri dati statistici.

4. Alcuni possibili utilizzi del database per la didattica

Il database GESTINV è utilizzato intensivamente in programmi di sviluppo professionale per i docenti in servizio nelle scuole italiane e in diversi percorsi universitari di formazione iniziale. Nel 2015 è stato utilizzato in diversi corsi di formazione realizzati in circa 70 scuole, in cui sono stati coinvolti più di 1.800 insegnanti e in corsi di studio dell'Università di Bologna e della Libera Università di Bolzano. Il suo utilizzo è comunque aperto e lo staff organizza webinar periodici di formazione.

L'impatto di questo progetto è stato valutato sia quantitativamente sia qualitativamente, attraverso indicatori stan-

² L'esperienza del database GESTINV di matematica è in via d'estensione alle prove di italiano. A oggi, nel database sono presenti 367 item delle prove standardizzate nazionali di italiano (sezione di grammatica) effettuate dal SNV per l'INVALSI dall'a.s. 2008-09 all'a.s. 2015-16.

dard come il numero di utenti registrati (più di 4.000), il numero di accessi (in media, 200 ogni giorno), il tempo di permanenza e altro.

Le modalità di utilizzo del database sono molteplici poiché al suo interno si possono infatti effettuare numerose ricerche. Entrando nella sezione di matematica si possono effettuare:

- ricerca per Indicazioni nazionali e Linee guida (è presente un'indicizzazione con Obiettivi di apprendimento al termine della classe terza della scuola primaria, Obiettivi di apprendimento al termine della classe quinta della scuola primaria, Traguardi per lo sviluppo delle competenze al termine della scuola primaria, Obiettivi di apprendimento al termine della scuola secondaria di I grado, Traguardi per lo sviluppo delle competenze al termine della scuola secondaria di I grado, Indicazioni nazionali per i licei, Linee guida per gli istituti tecnici e professionali, gli Assi culturali);
- ricerca per parole chiave (sono presenti circa 200 parole chiave che identificano i contenuti in gioco principali per ogni item);
- ricerca *full-text* (è possibile fare una ricerca nel testo di tutte le domande, processi, indicazioni curriculari, parole chiave);
- ricerca guidata (è possibile fare una ricerca incrociata, con connettori e/o, di tutti parametri in riferimento a ogni item e altre caratteristiche, come le percentuali di risposta nazionali).

Per esempio, tramite lo strumento Ricerca guidata si possono cercare tutti i quesiti delle prove INVALSI di matematica della classe seconda della scuola secondaria di II grado (grado 10), dell'ambito Numeri, che hanno avuto percentuali di risposte su scala nazionale corrette inferiori al 50% (fig. 2).

Fig. 2 – Schermata del database nella sezione di matematica, Ricerca guidata su <http://www.gestinv.it/RicercaGuidata.aspx>

The screenshot shows the 'ARCHIVIO PROVE INVALSI MATEMATICA' interface. The user is logged in as 'Federica Ferretti'. The search section is titled 'RICERCA GUIDATA' and contains three rows of filters. The first row has 'Livello' set to '10'. The second row has 'Ambito' set to 'NUMERI'. The third row has 'Perc. risp...' set to '< 50'. At the bottom are buttons for 'Aggiungi', 'Elimina', and 'Cerca'.

Il database GESTINV restituisce l'elenco di tutti gli item corrispondenti ai parametri indicati, come per esempio la domanda D05 del grado 10 dell'a.s. 2010-11 che ha avuto percentuali nazionali di risposte corrette del 10,2% (fig. 3). Le performance degli studenti in questo item sono state oggetto di studio in ambito di ricerca (Ferretti, 2015) e all'interno di materiale divulgativo destinato a docenti della scuola italiana (Maffia, 2013).

D5. L'età della Terra è valutata intorno ai $4,5 \times 10^9$ anni. L'Homo Erectus è comparso circa 10^6 anni fa. Qual è la stima che più si avvicina all'età che la Terra aveva quando è comparso l'Homo Erectus?

- A. $4,5 \times 10^9$ anni
- B. $3,5 \times 10^9$ anni
- C. $4,5 \times 10^6$ anni
- D. $4,5 \times 10^3$ anni

5. L'utilizzo del database come strumento di ricerca

In diversi corsi di formazione per insegnanti in servizio e in laboratori didattici in corsi universitari che abbiamo implementato negli ultimi anni³, i partecipanti hanno utilizzato il database come uno strumento di ricerca.

Molti corsi da noi effettuati hanno infatti come obiettivo quello di collegare le Indicazioni nazionali e/o Linee guida con le pratiche d'aula. I percorsi si basano infatti su tre elementi fondamentali:

- 1) le Indicazioni nazionali che delineano con chiarezza gli obiettivi, e quindi la direzione da intraprendere. Oltre a questa funzione, le Indicazioni descrivono anche, con precisione, il quadro di riferimento generale: quale matematica insegnare e perché, suggerendo anche diverse metodologie da poter adottare in aula (questo è un quadro di riferimento esplicito e lavorare su di esso può aiutare ciascuno a prendere consapevolezza del proprio quadro di riferimento personale, che spesso rimane implicito);
- 2) l'analisi delle prove INVALSI e delle risposte della scuola e dei propri allievi (quando lavoriamo con insegnanti in servizio) permette di rendere concreti gli obiettivi di apprendimento fissati dalle Indicazioni nazionali e/o Linee guida esplicitandone la difficoltà. Innanzitutto, essendo ogni domanda delle prove riferita a obiettivi e traguardi delle Indicazioni nazionali e/o Linee guida, i nostri percorsi utilizzano questi collegamenti spesso per costruire segmenti di curriculum verticale. Inoltre, l'analisi dei protocolli e delle scelte effettuate dagli studenti permette spesso di collegare tutto il lavoro alle proprie pratiche didattiche (attuali o future);
- 3) le ricerche in Didattica della matematica. Le Indicazioni ci forniscono la direzione, le prove INVALSI ci restituiscono informazioni sugli effettivi apprendimenti degli studenti; tutto questo ha bisogno però di essere interpretato alla luce dei risultati ottenuti dalla ricerca. Alcune idee chiave elaborate negli ultimi decenni (per esempio contratto didattico, ambienti di apprendimento, misconcezioni, registri semiotici, gestione delle difficoltà e dei disturbi specifici di apprendimento, inserimento di allievi stranieri ecc.) possono e devono diventare strumenti potenti in mano agli insegnanti e ai futuri insegnanti.

Per quanto riguarda la metodologia, uno degli strumenti che più facilmente permette di collegare questi tre elementi è il database GESTINV.

Solitamente si inizia il percorso analizzando alcuni fenomeni evidenziati dalla letteratura in Didattica della matematica e le evidenze che emergono dai risultati delle valutazioni standardizzate (con insegnanti in servizio, spesso analizziamo le criticità che emergono dalla lettura dei dati delle proprie classi); si individua così un contenuto o una specifica competenza su cui lavorare. Da qui si aprono diverse strade di indagine e, in base al tipo di lavoro che si vuole intraprendere assieme, si effettuano ricerche con il database GESTINV. Per esempio, inserendo le parole chiave che caratterizzano il contenuto individuato, il database restituisce tutti i quesiti che fanno riferimento a esso. Si possono così individuare gli Obiettivi di apprendimento e i Traguardi per lo sviluppo delle competenze delle Indicazioni nazionali che fanno riferimento ai contenuti in gioco (ogni quesito è indicizzato con le Indicazioni nazionali e/o Linee guida).

³ La maggior parte dei corsi di formazione a cui ci riferiamo sono stati progettati, condotti e realizzati da ForMATH Project (www.formath.it).

Una volta individuato l'obiettivo, la ricerca con il database GESTINV permette di cercare di "raggiungerlo" costruendo percorsi in verticale; infatti, senza restringere la ricerca a un determinato livello, il sistema ricerca tutte le domande che fanno riferimento al parametro inserito e questo permette di scandire il raggiungimento di un determinato obiettivo con delle tappe in verticale.

In funzione del percorso che si sta intraprendendo si scelgono le ricerche e le relative letture dei dati restituite dal database GESTINV.

6. Conclusioni e direzioni future

Questo strumento di ricerca si inserisce certamente all'interno di un forte dibattito didattico ed epistemologico internazionale sull'utilizzo delle valutazioni standardizzate in ottica formativa. Da molti punti di vista il database GESTINV è uno strumento funzionale a diverse ricerche e in ambito di formazione insegnanti. Le analisi dei dati delle valutazioni standardizzate forniscono informazioni, globalmente, circa il sistema scolastico e, puntualmente, circa le prestazioni dei propri studenti. Questi dati, organizzati e indicizzati nel database GESTINV che ne facilita le ricerche su diversi piani, ancora di più possono essere oggetto di ricerche e fornire elementi utili per un miglioramento delle pratiche didattiche.

Riferimenti bibliografici

- Black P., William D. (1998), "Assessment and classroom learning", *Assessment in Education: Principles, Policy & Practice*, 5: 7-74.
- Bolondi G., Ferretti F., Spagnuolo A. (2016), "Le prove INVALSI con Geogebra: trasformare la valutazione standardizzata in valutazione formativa", in Robutti O. (a cura di), *La formazione docenti con Geogebra. Atti del IV GeoGebra Italian Day 2014*, Ledizioni, Milano: 99-108.
- Ferretti F. (2015), *L'effetto "età della Terra". Contratto didattico e principi regolativi dell'azione degli studenti in matematica*, tesi di Dottorato di ricerca in Matematica, Alma Mater Studiorum Università di Bologna, testo disponibile al sito: http://amsdottorato.unibo.it/7213/4/Ferretti_Federica_Tesi.pdf, data di consultazione: 10 maggio 2017.
- Ferretti F., Lemmo A., Maffia A. (2016), "Confrontare decimali e frazioni: analisi delle concezioni degli studenti a partire da una domanda INVALSI", *L'insegnamento della matematica e delle scienze integrate*, 39, 5: 451-464.
- Ferretti F., Lemmo A., Maffia A. (2015), "Half of something": how students talk about rationals", in K. Beswick, T. Muir, J. Wells (eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education*, vol. 1, PME, Hobart, Australia: 159.
- Ferretti F., Lovece S. (2015), "La valutazione formativa per la didattica della matematica nell'ambito del progetto FAMT&L. Le concezioni degli studenti di scuola media nei confronti degli strumenti di verifica utilizzati in classe", *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education*, 10, 2: 39-68.
- Looney J.W. (2011), "Integrating formative and summative assessment: progress toward a seamless system?", *OECD Education Working Papers*, 58, testo disponibile al sito: [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=edu/wkp\(2011\)4&doclanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=edu/wkp(2011)4&doclanguage=en), data di consultazione: 20 giugno 2017.
- Maffia A. (2013), *I 10 quesiti più difficili delle prove INVALSI di matematica, nella classe seconda della scuola secondaria di secondo grado*, Zanichelli, Bologna.
- Martignone F. (2016), "Un'attività di formazione per insegnanti di scuola secondaria di primo grado: analisi di prove INVALSI di matematica", *Form@re-Open Journal per la Formazione in Rete*, 16, 1: 70-86.
- MIUR (2012), *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione*, Roma, testo disponibile al sito: http://www.indicazioninazionali.it/documenti_Indicazioni_nazionali/indicazioni_nazionali_infanzia_primo_ciclo.pdf, data di consultazione: 20 giugno 2017.
- Perelli D'Argenzio M.P. (2006), "La valutazione esterna degli apprendimenti: le prove di valutazione INVALSI. Le prove di matematica", *L'insegnamento della matematica e delle scienze integrate*, 29, 1: 31-46.
- Van der Linden W.J., Hambleton R.K. (eds.) (1997), *Handbook of Modern Item Response Theory*, Springer, New York.
- Vannini I. (2009), *La qualità nella didattica. Metodologie e strumenti di progettazione e valutazione*, Erickson, Trento.
- Vergnaud G. (1988), "Multiplicative structures", in J. Hiebert, M. Behr (eds.), *Number Concepts and Operations in the Middle Grades*, National Council of Teachers of Mathematics, Reston, VA; Lawrence Erlbaum Associates, Hillsdale (NJ): 141-161.

5. Un'analisi qualitativa delle prove di matematica

A qualitative analysis of Mathematics tests

di Federica Ferretti, Alice Lemmo, Andrea Maffia

La letteratura mostra come l'uso della *Latent Class Analysis* permetta di individuare quesiti che presentano particolari difficoltà per gli studenti “poveri di conoscenza”. A partire da ciò abbiamo portato avanti un'ulteriore riflessione sulle prove di valutazione standardizzate nazionali allo scopo di fornire strumenti significativi utilizzabili nella pratica di insegnamento. La ricerca mira a indagare le strategie che portano lo studente all'errore e le motivazioni che lo inducono. Fra i contenuti matematici nei quali gli studenti mostrano particolare difficoltà, si sono selezionati i numeri razionali. In particolare si è concentrata l'attenzione sulla gestione delle loro diverse rappresentazioni. Abbiamo selezionato e somministrato alcuni quesiti significativi a studenti di classi quinte della scuola primaria e classi prime della scuola secondaria di I grado. Gli studenti del nostro campione provengono da due diverse città italiane (Bologna e Barletta) e da scuole con differenti background socio-economici. Le risposte degli studenti sono state analizzate individuando il tipo di conversione utilizzata (se presente), la direzione di conversione (da decimale a frazione o viceversa) e la correttezza della risposta. Si sono identificate quattro strategie ed è stato ipotizzato un possibile legame con diverse concezioni dei numeri razionali. In particolare, l'uso del registro colloquiale sembra essere più adatto di una qualunque manipolazione sintattica degli oggetti in ambito numerico. Fra i dati raccolti è possibile individuare entrambe le direzioni di conversione: la conversione da frazioni a decimali è stata scelta più frequentemente rispetto a quella da decimale a frazione nonostante questa strategia porti a un numero maggiore di errori. Infine, alcuni studenti fanno uso di rappresentazioni iconografiche che risultano essere poco efficaci: meno della metà degli studenti che le adottano risponde correttamente. In particolare, la difficoltà più comune incontrata dagli studenti che usano questa strategia consiste nel trasformare il decimale in una rappresentazione grafica. Inoltre, nonostante non fosse uno degli obiettivi di questo studio, si sono osservate diverse tipologie di errori nella conversione da frazioni a decimali e viceversa. Analisi di questo tipo, a partire dai dati delle Rilevazioni nazionali, possono suggerire diverse piste di intervento per migliorare la pratica didattica e incidere sui processi di insegnamento/apprendimento.

Latent Class Analysis allows to identify items which are particularly difficult for “students poor in knowledge”. Drawing on these results, we developed a further reflection on standardised national tests. The aim of this work is to propose some tools for teaching practice. In particular, this research aims to investigate students' strategies, highlighting those that conduct easily to errors. Among the mathematical contents in which students show particular difficulties, we selected rational numbers. We focus our attention on the ways of managing different representations. According with this purpose, we select and administer some significant tasks to students of grade 5 and 6. Our sample includes students from two different Italian cities (one from the North and one from the South) with different socio-economic background. Each author analysed the data identifying the type of conversion (if there is one), the conversion direction (from decimal to fractions or vice versa) and the correctness of the answers. We identify four strategies and conjecture that each one is linked with a different conception of rationals. In particular, the use of colloquial register seems to be more suitable than any syntactical manipulation of the mathematical objects involved in the task. Within the data, it is possible to identify both the directions of conversion: conversion from fraction to decimal is more frequently chosen than conversion from decimal to fraction, even if this strategy leads to a larger amount of mistakes. A possible interpretation is that, in primary school, decimal numbers are generally introduced before fractions, differently than the historical process. Finally, some students use

iconographic representation; they result as ineffective: less than half of the students give a correct answer. In particular, the most frequent difficulty is in transforming the decimal representation in a graphic one. Furthermore, even if it was not one of the aims of this study, we observed many different typologies of errors in conversion from fractions to decimal and vice versa. This kind of analysis, drawing on data from national surveys, can suggest different lines of action with the goal of developing educational practices and affecting significantly the educational processes.

1. Introduzione

Negli ultimi anni, le prove standardizzate INVALSI stanno acquisendo un ruolo crescente all'interno della ricerca in Didattica della matematica. I primi articoli riguardo alle prove che si possono trovare sulle riviste di Didattica della matematica si occupano soltanto di presentare il quadro di riferimento alla comunità italiana (Perelli D'Argenzio, 2006; Bolondi, 2010), mentre i più recenti integrano le prove e i loro risultati presentandoli anche al contesto internazionale (Branchetti *et al.*, 2015; Ferretti, Lemmo e Maffia, 2015; Lemmo *et al.*, 2015; Giberti, Zivelonghi e Bolondi, 2016; Jakobsen *et al.*, 2016).

Generalmente negli articoli di ricerca in Didattica della matematica non si ricavano risultati a partire da analisi statistiche dei dati raccolti dall'INVALSI. Quando vengono citati i risultati, questi sono utilizzati come stimolo per ripensamenti sull'azione didattica (De Virgilis e Pesci, 2014); altrimenti si utilizzano i quesiti come consegne per attività in classe (*ibid.*) o nella formazione degli insegnanti (Jakobsen *et al.*, 2016; Martignone, 2016).

Tuttavia, i risultati raccolti nelle diverse rilevazioni che si sono succedute negli anni, forniscono numerose informazioni circa le conoscenze degli studenti italiani, le loro abilità in matematica e gli errori più frequentemente commessi. Molti di questi risultati corroborano evidenze che la ricerca in Didattica della matematica ha ottenuto utilizzando metodologie di tipo qualitativo. Lo scopo di questo contributo è quello di mostrare un possibile intreccio di metodi finalizzato a mettere in comunicazione i dati INVALSI con i risultati qualitativi che si ottengono con i metodi di analisi più tipici della ricerca in Didattica della matematica.

In pubblicazioni precedenti si è mostrato come l'uso della *Latent Class Analysis* permetta di individuare quesiti che presentano particolari difficoltà per gli studenti “poveri di conoscenza” (Bolondi *et al.*, 2016; Branchetti *et al.*, 2015).

L'analisi quantitativa, presentata nel prossimo paragrafo, fa uso sia dei dati INVALSI, sia di strumenti che consentono di analizzare il comportamento degli studenti nelle singole domande individuando i gruppi di studenti più deboli.

A partire da questi risultati abbiamo portato avanti un'ulteriore riflessione sulle prove di valutazione standardizzate nazionali effettuando una successiva sperimentazione. In dettaglio, l'analisi qualitativa (sviluppata successivamente) mira a indagare le strategie che portano lo studente all'errore e le motivazioni che lo inducono. Fra i contenuti matematici nei quali gli studenti mostrano particolare difficoltà, si sono selezionati (a titolo di esempio) i numeri razionali e in particolare si è concentrata l'attenzione sulla gestione delle diverse rappresentazioni e procedure che gli studenti associano a una particolare situazione.

2. Metodo quantitativo: uso della *Latent Class Analysis* per individuare i quesiti

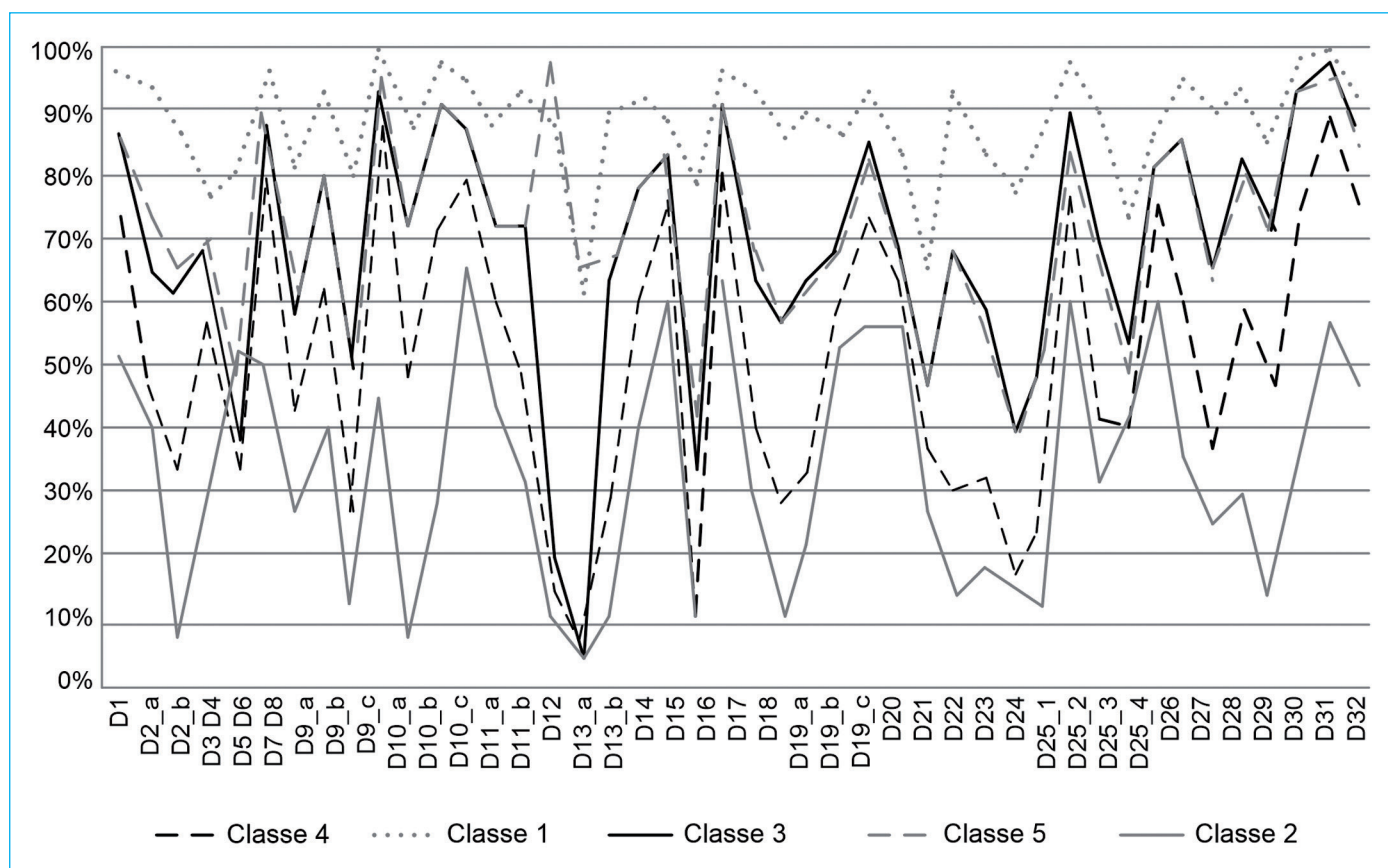
Per ogni rilevazione nazionale, l'INVALSI verifica la consistenza e l'affidabilità dei test somministrati a livello campionario, utilizzando strumenti della *Classical Test Theory* (Alagumalai e Curtis, 2005) quali l'*alpha* di Cronbach e il *coefficiente di correlazione punto biseriale*¹. A partire da queste prime elaborazioni, vengono successivamente stimati i parametri che descrivono le caratteristiche dei singoli item attraverso i modelli di *Item Response Theory* (van der Linden e Hambleton, 1997). Queste procedure standard permettono di caratterizzare non solo la prova nel suo complesso ma anche ogni singolo quesito in essa presentato.

¹ Per ogni rilevazione, questi valori sono presentati all'interno del Rapporto tecnico delle rilevazioni nazionali degli apprendimenti consultabili all'indirizzo www.invalsi.it.

Partendo dai dati elaborati attraverso tali procedure, sono stati classificati gli studenti coinvolti nell'indagine attraverso la tecnica della *Latent Class Analysis* (Lazarsfeld e Henry, 1968). Tale tecnica è stata applicata dopo aver dicotomizzato gli item, ossia considerando unicamente il caso di risposta corretta e risposta errata; questa procedura è stata necessaria in quanto la maggior parte degli item inclusi nel test sono a scelta multipla quindi di tipo categoriale non ordinato (nominale).

La *Latent Class Analysis* permette di suddividere gli studenti in un numero fissato di gruppi (nel nostro caso 5) caratterizzati da diversi livelli di prestazione relativi alla prova. In particolare, ogni livello di prestazione è stato definito sulla base delle probabilità stimate di risposta corretta per ogni item. In base a queste scelte, ogni gruppo può essere interpretato come l'insieme degli studenti con prestazioni "simili" di risposta sui singoli item all'interno dell'intera prova. La fig. 1 riporta i risultati di questa analisi per la prova di quinta primaria del 2010. Partendo da questa suddivisione, è possibile identificare gli item in cui si registrano dei comportamenti di risposta particolari relativi ai singoli gruppi. In altre parole, è possibile individuare i quesiti in cui i gruppi hanno prestazioni diverse.

Fig. 1 – Probabilità di risposta corretta per le cinque classi identificate (campione nazionale del 2010 di circa 40.000 studenti del livello 5)



L'analisi dei dati sul campione nazionale ha mostrato la presenza di gruppi/classi di studenti con probabilità di risposta corretta su tutti gli item molto inferiore rispetto ai risultati complessivi (Branchetti *et al.*, 2015). Analizzando la fig. 1, si nota che la classe 4 e la classe 5 identificano gli studenti con le più basse probabilità di risposta corretta. Di conseguenza, si possono considerare questi gruppi come composti dagli alunni che hanno mostrato più difficoltà nell'affrontare la prova. Attraverso il confronto delle performance delle varie classi sulle singole domande, si può osservare l'esistenza di un insieme di domande per le quali solo gli studenti delle classi 4 e 5 hanno probabilità di risposta basse (fig. 1). In particolare, si possono evidenziare gli item la cui probabilità di successo per gli studenti di queste due classi è molto inferiore alla stessa probabilità per gli studenti delle altre classi. Per esempio, nella domanda D27 (fig. 2) la probabilità di successo nelle classi 4 e 5 è inferiore al 40% mentre nelle altre classi è sempre superiore al 60%. Di conseguenza, questo item risulta interessante per studiare i possibili ostacoli incontrati dagli studenti in difficoltà nella prova

INVALSI. In aggiunta, si tratta di un quesito interessante anche per quanto riguarda l'ambito di contenuto e i processi coinvolti (relativi al confronto fra numeri razionali e la gestione delle loro rappresentazioni) ed è per questo che è stato selezionato per l'analisi qualitativa presentata nel paragrafo successivo.

3. Metodo qualitativo: analisi del campo concettuale

La ricerca in Didattica della matematica fa uso di una molteplicità di quadri teorici che fungono da strumenti per l'analisi qualitativa. Il nostro interesse principale è quello di indagare le modalità in cui concetti appresi dagli studenti evolvono nel tempo; a questo scopo si rende necessaria una definizione di "concetto" e un'individuazione degli osservabili che la determinano. Secondo Vergnaud:

Lo studio dello sviluppo di un concetto richiede ai ricercatori di vedere il concetto come una tripletta di insiemi: $C = (S, I, S)$, in cui S è un insieme di situazioni che rende il concetto significativo, I è un insieme di invarianti (oggetti, proprietà e relazioni) che possono essere riconosciuti e usati dai soggetti per analizzare e padroneggiare queste situazioni, e $d S$ è un insieme di rappresentazioni simboliche che possono essere usate per indicare e rappresentare questi invarianti e quindi per rappresentare le situazioni e le procedure per affrontarle (Vergnaud, 1988, p. 85).

Nel caso particolare dell'analisi dei quesiti delle prove standardizzate, la situazione è fornita dall'item oggetto di analisi, e rimangono quindi da indagare gli invarianti associati a una particolare situazione e le rappresentazioni simboliche. Laddove la domanda analizzata sia a risposta chiusa, a risposta univoca o di tipo *cloze*, il numero di invarianti e di rappresentazioni che lo studente può mettere in gioco è limitato. Nel caso di domande aperte, invece, lo studente può generalmente associare alla situazione proposta qualsiasi procedura o rappresentazione. Pertanto, onde evitare che sia la struttura di presentazione del quesito a indurre la scelta sugli elementi di I e S , si ritiene opportuno svincolare il quesito dalla modalità in cui è presentato. In particolare, si sceglie di somministrare i quesiti a un nuovo campione di studenti della stessa età, mantenendo invariato lo stimolo e proponendo la domanda come quesito aperto. In fig. 2 è mostrata una domanda nella versione originale (chiusa), mentre la fig. 3 presenta la versione che è stata utilizzata nella nuova somministrazione. Nell'esempio riportato nel paragrafo successivo, il numero di studenti coinvolti è pari a 231 e comprende classi di città del Nord e del Sud Italia (Bologna e Barletta). Nella scelta del campione si è cercato inoltre di selezionare, a parità di città, scuole situate in quartieri con background socio-economico differente in base all'indice ESCS rilevato dall'INVALSI.

Fig. 2 – Domanda D27 – livello 5, a.s. 2009-10

D27. $\frac{4}{8}$ e 0,5 indicano la stessa quantità?

- A. No, perché $\frac{4}{8}$ indica una quantità minore di 0,5
- B. No, perché 0,5 indica una quantità minore di $\frac{4}{8}$
- C. No, perché la prima è una frazione, il secondo è un numero decimale
- D. Sì, perché valgono entrambi la metà di un intero

Le risposte ottenute attraverso questa nuova somministrazione sono confrontate, in termini di numero di risposte corrette, con quelle ottenute dall'INVALSI nella rilevazione censuaria. Laddove vi sia corrispondenza fra le percentuali, si

procede con l'analisi più dettagliata delle risposte aperte finalizzata a far emergere le procedure che ricorrono in diversi protocolli (invarianti) e le rappresentazioni simboliche utilizzate dai diversi studenti. Tale analisi viene effettuata separatamente da almeno tre ricercatori; i risultati ottenuti sono stati messi a confronto realizzando quindi una triangolazione. Nel caso di classificazioni diverse si procede a una discussione volta a raggiungere un accordo. Le rappresentazioni e gli invarianti osservabili possono poi essere messi a confronto col numero di risposte corrette ottenute utilizzandoli, individuando così quelli che generalmente portano al successo e quelli che gli studenti manipolano con maggiore difficoltà. Un esempio di analisi di questo tipo, riguardante la domanda mostrata in fig. 3 e tratta da Ferretti, Lemmo e Maffia (2016), viene presentata nel paragrafo successivo.

Nei paragrafi successivi si presentano i risultati della versione della domanda D27 (fig. 3) somministrata nella nostra sperimentazione (Ferretti, Lemmo e Maffia, 2016).

Fig. 3 – Domanda somministrata nella sperimentazione

<p>D27. $\frac{4}{8}$ e 0,5 indicano la stessa quantità?</p> <p>Sì, perché _____</p> <p>_____</p> <p>_____</p> <p>_____</p> <p>No, perché _____</p> <p>_____</p> <p>_____</p> <p>_____</p>
--

In termini generali, la percentuale di risposte corrette ottenute è in linea con quelle della Rilevazione nazionale del 2010. Circa il 50% degli studenti risponde correttamente; si rileva un incremento nella percentuale di risposte mancanti probabilmente dovuto alla trasformazione del quesito in domanda a risposta aperta.

4. Analisi delle rappresentazioni

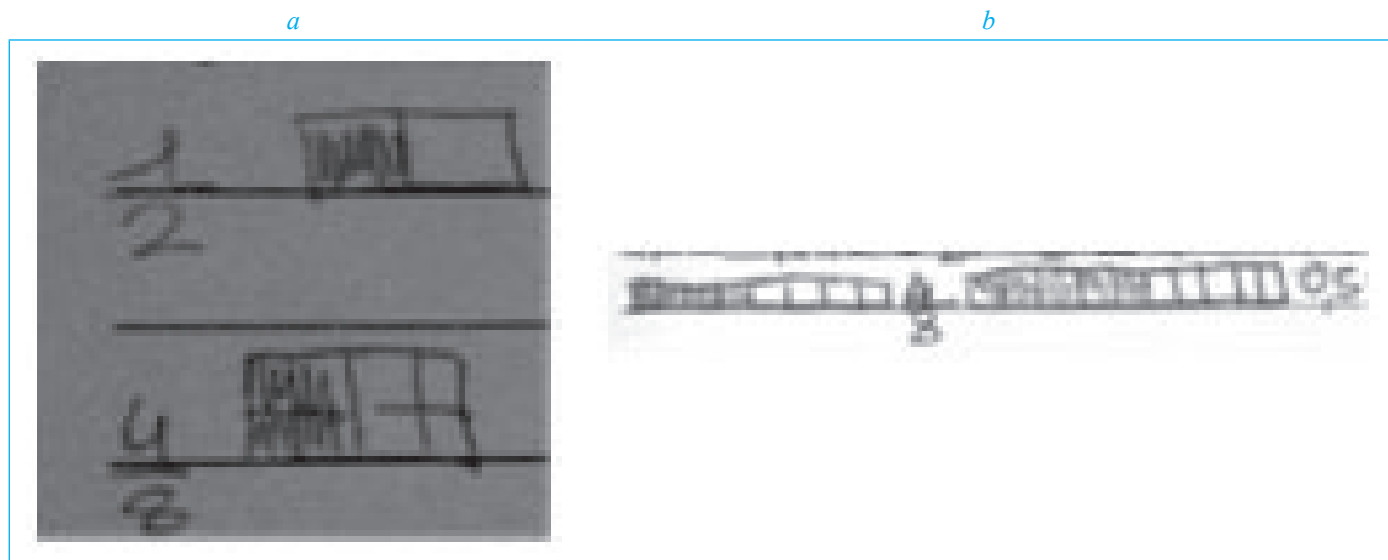
Secondo Duval (2006) è possibile classificare le diverse rappresentazioni di un oggetto matematico in registri, ovvero insiemi di segni e regole per manipolarli. Tali registri possono essere a loro volta classificati come discorsivi (il linguaggio naturale scritto o parlato, i simboli matematici) o non discorsivi (diagrammi e figure). Ancora, è possibile distinguere all'interno di ciascuna categoria quelli che sono registri multifunzionali, ovvero adatti a spiegare processi che non possono essere messi in forma algoritmica, da quelli mono-funzionali, ovvero dedicati soprattutto ai processi algoritmici. Nella prima categoria ricade sicuramente il linguaggio naturale, nella seconda i simboli aritmetico-algebrici.

Analizzando i protocolli relativi alla domanda D27 nella versione mostrata in fig. 3, è possibile osservare esempi di tutte queste tipologie di registri. Alcuni studenti fanno uso di *rappresentazioni di tipo iconografico* (fig. 4). Queste rappresentazioni sono molto comuni nelle prassi di scuola primaria specialmente quando si parla di frazioni. Tuttavia, nel caso di questa domanda, risultano poco efficaci: meno della metà degli studenti che le adottano risponde correttamente. In particolare, la difficoltà più comune incontrata dagli studenti che usano questa rappresentazione consiste nella con-

versione del decimale 0,5 in una rappresentazione iconica. Altri hanno difficoltà a confrontare le due rappresentazioni ottenute per i due numeri perché partono da “interi” differenti (fig. 4b).

Per quanto riguarda i registri discorsivi, vengono impiegati largamente il linguaggio naturale e quello simbolico suggerito dalla rappresentazione dei numeri all’interno dello stimolo della domanda. I simboli $4/8$ e $0,5$ sono convertiti da molti studenti nell’espressione colloquiale “metà” inserita in frasi del tipo “entrambi sono la metà dell’intero” oppure “ $4/8$ è la metà di $8/8$ e $0,5$ è la metà di 1 ”. L’uso del registro colloquiale sembra essere più adatto di una qualunque manipolazione sintattica degli oggetti matematici dato che è la rappresentazione che più frequentemente porta a risposte corrette (Ferretti, Lemmo e Maffia, 2015).

Fig. 4 – Rappresentazioni iconografiche nelle risposte alla domanda D27



Infine, andando ad analizzare i registri di tipo simbolico, si nota che alcuni studenti preferiscono utilizzare la *notazione decimale* dei numeri, altri ricorrono alla rappresentazione tramite *frazioni*. Chiaramente, essendo i numeri nello stimolo rappresentati in due notazioni diverse, è necessaria la conversione di almeno uno dei due. Le frequenze di errori nell’uso dell’uno o dell’altro registro simbolico dipendono dalla procedura utilizzata per la conversione. Tali procedure sono analizzate nel paragrafo successivo.

5. Analisi delle invarianti

Come già evidenziato nel paragrafo precedente, fra i dati raccolti è possibile individuare diverse procedure di conversione sia nella direzione da decimale a frazione sia nella direzione da frazione a decimale. La conversione da frazioni a decimali viene effettuata generalmente calcolando la divisione $4:8$. La maggior parte degli studenti riporta semplicemente il risultato del calcolo senza specificare il modo in cui è stato svolto; tuttavia in alcuni protocolli è possibile osservare il calcolo algoritmico “in colonna”. In questi casi si nota come alcuni errori di conversione da frazione a decimale dipendano proprio da errori nello svolgimento del calcolo.

Fra le risposte non corrette risultano particolarmente frequenti conversioni del tipo $4/8 \rightarrow 4,8$ oppure $4/8 \rightarrow 0,4$. Questo tipo di conversione sembra basarsi soltanto sulla sintassi del numero, così come è già stato osservato in passato da Markovits e Sowder (1991).

Anche la conversione da decimali a frazioni viene effettuata in modi diversi. Alcuni studenti ricorrono a una doppia conversione da decimale a linguaggio naturale e poi dal linguaggio naturale alla frazione: identificando che $0,5$ rappresenta la “metà” di 1 , lo ritengono equivalente a $1/2$. Semplificando $4/8$ verificano che anche questo è equivalente a $1/2$ e quindi deducono l’uguaglianza fra i due numeri rappresentati nello stimolo. Altri studenti invece ricorrono a una lettura del simbolo $0,5$ come “cinque decimi” il che implica la possibilità di scriverlo nella forma $5/10$. Ancora una

volta, si verifica che sia $\frac{4}{8}$ sia $\frac{5}{10}$ sono equivalenti a $\frac{1}{2}$ e quindi, per transitività, sono equivalenti fra loro. Queste strategie di conversione portano spesso a risposte corrette. Gli errori più frequenti sono, ancora una volta, quelli che si basano soprattutto su manipolazioni di tipo esclusivamente sintattico. Non sono rari i casi in cui 0,5 è convertito nella frazione $\frac{0}{5}$.

Gli studenti scelgono più frequentemente di ricorrere alla conversione da frazione a decimale nonostante questa strategia porti a un numero maggiore di errori. Di fatto solo il 48% di coloro che operano una conversione da frazione a decimale fornisce la risposta corretta; tale percentuale sale al 72% nel caso della conversione da decimale a frazione. Una possibile interpretazione di questo comportamento risiede nel fatto che, nella scuola primaria, la notazione decimale per i numeri razionali non interi è generalmente introdotta prima delle frazioni, diversamente da quello che è stato il processo storico che ha visto emergere l'uso delle frazioni in civiltà molto lontane da noi nel tempo mentre i decimali sembrano essere un'invenzione più recente (Ferretti, Lemmo e Maffia, 2016). Pertanto gli studenti preferiscono ricorrere alla rappresentazione che conoscono da più tempo e che gli è, quindi, più familiare.

6. Conclusioni

Uno degli obiettivi di questo lavoro è stato quello di fornire un modello di indagine e degli strumenti significativi di analisi utilizzabili anche nelle pratiche di insegnamento. Si sono analizzati fenomeni critici, evidenziati da analisi statistiche, con quadri di riferimenti consolidati in ricerche di Didattica della matematica, unendo diversi approcci e indagando le situazioni con lenti teoriche differenti. L'analisi qualitativa conferma i risultati mostrati dalla letteratura internazionale, evidenziando le difficoltà degli studenti nell'attribuire un significato condiviso ai concetti matematici; si è infatti evidenziato come essi vedano le frazioni principalmente come oggetto sintattico senza attribuirgli il significato di numero razionale che ci si attenderebbe. Indubbiamente, l'assenza dell'attribuzione del significato atteso incide in modo negativo su tutte le manipolazioni inerenti al concetto e quindi anche sulla gestione delle sue diverse rappresentazioni. Indagare il significato che gli studenti attribuiscono ai concetti può certamente essere utile ai fini di una maggior comprensione delle loro difficoltà da parte degli insegnanti. Questo tipo di analisi può permettere al docente di dare un significato ai risultati delle prove INVALSI della propria classe: invece di fermarsi al “chi?” e “quanto?” dell'errore, può passare al “come?” e “perché?” interrogandosi anche su come, eventualmente, modificare le proprie pratiche didattiche.

Utilizzando il nostro esempio, per quanto concerne l'insieme di rappresentazioni simboliche, lo studio dei protocolli mostra che le rappresentazioni generalmente utilizzate nelle prassi didattiche italiane non permettono sempre agli studenti di fornire una risposta corretta al quesito proposto. Per esempio, gli studenti che hanno scelto la rappresentazione iconografica, tradizionalmente utilizzata per introdurre le frazioni, nella maggior parte dei casi non forniscono una risposta corretta.

Lo stesso si può dire a partire dall'analisi degli invarianti; si vede infatti che nonostante la conversione da decimale a frazione porti maggiormente alla soluzione corretta, è scelta da pochissimi studenti. La maggior parte degli studenti che sceglie come strategia risolutiva la trasformazione tra decimale e frazione predilige la conversione tra frazione e decimale; questo fenomeno è molto probabilmente riconducibile alle prassi didattiche e si può interpretare con alcune categorie di contratto didattico (D'Amore *et al.*, 2010).

In definitiva, la metodologia utilizzata permette di individuare alcuni quesiti significativi e di indagare rappresentazioni e strategie risolutive che più frequentemente conducono all'errore. Pertanto le analisi, effettuate a partire dai dati delle rilevazioni nazionali, possono suggerire agli insegnanti diverse piste di intervento per migliorare la propria pratica didattica e incidere (probabilmente in modo più significativo) sul processo di insegnamento/apprendimento.

Riferimenti bibliografici

- Alagumalai S., Curtis D.D. (2005), “Classical test theory”, in S. Alagumalai, D.D. Curtis, N. Hungi (eds.), *Applied Rasch Measurement: a Book of Exemplars*, Springer, Dordrecht (The Netherlands): 1-14.
- Bolondi G. (2010), “Come usare in classe le prove INVALSI”, *L'insegnamento della matematica e delle scienze integrate*, 33, 6A-B: 686-702.

-
- Bolondi G., Branchetti L., Ferretti F., Lemmo A., Maffia A., Martignone F., Matteucci M., Mignani S., Santi G. (2016), “Un approccio longitudinale per l’analisi delle prove INVALSI di matematica: cosa ci può dire sugli studenti in difficoltà”, in P. Falzetti (a cura di), *Concorso di idee per la ricerca*, Cleup, Padova: 81-102.
- Branchetti L., Ferretti F., Lemmo A., Maffia A., Martignone F., Matteucci M., Mignani S. (2015), “A Longitudinal analysis of the Italian national standardized mathematics tests”, in K. Krainer, N. Vondrová (eds.), *CERME 9: Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education*, Charles University in Prague, Faculty of Education, ERME, Prague, Czech Republic: 1695-1701.
- D’Amore B., Fandiño Pinilla M.I., Marazzani I., Sarrazy B. (2010), *Didattica della matematica. Alcuni effetti del “contratto”*, Archetipolibri, Bologna.
- De Virgilis R., Pesci A. (2014), “I quesiti di matematica INVALSI 2013 sulle percentuali: dall’analisi degli errori al ripensamento dell’azione didattica”, *L’insegnamento della matematica e delle scienze integrate*, 37, 2B: 139-158.
- Duval R. (2006), “A cognitive analysis of problems of comprehension in a learning of mathematics”, *Educational Studies in Mathematics*, 61: 103-131.
- Ferretti F., Lemmo A., Maffia A. (2015), “‘Half of something’: how students talk about rationals”, in K. Beswick, T. Muir, J. Wells (eds.), *Proceedings of the 39th Conference of the International Group for the Psychology of Mathematics Education*, vol. 1, PME, Hobart, Australia: 159.
- Ferretti F., Lemmo A., Maffia A. (2016), “Confrontare decimali e frazioni: analisi delle concezioni degli studenti a partire da una domanda INVALSI”, *L’insegnamento della matematica e delle scienze integrate*, 39, 5: 451-464.
- Giberti C., Zivelonghi A., Bolondi G. (2016), “Gender differences and didactic contract: analysis of two INVALSI tasks on powers properties”, in C. Csikos, A. Rausch, J. Sztányi (eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education*, vol. 2, PME, Szeged, Hungary: 275-282.
- Jakobsen A., Mellone M., Ribeiro M., Tortora R. (2016), “Discussing secondary prospective teachers’ interpretative knowledge: a case study”, in C. Csikos, A. Rausch, J. Sztányi (eds.), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education*, vol. 3, PME, Szeged, Hungary: 35-42.
- Lazarsfeld P.F., Henry N.W. (1968), *Latent Structure Analysis*, Houghton Mifflin, Boston (MA).
- Lemmo A., Branchetti L., Ferretti F., Maffia A., Martignone, F. (2015), “Students’ difficulties dealing with number line: a qualitative analysis of a question from national standardized assessment”, *Quaderni di Ricerca in Didattica (Mathematics)*, 25, 2: 143-150.
- Markovits Z., Sowder J.T. (1991), “Students’ understanding of the relationship between fractions and decimals”, *Focus on Learning Problems in Mathematics*, 13: 3-11.
- Martignone F. (2016), “Un’attività di formazione per insegnanti di scuola secondaria di primo grado: analisi di prove INVALSI di matematica”, *Form@re-Open Journal per la Formazione in Rete*, 16, 1: 70-86.
- Perelli D’Argenzio M.P. (2006), “La valutazione esterna degli apprendimenti: le prove di valutazione INVALSI. Le prove di matematica”, *L’insegnamento della matematica e delle scienze integrate*, 29, 1: 31-46.
- Van der Linden W.J., Hambleton R.K. (eds.) (1997), *Handbook of Modern Item Response Theory*, Springer, New York (NY).
- Vergnaud G. (1988), “Multiplicative Structures”, in J. Hiebert, M. Behr (eds.), *Number Concepts and Operations in the Middle Grades*, National Council of Teachers of Mathematics, Reston (VA); Lawrence Erlbaum Associates, Hillsdale (NJ): 141-161.

6. Le prove INVALSI quale strumento di miglioramento

INVALSI tests as a tool for improvement

di Maria Brutto

L'obiettivo primario della ricerca-azione è stato accrescere la consapevolezza che le prove INVALSI siano uno strumento efficace per avviare azioni di miglioramento sugli esiti e sui processi didattici in italiano e matematica. I destinatari sono stati i docenti di italiano e matematica di tutte le scuole della regione Calabria in due seminari per provincia, per complessive sei ore di formazione¹.

L'azione operativa è stata organizzata in sei format di lavoro, nella fattispecie: report di scuola su autovalutazione dei risultati; analisi di un quesito critico; costruzione di un quesito ispirandosi a prove già somministrate che facciano da modello; analisi delle risposte aperte sul fascicolo di classe; attività didattica di rinforzo; somministrazione di una prova.

Ciascuno di essi è stato inquadrato in un percorso metodologicamente supportato da ulteriori strumenti di lavoro fra cui www.gestinv.it, il database delle prove INVALSI, già sperimentati nei laboratori di formazione paralleli in scuole della Calabria per avviare un dibattito sull'analisi delle prove, nonché azioni mirate di didattica metacognitiva, compensativa e di mantenimento. Per avviare la riflessione si è ragionato su ciò: perché analizzare gli errori anche ricorsivi dei propri allievi sui fascicoli; come dare importanza ai processi e alle competenze o come circoscrivere l'ambito di apprendimento a singoli obiettivi; come potenziare la didattica su aree di criticità (*exemplum*: come far superare le difficoltà palesate sulla lemmatizzazione?); come considerare l'errore (ove non si tratti di misconcezione del docente) quale ipotesi di soluzione.

Si è ragionato su alcune parole chiave quali significatività e acquisizione di un metodo di lavoro basato su *problem posing/solving*, *task analysis*, ricerca-azione, studi di caso, *cooperative learning* e *cooperative teaching*, *prompting*, *orienting*, così da innescare nei docenti interessati processi di didattica orientata al raggiungimento di micro-obiettivi specifici.

I risultati raggiunti sono stati i seguenti: favorire processi di innovazione degli apprendimenti di base nell'ambito dell'educazione linguistica e matematica, in sintonia con le finalità esplicitate dalle Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione; favorire la costituzione di una comunità di pratica di docenti che, attraverso la condivisione di conoscenze ed esperienze, giungano a sperimentare piste operative di insegnamento intenzionale e ad applicare buone pratiche; sviluppare forme di didattica riflessiva nel montaggio e smontaggio delle prove INVALSI, nella costruzione di prove per il *modeling* (apprendimento imitativo), nell'analisi dell'errore; acquisire un metodo di lavoro che permetta di procedere dalla lettura dei risultati alla progettazione di una didattica orientata e allineata ai saperi imprescindibili ben delineati dalle Indicazioni nazionali del 2012; condurre *task analysis* (analisi del compito) su quesiti e progettare azioni didattiche compensative rispetto alle criticità e di mantenimento/valorizzazione delle eccellenze.

Un *cloud* ha infine raccolto i materiali prodotti da esperti e corsisti.

The primary objective of this action research is to heighten the awareness that the INVALSI tests are an effective tool to set up actions aimed to improve students' outcomes and the processes of teaching and learning

¹ *Prove INVALSI quale strumento di miglioramento*, Progetto di formazione finanziato dall'USR Calabria ai sensi della Nota MIUR prot. 11171 del 9/11/2015 all. 2 (30 ore in 10 seminari) realizzato da Maria Brutto, in qualità di esperta di Italiano e dal collega prof. Nicola Chiriano, esperto di Matematica.

both in Italian and in Mathematics. The addressees are Italian and Mathematics teachers coming from all the Calabrese schools divided into two seminars for each province (county), for a total of six hours of training². The training workshops consisted in the following six formats: a school's report on its self-evaluation results; analysis of critical questions; construction of a model question from the class booklet; analysis of the individual test about open answers; educational reinforcement activities; administration of a test.

Each of these sections has been framed in a methodological path supported by some tools such as *www.gestinv.it*, already tested in the parallel training laboratories in some Calabrese schools whose aim was, on the one hand, to start a debate on the analysis of the INVALSI test and, on the other, to perform targeted actions of compensatory, metacognitive and maintenance teaching.

Some useful questions have been raised about: the relevance of analyzing recurring errors in our students' work on individual tests; how to give importance to processes and skills or how to circumscribe the area of learning to individual objectives; how to reinforce teaching practices in critical areas (e.g. overcoming the difficulties highlighted on lemmatization); considering the error (if it is not a teacher's misconception) as the starting point towards a possible solution.

Some keywords like significance and acquisition of a working method based on problem posing/solving, task analysis, action research, case studies, cooperative learning and teaching, prompting, orienteering, have been discussed in order to let teachers activate learning processes oriented towards definite micro-targets.

The results achieved are the following: to encourage innovation of basic learning processes in language and mathematics education, in line with the aims indicated by the 2012 National Guidelines for primary and secondary school curriculum; to promote the establishment of a community of practice of teachers who, thanks to the sharing of knowledge and experiences, aim to try operational paths of intentional teaching and to apply good practices; to develop forms of reflective teaching in assembling and disassembling the INVALSI tests, in designing tests for modeling, and in analyzing the error; to acquire a work method that allows to move from the mere reading of the results to the project of a teaching intended to fit the essential knowledge well defined by the 2012 National Guidelines; to conduct task analysis of the questions and to plan educational actions apt to compensate for students' weaknesses and to promote educational excellence.

The materials produced by experts and teachers during the training activities have finally been gathered in a cloud.

Le prove INVALSI quale strumento di miglioramento è il titolo di un progetto di formazione, finanziato dall'USR Calabria³, articolato in dieci seminari abbinati tra loro nella forma di un seminario "di andata" e uno "di ritorno", per ciascuno dei cinque diversi gruppi interprovinciali.

Interessante è stato il modello di formazione e ricerca che ha contemplato una fase teorica, una fase di ricerca *stricto sensu*, una fase di produzione di materiali e un'ultima fase laboratoriale *on-site* e a distanza.

L'obiettivo principe del progetto è stato far accrescere la consapevolezza che le prove INVALSI siano uno strumento efficace per avviare azioni di miglioramento sugli esiti e sui processi didattici in italiano e matematica. I destinatari sono stati i docenti di italiano e matematica di tutte le scuole della regione Calabria in due seminari per provincia, per complessive sei ore di formazione, selezionati fra coloro che avessero dato disponibilità a farsi portavoce nell'ambito dei dipartimenti per disseminare i contenuti della formazione in modalità di *peer tutoring*.

L'azione teorica ha necessariamente fornito uno *scaffolding* concettuale ad ampio spettro: literacy e Quadri di riferimento; tipo di competenza misurabile; le misurazioni internazionali e quelle nazionali previste dal Sistema nazionale di valutazione (in seguito chiamato SNV); obiettivi educativi 2020; analisi di una prova; SNV come strumento di misurazione; la storia di una prova e i riferimenti alle Indicazioni nazionali; suggerimenti e proposte operative⁴. Entrando

² The *INVALSI tests as a tool for improvement* is a training project funded by the USR Calabria under the Note of the Ministry of Education prot. 11171 11/09/2015 at. 2 (30 hours in ten seminars) carried out by Maria Brutto, as an expert in the Italian language and accompanied by the colleague prof. Nicola Chiriano, an expert in Mathematics.

³ Come da Circolare MIUR. AOODRCAL. Registro ufficiale (U). 0003738.22-03-2016, ai sensi della Nota MIUR prot. 11171 del 9/11/2015 all. 2, il modello di formazione è stato ideato e realizzato nelle diverse fasi dai professori Maria Brutto, docente di materie letterarie, e Nicola Chiriano, docente di Matematica sulle seguenti aree tematiche: Quadri di riferimento INVALSI di italiano e matematica; lettura della restituzione dei dati sulle prove; analisi delle prove di italiano e matematica; Servizio nazionale di valutazione (SNV) quale strumento di miglioramento.

⁴ I docenti corsisti hanno avuto la possibilità di fruire di un *cloud* per l'invio di riflessioni operative scaturite dagli input dei format.

maggiormente nel dettaglio, l'azione di formazione si è avviata incardinandola nel quadro più ampio di una didattica per competenze dell'italiano (reading literacy) e della matematica (problem solving), il cui scopo precipuo è aprire la mente (Gardner, 1991), nello spirito delle indagini OCSE-PISA, IEA-PIRLS/TIMSS e SNV, delle cui prove è stato palesato il disegno e le finalità.

Il tutto è stato inquadrato all'interno del concetto di *accountability*, quale modalità di rendicontazione trasparente del sistema scuola. A riguardo si è voluto focalizzare l'attenzione sulle prove standardizzate a livello internazionale, quale chiaro presupposto teorico e metodologico delle Rilevazioni nazionali.

Si è ripercorso concettualmente l'iter che ha condotto al *framework* SNV, il Quadro di riferimento che presenta le idee chiave sottese alla progettazione delle prove, dagli ambiti alle modalità di valutazione, cioè le caratteristiche degli strumenti e i criteri seguiti nella costruzione delle prove stesse. Ci si è soffermati sulla definizione di competenza in senso lato, il *cum petere* latino che ha in sé il dinamismo del *rivolgerci a* e che rintraccia la componente motivazionale ed emotiva *dell'impegnarsi a* (*l'engaging with*, così come nel *framework* del 2009) mobilitare le proprie conoscenze in contesti e situazioni nuove, come si desume dagli studi internazionali, riflettendo altresì su elementi quali quesiti, ambiti, livelli di difficoltà, tipologie di testo. A scopo esemplificativo, sono state presentate alcune prove internazionali tra quelle rilasciate, così da entrare nel merito degli aspetti della competenza in lettura e matematica.

La ricerca educativa si offre, così, a supporto di una didattica riflessiva volta al miglioramento dei processi e dei risultati. Sono stati esaminati, quindi, i sottili legami che intercorrono tra autovalutazione, valutazione esterna e miglioramento, evidenziando gli aspetti importanti da osservare per una lettura critica dei risultati in prospettiva di un innalzamento del livello di qualità dell'offerta formativa.

Si è evidenziato come sia necessario che l'autovalutazione rappresenti un processo interno delle scuole più che la produzione di un documento come mero adempimento, spiegando che la misurazione INVALSI, a partire dal 2008, si è incentrata sulla restituzione dei dati alle scuole per l'innescio di processi di analisi e miglioramento al loro interno.

Sono state fornite le coordinate per una corretta lettura dei dati tabellari e dei grafici, sia nello schema generale dei punteggi sia in quello dettagliato per quanto concerne il *cheating*⁵, l'allineamento tra valutazione interna ed esterna, la variabilità tra le classi, i punti di forza e punti di debolezza sulle dimensioni oggetto di misurazione.

Successivamente, si è posto l'accento sulla necessità di una corretta somministrazione e correzione delle prove, nonché di una lettura attrezzata, perché consapevole e approfondita, dei risultati, che ora vengono offerti per le quinte classi della scuola primaria in un'ottica di continuità come dati in ingresso per la classe iniziale dell'ordine successivo.

La seconda parte del progetto ha avuto un taglio laboratoriale ed è stata strutturata nell'eventuale adozione di uno o più format di lavoro, nella fattispecie:

- report di scuola su autovalutazione e risultati;
- analisi di un quesito critico;
- costruzione di un quesito da modello;
- analisi su fascicolo di classe delle risposte aperte autografe;
- attività didattica di rinforzo;
- somministrazione di una prova.

Ogni format ha focalizzato l'attenzione sui processi prima che sui risultati e ha fornito ai docenti metodologie e strumenti di lavoro atti ad avviare itinerari di ricerca-azione da esperire individualmente e/o in gruppo.

È importante rendersi conto che quanto proposto non è un "addestramento" sulla prova volto a potenziare lo sviluppo della competenza, ma la comprensione dei modelli di riferimento, sottesi alla stessa, di tipo incrementale e dinamico. Sarebbe riduttivo, per esempio, imparare la sillabazione di dieci parole rispetto alla competenza della consultazione dello strumento, il vocabolario, che la indica per migliaia di lemmi.

Al fine di avviare forme di didattica metacognitiva che, fotografando la situazione reale, inneschino processi di ri-orientamento, i format sono stati offerti quale guida, in ordine diversificato di difficoltà, dal più complesso *Report di istituto* sugli esiti, alla *Somministrazione di una prova*, meglio se già strutturata, per valutarne le dinamiche nelle diverse fasi (selezione, somministrazione, svolgimento, analisi dell'errore, comparazione con i dati, criticità osservate, ipotesi di miglioramento/compensazione o mantenimento).

⁵ Con il termine *cheating* si intende l'azione del barare, che nel contesto etico-pedagogico consiste nell'imbroglio scolastico, posto in essere dagli studenti che copiano o dai docenti che in fase di correzione o trasmissione dati operano delle manipolazioni.

Fig. 1 – Format 1a



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016

Report di Scuola Autovalutazione Prove INVALSI

Nome Istituzione scolastica: _____

Codice meccanografico Istituzione scolastica: | | | | | | | | | | | | |

Indirizzo: _____ Provincia: _____

1. Punteggio generale

- Quali risultati raggiunge la scuola nelle prove standardizzate nazionali di Italiano e Matematica?
- Quale punteggio generale la scuola raggiunge rispetto al dato nazionale, regionale, d'area?
- Quale valore si riporta in termini di *cheating*? (trascurabile se <10%)
- Quale trend si rileva rispetto all'anno precedente? (scarsamente significativo se ≤5%)

Livello	Area	Italiano			Matematica		
		Limite inf	Media	Limite sup	Limite inf	Media	Limite sup
Primaria L02 L05	Scuola						
	Calabria						
	Sud e isole						
	Italia						
Secondaria L08	Scuola						
	Calabria						
	Sud e isole						
	Italia						

- Dalla lettura dei dati riportati in tabella emerge quanto segue:

.....
.....

Fig. 2 – Format 1b

-
- Il livello raggiunto dagli studenti nelle prove INVALSI è ritenuto affidabile - conoscendo l'andamento abituale delle classi - oppure c'è il sospetto di comportamenti opportunistici (*cheating*)?
-
-

2. Varianza classi

- Quali esigenze ambientali incidono sulla formazione delle classi?
- La scuola riesce ad assicurare esiti uniformi tra le varie classi?

Varianza	
Punti di forza	Punti di debolezza

3. Allineamento voti

- Cosa si rileva circa l'allineamento delle valutazioni interne ed esterne?

Allineamento voti	
Punti di forza	Punti di debolezza

4. Risultati scolastici e prove SNV

- Quanti studenti non sono ammessi alla classe successiva e perché? Ci sono concentrazioni di non ammessi in alcuni anni di corso?
- I criteri di valutazione adottati dalla scuola sono adeguati a garantire il successo formativo degli studenti?
- Quali considerazioni si possono fare analizzando la distribuzione degli studenti per fascia di voto e fascia di livello? Cosa emerge dal confronto con il dato del SNV?

Risultati scolastici e prove SNV	
Punti di forza	Punti di debolezza

Fig. 3 – Format 1c

5. Fasce di livello

- Le disparità a livello di risultati tra gli alunni meno dotati e quelli più dotati sono in aumento o in regressione nel corso della loro permanenza a scuola?
- Queste disparità sono concentrate in alcune sedi, indirizzi o sezioni?

Fasce di livello	
Punti di forza	Punti di debolezza

6. Risultati a distanza

- Per le classi che hanno effettuato le prove cosa si rileva?
- Quali misure si adottano per promuovere una didattica orientata al miglioramento?

Risultati a distanza	
Punti di forza	Punti di debolezza

7. Risultati in prospettiva

- In relazione ai punti di criticità rilevati quali azioni mirate si intendono perseguire nella didattica?
- Vengono individuati particolari obiettivi, abilità/competenze da raggiungere attraverso i progetti di ampliamento dell'offerta formativa?

Curricolo e offerta formativa	
Punti di forza	Punti di debolezza

Proposte

8. Autovalutazione di Istituto

- Nella scuola vi sono strutture di riferimento (es. dipartimenti) per la autovalutazione di Istituto?
- I docenti effettuano una riflessione iniziale e/o periodica legata ai risultati delle Prove?

Fig. 4 – Format 1d

- In che modo avviene l'analisi delle scelte adottate e la revisione della programmazione a breve termine e annuale?

Progettazione didattica	
Punti di forza	Punti di debolezza

9. Collaborazione tra insegnanti

- La condivisione di strumenti e materiali tra i docenti è ritenuta adeguata?

Collaborazione tra insegnanti	
Punti di forza	Punti di debolezza

10. Riflessione finale sul processo di autovalutazione

- Nella fase di lettura degli indicatori e di raccolta e analisi dei dati della scuola quali problemi o difficoltà sono emersi?

.....
.....
.....
.....

- Quali indicazioni pratiche su come orientare la didattica a fini correttivi, migliorativi o di mantenimento si propongono?

.....
.....
.....
.....

I docenti compilatori:
(nominativi)

La formazione si è ispirata al principio della significatività, non essendo pensabile l'esaustività, per offrire un metodo di lavoro basato sulla ricerca-azione utile a innescare nei docenti interessati processi di didattica orientata nonché maggiore familiarità con le azioni di autovalutazione.

Il Format 1 esamina, a questo riguardo, i sottili legami che intercorrono tra autovalutazione, valutazione esterna e miglioramento, evidenziando gli aspetti importanti da osservare per una lettura critica dei risultati in prospettiva di un innalzamento del livello di qualità dell'offerta formativa.

I dieci descrittori di cui il format consta, da "punteggio generale", primo descrittore, all'ultimo, "riflessione finale sul processo di autovalutazione", forniscono delle coordinate per una corretta lettura dei dati tabellari e dei grafici.

Si è posto l'accento sulla necessità di una corretta somministrazione e correzione delle prove, nonché di una lettura attrezzata dei risultati per fare tesoro della restituzione di dati che, grazie al confronto regionale e nazionale, danno la misura di come ci si situi rispetto all'oggetto della misurazione.

Nel caso della somministrazione di una prova, per esempio, l'analisi dell'errore è un momento particolarmente delicato. Esso permette di puntare l'accento sui processi di *transfer*, ovvero il trasferimento di competenze pregresse già acquisite trasferite nei processi di apprendimento di una lingua, partendo da un sostrato linguistico di una lingua madre rispetto alla quale la lingua standard a volte risulta non familiare alla stregua di una seconda lingua. I modelli linguistici, creatisi con l'acquisizione della lingua materna nell'area neuronale preposta, condizionano l'apprendimento di una lingua standard, non del tutto o magari per nulla coincidente con la prima. È quello che viene definito interferenza, anche se questa, oggi, non è più considerata come un momento ostativo all'apprendimento, ma piuttosto come una fase fondamentale nel processo di apprendimento che tende a una maggiore consapevolezza dell'apprendere, un elemento aggiuntivo nell'apprendimento linguistico. Le strutture pregresse costituiscono quello che viene definito sapere interlinguistico (Mariani, 2016). Tale sapere permette agli apprendenti di formulare e verificare ipotesi sulle regolarità di una (nuova?) lingua d'arrivo. Nel far ciò, ricorrono al *transfer* già disponibile laddove sono stati attivati precedentemente processi mentali, costruendo in tal modo una grammatica spontanea e ipotetica (Curci, 2004).

Alla luce di tutto questo si è valorizzata l'analisi dell'errore, quale ipotesi di soluzione, e la metacognizione, quale processo di consapevolezza dell'apprendente.

Forme di didattica compensativa, per il superamento di criticità, o di mantenimento, per rafforzare competenze già presenti, possono essere adottate dopo l'analisi di un quesito critico, come indicato dal Format 2 (fig. 5).

Il Format 3 (fig. 6) stimola i docenti a formulare un quesito che tenga conto di parametri quanto più simili a quelli cui si ispirano le prove INVALSI, archiviate con un prezioso e puntuale *labelling* in www.gestinv.it, la banca dati dei risultati delle prove INVALSI che permette una consultazione interattiva, con ricerca per filtro, e si presenta assai utile per il *download* dei formati testuali e fotografici dei singoli quesiti. Il *modeling*, non da intendersi come automatica *mimicry* (imitazione), ma trasposizione in compiti sempre nuovi per generazione da paradigma, in questo caso, aiuta a ideare buone batterie di domande.

A riguardo sono stati dati dei suggerimenti che aiutino a ben formulare un quesito, puntando l'attenzione sulla domanda a scelta multipla strutturata nei quattro item, dove l'equilibrio fra risposta esatta e distrattori va curato per ottimizzare la capacità misuratoria, vagliata appunto dai pre-test nel caso delle prove standardizzate. La storia di una prova, trattata propedeuticamente, dà una chiara idea di come nasca e trovi una sua forma definitiva e scientificamente accettabile.

Il Format 4 (fig. 7) stimola una riflessione sulla didattica compensativa, per il superamento di criticità, o di mantenimento, per rafforzare competenze già presenti. Il suggerimento in tal senso può essere colto dopo l'analisi di un quesito critico.

L'analisi delle risposte autografe da fascicolo suggerisce un'altra strategia didattica migliorativa che consiste in un'azione orientata e circoscritta a singoli obiettivi. Qualora non fosse chiara agli apprendenti la differenza tra modi finiti e indefiniti, volendo esemplificare un caso ricorrente, il docente avrà cura di focalizzare l'attenzione solo su questo oggetto di analisi spiegando agli studenti solo quanto strettamente attiene al focus in dettaglio, tralasciando, sul momento, eventuali imperfezioni a esso non afferenti. L'attenzione selettiva concentra, infatti, gli sforzi e incide efficacemente sulla creazione di nuovi costrutti di apprendimento, limitando il numero di input al fine di evitare un sovraccollamento di informazioni che non facilitano la permanenza a lungo termine.

Fig. 5 – Format 2



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016



Analisi di un quesito critico tratto da un fascicolo SNV



FOGLIO DI LAVORO

SITUAZIONE: _____

FORMATO DEL TESTO: _____

TIPO DI TESTO: _____

ASPETTO: _____

FORMATO DELLA DOMANDA: _____

SCALA DIFFICOLTÀ: _____

Quali difficoltà gli studenti potrebbero incontrare nel rispondere

Quali attività d'aula possono contribuire a sviluppare la capacità di rispondere a queste domande?

Fig. 6 – Format 3



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016




Costruzione di quesiti


Quesito di riferimento

Domanda
ITEM
Aspetto della competenza di lettura/matematica valutato Difficoltà (-4 0 +4).....
Riflessione

Fig. 7 – Format 4



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016



Analisi di un quesito su fascicoli di classe

CLASSE	Analisi risultati degli studenti	
QUESITI RELATIVI AI PROCESSI	RISULTATI	ERRORI
<i>In questa colonna va messo il N° del quesito e l'ambito</i>	<i>Quanti alunni hanno risposto correttamente e quanti hanno sbagliato? quante risposte omesse? (valore assoluto e percentuale) Confronto con i dati regionali (rapporto) e nazionali (dati scuole)</i>	<i>Quali sono stati gli errori più frequenti? Quale distrattore (risposta errata) è stato scelto con maggior frequenza dagli alunni? Che ipotesi si possono fare?</i>
OSSERVAZIONI		
<p>.....</p> <p>.....</p> <p>.....</p> <p>.....</p>		

Si prendano in esame i seguenti ambiti di misurazione dell'SNV:

- competenza testuale: comprendere, interpretare e valutare un testo;
- competenza grammaticale: formazione delle parole, morfologia, ortografia, sintassi, testualità e pragmatica;
- competenza semantico-lessicale: lemmatizzazione, usi denotativi e connotativi dei lemmi, rapporti semantici, rapporti sintagmatici e unità polirematiche.

Partendo da essi, l'enucleare le relative aree di debolezza del curriculum linguistico permette di intervenire, grazie all'analisi del compito (*task analysis*) – da intendersi come descrizione delle operazioni decisionali ottimali per la risoluzione del compito – attraverso il *prompting*, cioè l'accesso facilitato all'area critica. Ciò avviene offrendo esempi ed esercizi di rinforzo come il pronunciare più lentamente il soggetto nella frase che lo vede in posizione finale, fino a non averne più bisogno (*fading*) in un progressivo raggiungimento dell'autonomia dell'apprendente.

Quale importanza riveste l'analisi degli errori nei propri allievi? Intanto se ne possono evidenziare di ricorsivi e proprio a questi va rivolta un'attenzione particolare: la domanda da porsi non è tanto cosa non funziona, ma cosa sta funzionando. Quale ipotesi di soluzione? L'errore va infatti analizzato per ripercorrere il processo di risposta al problema, comunque positivo, seguito dallo studente, e scoprire quanti passi sono stati percorsi nel suo tentativo di soluzione verso la giusta direzione.

Non dobbiamo escludere che il docente applichi dei suoi modelli soggettivi alla correzione dell'errore, agendo in un atteggiamento di misconcezione tale da non accogliere ciò che non rientra nei suoi schemi mentali, in taluni casi meno aperti del dovuto ed erroneamente stereotipici.

Nei casi in cui ciò non accade, la riflessione sui fascicoli dei nostri studenti ci chiarisce passaggi importanti dei processi mentali e colloca l'errore in un momento preciso dell'atto risolutorio, identificato il quale è possibile attuare un intervento mirato.



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016



Attività didattica di rinforzo

Sulla base di una criticità emersa organizzare delle attività correttive.
Motivazione della scelta e breve diario di bordo sulle fasi condotte.

Criticità emersa

.....
.....

Riferimenti al contesto di rilevazione

.....
.....

Azione didattica

.....
.....

Livello

.....
.....

Contenuti

.....
.....

Competenza

.....
.....

Fig. 9 – Format 5b



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016



Tempi

.....
.....

Attività

.....
.....

Rapporti con le Indicazioni Nazionali

.....
.....

Risultati attesi

.....
.....

Risultati raggiunti

.....
.....

Osservazioni

.....
.....
.....
.....

Fig. 10– Format 6a



Prove INVALSI, dalla Valutazione al Miglioramento
- Didattica riflessiva -
A.S. 2015-2016



Prova di verifica

Costruire una prova di verifica che tenga conto degli aspetti del quadro di riferimento INVALSI o adottare una prova INVALSI/IEA/PISA e applicarla ad una ordinaria attività didattica di comprensione del testo/soluzione di un problema.

Criticità emersa

.....
.....

Riferimenti al contesto di somministrazione della prova

.....
.....

Sottocompetenza verificata

.....
.....

Tempi

.....
.....



Struttura della prova

.....
.....

Rapporti con le Indicazioni Nazionali

.....
.....

Fig. 11 – Format 6b

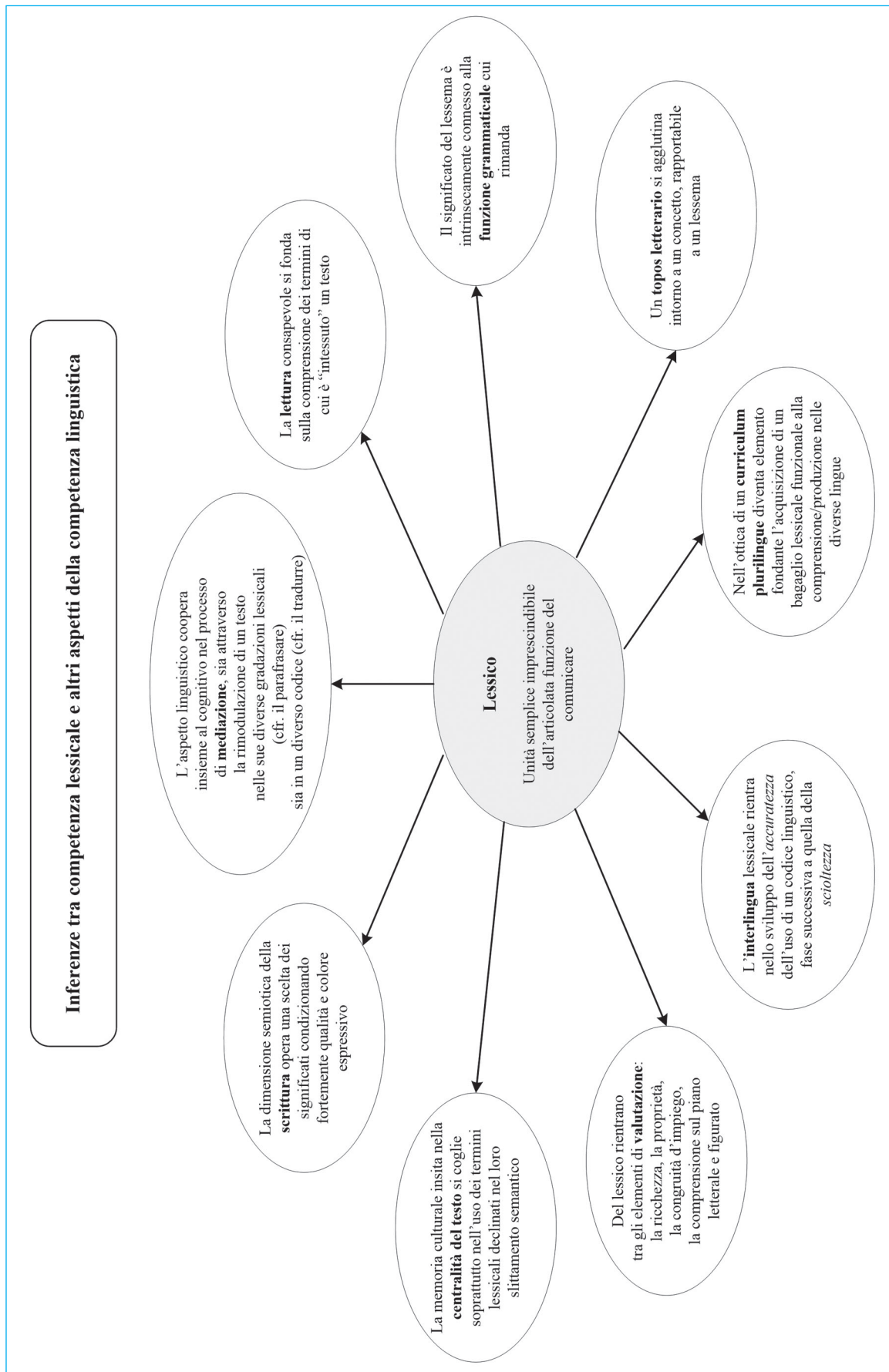
	<p>Prove INVALSI, dalla Valutazione al Miglioramento - Didattica riflessiva - A.S. 2015-2016</p>	
Risultati attesi		
.....		
Risultati raggiunti		
.....		
Osservazioni		
.....		

L'errore ricorsivo ed esteso a un gran numero di studenti ci pone di fronte a un interrogativo: i tempi di acquisizione del concetto sono stati sufficienti? La didattica è stata appropriata in termini di esercizio all'applicazione contestualizzata? La comprensione non coincide *tout court* con la capacità di applicazione in contesti nuovi, come sappiamo. Ciò può convincerci che sia utile adottare, in una didattica per competenze, una metodologia di *problem posing* e *problem solving*. Stando alle differenze cognitive che una classe presenta, il *cooperative learning* aiuta, inoltre, a valorizzare il potenziale cognitivo di chi non ha ancora raggiunto una modellizzazione autonoma e può in questo modo imparare a strutturarla con maggiore facilità nel confronto fra pari, riconoscendo la necessità di operare un miglioramento di sé, per aumentare il sodalizio di gruppo e al contempo innalzare la propria autostima. Se il soggetto di una frase non viene identificato in anastrofe da un gran numero di allievi e l'errore ricorre più volte, l'intervento compensativo sta nello sfatare l'errato convincimento che esista nella frase italiana un *ordo verborum*, un ordine delle parole. Giochi di montaggio e rimontaggio di frasi, si capisce bene, sostituiscono vecchi modelli mentali con altri più corretti e dinamici: la funzione logica si scardina dalla posizione, che semmai acquista una valenza enfatica (Serianni, 1988). Ancora sul rinforzo e l'attenzione al processo di insegnamento-apprendimento si concentra il Format 5.

Al fine di esperire le dinamiche sottese alla somministrazione di una prova, durante la quale bisogna fare i conti con il fattore tempo, l'autonomia di lavoro, i labirinti cognitivi in cui spesso i più bravi si perdono, le stereotipie da superare rispetto alla nuova contestualizzazione di situazioni problematiche già affrontate, si è proposto il Format 6.

Il lessico in una lingua si offre quale ventaglio di varianti comunicative (sinonimo, antonimo, iponimo e iperonimo) ed espressioni polirematiche declinato secondo la grammatica, quale struttura fondante della lingua codificata e socialmente identitaria. In particolare il lessico è stato trattato nella sua centralità rispetto al sistema linguistico e a quegli elementi sottesi al suo apprendimento. La mappa seguente ne esemplifica i nodi concettuali.


Fig. 12 – Il lessico al centro della competenza lessicale



Come dare importanza ai processi e alle competenze piuttosto che alle procedure e agli esiti? Qualcuno insiste sull'adozione di nuove forme di comunicazione (Alessandrini, 2005) atta a risvegliare l'intelligenza pratica (Sennet, 2008), situata e contestualizzata piuttosto che astratta, stereotipata e autoreferenziale.

Volendo infine affrontare gli studi di caso, le prove INVALSI si prestano assai bene a ciò. Come superare le difficoltà palesate sulla lemmatizzazione? Un caso affrontato è quello delle unità polirematiche, espressioni intraducibili da una lingua all'altra, che sono portatrici di un'unità di significato. Significativo è il fatto che con molta grammatica e poco lessico non si comunichi, ma con molto lessico e poca grammatica lo si faccia. Allora, è forse opportuno porsi le giuste domande, rimandando a un momento successivo le risposte, frutto di meditata riflessione.

Fig. 13 – Metadidattica sulla competenza semantico-lessicale



Competenza lessicale: metadidattica

- - - -

<ul style="list-style-type: none">➤ Quanta parte del curriculum destini ad uno studio sistematico del lessico?➤ Ritieni carenti gli alunni nella conoscenza del lessico?➤ Quali cause sono da addurre alla presenza di carenze lessicali?➤ Hai mai fatto una indagine sulla conoscenza lessicale degli allievi? Come l'hai realizzata?➤ Dai allo studio del lessico un'attenzione minore o maggiore rispetto allo studio grammaticale?	<ul style="list-style-type: none">➤ Ritieni possa essere utile ricavare degli spazi dedicati alla presentazione di un lessico ragionato?➤ Quali sono i problemi di lessico che riscontri più frequentemente nei tuoi allievi oppure che ritieni significativi a fronte della tua esperienza didattica?➤ Sui problemi di lessico riscontrati nella pratica didattica come sei intervenuto, con quali strategie?
--	--

La metadidattica ci induce a valorizzare la figura del professionista riflessivo capace di trovare ispirazione in una nuova epistemologia pedagogica della riflessione del suo agire in situazione, nella concreta prassi didattica.

A esperienza conclusa, si ritiene che i risultati attesi siano stati positivi, nella misura in cui ciascun docente corsista ha colto le suggestioni offerte rispondenti alle proprie aspettative relative alla promozione di processi di innovazione degli apprendimenti di base, nell'ambito dell'educazione linguistica e matematica, in linea con quanto previsto dai documenti programmatici ministeriali. Il modello collaborativo si è mosso nella direzione di auspicabili comunità di pratica di docenti che, attraverso la condivisione di conoscenze ed esperienze di ricerca, giungano a sperimentare piste operative di insegnamento intenzionale e ad applicare buone pratiche (Alessandrini e Pignalberi, 2012). Sono state avviate forme di didattica riflessiva nel "montaggio e smontaggio" delle prove INVALSI, nella costruzione di prove per *modeling* (Bondioli, 1995) nell'accezione di attività produttiva e creativa, nell'analisi dell'errore valorizzato e non demonizzato.

L'aver offerto un metodo di lavoro, che permetta di procedere dalla lettura dei risultati alla progettazione di una didattica orientata e allineata ai saperi imprescindibili inquadrati in standard nazionali, libera, in prospettiva, dagli angusti limiti dell'autoreferenzialità di istituto. Infine, la conduzione di ragionate *task analysis* (Fagetti, 1990) su quesiti ha permesso di imparare a progettare azioni didattiche compensative rispetto alle criticità e di mantenimento/valorizzazione delle eccellenze, nell'ottica del miglioramento, perché ciascun apprendente sappia realizzarsi al meglio.

Riferimenti bibliografici

- Alessandrini G. (2005), *Manuale per l'esperto dei processi formativi*, Carocci, Roma.
- Alessandrini G., Pignalberi C. (2012), *Le sfide dell'educazione oggi. Nuovi habitat digitali, reti e comunità*, Pensa Multimedia, Lecce.
- Bondioli A. (1995), "L'esempio tra pedagogia e psicologia: modeling, tutoring, scaffolding", *Mélanges de l'Ecole française de Rome. Italie et Méditerranée*, 107, 2: 433-457.
- Curci A.M. (2004), *Educazione plurilingue: percorsi*, testo disponibile al sito: <http://digilander.libero.it/dibiasio.neoassunti/TEMATICA7/Promuovere/competenza.pdf>, data di consultazione: 14 giugno 2017.
- Fagetti M.A. (1990), "La task analysis: quali usi?", *HD – Handicap e disabilità di apprendimento*, 33: 29-38.
- Gardner H. (1991), *Aprire le menti. La creatività e i dilemmi dell'educazione*, Feltrinelli, Milano.
- INVALSI (2006), *Rapporto nazionale OCSE-PISA 2003. Il livello di competenza dei quindicenni italiani in matematica, lettura, scienze e problem solving*, Armando, Roma.
- INVALSI (2008a), *Le competenze in scienze, lettura e matematica degli studenti quindicenni. Rapporto nazionale PISA 2006*, Armando, Roma.
- INVALSI (2008b), *Ricerca internazionale IEA PIRLS 2006. La lettura nella scuola primaria. Rapporto nazionale*, Armando, Roma.
- INVALSI (2009), *Prove INVALSI. Quadro di riferimento di italiano*, testo disponibile al sito: http://www.invalsi.it/snv0910/documenti/Qdr_Italiano.pdf, data di consultazione: 26 aprile 2017.
- Mariani L. (2016), *La sfida della competenza plurilingue. Per un'educazione linguistica trasversale ai curricoli*, Learning Paths, Milano.
- Sennet R. (2008), *L'uomo artigiano*, Feltrinelli, Milano.
- Serianni L. (1988), *Grammatica italiana. Italiano comune e lingua letteraria. Suoni, forme, costrutti*, UTET, Torino.

7. *Somministrazione delle prove INVALSI dal 2009 al 2015:
un patrimonio d'informazioni tra evidenze psicometriche e didattiche*
INVALSI tests' administration from 2009 to 2015.
Big data assets between Psychometrics and Didactics

di Giorgio Bolondi, Clelia Cascella

Nell'ambito dei modelli di risposta all'item (IRT), sebbene siano stati proposti diversi modi per la valutazione del *fit*, ossia la congruenza tra dati raccolti e modello statistico utilizzato per l'analisi, questo settore lascia ancora aperti importanti interrogativi, soprattutto per campioni di grandi dimensioni. In questo lavoro, il controllo del *fit* si è basato sull'ispezione grafica delle curve caratteristiche degli item (*Item Characteristic Curve* – ICC), stimate con il modello di Rasch, che, diversamente dai tradizionali indici di *fit*, consente di osservare tali scostamenti per specifici livelli di abilità, indirizzando verso la formulazione di alcune ipotesi che possano spiegare lo scostamento tra osservato e teorico. A tale scopo, sono stati utilizzati i dati raccolti, dal 2009 al 2015, dall'INVALSI, mediante la somministrazione di test psicometrici tesi alla rilevazione dell'abilità matematica di campioni composti, in media, ogni anno, da più di 30.000 studenti della classe seconda della scuola secondaria di II grado.

L'analisi dei dati raccolti ha consentito di individuare e catalogare tutti gli scostamenti tra le risposte osservate e le previsioni del modello di Rasch, facendo emergere una relazione sistematica tra caratteristiche degli item e tipo di violazione, alcune delle quali consentirebbero un'interpretazione dello scostamento alla luce dei più diffusi paradigmi didattici. In questo lavoro, presentiamo un approfondimento su un particolare tipo di violazione, l'*over-discrimination*, che consente di mettere in luce alcune caratteristiche delle domande che, da un lato, chiariscono la funzionalità del modello statistico utilizzato, nei confronti del fenomeno osservato, e, dall'altro, offrono interessanti spunti di riflessione sulla natura del fenomeno stesso.

Within the *Item Response Theory*, although a lot of different methods and techniques have been proposed to assess item fit, this topic rises up relevant questions in relation to which not completely satisfactory answers have been given especially for big data. In this paper, fit control was based on the graphical inspection of *Item Characteristic Curves* (ICCs), estimated by using the Rasch model (one of the most used tools in educational research to estimate students' ability) that allows observing deviations between observed and expected values for specific ability levels. This can be particularly useful in order to formulate specific hypotheses to understand and pick out at least some possible causes of violations. In order to do this, we analyzed data collected from 2009 to 2015 by the Italian National Institute for the Evaluation of Educational System (INVALSI), by administering Math achievement test to more than 30.000 students attending the 2nd grade level of high school, per year. Data analysis allowed picking out, catalogue and classify all deviations between observed and expected values, fostering the emergence of a systematic relationship between item characteristics and violation types, some of which could suggest interpretations based on some of the most popular didactic paradigms. In this paper, we presented a close examination of a particular violation type, i.e. the *over-discrimination* that allows highlighting some characteristics of items that, on the one hand, clarify the functionality of the employed statistical model and, on the other hand, offer some interesting avenues to explore regarding the nature of the phenomenon.

1. Introduzione

A partire dall'anno scolastico 2007-08, l'Istituto nazionale per la valutazione del sistema di istruzione e formazione (INVALSI) somministra, a livello censuario, prove per la valutazione degli apprendimenti sia in italiano sia in matematica, in II e V primaria e in III secondaria di I grado. Attualmente, i livelli scolastici coinvolti nell'attività di rilevazione sono attualmente quattro: II e V primaria (indicate con L02 e L05 nella prassi INVALSI), III secondaria di I grado (L08), II secondaria di II grado (L10); fino all'anno scolastico 2012-13 la rilevazione era effettuata anche nella classi I della secondaria di I grado (L06). Per ciascun livello, sono mediamente coinvolti circa 500.000 studenti, in relazione ai quali l'Istituto raccoglie anche informazioni di contesto, come per esempio la regolarità dello studente rispetto al percorso di studio, la provenienza geografica, il livello di istruzione e lo status professionale dei genitori, il voto ottenuto al primo quadrimestre in italiano e in matematica, il luogo di residenza, e la frequenza alla scuola pre-primaria (asilo nido e scuola dell'infanzia). Inoltre, per gli studenti del livello 5 (nella scuola primaria) e del livello 10 (nella scuola secondaria di II grado) viene predisposto e somministrato un questionario studente, attraverso cui rilevare anche il livello di interesse/disinteresse verso le materie oggetto di rilevazione e il livello di ansia nello studio di tali discipline. Inoltre, dall'anno scolastico 2013-14, il questionario studente è stato infine arricchito inserendo alcune domande sulla rete sociale degli studenti all'interno della classe.

Questo patrimonio di informazioni viene utilizzato dall'Istituto innanzitutto per offrire una fotografia della qualità del sistema di istruzione e formazione italiano, ma anche, per esempio, per studiare le cause dell'insuccesso e della dispersione scolastica con riferimento al contesto sociale e alle tipologie dell'offerta formativa; valutare il valore aggiunto realizzato dalle scuole ecc.

I dati raccolti dall'INVALSI possono però essere utilizzati, come si propone in questo lavoro, anche per inquadrare e impostare sulla base di evidenze quantitative un'analisi qualitativa *ex post* sugli effetti delle pratiche didattiche poste in essere nelle classi. In particolare, attraverso l'analisi psicometrica degli item che compongono le prove, è infatti possibile studiare il comportamento di risposta degli studenti a ciascun quesito e formulare ipotesi circa le pratiche didattiche che possono aver concorso a determinarlo. Questa operazione ha certamente una rilevanza strategica: le rilevazioni effettuate dall'INVALSI, oltre a essere il modo attraverso cui il MIUR può verificare e misurare, almeno per quanto riguarda i livelli di apprendimento degli allievi, l'efficacia e l'efficienza del sistema educativo italiano, ha infatti anche l'obiettivo di segnalare le aree di criticità e individuarne le cause in modo da predisporre interventi mirati che possano contenerne e possibilmente annullarne gli effetti.

Tutti questi obiettivi possono essere raggiunti solo attraverso la predisposizione di uno strumento unico (la prova) per la valutazione dell'abilità, uguale per tutti gli studenti, indipendentemente dalla regione di appartenenza: la prova (test psicometrico) funziona quindi come un termometro, che consente di individuare un livello medio di abilità rispetto al quale "misurare" non solo la resa del singolo studente, ma anche, a livelli via via crescenti, della classe, della scuola, e dell'intera area territoriale.

Seppure nella chiara consapevolezza che molti sono i fattori che possono concorrere a spiegare differenze nelle performance scolastiche (come per esempio lo status socio-economico e culturale non solo del singolo studente ma anche del contesto sociale in cui è inserito, dentro e fuori la scuola), questo studio focalizza l'attenzione sulle potenzialità interpretative che le informazioni raccolte dall'INVALSI possono offrire rispetto alle pratiche didattiche poste in essere nelle scuole, e sul modo in cui queste orientano (o possono concorrere a orientare) le strategie di soluzione messe in campo dagli studenti nel rispondere agli item contenuti nella prova.

2. Il modello di Rasch per la stima dell'abilità matematica

Le risposte fornite dagli studenti agli item contenuti nella prova sia di italiano sia di matematica sono state analizzate con il modello di Rasch, il quale ipotizza che la risposta di un soggetto a un item dipenda dalla sua abilità relativa, cioè dall'abilità intrinseca del soggetto rispetto alla difficoltà dell'item cui risponde.

La diffusione di questo modello in ambito psicometrico è dovuta, oltre che alla sua intrinseca semplicità, anche alle proprietà di cui gode. Tra queste, una delle più importanti è, senza dubbio, l'invarianza della misurazione la quale consente di poter confrontare i soggetti tra loro, gli item tra loro, e i soggetti con gli item. Sebbene si tratti solo di una

caratteristica matematica del modello che, di fatto, non azzerava i condizionamenti reciproci tra difficoltà degli item e abilità degli studenti, essa consente comunque di eliminare, almeno dal punto di vista tecnico/computazionale, le possibili distorsioni derivanti da tali condizionamenti. La meccanica del modello è tale, infatti, da consentire che:

- 1) l'abilità dei soggetti sia stimata utilizzando solo i relativi profili di risposta, senza che nessun contributo sia dato a tale stima dal numero di volte in cui gli item sono stati correttamente superati da tutti gli studenti che compongono il campione;
- 2) la difficoltà degli item che compongono la prova sia stimata indipendentemente dai profili di risposta individuali, cioè in funzione del solo numero di volte in cui sono stati superati. Da questa proprietà deriva quindi la possibilità di instaurare confronti robusti tra sottogruppi di studenti, clusterizzati in funzione di variabili rilevanti, come per esempio l'area geografica di riferimento, il genere, la provenienza socio-economica e culturale, e così via.

Perché l'invarianza della misurazione possa essere garantita, occorre che la funzionalità psicometrica degli item contenuti nella prova, e del test nel suo complesso, sia coerente con le assunzioni teoriche del modello di Rasch (Hambleton e Swaminathan, 1985), che sono:

- 1) l'unidimensionalità (tutti gli item si riferiscono a un'unica dimensione – almeno prevalente – detta abilità latente dello studente);
- 2) l'indipendenza locale (la probabilità di rispondere correttamente a un quesito è stocasticamente indipendente dalla probabilità di una risposta corretta a tutti gli altri item della prova);
- 3) la monotonicità (la probabilità di rispondere correttamente a un item è una funzione monotona dell'abilità, cioè cresce al crescere dell'abilità intrinseca dello studente).

Il controllo empirico della congruenza (*fit*) tra i dati raccolti e gli assunti teorici del modello avviene, in prima battuta, in fase di pre-test, e, successivamente, dopo la somministrazione nazionale, in main study, per controllare la tenuta, sui dati censuari, delle proprietà psicometriche che la scala deve avere per garantire adeguati livelli di *fit* e, quindi, la tenuta delle proprietà del modello.

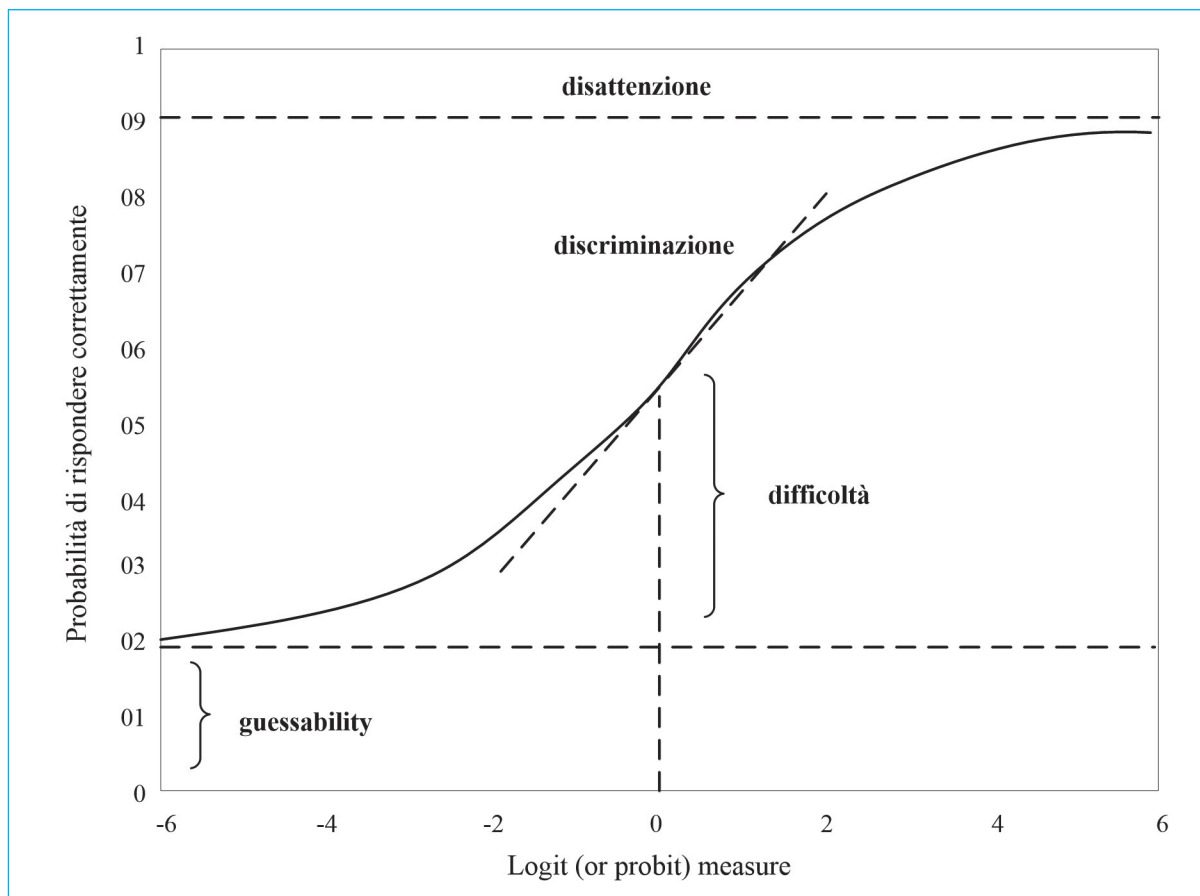
Com'è però ben noto, la verifica empirica del *fit* pone problemi metodologicamente assai rilevanti con big data, quali certamente sono quelli che l'INVALSI raccoglie annualmente per ciascun grado scolastico coinvolto nella rilevazione. Anzi, c'è ormai, da tempo, un accordo pressoché unanime sul fatto che tutti i dati empiricamente raccolti si discostano, in qualche misura, dalle attese del modello (Wright *et al.*, 1994). Inoltre, con campioni di grandi dimensioni (quali certamente sono quelli cui l'INVALSI somministra le prove cognitive), livelli di significatività ordinaria ci imporrebbero di rifiutare qualsiasi modello solo perché il numero delle osservazioni è troppo ampio. In tali casi, abbiamo piuttosto bisogno di una misura che quantifichi lo scostamento accettabile tra il modello e i dati (Gustafson, 1980). Alla luce di quanto detto, la vera domanda, quindi, da porsi almeno per dataset di grandi dimensioni, non è se i dati sono coerenti con gli assunti teorici del modello, ma piuttosto con quale misura può essere tollerata la presenza di “rumore” nei dati, e cioè lo scollamento tra i dati raccolti e gli assunti teorici del modello affinché quest'ultimo continui a conservare le desiderate proprietà misuratorie.

In risposta a questo quesito, sono stati definiti, in funzione della finalità del test, specifici range di tolleranza (Wright *et al.*, 1994). In particolare, Linacre scrive che studi basati su dati simulati e l'esperienza maturata nell'analizzare centinaia di dataset indicano che sono produttivi per la misura valori di *fit* compresi tra 0,5 e 1,5 (*ibid.*).

Ripetendo le analisi condotte dagli autori sui tutti i dataset raccolti dall'INVALSI, dal 2009 a oggi, per ciascun livello scolastico, abbiamo non soltanto potuto verificare che gli indici di *fit* sono sempre all'interno di tali range, ma abbiamo potuto anche definire, per le prove INVALSI, su base empirica invece che simulata, un nuovo intervallo di tolleranza, più piccolo rispetto a quello proposto da Wright.

Questa operazione è stata possibile utilizzando, oltre ai tradizionali indici di *fit*, anche l'osservazione della curva caratteristica degli item (fig. 1).

Fig. 1 – Effetti di interazione tra genere e status socio-culturale osservato per la prova di matematica del 2015



Fonte: adattamento da Hays, Morales e Reise (2000).

La curva caratteristica degli item (ICC) esprime la probabilità di dare una risposta corretta a un certo item in funzione del livello di abilità posseduto dallo studente. Secondo la meccanica del modello di Rasch, tale probabilità è influenzata esclusivamente dall'abilità relativa dello studente e, quindi, qualsiasi altro fattore (per esempio legato alle caratteristiche personali e/o socio-demografiche dello studente, la fortuna nel dare a caso la risposta corretta, la distrazione ecc.) è da considerarsi come un fattore di disturbo, non considerato nell'ipotesi di unidimensionalità (Embretson e Reise, 2000). Cionondimeno, da ciascuna ICC, possono essere tratte molte informazioni utili per la corretta interpretazione della funzionalità dell'item cui si riferiscono. Per esempio, è possibile quantificare l'effetto del caso (*guessability*, cfr. fig. 1), cioè capire qual è la probabilità di dare una risposta corretta anche da parte di studenti che, in base al profilo di risposta dello studente, il modello "ritiene" totalmente privi della quantità di abilità necessaria per poter rispondere correttamente. È inoltre possibile valutare la capacità discriminativa dell'item, e cioè la sua capacità di separare correttamente gli studenti in funzione del livello di abilità da ciascuno posseduto. La valutazione della discriminatività avviene attraverso l'osservazione della pendenza, cioè dell'inclinazione della curva teorica (ICC). La distanza tra la spezzata empirica (data dall'insieme di tutte le risposte osservate) e la curva teorica consente di vedere lo scostamento tra le risposte effettivamente date dagli studenti alla prova e le attese del modello, e quindi di formulare alcune prime ipotesi qualitative sulle relazioni intercorrenti tra l'item e il costrutto misurato complessivamente dalla prova.

L'ispezione grafica della curva caratteristica e della sua corrispondenza con la spezzata empirica consente quindi di controllare la coerenza tra i dati raccolti e, in particolare, l'assunto di cumulatività, secondo cui gli studenti che, sulla base delle risposte da questi dati a tutti gli item contenuti nella prova, sono quelli che ottengono i punteggi più alti su scala di Rasch sono anche quelli che hanno una maggiore probabilità di rispondere correttamente al maggior numero di item.

Dopo aver accertato, attraverso il controllo del *fit*, che la batteria di item che compone la prova costituisca una scala nel senso di Rasch, l'ispezione grafica della curva caratteristica di ciascun item diventa uno strumento interpretativo di grande utilità: qualsiasi scostamento nella pendenza della spezzata empirica rispetto alla curva teorica, osservato per uno o più item, segnala la presenza, per tali item, di uno o più fattori di disturbo, cioè di fattori non ipotizzati dal modello di Rasch, che modificano, nella realtà, la probabilità di una risposta corretta stimata ex ante dal modello di Rasch. In questo senso è come se i piani della rilevazione si separassero: per questi item, a governare la probabilità di una risposta corretta non è più il livello di abilità (intrinseca) dello studente ma sono piuttosto fattori esterni, di cui il modello non può tenere conto.

Quali sono questi fattori? E che legame hanno con le strategie di risposta degli studenti agli item? L'analisi che proponiamo in questo lavoro parte da una base empirica forte, resa possibile dall'immenso patrimonio informativo raccolto, negli anni, dall'INVALSI. Sulla base di questa, abbiamo individuato un set di item che, negli anni, hanno presentato caratteristiche di funzionamento simile. Questa operazione di classificazione *ex post* ha seguito diversi criteri, uno dei quali si pone nel filone degli studi della didattica della matematica.

In sintesi, tra le domande somministrate in questi anni ce ne sono diverse che si presentano corrette come formulazione, coerenti con la pratica didattica della disciplina, ammissibili tra quelle che una indagine sugli apprendimenti in matematica potrebbe voler considerare, e che però per qualche aspetto si discostano, nel comportamento osservato, da quello previsto dal modello. Un'analisi specifica delle caratteristiche delle domande, e dei risultati ottenuti, permette di affermare che anche queste domande forniscono informazioni interessanti, in due direzioni speculari. Da un lato, la presenza di caratteristiche comuni a blocchi di domande per qualche aspetto "anomale", permette di definire e caratterizzare meglio il carattere latente misurato dalla prova. Dall'altro, il confronto tra il risultato empirico rilevato e il comportamento teorico atteso permette di formulare ipotesi riguardo ai comportamenti degli studenti di fronte a queste domande.

3. Un esempio di analisi

L'analisi effettuata ha coinvolto i 211 item somministrati nel livello 10 dal 2011 al 2014. Il tipo di comportamento anomalo analizzato è stata l'*over-discrimination*: sono stati cioè considerati item per i quali la pendenza della curva interpolante la spezzata empirica è superiore alla pendenza della curva teorica attesa (in generale, sempre nei limiti di tollerabilità previsti dalla letteratura). In particolare (ma non solo), questo comportamento si può verificare quando il modello sovrastima la probabilità di risposta corretta degli studenti deboli, e sottostima la probabilità di risposta corretta degli studenti più abili. In misura più o meno forte, questo comportamento è stato rilevato in 57 degli item analizzati (presenti in tutte le prove considerate) e l'analisi si è poi concentrata su quelli in cui il fenomeno era particolarmente rilevante e quindi tale da segnalare un comportamento sensibilmente diverso da quello atteso.

Un caso tipico è rappresentato dall'item a della domanda D8 della prova (di livello 10) del 2011.

Fig. 2 – Item D8a della prova somministrata al livello 10 nel 2011

D8. La dimensione di un televisore è la misura della diagonale dello schermo espressa in pollici (1 pollice = 2,54 cm). Nei televisori di nuova generazione il rapporto tra la larghezza e l'altezza dello schermo è 16:9.

a. **Se la larghezza dello schermo di uno di questi televisori è circa 57,5 cm, qual è all'incirca la sua altezza?**

Risposta: cm

L'*over-discrimination* emerge in maniera molto evidente dalla lettura del grafico in fig. 3.

Notiamo una forte incidenza di risposte mancanti, soprattutto nella parte bassa della scala di abilità (come è naturale aspettarsi, trattandosi di una domanda aperta), fattore che spiega l'azzeramento della *guessability*.

Fig. 3 – Curva caratteristica dell'item D8a della prova somministrata al livello 10 nel 2011

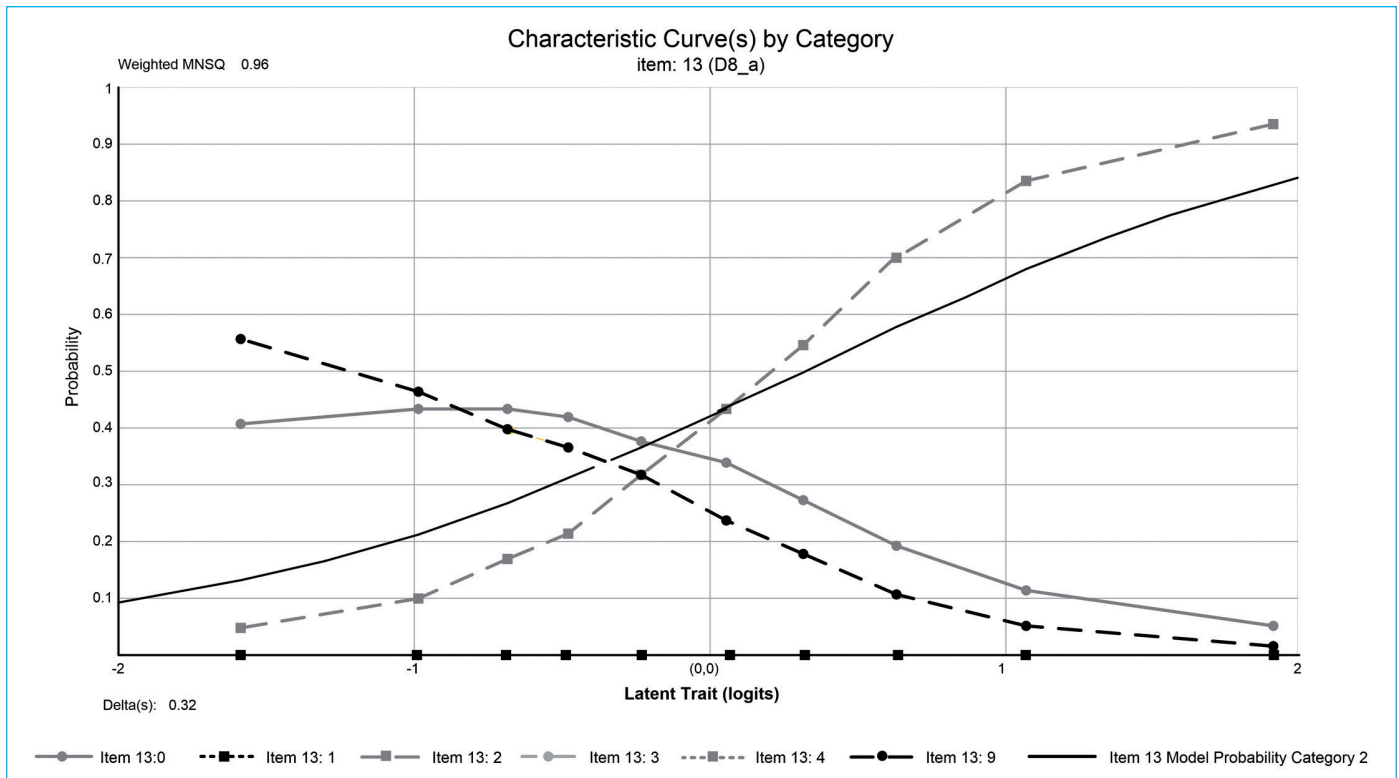
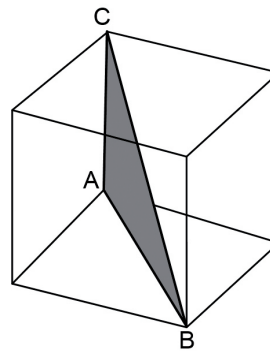


Fig. 4 – Item D9b somministrato al livello 10 nel 2011

D9. Nella figura è rappresentato un cubo.



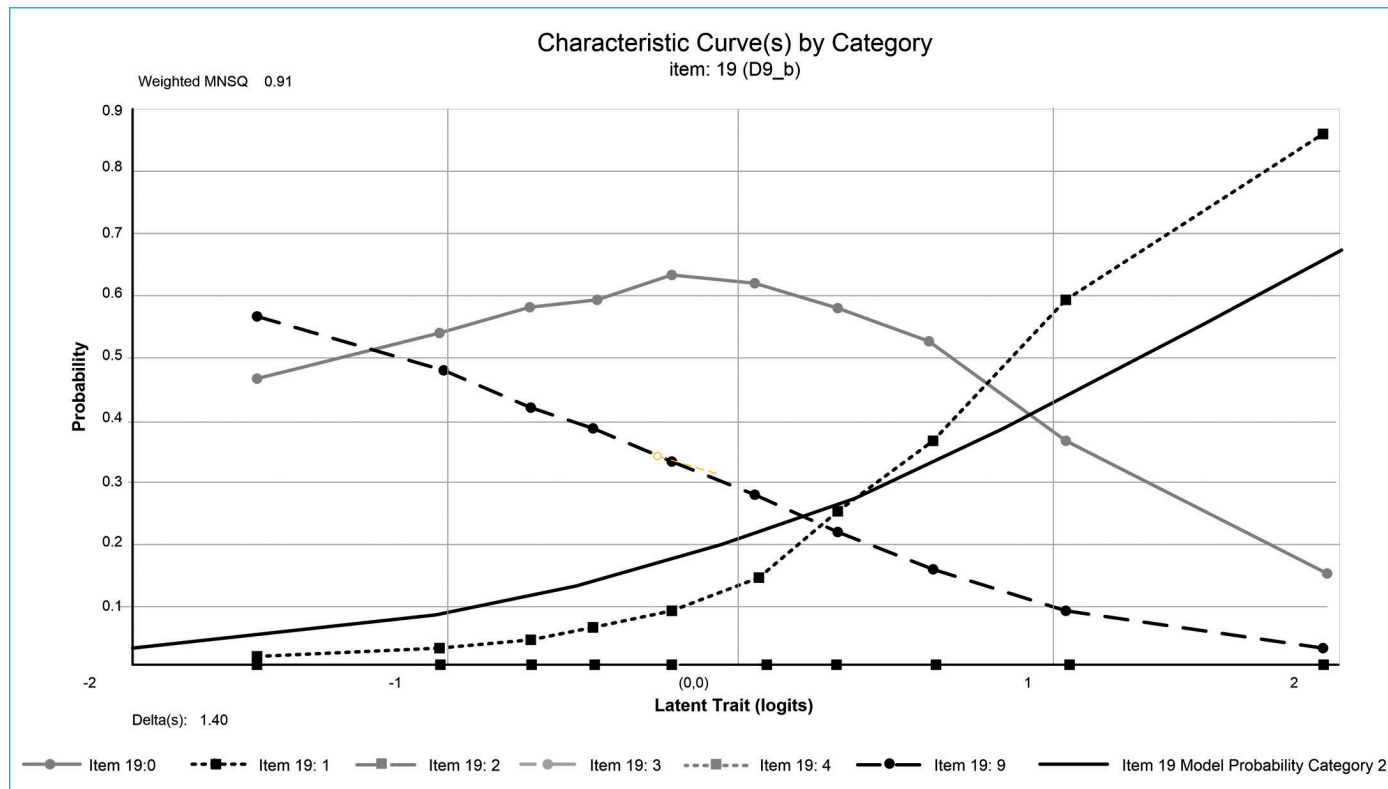
Il triangolo ABC ha come lati uno spigolo del cubo, la diagonale di una sua faccia e una diagonale del cubo.

- b.** Se lo spigolo del cubo misura 1 m, quanto misurano i lati del triangolo ABC?
- AC = m
- AB = m
- BC = m

Un comportamento del tutto simile si ha anche nella domanda successiva della stessa prova, la D09b.

Anche in questo caso, il modello sovrastima gli studenti deboli e sottostima quelli più forti. Come nel caso precedente, si ha un'altissima percentuale di risposte mancanti, per studenti deboli o medi, e una *guessability* inferiore alla stimata.

Fig. 5 – Curva caratteristica dell'item D9b somministrato al livello 10 nel 2011



Altre domande dove si rilevano gli stessi fenomeni sono le seguenti, selezionate considerando in ogni anno la domanda con la maggiore *over-discrimination*.

Fig. 6 – Item D11b somministrato al livello 10 nel 2011

D11. La relazione seguente esprime la spesa annuale per l'automobile, composta da una parte fissa e da una parte proporzionale al numero di km percorsi:

$$S = F + c \cdot k$$

dove F sono le spese fisse, c è il costo al km e k è il numero di km percorsi.

Nella tabella sono riportate le spese fisse e il costo al km per alcuni tipi di automobile.

	Auto A	Auto B	Auto C	Auto D
Spese fisse F	900 euro	580 euro	650 euro	1.200 euro
Costo al km c	0,25 euro/km	0,33 euro/km	0,27 euro/km	0,31 euro/km

b. Il proprietario di un'auto di tipo A ha speso 3.000 euro in un anno. Quanti km ha percorso?

Fig. 7 – Curva caratteristica dell'item D11b somministrato al livello 10 nel 2011

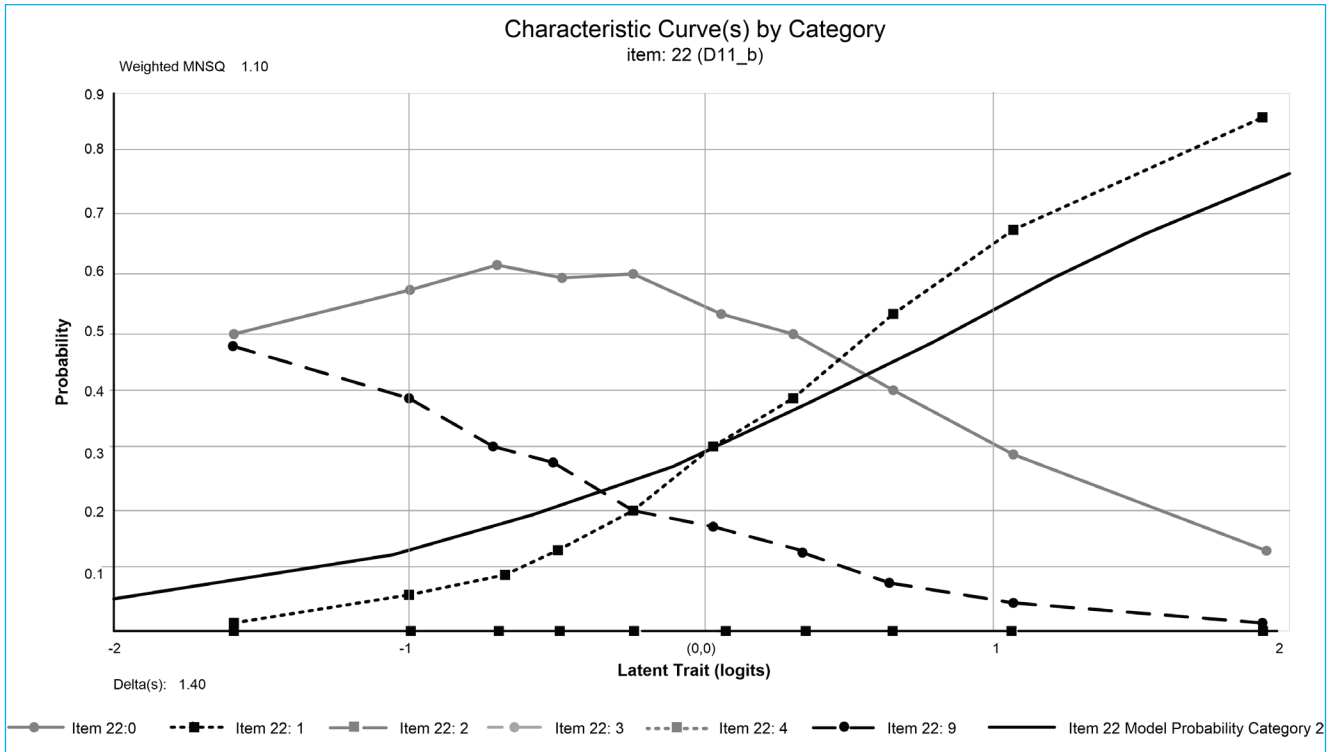
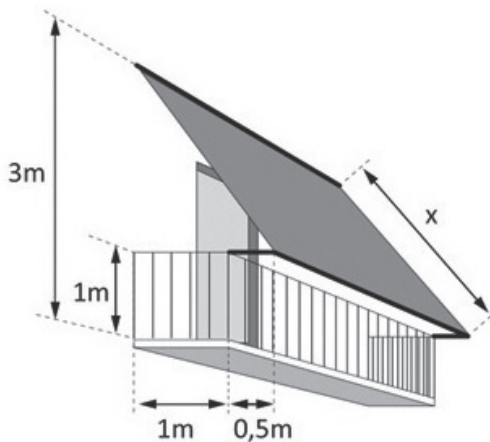


Fig. 8 – Item D24 somministrato al livello 10 nel 2012

D24. Occorre confezionare una tenda da sole per il balcone in figura.

La tenda deve essere fissata al muro a 3 m di altezza dal pavimento del balcone, che è largo 1 m. La tenda deve sporgere 0,5 m dalla ringhiera che è alta 1 m.



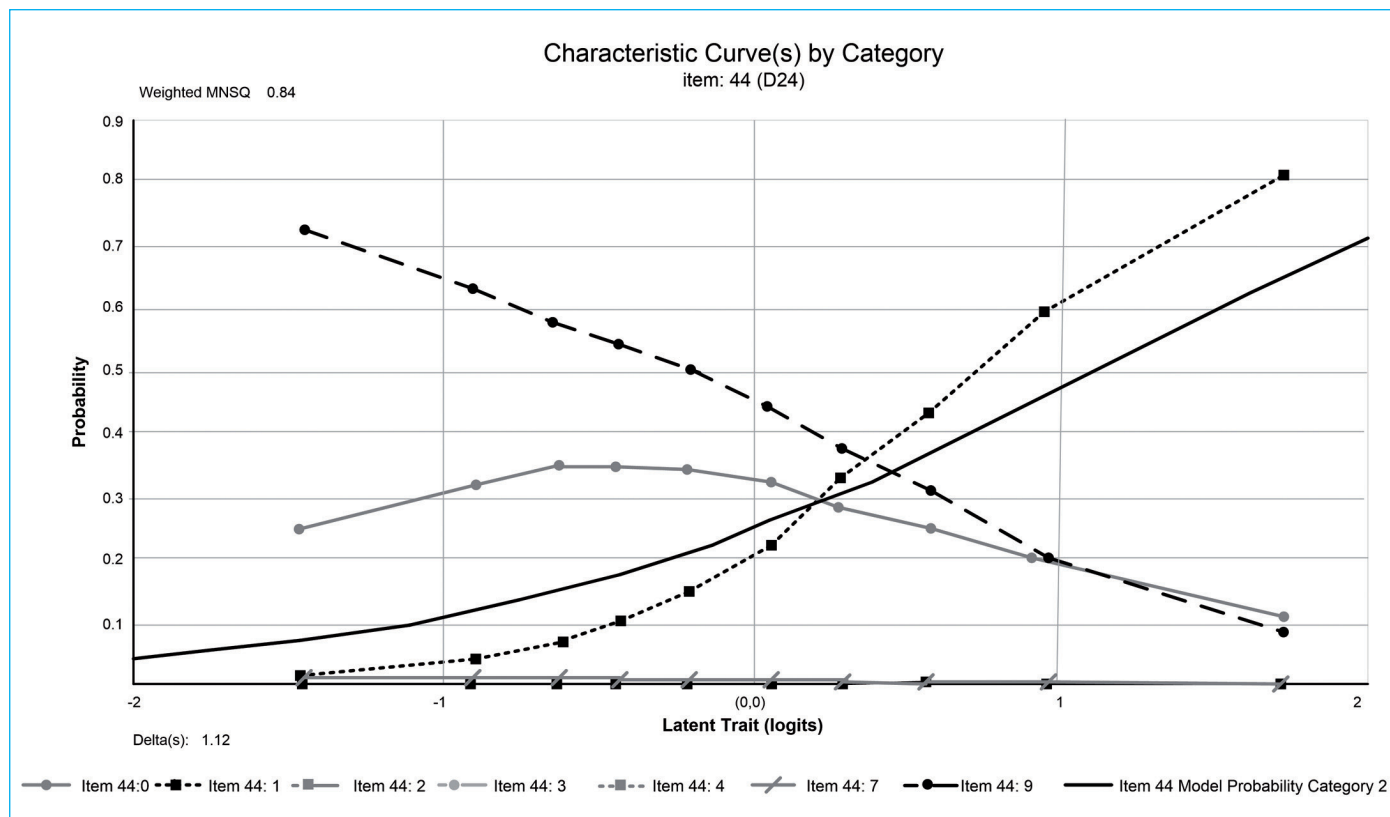
Scrivi i calcoli che fai per trovare la lunghezza x della tenda e infine riporta il risultato.

.....

.....

Risultato: $x = \dots\dots\dots$ metri

Fig. 9 – Distractor plot dell'item D24 somministrato al livello 10 nel 2012



In questo item, in particolare, la percentuale di risposte omesse supera il 50% per gli studenti del primo decile di abilità, con un contemporaneo azzeramento delle risposte corrette per gli stessi studenti.

Come negli altri casi di *over-discrimination* riportati prima, anche per l'item D24 (livello 10 del 2012), abbiamo una forte presenza di risposte mancate e un azzeramento della *guessability*.

In particolare, per tutti i decili della metà meno abile della popolazione rileviamo una percentuale di risposte mancate superiore o all'incirca uguale al 50%, e di conseguenza una bassissima percentuale di risposte corrette.

Di seguito, proponiamo altri item, somministrati a diversi livelli e in diversi anni, in cui si possono rilevare gli stessi elementi concomitanti: un'*over-discrimination* complessiva risulta caratterizzata da un'elevata presenza di risposte omesse nella parte meno abile della popolazione; questo implica che sono presenti in misura molto ridotta (in certi casi assenti) le risposte corrette fornite dai primi decili della popolazione, e che vengono abitualmente attribuite anche a risposte casuali (determinando quindi la soglia di *guessability*). L'indice di difficoltà dell'item, e quindi la curva stimata delle risposte, viene determinato dal modello in base alla percentuale di risposte corrette, e in questa percentuale vengono a mancare quelle fornite per caso dal segmento meno abile della popolazione. Sembra quindi ragionevole aspettarsi che questo indice risulti sovrastimato per il segmento più abile della popolazione, in cui le risposte mancate si azzerano (come risulta dai grafici proposti).

La selezione di item che proponiamo, con i relativi grafici, fornisce una prima raccolta di casi esemplari di questo comportamento tipico.

Fig. 10 – Item D18 somministrato al livello 10 nel 2013

D8. In un quartiere di una città, il calendario della raccolta differenziata (carta, vetro e plastica) prevede che la raccolta della carta avvenga ogni 28 giorni, quella del vetro ogni 21 giorni e quella della plastica ogni 14 giorni. Oggi sono state effettuate le raccolte di carta, vetro e plastica.

La prossima volta in cui la raccolta di carta, vetro e plastica verrà fatta contemporaneamente sarà tra giorni.

Fig. 11 – Curva caratteristica dell'item D18 somministrato al livello 10 nel 2013

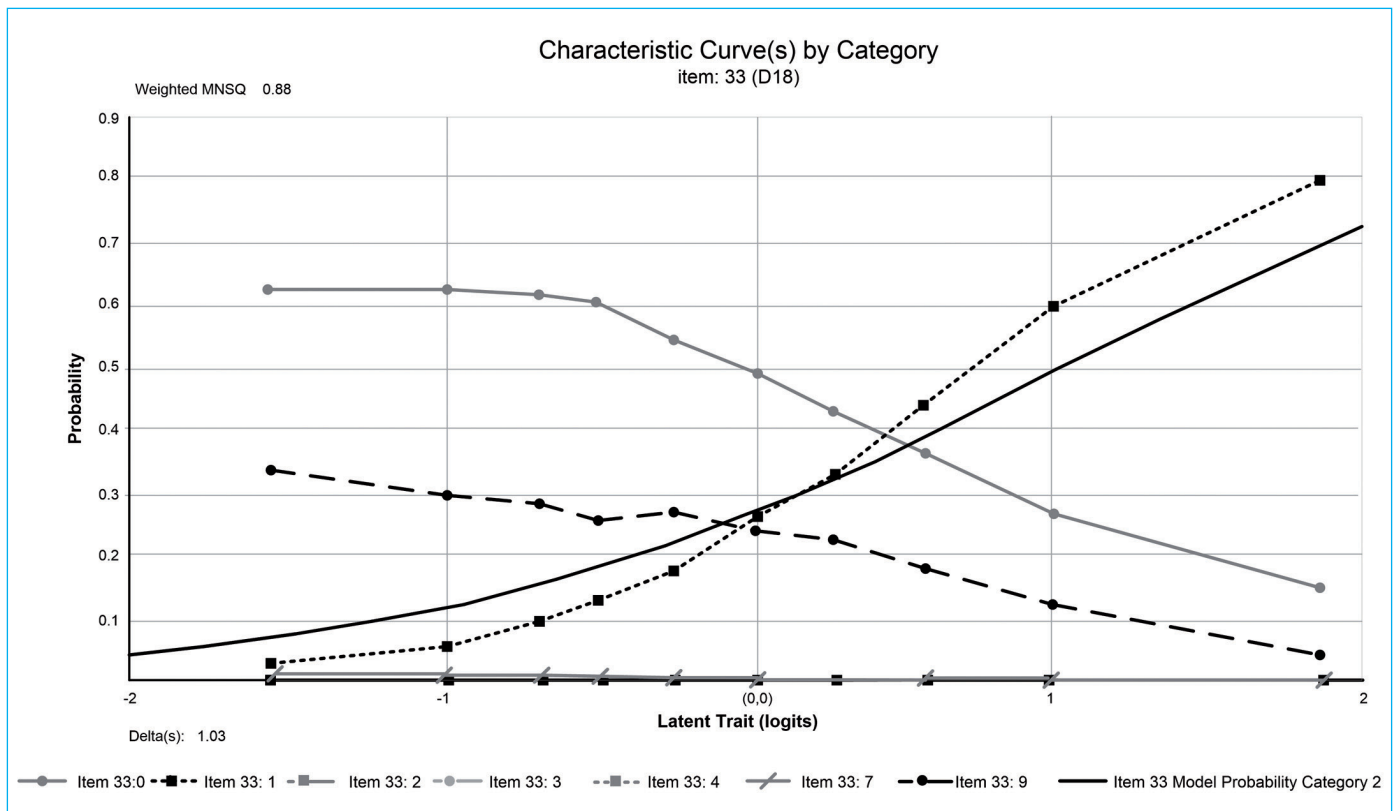


Fig. 12 – Item D17 somministrato al livello 10 nel 2014

D17. È data l'equazione $(2k-3)x + 1 - k = 0$, in cui x è l'incognita e k è un numero reale.

La soluzione dell'equazione è 1 per $k = \dots\dots\dots$

Fig. 13 – Curva caratteristica dell'item D17 somministrato al livello 10 nel 2014

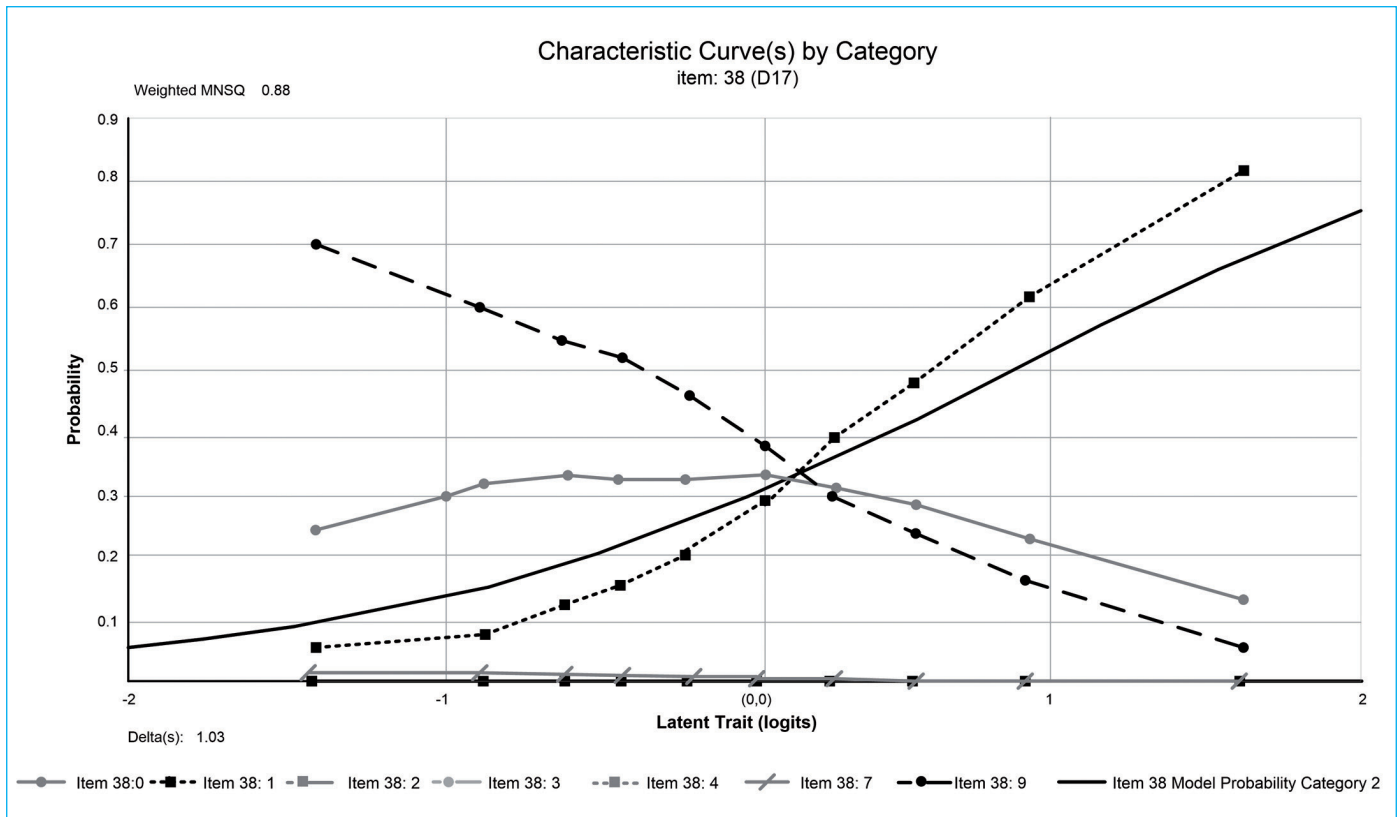


Fig. 14 – Item D20 somministrato al livello 10 nel 2014

D20. Da un controllo di qualità è emerso che una macchina ha prodotto 14 pezzi difettosi su una produzione di 1200 pezzi. Che stima è ragionevole fare del numero di pezzi difettosi su una produzione di 2150 pezzi?

Scrivi i calcoli che hai fatto per trovare la risposta e poi riporta il risultato approssimandolo all'unità.

.....

.....

.....

Risultato (approssimato all'unità):

Fig. 15 – Curva caratteristica dell'item D20 somministrato al livello 10 nel 2014

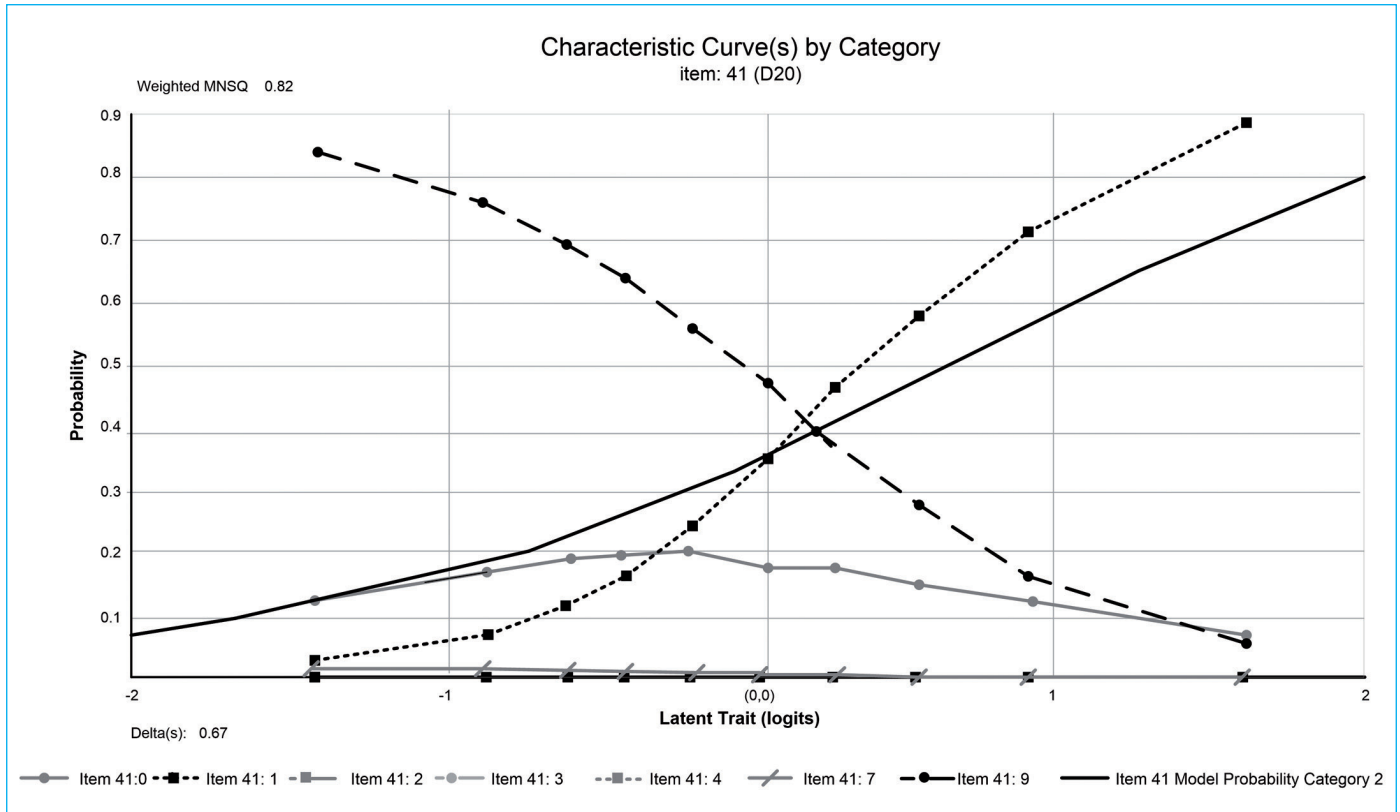


Fig. 16 – Item D22b somministrato al livello 10 nel 2014

D22. Da un controllo di qualità è emerso che una macchina ha prodotto 14 pezzi difettosi su una produzione di 1200 pezzi. Che stima è ragionevole fare del numero di pezzi difettosi su una produzione di 2150 pezzi?

Scrivi i calcoli che hai fatto per trovare la risposta e poi riporta il risultato approssimandolo all'unità.

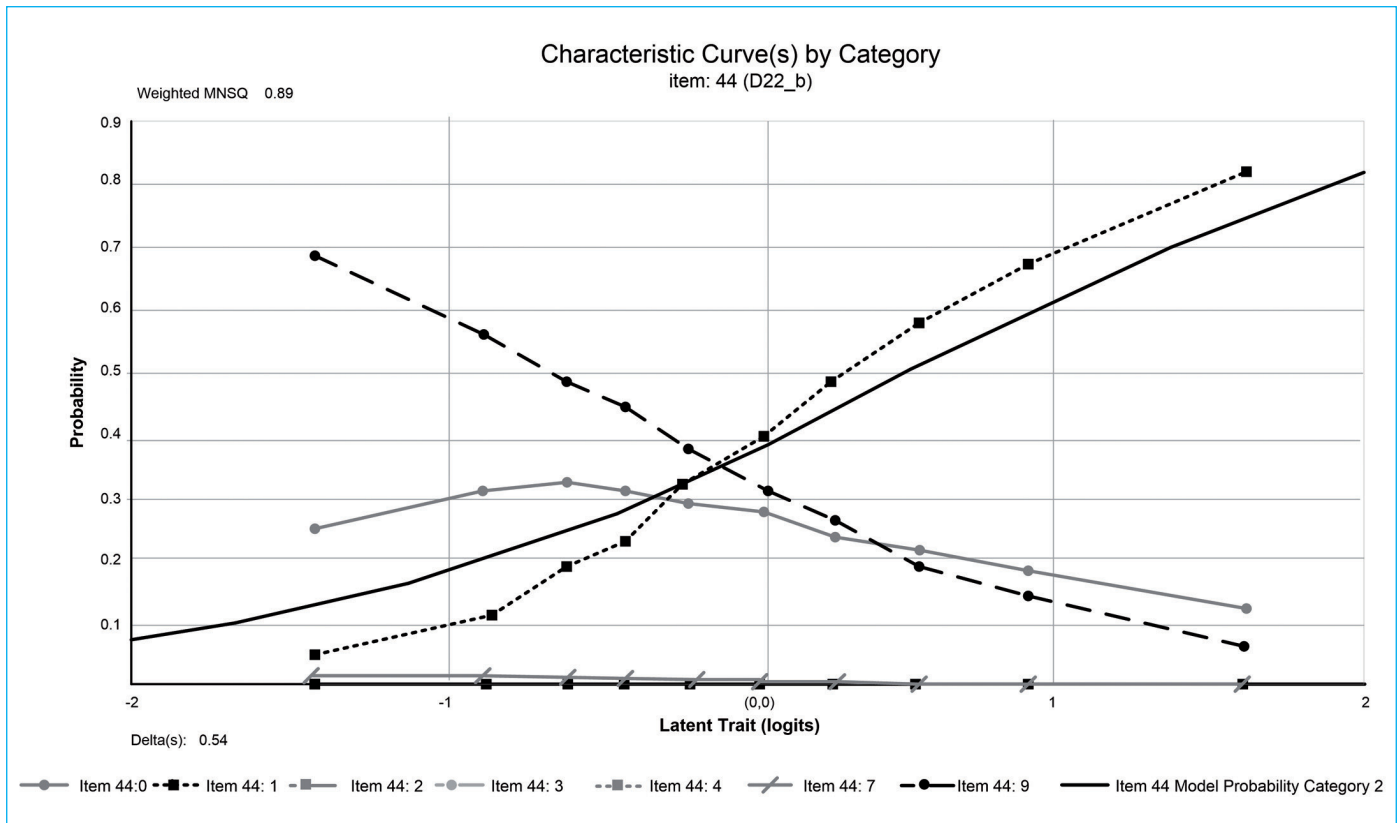
.....

.....

.....

Risultato (approssimato all'unità):

Fig. 17 – Item D22b somministrato al livello 10 nel 2014



4. Conclusioni

Le domande mostrate in questo lavoro sono quelle che presentano l'*over-discrimination* più marcata tra le 211 analizzate, e hanno alcune caratteristiche comuni, sia nell'input (formulazione tipologia di domanda, contestualizzazione) sia nell'output (i risultati degli studenti). In primo luogo, sono caratterizzate da una forte percentuale di risposte mancanti, soprattutto per i rispondenti di bassa abilità, e questo porta ad azzerare la *guessability* prevista dal modello. Sono domande di difficoltà media o medio alta, con parametro di difficoltà compreso tra 0,33 e 1,44 nelle rispettive prove. Dal punto di vista della formulazione dell'item, possiamo osservare che sono tutte domande a risposta aperta, o (secondo la terminologia adottata dall'INVALSI), "aperte univoche", cioè item aperti in cui però la risposta è univocamente determinata. Sono inoltre tutte domande in contesto, reale o matematico, in cui la risposta non dipende dall'applicazione diretta di una conoscenza o di un'abilità, ma dalla capacità di riconoscere in una situazione inusuale la presenza di concetti generalmente ben conosciuti. Una interpretazione didattica naturale è che i rispondenti di bassa abilità, di fronte a una situazione in cui i concetti appresi (e di conseguenza l'abilità misurata) non sono presenti in maniera immediatamente riconoscibile, rimangono spiazzati, e tendono a non rispondere. Per i rispondenti più abili questa difficoltà non sussiste, e le domande risultano di fatto più facili di quanto previsto dal modello. Sembra quindi che l'*over-discrimination* sia correlata, in questi casi, a uno "spaesamento" degli studenti deboli di fronte a situazioni di matematizzazione. In altre parole, gli studenti deboli, di fronte a una situazione in cui devono tradurre un problema in contesto in una formulazione matematica, non sapendo come procedere omettono la risposta. Simmetricamente, la capacità di matematizzare situazioni in contesto sembra essere un elemento di forte discriminazione delle abilità degli studenti. Naturalmente, queste interpretazioni vanno poi eventualmente confermate o raffinate mediante un'opportuna analisi di tipo qualitativo.

In conclusione si può comunque affermare che nei casi di *over-discrimination* analizzati la sottostima delle risposte degli studenti di bassa abilità dipende dall'alta incidenza delle risposte mancanti, che determina nel contempo

l'azzeramento della *guessability*. Questo ovviamente porta il modello a stimare un coefficiente di difficoltà più alto di quanto non si avrebbe, per esempio, con una formulazione a risposta chiusa, e quindi a sottostimare le risposte degli studenti di abilità più alta. Il comportamento degli studenti è a sua volta spiegabile in termini di caratteristiche della domanda: la sua tipologia, come si è detto, ma anche e soprattutto il tipo di processo richiesto, meno frequente nella pratica didattica rispetto ad altri.

Riferimenti bibliografici

- Embretson S.E., Reise S.P. (2000), *Item Response Theory for Psychologists*, Lawrence Erlbaum Associates Publishers, Mahwah (NJ).
- Gustafson J.E. (1980), "Testing and obtaining fit of data to the rasch model", *British Journal of Mathematical and Statistical Psychology*, 33, 2: 205-233.
- Hambleton R.K., Swaminathan H. (1985), *Item Response Theory: Principles and Applications*, Kluwer-Nijho, Boston.
- Hays R.D., Morales L.S., Reise S.P. (2000), "Item response theory and health outcomes measurement in the 21st century", *Medical Care*, 38, 9 Suppl.: II 28-II 42.
- Wright B.D., Linacre J.M., Gustafson J.E., Martin-Lof P. (1994), "Reasonable mean-square fit values", *Rasch Measurement Transactions*, 8, 3: 370, testo disponibile al sito: <http://www.rasch.org/rmt/rmt83b.htm>, data di consultazione: 11 maggio 2017.

8. “Questione di feedback”: dati INVALSI e pratiche di valutazione in classe “A matter of feedback”: INVALSI data and classroom assessment practices

di Serafina Pastore, Michela Freddano

L'enfasi sull'uso dei dati per il miglioramento dei livelli di apprendimento degli studenti rappresenta uno dei trend più evidenti in ambito scolastico. I movimenti di *school effectiveness* e *school improvement* hanno esercitato una notevole pressione sugli insegnanti alimentando, spesso, identificazioni immediate e non sempre corrette: se gli studenti ottengono punteggi alti nelle prove abbiamo insegnanti di successo, se gli studenti hanno risultati di livello inferiore, la conclusione è opposta. Performance di livello superiore si possono però ottenere anche attraverso metodi “alternativi” come il *teaching to the test* e il *cheating* (Shepard, 2000). Le modalità utilizzate dagli insegnanti per prendere decisioni didattiche rispondenti al contesto classe non sono ancora ben definite e sufficientemente indagate: utilizzare i dati delle rilevazioni su vasta scala per informare le azioni didattiche sembra una pratica incoativa (Wyatt-Smith, 2000). Opportuno, allora, interrogarsi su quanto i dati delle Rilevazioni nazionali INVALSI supportino gli insegnanti a:

- stabilire le priorità dell'insegnamento, anche rispetto a chi presenta difficoltà o ritardi nell'apprendimento;
- ridefinire i metodi didattici;
- considerare se, e come, adattare le pratiche didattiche.

Con un richiamo forte all'orientamento internazionale dell'*Assessment for learning* (Wiliam, 2010), i dati provenienti dalle indagini su larga scala possono essere utilizzati dagli insegnanti, non solo in ottica sistemica, per la progettazione e implementazione di percorsi di miglioramento, ma per restituire agli studenti, attraverso la pratica del *formative assessment*, informazioni dettagliate sul loro apprendimento e costruire così percorsi riflessivi e meta-cognitivi funzionali a una reale didattica per competenze.

Dal punto di vista metodologico, il presente contributo si concentra sui dati delle Rilevazioni nazionali INVALSI condotte nell'a.s. 2013-14 e nell'a.s. 2014-15 e ricorrendo alle informazioni raccolte con il Questionario insegnante si focalizza sulle attività dei docenti di italiano delle classi quinte di scuola primaria.

I dati raccolti mostrano come sia profonda la frattura tra le forme del *classroom assessment* e del *large-scale assessment*. Quest'ultimo appare quasi “rimosso” dal processo di istruzione, al punto che spesso il senso e la finalità di una simile forma di valutazione rimangono “oscuri” per gli insegnanti.

The emphasis on data gathered through large-scale assessment programs to improve students' achievement levels represents one of the most evident trends within the educational field.

The school effectiveness and school improvement movements have exerted a great pressure on teachers. These movements have also reinforced identifications that are not always correct: if students' scores on such tests are high, educators are regarded as successful. If students' scores are low, the opposite conclusion is reached. However, high scores in high-stakes testing environments can be achieved using “alternative” methods such as excessive teaching to the test and cheating (Shepard, 2000).

In spite of the attention on large-scale assessment systems (both at national and international levels) how teachers use data for decision-making in their daily classroom activity is not well defined and represents a research object not sufficiently considered. Using data gathered through large-scale programs seems to be an inchoative practice (Wyatt-Smith, 2000).

In this vein, it is opportune to wonder about how much data gathered through INVALSI can help teachers to:

-
- articulate teaching priorities and define actions for students with learning difficulties;
 - refine teaching methods;
 - consider if, and how, to adapt the instructional practices.

Recalling the *Assessment for learning* perspective (William, 2010), data of large-scale assessment can be used by teachers not only to design and implement school improvement actions, but also, in a formative assessment framework, to give students detailed information on their achievement. This paper focuses on the INVALSI test (“Rilevazioni nazionali”) conducted in 2013-14 and 2014-15, and on the results obtained through the Teacher Questionnaire, an instrument designed to gain information from teachers selected within the sample-classes during the INVALSI test administration. More specifically, we focus on Italian teachers of the 5th level of the primary school.

Data gathered confirm how deep is the fracture arisen between classroom assessment and large-scale assessment, as to such an extent that teachers do not understand the meaning and aims of large-scale assessment.

1. Introduzione

Il tema della valutazione in ambito scolastico costituisce un aspetto di cruciale rilevanza. Sullo sfondo della *learning society* e dei rapidi e profondi cambiamenti che hanno pervaso i sistemi educativi attuali, la valutazione ha finito con il catalizzare e monopolizzare una crescente attenzione, tanto da configurarsi quasi come un passaggio obbligato dei processi di cambiamento e miglioramento. Entrata a far parte della *governance* dei sistemi educativi, la valutazione, a servizio della collettività e dei diversi attori coinvolti nel mondo della scuola, ha cominciato a facilitare le scelte educative, l’identificazione di punti di forza e di debolezza del servizio scolastico, la definizione di programmi per un agevole conseguimento dei risultati auspicati. Si è così posta in termini di funzionalità strategica per comprendere le criticità di un sistema di istruzione e per migliorare le politiche scolastiche individuandone i punti deboli e definendone gli investimenti e i margini di azione. Le riforme che negli ultimi 20-25 anni hanno cercato di rispondere, in modo più o meno coordinato, alle trasformazioni sociali, politiche, economiche, hanno alimentato un inedito e più vigoroso interesse per la valutazione, tanto nelle scelte a livello di *policies* educative, tanto a livello pratico-operativo rispetto alle modalità di utilizzo e implementazione delle stesse politiche per raggiungere buoni risultati.

L’espansione della scuola di massa, il riconoscimento della valenza del capitale umano per la produzione di ricchezza nella società della conoscenza, la decentralizzazione del potere e l’aumento dell’autonomia scolastica, hanno reso la valutazione un’enorme cassa di risonanza, in grado di assorbire e amplificare l’interesse politico e sociale per l’analisi dei sistemi organizzativi dell’istruzione. Produttività, concorrenza, efficienza, misurabilità dei risultati sono, in breve tempo, divenuti aspetti rappresentativi di una valutazione tesa ad assolvere sia le funzioni di controllo, garanzia, regolamentazione del sistema di istruzione, sia quelle di promozione e sostegno dell’innovazione (Laveault, 2016; Pastore, 2015; Faggioli, 2014; Bottani, 2013; Allulli, Farinelli e Petrolino, 2013). Nell’ottica di una maggiore qualità, trasparenza e disponibilità alla rendicontazione di quanto realizzato si è rimarcato il valore della scuola in termini di apprendimento e di risultati attraverso prove strutturate con standard di riferimento nazionali (INVALSI) e comparazioni internazionali (OCSE-PISA; IEA-TIMSS e PIRLS).

L’enfasi sulla dimensione dell’*accountability* spesso, però, ha indotto a declinare la valutazione solo nei termini del controllo e delle indagini campionarie su larga scala sul profitto degli studenti.

Tali indagini sono finalizzate a:

- mettere a punto indicatori delle prestazioni degli studenti comparabili (e dunque dei risultati dei sistemi di istruzione), in riferimento ai giovani in uscita dalla scuola dell’obbligo;
- individuare i fattori che spieghino gli esiti delle prove e gli elementi che caratterizzano i sistemi (o le scuole) che hanno ottenuto risultati migliori, in termini di livello medio e di omogeneità ed equità dei risultati, in modo da trarre indicazioni relative a scelte politiche e gestionali efficaci;
- fornire dati sui risultati del sistema di istruzione, in modo regolare e prevedibile, così da consentire un monitoraggio del sistema di istruzione che ne segua gli sviluppi e rilevi l’impatto di provvedimenti innovativi e di interventi di riforma (Bolletta e Siniscalco, 2008).

È difficile che tali finalità siano immediatamente attribuite alle pratiche di *testing* su larga scala; più frequentemente, queste sono percepite come imposte dall'alto e pertanto da dover gestire (quando non è possibile evitarle del tutto), anziché “illuminating, helpful, or even essential to better learning” (Popham, 2003, p. VI).

I dati raccolti attraverso la valutazione dovrebbero servire a dare riscontro rispetto all'allineamento tra i processi di istruzione, apprendimento e valutazione. Tuttavia negli anni si è consumata una progressiva rimozione di questo tipo di valutazione dal contesto di insegnamento-apprendimento (Darling-Hammond e Adamson, 2014; Popham, 2003 e 2008; Shepard, 2000; Wyatt-Smith, 2000). Per via della loro impostazione tecnica, test e misurazioni nell'ambito dell'istruzione (e nello specifico in riferimento al livello di apprendimento degli studenti) sono stati visti come dominio esclusivo degli statistici, più che degli insegnanti e degli specialisti del settore educativo. La dimensione più tecnica della valutazione è, così, stata quasi scorporata dalla pratica degli insegnanti. La valutazione agita in classe ha finito con il differenziarsi in maniera progressiva dalle valutazioni effettuate mediante indagini su larga scala (Bottani e Checchi, 2012) tanto da far apparire le due forme di valutazione, del *classroom assessment* e del *large-scale assessment*, come inconciliabili. Due aspetti hanno, inoltre, esercitato un peso notevole in questo processo di separazione: da un lato, la diffusione del concetto di competenza e, dall'altro, il radicarsi della cultura dell'evidenza. Nel primo caso, le attività di insegnamento-apprendimento, abbandonando un'impostazione didattica di tipo tradizionale (trasmissione di nozioni e contenuti), si sono orientate, in ottica socio-costruttivista, alla promozione di studenti cosiddetti competenti. La definizione di apprendimento in termini di competenza ha investito non soltanto il processo di insegnamento in senso stretto (ciò che l'insegnante fa in classe), ma anche i processi di progettazione e valutazione; ha, inoltre, orientato lo sviluppo di pratiche didattiche in grado di sostenere e incoraggiare lo sviluppo di processi cognitivi, riflessivi, meta-cognitivi e di auto-regolazione dell'apprendimento degli studenti. Ne deriva per gli insegnanti, e siamo al secondo aspetto, l'opportunità di avvalersi dei risultati desunti dalla valutazione degli apprendimenti condotta mediante indagini su larga scala, al fine di migliorare tanto l'insegnamento, quanto l'apprendimento stesso.

Nella prospettiva dell'*evidence-based education* e abbandonando la stagione dell'improvvisazione estemporanea, dell'autoreferenzialità, degli apriorismi ideologici che spesso hanno condizionato, specialmente in Italia, la pratica educativa e didattica, i dati raccolti, anche attraverso le indagini su larga scala, si offrono come base e opportunità per assumere decisioni consapevoli e rispondenti alle esigenze degli studenti (Calvani, 2013). L'enfasi sull'uso dei dati e delle evidenze di apprendimento per informare le pratiche dell'istruzione rappresenta un trend ormai decennale (Stiggins, 1995): un orientamento che ha progressivamente acquisito visibilità tanto nell'ambito delle politiche educative, quanto nella ricerca, agganciandosi al tentativo di incrementare i processi di cambiamento nei diversi sistemi di istruzione. In tale ottica, Darling-Hammond e Adamson (2014) prospettano un nuovo approccio all'*accountability* più rispondente alle esigenze del mondo della scuola e in grado di consentire agli insegnanti, agli operatori e ai *policymakers* di migliorare e supportare i processi decisionali. Un simile cambiamento diviene possibile se integra il nuovo paradigma di *accountability* con la visione dell'apprendimento, allineandolo coerentemente ai cambiamenti sistemici implicati dall'obiettivo stesso.

Tuttavia, come gli insegnanti utilizzino i dati (tanto quelli ricavati dalla valutazione in classe, quanto quelli relativi alle indagini su larga scala) in modo da supportare e orientare la loro azione didattica rappresenta un tema di indagine ancora poco approfondito (De Luca, 2012; Brookhart, 2011). A partire da simile considerazione il presente contributo intende soffermarsi su quanto i dati delle Rilevazioni nazionali condotte dall'INVALSI siano di supporto agli insegnanti in termini di implementazione e modulazione dell'attività didattica e valutativa in classe. I dati provenienti dalle indagini su larga scala potrebbero essere utilizzati dagli insegnanti, non solo per definire e implementare percorsi di miglioramento dell'istituzione scolastica, ma anche per sostenere e accompagnare, attraverso la pratica del *formative assessment*¹, gli studenti nell'apprendimento per competenze.

Nei paragrafi seguenti, alla rassegna della letteratura sul feedback e sulle modalità più efficaci che possono consentire agli insegnanti di restituire informazioni costruttive e di supporto all'apprendimento degli studenti è affiancata un'analisi delle pratiche di valutazione degli insegnanti italiani così come emerse dalla somministrazione del Questionario insegnante.

¹ Il *formative assessment* (che può essere reso in italiano anche con *valutazione per l'apprendimento* o *valutazione a sostegno dell'apprendimento*) corrisponde a un processo attivo e intenzionale per promuovere ulteriore apprendimento negli studenti (Pastore e Heritage 2015). Tale valutazione è realmente funzionale se consente agli studenti di esercitarsi all'uso di una postura riflessiva (meta-cognizione).

2. Al cuore della valutazione: il feedback

L'intenzione di non ridurre la valutazione degli apprendimenti alla dimensione sommativa, oltre a comportare la valorizzazione del ruolo e delle finalità della valutazione stessa, investe le modalità attraverso cui poter realisticamente gestire tale processo rispetto alla pratica didattica. La valutazione non è solo finalizzata a determinare quanto gli studenti abbiano acquisito in termini di contenuto a conclusione di un'unità di apprendimento o di un segmento didattico (visione tradizionale e, in parte, strumentale della valutazione dell'apprendimento), ma dovrebbe, in un'ottica di miglioramento continuo, consentire agli insegnanti di fornire agli studenti informazioni sul loro apprendimento al fine di gestire, autonomamente, il proprio apprendimento e divenire, in tal senso, *self-directed learners* (Angelo e Cross, 1993).

La valutazione, in quest'ottica, assume un ruolo fondamentale nell'architettura del processo didattico in quanto consente di stabilire ciò che gli studenti sanno già, individuare gli elementi di criticità e lavorare su tali aspetti (Pastore, 2014; Pastore e Salamida, 2013). La corretta e chiara definizione dell'oggetto di apprendimento permette all'insegnante di monitorare il progresso degli studenti, di comprendere a che punto dell'apprendimento essi si trovino e di progettare azioni didattiche che siano rispondenti al contesto della classe. Le strategie di valutazione devono pertanto essere significative per gli studenti e situate rispetto al contesto in modo da poter garantire informazioni precise e attendibili sul livello di apprendimento maturato. Attraverso le evidenze raccolte, con modalità formali e strutturate (per es. prove scritte o interrogazioni) e modalità informali (per es. osservazioni, conversazioni in aula), l'insegnante può scandire al meglio il ritmo della sua azione didattica e rimodulare e mediare le sue interpretazioni perché siano funzionali a sostenere gli studenti nell'apprendimento.

Numerosi studi mostrano quanto la valutazione influenzi e direzioni gli studenti verso l'apprendimento dei contenuti formativi che gli insegnanti indicano essere oggetto di valutazione (Kingston e Nash, 2011; Hattie e Timperley, 2007; Black e Wiliam, 1998; Crooks, 1988; Natriello, 1987). È anche vero però che, spesso, nella pratica le cose vanno diversamente: la valutazione può non essere di sostegno per docenti e studenti, e rivelarsi irrilevante (Brown, 2006; Warren e Nisbet, 1999; Torrance e Pryor, 1998).

Diversi gli interrogativi che, a questo punto, emergono: quanto la valutazione migliora l'apprendimento degli studenti? I docenti forniscono un feedback utile, adeguato e tempestivo? Gli studenti sono in grado di riconoscere e comprendere gli elementi che possono indurre un miglioramento nella loro performance? (Brookhart e Bronowicz, 2003; Elwood e Klenowski, 2002). Talvolta la valutazione ha un'influenza più forte dello stesso insegnamento sull'apprendimento degli studenti, perché:

- orienta la selezione dei contenuti di insegnamento/apprendimento;
- ha un potente effetto su *cosa* e su *come* gli studenti apprendono;
- consolida lo sviluppo di strategie di apprendimento;
- influenza il valore che il soggetto attribuisce alla formazione, così come il senso di realizzazione personale e la volontà di portare a termine determinati compiti di apprendimento;
- contribuisce a definire cosa gli studenti associano, in generale, all'esperienza della valutazione in ambito scolastico e formativo (Price *et al.*, 2010).

Se la valutazione ha un tale impatto, meriterebbe una maggiore attenzione da parte degli stessi docenti che, spesso, prestano eccessiva attenzione alla definizione di voti e punteggi e non considerano gli effetti che simili pratiche possono provocare (abbassamento del livello di stima personale, incidenza sui livelli di performance, mancato incoraggiamento a migliorare) (Pastore, 2012, p. 64).

Gli studi attuali, specie a livello internazionale, muovono verso la revisione delle modalità tradizionali di *testing*, l'individuazione di forme alternative di valutazione, l'analisi delle rappresentazioni e delle percezioni che della valutazione hanno insegnanti e studenti (Brown, 2006) e la comprensione dell'impatto del feedback sulla qualità dei processi di insegnamento e apprendimento. È all'interno dell'ultimo orientamento che si colloca il presente contributo. L'attenzione al feedback nei processi valutativi in ambito didattico si aggancia così al dibattito sul miglioramento della qualità dell'istruzione.

L'esigenza di affiancare alla tradizionale valutazione dell'apprendimento una valutazione pensata e agita per promuovere l'apprendimento comporta una revisione del ruolo e delle pratiche dell'insegnante in classe. Predisporre, infatti, una valutazione che, attraverso feedback efficaci e tempestivi, guidi l'alunno verso un'accurata esplorazione delle proprie conoscenze e abilità rappresenta, in termini didattici, un'impresa di non poco conto.

Per quanto la letteratura ribadisca l'importanza e la centralità del *feedback* (Popham, 2011; Black e Wiliam, 2009), specie nel *formative assessment* (di quella valutazione agita nel contesto classe per supportare l'apprendimento degli studenti attraverso una serie di strategie che ruotano attorno alle dimensioni del coinvolgimento, della metacognizione e dell'auto-apprendimento), sono ancora poche le evidenze di un cambiamento reale nelle pratiche valutative degli insegnanti (Hamilton *et al.*, 2009; Miller, 2009; Herman e Gribbons, 2001).

Il feedback si presenta come fattore chiave per la promozione dell'apprendimento (Black e Wiliam, 1998). Come componente essenziale del *formative assessment*, il feedback aiuta chi apprende a essere più consapevole dei gap che esistono tra gli obiettivi stabiliti dal docente e il livello di apprendimento acquisito dagli alunni (Hattie e Timperley, 2007). Il feedback più utile è quello che fornisce commenti sugli errori e suggerimenti specifici per migliorare e incoraggiare gli studenti a guardare in modo diverso al compito. Inoltre, a seconda di come è praticato, il feedback esercita un'influenza positiva o negativa sulle emozioni, sull'autostima e sulla motivazione degli alunni, contribuendo ad accrescere la fiducia in se stessi. Per questo è importante che si riveli:

- *realistico* e, dunque, rispondente a dati e situazioni concrete ed evidenti;
- *specifico*, centrato cioè sul compito e orientato in modo diretto sugli obiettivi formativi coinvolti in un determinato iter di apprendimento;
- *tempestivo*, immediato;
- *incline* all'incoraggiamento e, pertanto, al suggerimento dei passi successivi per raggiungere gli obiettivi di apprendimento;
- *descrittivo*, in grado di esporre giudizi espliciti (e comprensibili) e di fornire informazioni chiare e dettagliate sullo stato di apprendimento raggiunto e sulle possibili strategie da adottare;
- *diretto*, volto a individuare errori specifici e un insoddisfacente utilizzo delle strategie di apprendimento e a fornire consigli puntuali su come migliorare (Pastore e Beccia, 2017, p. 41).

In quest'ottica, anche le prove INVALSI, talvolta oggetto di contestazione (Castoldi, 2014), possono essere utilizzate dagli insegnanti per restituire agli alunni un feedback tale da consentire l'individuazione del gap di apprendimento, la correzione degli eventuali errori e *misconception*, e agevolare così tanto la motivazione quanto la loro partecipazione attiva.

Differenziandosi notevolmente dalle cattive pratiche che interessano il *testing* e diffusesi anche nel nostro sistema scolastico (a cominciare dai comportamenti opportunistici come il *cheating* e il *teaching to the test*) con l'avvio di indagini sistematiche sugli apprendimenti (Freddano, 2016), si possono suggerire agli studenti i passi da compiere per migliorare il proprio lavoro, fornendo loro un prezioso aiuto e favorendo la comprensione delle modalità e delle strategie atte a colmare il divario tra il loro rendimento effettivo e i risultati di apprendimento attesi.

3. Lo studio: obiettivi, analisi e risultati

Cosa pensa l'insegnante della valutazione? Quali finalità persegue attraverso essa? Che incidenza ha la valutazione nella sua pratica didattica? A quali strategie ricorre l'insegnante e quali difficoltà incontra nel valutare? Il docente è in grado di promuovere l'apprendimento dei suoi studenti attraverso la valutazione?

Queste le domande che scandiscono la riflessione di seguito presentata. A tal fine riprenderemo alcuni risultati del Questionario insegnante rivolto ai docenti delle classi quinte di scuola primaria campionate nelle Rilevazioni nazionali dell'a.s. 2013-14 e dell'a.s. 2014-15². Complessivamente sono state raccolte le risposte degli insegnanti di italiano di 1.188 delle classi quinte di scuola primaria campionate nell'a.s. 2013-14 e di 1.079 classi quinte di scuola primaria campionate nell'a.s. 2014-15.

In ottica *evidence-based*, presentiamo i primi risultati di un progetto di ricerca più ampio teso a indagare il rapporto tra gli esiti delle prove di apprendimento, gli atteggiamenti degli studenti e le pratiche di feedback dei docenti, ovvero, a esplicitare se e quanto il feedback (valutazione agita in classe) incida sui livelli di apprendimento degli studenti.

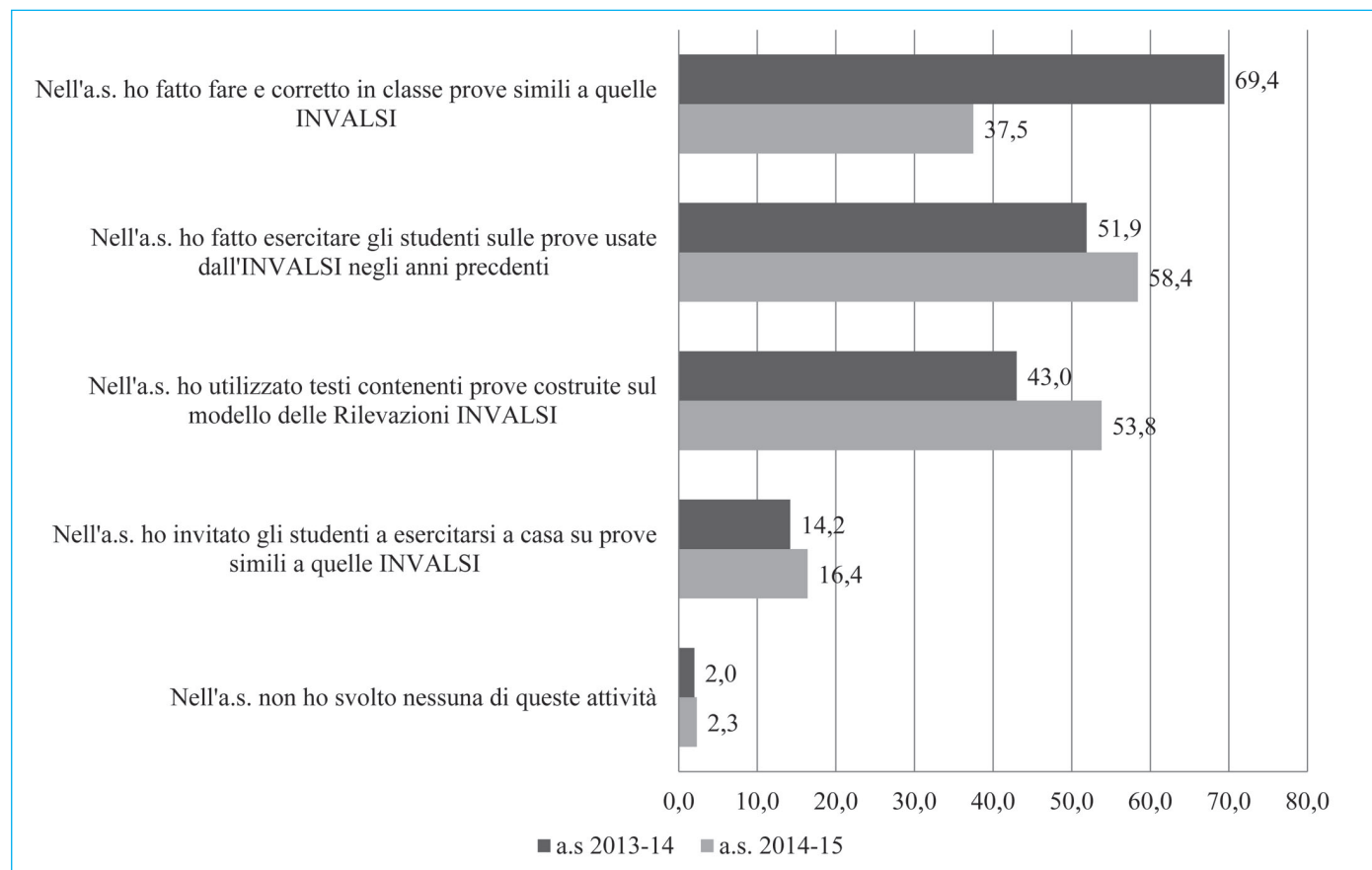
Nello specifico sono state considerate alcune domande, contenute nel Questionario insegnante, riguardanti le pratiche didattiche e valutative in classe e che fanno specifico richiamo all'uso, da parte dei docenti, di materiale concernente

² I testi del Questionario insegnante proposti nell'a.s. 2013-14 e nell'a.s. 2014-15 sono disponibili sul sito dell'INVALSI al link <http://www.invalsi.it/invalsi/ri/sis/questionstins.php>.

le Rilevazioni nazionali INVALSI, relativi dunque non soltanto ai risultati di apprendimento raggiunti dagli studenti ma anche ad altri documenti, quali per esempio le stesse prove standardizzate nazionali in qualità di strumento didattico da utilizzare in classe.

Rispetto all'uso dei testi contenenti prove costruite sul modello delle Rilevazioni INVALSI (fig. 1) emerge come gli insegnanti si siano decisamente orientati alla pratica del *teaching to the test*.

Fig. 1 – Percentuale di classi quinte di scuola primaria di cui gli insegnanti di italiano rispondono in modo affermativo alla seguente domanda: “Nel corso dell’anno scolastico, ha proposto agli studenti della classe qualcuna delle attività sottoelencate?”. Confronto tra a.s. 2013-14 e a.s. 2014-15

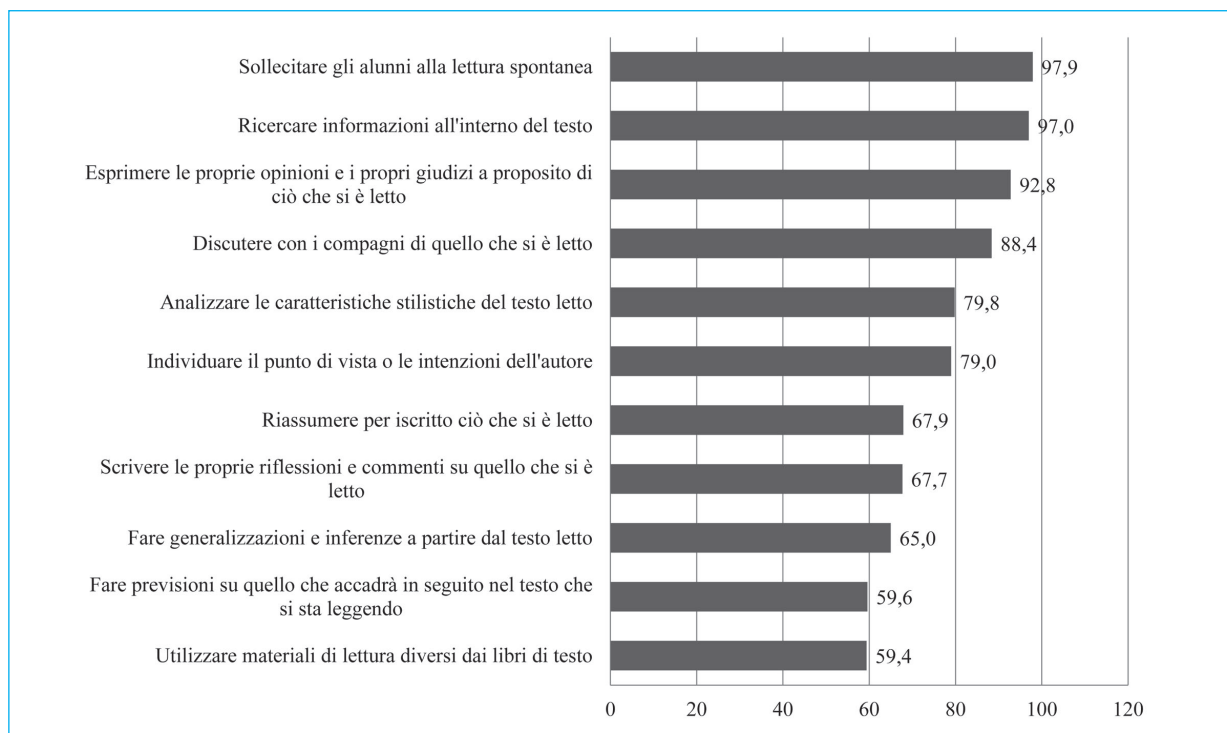


Fonte: elaborazioni su dati INVALSI Questionario insegnante di italiano delle classi quinte di scuola primaria campionate nelle Rilevazioni nazionali a.s. 2013-14 e a.s. 2014-15.

Nelle due annualità esaminate si nota che il ricorso a prove simili a quelle delle Rilevazioni nazionali è, infatti, considerevolmente diminuita da un'annualità all'altra (da un valore sostanzialmente alto di 69,4% delle classi che praticavano questa attività a un valore di 37,5%). Aumenta invece la pratica di far esercitare gli studenti sulle prove usate dall'INVALSI negli anni passati di circa 8 punti percentuali da un'annualità all'altra. Poco utilizzato l'esercizio a casa su prove simili a quelle INVALSI, mentre è pressoché nullo il fatto di non svolgere alcuna tra le attività previste.

Rispetto alle pratiche didattiche utilizzate per la promozione della competenza di italiano, gli insegnanti delle quinte classi di scuola primaria rispondenti nell'a.s. 2013-14 non solo affermano di ricorrere a una pluralità di attività legate ai testi letti, ma ritengono di farlo in maniera frequente, se non addirittura costante (fig. 2).

Fig. 2 – Percentuale di classi quinte di scuola primaria di cui i docenti di italiano rispondono “Spesso” o “Sempre” alla seguente domanda: “Con quale frequenza Le capita di far svolgere in classe agli alunni le seguenti attività in relazione ai testi letti per la scuola?” – a.s. 2013-14



Fonte: elaborazioni su dati INVALSI Questionario insegnante di italiano delle classi quinte di scuola primaria campionate nelle Rilevazioni nazionali a.s. 2013-14.

L'analisi dei dati mostra che nella maggior parte delle classi quinte di scuola primaria si svolgono le attività proposte in relazione ai testi letti; tra le attività meno svolte, si nota il fare generalizzazioni, inferenze o previsioni a partire dai testi letti (che richiede livelli di competenza superiori, e dunque una maggiore abilità linguistica). Inoltre, l'attività meno condotta in assoluto nelle classi quinte è quella di utilizzare materiali di lettura diversi dai libri di testo: nel 40% circa delle classi questa attività viene condotta soltanto occasionalmente o per nulla. Di riflesso emerge l'importanza dedicata all'uso del libro di testo, che permane come il dispositivo prediletto nelle attività didattiche, nonostante nel tempo si siano sviluppati altri canali in grado di veicolare con facilità ed efficacia la conoscenza. Questo aspetto trova conferma laddove si chiede ai docenti quali strumenti di valutazione utilizzino e con quale temporalità (tab. 1).

È evidente, infatti, come gli insegnanti si orientino prevalentemente su prove di taglio tradizionale, come, per esempio, il componimento scritto (in più dell'80% delle classi vi è un uso regolare) e le prove strutturate a risposta aperta, utilizzate regolarmente in circa il 70% delle classi.

L'attività meno utilizzata nelle classi quinte di scuola primaria è l'interrogazione orale individuale programmata; infatti, nell'a.s. 2013-14 ben nel 17,5% delle classi non viene utilizzata e nel 38,3% delle classi se ne fa un uso saltuario. Aumenta leggermente la percentuale di classi che utilizza questa tipologia di valutazione nell'a.s. 2014-15.

Nell'a.s. 2013-14, le prove strutturate a risposta chiusa così come le prove attinte da un manuale o da un libro di testo sono utilizzate in modo regolare nel 62% delle classi, nelle restanti questi dispositivi sono utilizzati saltuariamente. Nell'a.s. 2014-15, si nota che aumenta l'utilizzo delle prove strutturate a risposta chiusa, mentre diminuisce, seppur di poco, l'uso di prove attinte da manuali o libri di testo.

La consuetudine all'uso di prove strutturate a risposta chiusa trova conferma con quanto emerge dalle analisi condotte sulle informazioni pervenute con la compilazione censuaria da parte delle scuole del Questionario scuola nell'a.s. 2014-15³, nell'ambito del Sistema nazionale di valutazione.

³ Per eventuali approfondimenti rimandiamo al rapporto *I processi e il funzionamento delle scuole – Dati dal Questionario scuola INVALSI e dalle sperimentazioni VALES e VM* all'indirizzo http://www.invalsi.it/snv/docs/141016/Rapporto_Processi_2016.pdf.

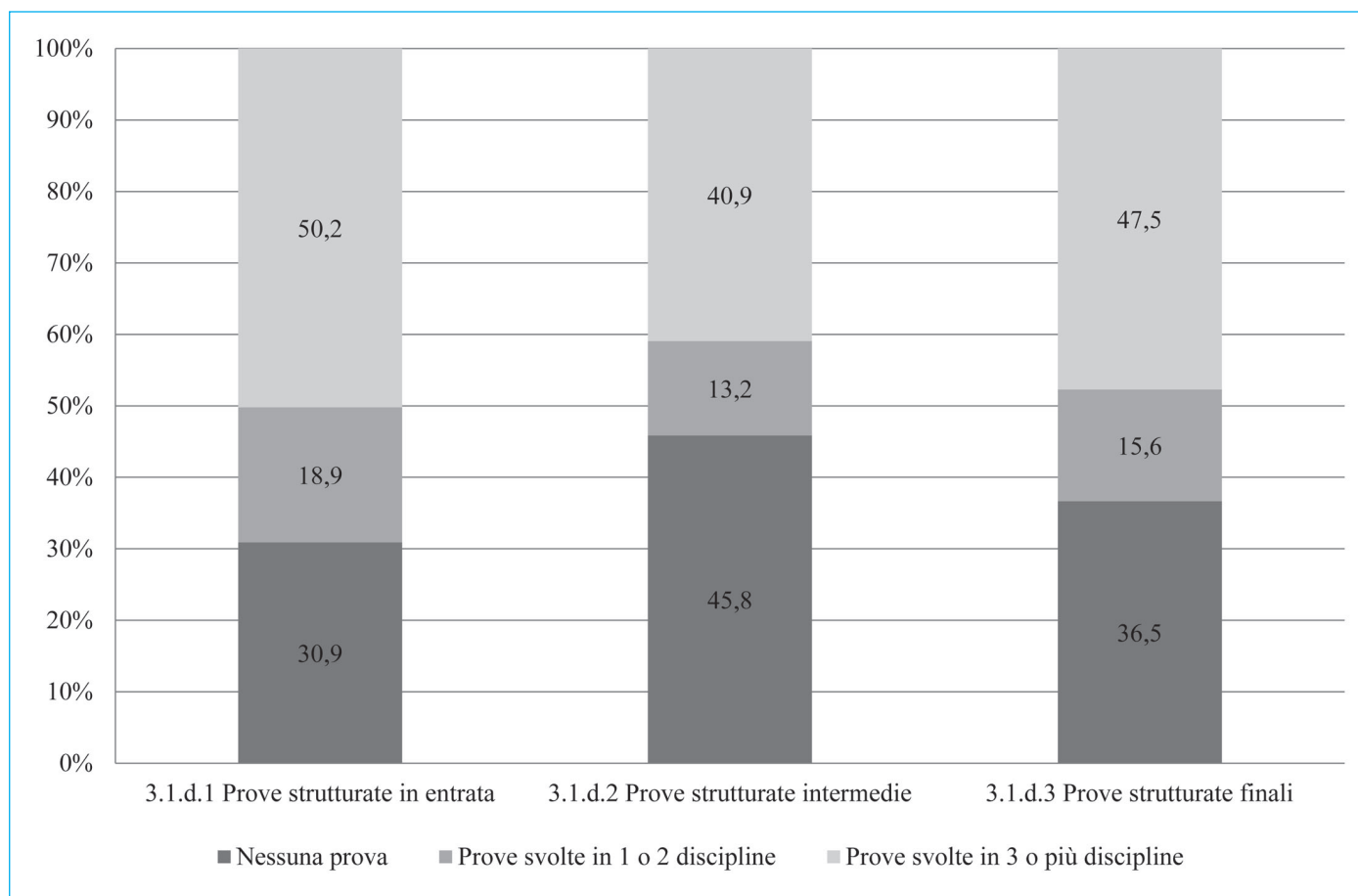
Tab. 1 – Percentuali di classi quinte di scuola primaria campione nelle Rilevazioni nazionali per uso di strumenti di valutazione

	a.s. 2013-14			a.s. 2014-15		
	Uso regolare	Uso saltuario	Non uso	Uso regolare	Uso saltuario	Non uso
Interrogazione orale individuale programmata	44,2	38,3	17,5	50,4	37,6	12,0
Interrogazione orale individuale non programmata	62,4	28,9	8,8	56,7	31,6	11,7
Interrogazione in gruppo o all'intera classe	55,1	37,7	7,2	59,5	33,3	7,2
Compito scritto con svolgimento aperto (componimento scritto, tema ecc.)	86,4	13,0	0,6	83,1	15,6	1,3
Prove strutturate... a risposta chiusa	62,0	36,0	1,9	67,3	29,9	2,8
Prove strutturate ...a risposta aperta	70,4	28,3	1,3	68,6	29,6	1,9
Prova attinta dal manuale/libro di testo (o da altra fonte)	63,0	33,0	4,0	60,7	33,0	6,3

Fonte: elaborazioni su dati INVALSI Questionario insegnante di italiano delle classi quinte di scuola primaria campionate nelle Rilevazioni nazionali a.s. 2013-14 e a.s. 2014-15.

La fig. 3 mostra la percentuale di scuole primarie che nel Questionario scuola dichiara di utilizzare prove strutturate in ingresso, intermedie e finali e per quante discipline, confermando che più della metà delle scuole è solita svolgere prove strutturate, con un'incidenza maggiore per quanto riguarda le prove in ingresso e le prove in uscita, per le quali circa il 47-50% delle scuole primarie propone prove strutturate in tre o più discipline, seguito da una più contenuta percentuale di scuole che svolge questo tipo di attività al più per due discipline. Le prove meno comuni sono quelle intermedie (Pandolfini, Freddano e Siri, 2014; Freddano e Siri, 2012) che non sono utilizzate da circa il 46% delle scuole primarie.

Fig. 3 – Percentuale di scuole primarie che dichiara di utilizzare prove strutturate in ingresso, intermedie e finali per numerosità di discipline

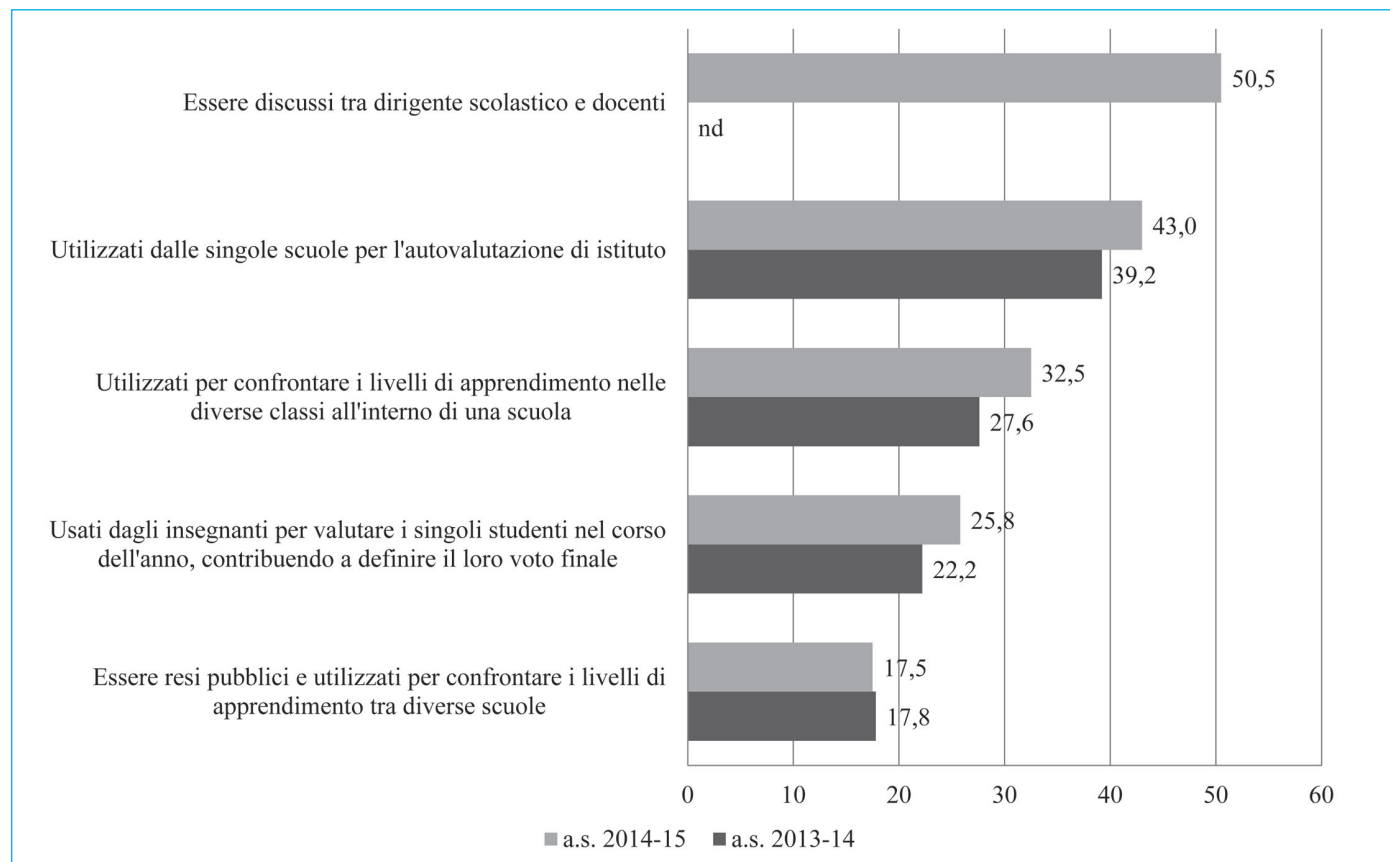


Fonte: elaborazioni su dati Questionario scuola a.s. 2014-15 – scuola primaria.

Rispetto all'uso dei risultati delle prove INVALSI, sono diverse le attività promosse al fine di diffondere un uso consapevole e contestualizzato dei risultati delle prove in un'ottica di miglioramento dei processi educativi e didattici all'interno delle singole scuole (Barone e Serpieri, 2016; Landri e Maccarini, 2016; Agasisti, Falzetti e Freddano, 2015; Faggioli, 2014; Palumbo, 2014).

Dall'analisi dei dati del Questionario insegnante rivolto agli insegnanti di scuola primaria delle classi quinte campione (fig. 4), la metà dei rispondenti ritiene che i risultati delle prove INVALSI dovrebbero essere discussi tra dirigente scolastico e docenti⁴.

Fig. 4 – Percentuale di classi quinte di scuola primaria di cui gli insegnanti di italiano rispondono “Abbastanza” o “Molto” alla seguente domanda sui risultati delle prove INVALSI: “Dica ora quanto è d'accordo con le seguenti affermazioni. I risultati delle prove INVALSI dovrebbero essere...”



Fonte: elaborazioni su dati INVALSI Questionario insegnante di italiano delle classi quinte di scuola primaria campionate nelle Rilevazioni nazionali a.s. 2013-14 e a.s. 2014-15.

Complessivamente, i docenti mostrano una percezione critica sull'uso dei dati INVALSI, seppur nel confronto tra le due annualità considerate aumenti leggermente l'atteggiamento positivo nei confronti dell'uso di questi dati. Nello specifico, si mantiene costante e molto contenuta la percentuale di insegnanti che pensa che i risultati delle prove INVALSI debbano essere resi pubblici e utilizzati per confrontare i livelli di apprendimento tra le diverse scuole (circa il 17,5%). Sempre contenuta, ma in lieve aumento (di circa 3 punti percentuali), la percentuale dei docenti che pensa che i risultati delle prove INVALSI debbano essere utilizzati per valutare i singoli studenti e considerati nel calcolo del voto conclusivo a fine anno.

Sull'utilità percepita per la valutazione a livello di scuola, aumenta di 4 punti la percentuale di docenti che dichiara che i risultati delle prove INVALSI dovrebbero essere utilizzati per confrontare i livelli di apprendimento nelle diverse classi all'interno di una stessa scuola.

⁴ Il dato non è disponibile per l'a.s. 2013-14, perché l'item è stato aggiunto nel Questionario insegnante 2014-15.

Si distingue, tra le altre, la percentuale di docenti favorevoli all'uso dei dati INVALSI per l'autovalutazione di istituto, aspetto in linea con le sperimentazioni sul tema e l'avvio del Sistema nazionale di valutazione (DPR 80/2013), ove una delle aree della dimensione degli Esiti è dedicata proprio ai risultati della scuola nelle prove standardizzate nazionali (INVALSI, 2014b). La messa a sistema dell'autovalutazione delle istituzioni scolastiche ha favorito l'evolversi di un linguaggio valutativo specifico e diffuso tra le scuole una maggiore sensibilità all'uso dei dati INVALSI, non soltanto per scopi autovalutativi, ma anche per pianificare le proprie attività in modo più consapevole perché basate sulle evidenze (Alemanno e Freddano, 2015).

La tab. 2 mostra le percentuali di risposta degli insegnanti di italiano delle classi quinte di scuola primaria campionate nelle due annualità esaminate, rispetto all'utilizzo dei risultati delle prove INVALSI.

Tab. 2 – Percentuali di risposta alla domanda relativa all'uso dei risultati delle prove INVALSI

Nella scuola i risultati delle prove INVALSI sono...	a.s. 2013-14			a.s. 2014-15		
	Sì	No	Non so	Sì	No	Non so
Discussi dal dirigente con i singoli insegnanti	28,0	49,2	22,9	24,0	51,7	24,4
Comunicati nel collegio dei docenti	73,6	19,2	7,2	67,8	24,5	7,8
Commentati in una commissione apposita	27,4	53,1	19,4	26,4	53,3	20,3
Discussi fra colleghi della stessa disciplina	58,2	28,3	13,4	50,2	34,1	15,7
Presentati ai genitori durante le assemblee di classe	19,4	66,2	14,4	14,7	69,7	15,6
Presentati ad altri soggetti esterni alla scuola	4,9	73,2	21,9	12,7	53,9	33,4

Fonte: elaborazioni su dati INVALSI Questionario insegnante di italiano delle classi quinte di scuola primaria campionate nelle Rilevazioni nazionali nell'a.s. 2013-14 e nell'a.s. 2014-15.

Con riferimento, infine, al modo in cui i risultati INVALSI sono effettivamente utilizzati e pubblicizzati dalle scuole, prevale la condivisione nell'ambito del collegio dei docenti e con i colleghi della stessa disciplina per più della metà dei docenti in entrambe le annualità; tuttavia, tra un anno scolastico e l'altro, diminuisce rispettivamente di 6 e 8 punti la percentuale di docenti che affermano che i risultati INVALSI siano comunicati nel collegio dei docenti e con i colleghi della stessa disciplina. Gli insegnanti dichiarano che i risultati delle prove INVALSI non sono comunicati ai genitori durante le assemblee di classe: sarebbe interessante approfondire questo aspetto per vedere se sono predisposte occasioni *ad hoc* anche perché il dato si conferma e cresce nell'anno scolastico successivo. Inoltre aumenta la percentuale di docenti che non sa se i risultati delle prove INVALSI siano presentati o meno ad altri soggetti esterni alla scuola.

Questo aspetto meriterebbe un approfondimento per tre ordini di ragioni: anzitutto perché nei genitori è aumentato, nel tempo, l'interesse nei confronti delle prove standardizzate nazionali, interrogandosi, da una parte, sulla pertinenza o meno della loro somministrazione e, dall'altra, sull'accessibilità, interpretabilità e utilità dei risultati di apprendimento; perché i risultati delle prove INVALSI sono pubblicamente accessibili attraverso il sito "Scuola in chiaro" e il Rapporto di autovalutazione delle scuole; infine, perché la quarta fase del Sistema nazionale di valutazione prevede a ogni modo la rendicontazione sociale delle scuole e, quindi, una maggiore interlocuzione anche con gli *stakeholders* esterni alla scuola⁵.

4. Conclusioni

Il quadro che emerge dall'analisi dei dati raccolti non è particolarmente confortante se lo si proietta sullo sfondo del Sistema nazionale di valutazione, per diverse ragioni, tra cui:

- *in primis*, la pratica valutativa degli insegnanti realizzata in classe è ancora di tipo tradizionale e poco risponde ai cambiamenti e alle istanze della scuola e della società attuali;
- le trasformazioni del sistema educativo e le innovazioni nelle politiche scolastiche nazionali incidono, in minima parte, sulle azioni realizzate al livello micro della classe;
- permane un forte disallineamento tra le forme della valutazione esterna e della valutazione interna.

⁵ Per approfondimenti cfr. DPR 80/2013 e L. 107/2015.

Il consenso un po' *naïf* per le rilevazioni su vasta scala ha accompagnato l'investimento per lo sviluppo di sistemi validi e affidabili di valutazione; si è però venuta a creare una sorta di frattura e di inconciliabilità tra le forme del *classroom assessment* e del *large-scale assessment*.

La valutazione costituisce per l'insegnante un passaggio fondamentale perché consente di poter disporre di informazioni funzionali alla presa di decisioni e tali da guidare il processo didattico nella quotidianità. Come afferma D. Wiliam (2010), la valutazione dell'apprendimento può essere realizzata per finalità differenti: di diagnosi; di verifica; di accertamento; di sostegno e promozione di ulteriore apprendimento. A seconda delle finalità perseguite, la valutazione offre informazioni "prospettiche" che dovrebbero poi essere armonizzate in una rappresentazione chiara dei livelli di conoscenza e competenza raggiunti dallo studente durante un percorso formativo.

Le valutazioni in classe e quelle su vasta scala, come le Rilevazioni INVALSI, non dovrebbero escludersi a vicenda. Dovrebbero, piuttosto, per quanto possibile, essere integrate in un quadro coerente, coeso e continuo di valutazione.

Riferimenti bibliografici

- Agasisti T., Falzetti P., Freddano M. (2015), "L'uso dei risultati delle Rilevazioni nazionali per l'autovalutazione delle scuole", *Rassegna italiana di valutazione*, 61: 28-48.
- Alemanò I., Freddano M. (2015), "Autovalutazione e rendicontazione sociale per migliorare la scuola: un'esperienza di formazione e di empowerment", *Rassegna italiana di valutazione*, 61: 118-140.
- Allulli G., Farinelli F., Petrolino A. (2013), *L'autovalutazione di istituto. Modelli e strumenti operativi*, Guerini, Milano.
- Angelo T.A., Cross P.K. (1993), *Classroom Assessment Techniques*, Jossey-Bass Publishers, San Francisco.
- Barone C., Serpieri R. (a cura di) (2016), *Special Issue: Valutazione e miglioramento nei processi educativi: Valutazione e miglioramento nei processi educativi*, numero monografico di *Scuola democratica*, 2.
- Black P.J., Wiliam D. (1998), "Assessment and classroom learning", *Assessment in Education: Principles, Policy and Practice*, 5, 1: 7-74.
- Black P.J., Wiliam D. (2009), "Developing the theory of formative assessment", *Educational Assessment, Evaluation and Accountability*, 21, 1: 5-31.
- Bolletta S., Siniscalco M. (2008), *Le valutazioni internazionali e la scuola italiana*, Zanichelli, Bologna.
- Bottani N. (2013), *Requiem per la scuola. Ripensare il futuro dell'istruzione*, il Mulino, Bologna.
- Bottani N., Checchi D. (a cura di) (2012), *La sfida della valutazione*, il Mulino, Bologna.
- Brookhart S.M. (2011), "Educational assessment knowledge and skills for teachers", *Educational Measurement: Issues and Practices*, 30, 1: 3-12.
- Brookhart S.M., Bronowicz D.L. (2003), "I don't like writing. It makes my fingers hurt: Students talk about their classroom assessments", *Assessment in Education*, 10, 2: 221-242.
- Brown G.T.L. (2006), "Integrating teachers' conceptions: assessment, teaching, learning, curriculum and efficacy", in M.B. Klein (ed.), *New Teaching and Teacher Issues*, Nova Science Publishers, New York: 1-49.
- Calvani A. (2013), "Evidence Based (Informed?) Education: neopositivismo ingenuo o opportunità epistemologica?", *Form@re. Open Journal per la formazione in rete*, 13, 2: 91-101.
- Castoldi M. (2014), *Capire le prove INVALSI*, Carocci, Roma.
- Crooks T.J. (1988), "The impact of classroom evaluation practices on students", *Review of Educational Research*, 58, 4: 438-481.
- Darling-Hammond L., Adamson F. (2014), *Beyond the Bubble Test: How Performance Assessments Support 21st Century Learning*, Jossey-Bass & Pfeiffer Imprints, Wiley, San Francisco.
- De Luca C. (2012), "Preparing teachers for the age of accountability: toward a framework for assessment education", *Action in Teacher Education*, special issue: *Teacher Education Yearbook XXI (Part I): Issues in Education: Examining the Perspectives of Students, Candidates, Teachers, and Teacher Educators*, 34, 5-6: 576-591.
- Elwood J., Klenowski V. (2002), "Creating communities of shared practice: the challenges of assessment use in learning and teaching", *Assessment & Evaluation in Higher Education*, 27, 3: 243-256.
- Faggioli M. (a cura di) (2014), *Migliorare la scuola. Autovalutazione, valutazione e miglioramento per lo sviluppo della qualità*, Edizioni Junior, Bergamo.
- Freddano M. (2016), "La valutazione di sistema sugli apprendimenti come misura del curriculum appreso", in P. Landri, A. Maccarini (a cura di), *Uno specchio per la valutazione della scuola. Paradossi, controversie, vie d'uscita*, FrancoAngeli, Milano: 135-156.
- Freddano M., Siri A. (2012), "Teacher training for school self-evaluation", *Procedia – Social and Behavioral Sciences*, 69: 1142-1149.
- Hamilton L., Halverson R., Jackson S.S., Mandinach E., Supovitz J.A., Wayman J.C. (2009), *Using Student Achievement Data to Support Instructional Decision Making*, (NCEE 2009-4067), National Center for Education Evaluation and Regional Assistan-

-
- ce, Institute of Education Sciences, U.S. Department of Education, Washington, DC, testo disponibile al sito: https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/dddm_pg_092909.pdf, data di consultazione: 13 giugno 2017.
- Hattie J., Timperley E. (2007), "The power of feedback", *Review of Educational Research*, 77, 1: 81-112.
- Herman J., Gribbons B. (2001), *Lessons Learned in Using Data to Support School Inquiry and Continuous Improvement: Final Report to the Stuart Foundation* (CSE Tech. Rep. 535), Center for the Study of Evaluation (CSE), University of California, Los Angeles, testo disponibile al sito: <http://cresst.org/wp-content/uploads/TR535.pdf>, data di consultazione: 13 giugno 2017.
- INVALSI (2014a), *Nota metodologica. La validazione del Questionario insegnante*, testo disponibile al sito: http://www.invalsi.it/invalsi/ri/sis/Documenti/quest_insegnanti/12-13/Nota_metodologica_Q-Insegnante_2012-13.pdf, data di consultazione: 13 giugno 2017.
- INVALSI (2014b), *Mappa indicatori per Rapporto di autovalutazione*, testo disponibile al sito: http://www.istruzione.it/snv/allegati/Indicatori_24_11_2014_DEF.pdf, data di consultazione: 13 giugno 2017.
- INVALSI (2016), *I processi e il funzionamento delle scuole. Dati dal Questionario scuola INVALSI e dalle sperimentazioni VALES e VM*, testo disponibile al sito: http://www.invalsi.it/snv/docs/141016/Rapporto_Processi_2016.pdf, data di consultazione: 13 giugno 2017.
- Kingston N., Nash B. (2011), "Formative assessment: A meta-analysis and a call for research", *Educational Measurement: Issues and Practice*, 30, 4: 28-37.
- Landri P., Maccarini A. (a cura di) (2016), *Uno specchio per la valutazione della scuola. Paradossi, controversie, vie d'uscita*, FrancoAngeli, Milano.
- Laveault D. (2016), "Assessment policy enactment in education systems: a few reasons to be optimistic", in D. Laveault, L. Allal (eds.), *Assessment for Learning: Meeting the Challenge of Implementation*, Springer, Dordrecht.
- Miller M. (2009), *Achieving a Wealth of Riches: Delivering on the Promise of Data to Transform Teaching and Learning*, Alliance for Excellent Education, Washington (DC), testo disponibile al sito: <http://all4ed.org/wp-content/uploads/AchievingWealthOfRiches.pdf>, data di consultazione: 13 giugno 2017.
- Natriello G. (1987), "The impact of evaluation processes on students", *Educational Psychologist*, 22, 2: 155-175.
- Palumbo M. (2014), "Come utilizzare i dati della valutazione esterna e dell'autovalutazione per compiere scelte di miglioramento", in M. Faggioli (a cura di), *Migliorare la scuola. Autovalutazione, valutazione e miglioramento per lo sviluppo della qualità*, Edizioni Junior, Bergamo.
- Pandolfini V., Freddano M., Siri A. (2014), "I dati di sistema e gli esiti della valutazione esterna", in M. Faggioli (a cura di), *Migliorare la scuola. Autovalutazione, valutazione e miglioramento per lo sviluppo della qualità*, Edizioni Junior, Bergamo.
- Pastore S. (2012), "Silent assessment. Cosa pensano della valutazione gli studenti universitari", *Giornale italiano della ricerca educativa*, numero speciale: 62-73.
- Pastore S. (2014), "Formative assessment, mediazione didattica e regolazione dell'apprendimento", *Formazione e insegnamento. Rivista internazionale di Scienze dell'educazione e della formazione*, 12, 3: 153-164.
- Pastore S. (2015), "Valutare il sistema scuola", in G. Moro, S. Pastore, A.F. Scardigno, *La valutazione del sistema scuola. Contesti, logiche, modelli e principi operativi*, Mondadori, Milano: 37-74.
- Pastore S., Beccia V. (2017), *Valutazione per l'apprendimento. Guida didattica a fumetti*, DeAgostini, Milano.
- Pastore S., Heritage M. (2015), "Cambiare il sistema scuola a partire dalla valutazione in classe. Spunti per un formative assessment integrato", *Rassegna italiana di valutazione*, 61: 102-117.
- Pastore S., Salamida D. (2013), *Oltre il mito educativo? Formative assessment e pratica didattica*, FrancoAngeli, Milano.
- Popham W.J. (2003), *Test better, Teach better: The Instructional Role of Assessment*, ASCD, Alexandria (VA).
- Popham W.J. (2008), *Transformative Assessment*, ASCD, Alexandria (VA).
- Popham W.J. (2011), "Formative assessment – a process, not a test", *Education Week*, 30, 21: 35, testo disponibile al sito: <http://www.edweek.org/ew/articles/2011/02/23/21popham.h30.html>, data di consultazione: 14 giugno 2017.
- Price M., Handley K., Millar J., O'Donovan B. (2010), "Feedback: all that effort, but what is the effect?", *Assessment & Evaluation in Higher Education*, 35, 3: 277-289.
- Shepard L.A. (2000), "The role of assessment in a learning culture", *Educational Researcher*, 29, 7: 4-14.
- Stiggins R.J. (1995), "Assessment literacy for the 21st century", *Phi Delta Kappan*, 77, 3: 238-245.
- Torrance H., Pryor J. (1998), *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom*, Open University Press, Buckingham.
- Warren E., Nisbet S. (1999), "The relationship between the purported use of assessment techniques and beliefs about the uses of assessment", in J.M. Truran, K.M. Truran (eds.), *Making the Difference. Proceedings of the Twenty-second Annual Conference of the Mathematics Education Research Group of Australasia Incorporated, held at Adelaide, South Australia, 4-7 July, 1999*, MERGA, Sydney, New South Wales, Australia: 515-521.
- William D. (2010), "Standardized testing and school accountability", *Educational Psychologist*, 45, 2: 107-122.
- Wyatt-Smith C.M. (2000), "Exploring the relationship between large-scale literacy testing programs and classroom-based assessment: a focus on teachers' accounts", *Australian Journal of Language and Literacy*, 23, 2: 109-127.
-

9. Uno strumento per analizzare l'impatto di una variazione nella formulazione di un quesito INVALSI di matematica

A tool for analyzing the impact of a variation in the formulation of an INVALSI question in Mathematics

di Rebecca Boninsegna, Giorgio Bolondi, Laura Branchetti, Chiara Giberti, Alice Lemmo

In questo capitolo viene presentata una nuova metodologia che permette di misurare e analizzare l'impatto di una variazione nella formulazione di un quesito di matematica sulle risposte degli studenti. Esistono numerose ricerche che studiano in che modo la formulazione di un quesito possa influenzare le risposte degli studenti ma risulta molto complesso analizzare l'impatto di una singola variazione nella formulazione perché non è possibile somministrare a uno stesso studente due quesiti molto simili senza che si condizionino a vicenda. Lo strumento statistico presentato permette di superare questo ostacolo attraverso l'uso di prove standardizzate analizzate attraverso il modello di Rasch e i principali indici statistici. In particolare, nel capitolo viene descritto lo strumento statistico utilizzato e il relativo piano di validazione, basato su uno studio condotto su circa 800 studenti a partire da una prova INVALSI di livello 6 somministrata nell'anno 2012-13. Infine viene analizzato un quesito tratto dallo studio citato per mettere in luce le potenzialità della metodologia non solo per analizzare l'impatto di una variazione in termini di performance ma anche per trarre informazioni di natura didattica attraverso un approccio qualitativo.

In this paper we present a new methodology that allows to measure and analyse the impact of a variation in the formulation of a math question on students' responses. There are many researches on how the formulation of a question influences students' performances in solving a task; analysing the impact of a single variation in the formulation of a task is very complex in terms of students' resolution processes because it is not possible to administer two similar tasks to the same student without them affecting each other. The statistical tool presented allows to overcome this obstacle by means of standardized tests analysed through the Rasch model and the main statistical indices. Specifically, the article describes the statistical tool used and its validation plan, based on a study that involves about 800 students starting from an INVALSI test for grade 6 administered in 2012-13. Finally, we analyse an example of a task to highlight the potential of the methodology that we present. Such analysis is presented not only to analyse the impact of a variation in students' performance but also to obtain educational information through a qualitative approach.

1. Introduzione

Questo lavoro riporta la descrizione della parte metodologica di una ricerca più ampia presentata da tre degli autori (Branchetti, Giberti e Bolondi) per la discussione nel Topic Study Group 52 della tredicesima edizione dell'International Congress on Mathematical Education, svoltosi ad Amburgo dal 24 al 31 luglio 2016, con l'analisi dettagliata di un caso di particolare interesse. A partire da una ricerca precedente (Branchetti e Viale, 2015), prevalentemente qualitativa, sono stati sviluppati metodi di analisi quantitativa per analizzare l'impatto delle variazioni di formulazione del testo di un problema di matematica sulle performance di studenti di scuola secondaria di I grado. Si è scelto di lavorare sul testo di una prova INVALSI per due principali motivi: 1) la possibilità di fare un confronto tra le performance di studenti di una stessa classe su una prova originale – di cui si conoscono le caratteristiche statistiche rilevate su un campione nazionale molto numeroso e significativo – e una prova variata che ha una consistente base comune con la prova originale; 2) la qualità delle domande, già testate dall'INVALSI prima di essere proposte agli studenti e note dal punto di vista delle

caratteristiche fondamentali (*question intent*, analisi a priori delle opzioni di risposta nelle domande a risposta multipla e difficoltà relativa nella prova, misurata dal modello di Rasch).

2. Presentazione del problema

Il problema della formulazione dei quesiti di matematica ha sempre suscitato molto interesse nella ricerca in Didattica. Da molti anni, diversi studi hanno mostrato che i comportamenti e di conseguenza le prestazioni degli studenti coinvolti in una particolare attività matematica sono influenzati dalla formulazione della consegna. In particolare, Mayer (1982) e successivamente De Corte e Verschaffel (1985) hanno osservato che una parte delle difficoltà che gli studenti incontrano nel processo di problem solving è causata da un'interpretazione errata del testo del problema. Questo tema diventa molto rilevante quando gli allievi affrontano i quesiti di un test standardizzato, che non sono prodotti dal docente della classe. Solitamente nei test, specialmente quelli standardizzati, conoscenze e abilità degli studenti sono valutati tramite quesiti costituiti da uno stimolo iniziale (generalmente presentato in forma scritta) seguito da un certo numero di domande. Questa caratteristica dei quesiti li rende paragonabili a quelli che in letteratura vengono chiamati *word problems* (*problemi verbali*). In generale, un problema verbale di matematica viene definito come un esercizio in cui le informazioni sono presentate all'interno di una situazione descritta attraverso una forma verbale, arricchita eventualmente da immagini, tabelle o grafici.

Diversi autori si sono occupati della formulazione dei problemi verbali; in particolare Nesher (1982) ha analizzato alcuni dei fattori che potrebbero influenzare l'attività di risoluzione. Nello specifico, l'autrice elenca tre componenti che possono variare all'interno di un problema verbale: logica (operazioni, la mancanza o sovrabbondanza di dati ecc.), sintattica (posizione della domanda nel testo, numero di parole ecc.) e semantica (relazioni contestuali, suggerimenti impliciti ecc.). Recentemente, Daróczy, Wolska, Meurerse e Nuerk (2015) hanno proposto una panoramica dei fattori che influenzano la difficoltà dei problemi verbali, distinguendo tra tre componenti di difficoltà: la complessità linguistica del testo, la complessità numerica del problema aritmetico, la relazione tra la complessità linguistica e quella numerica. Alla luce di ciò, in accordo con D'Amore (2014), è ragionevole pensare che le modifiche nella formulazione di un testo, anche le più piccole, possano provocare dei cambiamenti nelle strategie risolutive che gli studenti mettono in atto per giungere alla soluzione di un problema. Duval (1991) ha definito queste modifiche nella formulazione usando il termine "variabili redazionali", che successivamente Laborde ha ridefinito al fine di includere anche variazioni di tipo non verbale, come l'introduzione di immagini e disegni (Laborde, 1995).

A fronte di questa abbondante letteratura, va peraltro osservato che, negli studi citati, l'effetto delle variazioni è stato studiato prevalentemente da un punto di vista qualitativo, e con impianti sperimentali che prevedevano fondamentalmente l'interazione del ricercatore con piccoli gruppi di studenti.

Non è facile indagare quantitativamente l'effetto che le variazioni hanno sulle prestazioni degli studenti poiché è difficile, se non impossibile, realizzare la situazione di osservazione ottimale, in cui uno stesso studente, a distanza di pochi minuti di tempo, risponde a due domande molto simili, senza che la risposta fornita alla prima interferisca e influenzi la risoluzione dell'altra. In questa situazione sarebbe necessario far "dimenticare" allo studente di aver affrontato la prima domanda rispondendo alla seconda, oppure il cambiamento dovrebbe essere così evidente da trasformare profondamente la natura stessa del quesito. Un eventuale studio qualitativo a posteriori, condotto mediante una discussione in aula e riguardante le strategie utilizzate dagli studenti nella risoluzione di un quesito, potrebbe suggerire interpretazioni a posteriori delle difficoltà incontrate in due quesiti simili ma diversi dal punto di vista della formulazione, ma risulta comunque complicato superare l'ostacolo dell'influenza reciproca tra i due quesiti (quello originale e quello variato).

Lo scopo di questa ricerca è quello di indagare sperimentalmente i seguenti problemi: in che modo le variazioni di formulazione del testo di un quesito influenzano le risposte degli studenti, o di particolari gruppi di studenti? Una variazione nella formulazione di un quesito può generare una distribuzione di risposte significativamente diversa?

Docenti e ricercatori coinvolti nella produzione e nell'analisi dei test standardizzati sono particolarmente interessati a questi aspetti, che sono cruciali nel momento in cui bisogna scegliere, tra diverse formulazioni di uno stesso quesito, quale somministrare. In questo caso, la metodologia qualitativa basata su un approccio interattivo non può essere considerata adeguata.

Branchetti e Viale (2015) hanno proposto una metodologia basata sull'IRT (*Item Response Theory*) e sul modello di Rasch (1960) per studiare tale problema. Gli autori hanno condotto uno studio pilota su una popolazione di circa 200 studenti di scuola secondaria di I grado (livello 6 e 7), nel quale sono stati indagati gli effetti delle variazioni linguistiche, soprattutto sintattiche, apportate ad alcuni quesiti della prova di matematica somministrata dall'INVALSI nell'anno 2009-10. Gli autori hanno confrontato le risposte degli studenti raccolte nella loro popolazione con le distribuzioni delle risposte ai quesiti originali presentati nel test del 2009-10 e hanno confrontato il punteggio di Rasch di questi ultimi quesiti, che sono stati variati, con il punteggio ottenuto dagli studenti della popolazione nella parte di test non variata; hanno poi analizzato i dati di risposta per iniziare a individuare quale percentuale di studenti era stata potenzialmente influenzata dalla variazione. In seguito hanno eseguito un'analisi qualitativa dei risultati, senza il supporto di alcun software. Lo studio presentato in questo report intende migliorare la metodologia qui descritta facendo uso di tecniche statistiche più sofisticate e validandola su una popolazione più ampia.

Alla luce di questo quadro, le nostre domande di ricerca sono le seguenti:

- Come si può misurare l'impatto di una variazione nella formulazione di un quesito sulle distribuzioni di frequenza di risposte di studenti classificati in base a caratteristiche potenzialmente rilevanti (abilità relativa manifestata nel test, appartenenza di genere ecc.)?
- Una tipologia di variazione di formulazione di un quesito (sintattica, semantica, di editing grafico) può causare cambiamenti significativi nelle distribuzioni di risposte di una popolazione analizzata o di un particolare gruppo di studenti?

Presentiamo qui la metodologia di ricerca e un esempio di analisi di una domanda variata (una variazione di tipo numerico: ordine di grandezza e tipo di numero) inserito nel quadro di una ricerca più ampia in cui abbiamo analizzato gli effetti di diversi tipi di variazione su 777 studenti.

3. Lo strumento statistico

I risultati delle indagini nazionali e internazionali, come per esempio le prove INVALSI e OCSE-PISA, vengono spesso analizzati facendo uso del modello di Rasch; tale modello si rivela particolarmente utile quando è necessario un confronto tra due diversi test o il confronto tra gruppi di studenti (Barbaranelli e Natali, 2005; INVALSI, 2013; OECD, 2013). Si tratta di un modello logistico a un parametro che appartiene alla categoria dell'*Item Response Theory* (IRT) e opera una stima congiunta di due tipologie di parametri: un parametro di difficoltà per ogni domanda del test e un parametro d'abilità per ogni studente. In particolare, il modello di Rasch consente di esprimere la probabilità di scegliere la risposta corretta in un item in funzione della difficoltà dell'item stesso e dell'abilità dello studente misurata sull'intera prova. La relazione tra l'abilità degli studenti sull'intero test e la probabilità di rispondere correttamente a un item è rappresentata da una curva chiamata *curva caratteristica dell'item* (ICC). In modo analogo è possibile utilizzare i parametri dell'output di Rasch per rappresentare i dati empirici e, in particolare, l'andamento di ciascuna delle alternative di risposta in funzione dell'abilità degli studenti. Questi specifici grafici, chiamati *distractor plots*, consentono di analizzare come gli studenti hanno risposto a una domanda in base al loro livello di abilità ottenuto sull'intero test, tenendo conto anche dell'andamento delle risposte sbagliate.

Le informazioni ricavate dall'uso del modello di Rasch sono significative e predittive, nel caso di nuove somministrazioni del medesimo test, a condizione che la numerosità del campione di studenti sia sufficientemente alta e siano rispettati i valori di alcuni indici statistici (p-value, *alpha* di Cronbach e altri). La possibilità di avere informazioni predittive rispetto all'andamento di un item all'interno di un test risulta essere preziosa, in quanto queste informazioni possono essere usate come indicatori di quale sarà la performance degli studenti ancora prima della somministrazione del test.

Il modello di Rasch sarà quindi il principale strumento per rispondere alla nostra prima domanda di ricerca: in che modo è possibile valutare l'impatto di una variazione nella formulazione di un item sulle prestazioni degli studenti?

La procedura che ci proponiamo di esporre e validare è la seguente. Partiamo da un test (T) composto da N domande già sottoposto a un campione di studenti. Nel nostro caso, questo campione è composto da circa 27.000 studenti che nel 2013 hanno svolto la prova INVALSI di livello 6 ed è quindi rappresentativo della popolazione degli studenti italiani frequentanti la classe prima della scuola secondaria di I grado in quell'anno. La robustezza del campione nazionale e

le analisi statistiche effettuate dall'INVALSI su questi dati permettono quindi di partire da un test che mostra ottime caratteristiche misuratorie, sia in termini di singoli item sia in termini globali.

Di questo test (T), abbiamo individuato un *core test* (CT) composto da N-m item che costituirà la parte del test che rimane invariata. Il *core test* deve essere tale da fornire una stima statisticamente robusta dell'abilità degli studenti sull'intero test; in questo modo è possibile applicare il modello di Rasch al *core test* e assegnare un livello di abilità a ogni studente a partire da questa parte invariata della prova.

Indichiamo quindi con A_1, A_2, \dots, A_m gli item rimanenti, non facenti parte del CT e che costituiranno l'oggetto del nostro studio. Abbiamo quindi modificato ognuno degli item A_1, A_2, \dots, A_m , effettuando su ciascuno di essi una singola variazione ben definita e ottenendo così nuovo set di item A'_1, A'_2, \dots, A'_m che, unito con gli N-m item del CT, costituiscono un nuovo test T'.

In questo modo sono stati creati due test T e T' con una parte consistente di item in comune (CT) e un set di m item differenti. In particolare, nel test T sono presentati A_1, \dots, A_m senza alcuna variazione rispetto al test INVALSI nazionale mentre nel test T' gli item A'_1, \dots, A'_m si presentano con determinate variazioni nella formulazione rispetto a A_1, \dots, A_m .

Abbiamo somministrato i due test T e T' in 40 classi. In ogni classe abbiamo somministrato il test T a metà degli studenti (scelti a caso) e il test T' alla restante metà degli studenti.

La prima analisi che abbiamo svolto ha riguardato la parte comune (CT) dei test T e T' a cui abbiamo applicato, separatamente, il modello di Rasch. Congiuntamente all'applicazione del modello di Rasch ai due test si è anche fatto uso di specifici indici statistici della *Teoria classica dei test*, tra i quali, per esempio, l'*alpha* di Cronbach che misura la coerenza interna del test. Tale scelta ha permesso di avere le prime informazioni riguardo alla comparabilità dei risultati dei due campioni di studenti a cui sono stati somministrati rispettivamente T e T' (per esempio confrontando le mappe di Wright) e, inoltre, ha reso possibile il confronto tra i risultati delle nuove sperimentazioni con i risultati del campione nazionale. Una volta confrontati i risultati degli studenti sulla parte comune del test (CT), abbiamo proseguito con l'analisi delle restanti domande che compaiono nei due test T e T' in due forme diverse.

Il *core test* CT permette di ancorare i risultati della nuova somministrazione dei test T e T' tra loro e con i risultati dell'indagine nazionale INVALSI. Per fare ciò l'abilità degli studenti viene quindi calcolata applicando il modello di Rasch esclusivamente agli item del CT, questa volta però unendo i dati delle due prove e collocando quindi tutti gli studenti dei due campioni sulla medesima scala di abilità. A questo punto è possibile quindi stimare la probabilità che uno studente di un determinato livello di abilità p (misurato come punteggio di Rasch sul CT) ha di rispondere correttamente agli item A_j e A'_j . Inoltre, in questo modo è possibile approfondire il confronto delle domande originali A_j con le relative domande variate A'_j attraverso l'uso dei *distractor plots* delle due domande. Infatti, ponendo sull'asse delle ascisse il punteggio di Rasch ottenuto dagli studenti sul CT, è stato possibile confrontare direttamente i *distractor plots* di A_j e A'_j e osservare possibili cambiamenti nell'andamento della risposta corretta e delle altre alternative di risposta dovute alla variazione nella formulazione della domanda.

Un ulteriore riscontro di quanto osservato in questa prima fase dell'analisi dei dati è stato possibile grazie all'ancoraggio delle due prove somministrate. Solitamente le tecniche di ancoraggio statistico (*test equating*) vengono applicate al fine di confrontare i punteggi di diversi gruppi di studenti che, anche in anni diversi, hanno risposto a due diversi test che misurano lo stesso tratto latente e che hanno un set di item in comune. Nel nostro caso il *test equating* ha lo scopo principale di ancorare i due test T e T', al fine di confrontare non tanto i risultati dei rispondenti, quanto i parametri relativi agli item. Questa procedura ha il compito di esprimere sulla stessa scala i risultati delle due prove, ancorando i due test grazie alla presenza di una parte consistente di item in comune (CT). In particolare, abbiamo utilizzato una procedura di *test equating* scegliendo di fare una calibrazione congiunta che consente stimare la difficoltà di ogni item e l'abilità di ogni studente considerando i risultati dei due test contemporaneamente e che risulta più precisa rispetto a una calibrazione separata (Kolen e Brennan, 1995). I parametri così stimati sono espressi sulla stessa scala e questo permette di confrontare i parametri di difficoltà degli item A_1, A_2, \dots, A_m con quelli dei rispettivi item variati A'_1, A'_2, \dots, A'_m .

L'applicazione dello strumento statistico descritto e l'analisi quantitativa dei risultati ci ha consentito di formulare congetture relative agli effetti di ogni specifica tipologia di variazione che potrà poi in un secondo momento essere validata attraverso un'indagine di tipo qualitativo.

4. Piano di validazione

Il piano di validazione della metodologia presentata in questo capitolo è il seguente.

Siamo partiti da un test INVALSI somministrato su scala nazionale nel maggio del 2013 a 590.728 studenti frequentanti la classe prima della scuola secondaria di I grado (livello 6). Il test originale (T) era composto da $N = 48$ domande. Le analisi statistiche dell'INVALSI sono state condotte su un campione rappresentativo di circa 27.000 studenti di cui un sottogruppo di 1.528 formava il campione rappresentativo della regione Emilia Romagna. Abbiamo quindi scelto $m = 7$ domande del test T e le abbiamo modificate secondo diversi criteri legati alle variabili redazionali descritte da Laborde (1995). Abbiamo somministrato il nuovo test T', contenente gli item variati, e il test originale T a 777 studenti della stessa età e della stessa regione (Emilia Romagna), assicurandoci che gli studenti non avessero già risposto alla prova del 2013. In particolare, in ciascuna delle 40 classi coinvolte nella ricerca, metà degli studenti hanno svolto il nuovo test T' (per un totale di 397 studenti) e il resto ha risposto alla prova originale T (per un totale di 380 alunni). Gli alunni di ogni classe sono stati suddivisi in modo casuale allo scopo di considerare paragonabili le due popolazioni così ottenute.

In primo luogo abbiamo confrontato i risultati globali dei nostri test con i risultati del campione nazionale e con quello dell'Emilia Romagna. Per fare ciò abbiamo applicato il modello di Rasch sia sui test interi T e T', sia sulle 41 (N-m) domande in comune del *core test* CT. In aggiunta al modello di Rasch, abbiamo utilizzato indici specifici provenienti dalla teoria classica dei test; per esempio, l'*alpha* di Cronbach ha permesso di verificare che fosse rispettata la coerenza interna del test e le principali caratteristiche psicometriche degli item. Inoltre, attraverso l'analisi delle mappe di Wright, è stato possibile confrontare le distribuzioni, relative alle diverse somministrazioni, delle 41 domande del CT in funzione dei parametri di difficoltà di Rasch e verificarne la corrispondenza.

Una volta svolte le prime analisi per appurare l'effettiva comparabilità dei campioni e dei risultati delle prove T e T', è stato possibile applicare il modello di Rasch e le procedure di *test equating* descritte nel paragrafo precedente per raccogliere le informazioni relative ai 7 item interessati dalle variazioni e andare quindi ad analizzare in che modo queste variazioni siano andate a impattare sulle performance degli studenti.

In aggiunta, abbiamo ripetuto le analisi scorporando gruppi di studenti in base a un criterio, per esempio il genere, per studiare se una certa tipologia di variazione avesse avuto un'influenza maggiore su una parte degli alunni.

5. Esempio di analisi

Di seguito presentiamo l'analisi di uno dei sette item modificati. In questo caso, al quesito originale è stata apportata una variazione numerica relativa all'ordine di grandezza dei numeri presentati, e di conseguenza anche alla tipologia dei numeri stessi. L'item originale, presentato in fig. 1, chiede di stimare il risultato della moltiplicazione di due numeri decimali.

Fig. 1 – Item D22 nella forma originale (test T)

D22. Quale dei seguenti numeri interi è più vicino al risultato di questa moltiplicazione?	
	$4,82 \times 9,95$
A.	<input type="checkbox"/> 36
B.	<input type="checkbox"/> 42
C.	<input type="checkbox"/> 48
D.	<input type="checkbox"/> 50

Nella versione modificata (fig. 2) la richiesta è la stessa, ma i numeri presentati sono interi e il loro ordine di grandezza è superiore. È importante notare che tutte le alternative di risposta nella forma variata sono analoghe a quelle del item originale.

Fig. 2 – Item D22 nella forma variata (test T')

D22. Quale dei seguenti numeri è più vicino al risultato di questa moltiplicazione?		
482 x 995		
A.	<input type="checkbox"/>	360.000
B.	<input type="checkbox"/>	420.000
C.	<input type="checkbox"/>	480.000
D.	<input type="checkbox"/>	500.000

La domanda riguarda la stima del risultato di un'operazione, in particolare una moltiplicazione. La consegna, infatti, esplicita di indicare quale tra i risultati presentati sia il "più vicino" al prodotto. Tuttavia, anche se il tipo di numeri in gioco non dovrebbe cambiare la natura del problema, ci si può aspettare che gli studenti siano guidati dalle abituali pratiche d'aula, legate all'approssimazione e al calcolo attraverso diverse procedure.

L'item è stato formulato come una domanda a scelta multipla; per questo motivo, la nostra analisi a priori si concentra solo sulle possibili scelte degli studenti in una rosa di quattro possibili opzioni. Sulla base di ciò, possiamo elaborare delle ipotesi interpretative sulle motivazioni che hanno guidato la scelta di una particolare alternativa. Una strategia comune determinerebbe la stessa risposta in entrambe le formulazioni, con l'unica differenza nell'ordine di grandezza. Una differenza significativa nelle percentuali di scelta di un distrattore rispetto a un altro è quindi segnale di strategie risolutive diverse.

Thevenot e Oakhill (2005) hanno messo in luce che le strategie messe in campo dagli studenti possono dipendere da fattori linguistici o da fattori numerici come, nel nostro caso, l'ordine di grandezza. Abbiamo deciso di analizzare l'impatto di questo tipo di variazione – che ci aspettavamo potesse causare un cambiamento significativo nella distribuzione di frequenza delle risposte, dal momento che tali risultati sono presentati in letteratura – per verificare che la metodologia usata lo facesse effettivamente emergere e, in tal caso, in che modo. È ragionevole infatti ipotizzare che il passaggio da numeri decimali a interi possa modificare le performance degli studenti, come già è stato messo in luce nelle ricerche citate. Gli strumenti di ricerca elaborati consentono di validare tale ipotesi e, inoltre, di indagare in che misura tale variazione possa aver influenzato la distribuzione delle risposte e su quali livelli di abilità abbia avuto una maggiore incidenza, grazie al confronto su un'unica scala di abilità delle risposte di tutti gli studenti. Inoltre questo strumento, applicato separando gli studenti in gruppi in base al genere o alla cittadinanza, permette anche di studiare quali categorie di studenti sono state maggiormente influenzate dalla variazione.

Conducendo un'analisi a priori delle possibili risposte alla domanda, emergono alcune possibili strategie associabili alle diverse opzioni proposte agli studenti come alternative nel quesito a scelta multipla:

- arrotondare entrambi i numeri all'intero più vicino;
- considerare solo la parte intera del numero decimale;
- approssimare entrambi i fattori per eccesso o per difetto;
- altro.

La variazione della tipologia di numeri può influenzare gli studenti e portarli a un cambio di approccio alla risoluzione del quesito e perciò a un'altra scelta. Gli studenti potrebbero risultare abili nell'approssimazione di numeri interi e non sapere come affrontare la stima del prodotto tra numeri decimali, il che evidenzerebbe una conoscenza parziale dei metodi di stima. Al contrario si potrebbe osservare che gli studenti con un punteggio di Rasch medio/alto non siano influenzati da questo tipo di cambiamento dal momento che la loro conoscenza è più completa.

L'analisi dei dati riportata in fig. 1 mostra che l'item variato presenta una percentuale di risposte corrette (opzione C) più elevata dell'item originale. Infatti la percentuale di risposta corretta passa dal 46% (item originale) al 59% (item variato). Come si può osservare nel primo grafico, l'opzione che subisce maggiormente la variazione apportata è la B. Infatti, se le opzioni A e D aumentano o diminuiscono solo di alcuni punti percentuali, la risposta B perde circa l'8% delle scelte a seguito della variazione numerica. Per quanto riguarda la percentuale di risposte non date, si può notare che essa non è particolarmente influenzata dalla variazione, ciò significa che, nonostante la difficoltà dei due item risulti differente, quasi tutti gli studenti si ritengono abbastanza sicuri per tentare di rispondere.

Il *test equating*, applicato a entrambi i test, ci permette di stimare i parametri della difficoltà di tutti gli item, includendo entrambe le versioni dei sette item, e di considerarli sulla stessa scala.

Tab. 1 – Percentuali di risposta per l'item D22 (in forma originale e variata). Risposta corretta: C

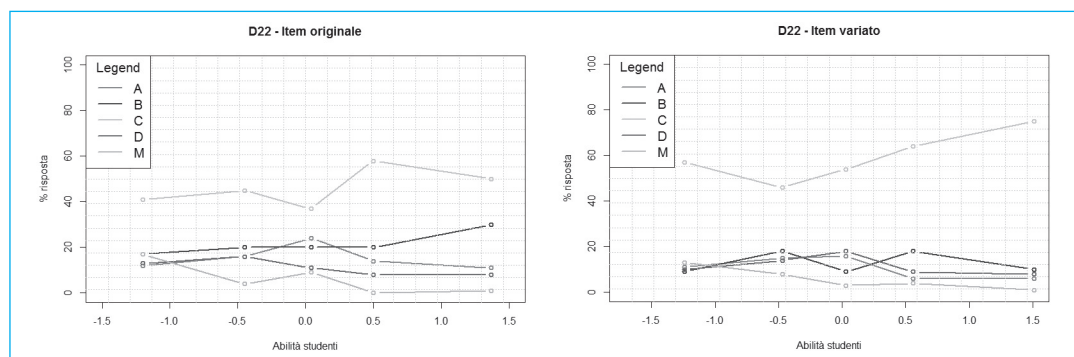
	Item originale	Item variato
A	15	11
B	21	13
C	46	59
D	11	12
Mancante	6	6

Il confronto tra i parametri di difficoltà stimati dalla tecnica di ancoraggio ci dà un'ulteriore prova che la variazione, in questo caso, renda l'item più facile. Infatti il valore di questo parametro è significativamente differente per l'una e l'altra formulazione: la difficoltà dell'item originale è 0,10 mentre la difficoltà di quello variato è -0,51, entrambi con un errore standard di 0,11.

A questo punto, può essere interessante analizzare i *distractor plots* per indagare se le differenze precedentemente identificate sono distribuite uniformemente su tutti gli studenti o se questi cambiamenti hanno influenzato maggiormente studenti con un certo livello di abilità. I *distractor plots* (fig. 3) sono stati realizzati come funzione dell'abilità degli studenti valutati sul CT.

Partendo dai dati raccolti per i 41 item comuni, attraverso il modello di Rasch, abbiamo stimato i parametri di abilità per ognuno dei 777 studenti. Si può notare che l'andamento della curva relativa alla risposta corretta risulta diverso nelle due versioni; in particolare, tale curva risulta più regolare (crescente a esclusione del primo quintile) nell'item variato.

Fig. 3 – D22 – Distractor plots: item originale e variato



In entrambe le forme si osserva che questa domanda non è molto discriminante, ovvero non distingue bene tra gli studenti con alti livelli di abilità e quelli con più basse abilità. Infatti, l'indice di discriminazione risulta essere 0,15 per l'item originale e 0,22 per l'item variato (quindi leggermente migliore). È interessante notare che la variazione ha migliorato le proprietà statistiche dell'item: l'andamento della curva relativa alla risposta corretta e la discriminazione risultano essere migliori nella forma variata.

Altri elementi interessanti emergono dall'analisi dell'andamento relativo alla scelta delle altre opzioni di risposta. Per esempio, l'opzione B mostra una variazione più sensibile della percentuale di risposta: viene scelta infatti dal 21% degli studenti nella versione originale e dal 13% degli studenti nella versione variata. Questo risultato può essere approfondito analizzando l'andamento della risposta nei *distractor plots*; si nota infatti che nella versione originale l'opzione B viene scelta maggiormente da studenti con un alto livello di abilità, cosa che non avviene per la versione variata. Dopo la modifica esso risulta molto meno appetibile per questi studenti, i quali optano invece per la risposta corretta, che in questo modo risulta crescente per livelli di abilità medi e alti.

Inoltre, l'analisi di questo item è molto interessante anche differenziando gli studenti in base al genere. Nella tabella sottostante sono presentate le percentuali relative a ogni opzione di risposta per entrambe le versioni dell'item e suddividendo la popolazione in maschi e femmine.

Tab. 2 – Percentuali di risposta all'item D22 in base al genere. Risposta corretta: C

	Maschi		Femmine	
	Item originale	Item variato	Item originale	Item variato
A	16	10	12	11
B	23	12	20	13
C	44	62	50	56
D	11	11	10	13
Mancante	5	4	8	7

Nell'item originale, che presentava i numeri decimali, le risposte corrette sono il 44% per i maschi e il 50% per le femmine. La variazione ha un enorme impatto sulle prestazioni dei maschi che, rispondendo all'item variato, guadagnano il 18% in più di risposte corrette. Per quanto riguarda le femmine, invece, si nota che la percentuale di risposte corrette aumenta solo di 6 punti percentuali. Questo fenomeno potrebbe essere spiegato ipotizzando che maschi e femmine applichino diverse strategie per risolvere questo tipo di problema e che le strategie utilizzate dalle femmine varino di meno in dipendenza dalla tipologia e dall'ordine di grandezza dei numeri.

6. Conclusioni

La metodologia messa a punto per indagare l'impatto di una variazione nella formulazione di un quesito sulla distribuzione di frequenza di risposte degli studenti si è rivelata efficace in quanto, come ci si attendeva, sono emerse alcune differenze nelle distribuzioni di risposte alle due domande relative alle caratteristiche del livello di abilità relativa manifestata nel *core test* e al genere. In particolare questo strumento statistico permette di evidenziare se la variazione ha influito sulle risposte degli studenti e su quali livelli di abilità l'impatto è stato più significativo. Inoltre questo approccio ha permesso di evidenziare differenze di performance tra diverse categorie di studenti e di indicare percorsi per ulteriori indagini sulle cause di queste differenze. Questa metodologia sembra adeguata per analizzare gli effetti di ulteriori categorie di variazioni. La metodologia quantitativa potrebbe contribuire, in future ricerche, a far emergere dei macro-fenomeni che possono successivamente essere investigati attraverso un'impostazione sperimentale qualitativa, con la quale è possibile verificare le ipotesi di un cambio di strategia indotto dal cambiamento nella formulazione delle domande, diventando un tassello fondamentale di una metodologia mista quantitativa e qualitativa (Johnson e Onwuegbuzie, 2004) che consenta di indagare anche qualitativamente nuovi fenomeni partendo da evidenze quantitative.

Riferimenti bibliografici

- Barbaranelli C., Natali E. (2005), *I test psicologici: teorie e modelli psicometrici*, Carrocci, Roma.
- Branchetti L., Viale M. (2015), "Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico", in M. Ostinelli (a cura di), *Didattica dell'italiano. Problemi e prospettive*, Dipartimento Formazione e apprendimento – Scuola universitaria professionale della Svizzera italiana (SUPSI), Locarno: 138-148.
- D'Amore B. (2014), *Il problema di matematica nella pratica didattica*, Digital Index, Modena.
- Daroczy G., Wolska M., Meurers W.D., Nuerk HC. (2015), "Word problems: a review of linguistic and numerical factors contributing to their difficulty", *Frontiers in Psychology*, 6: 1-13.
- De Corte E., Verschaffel L. (1985), "Beginning first graders' initial representation of arithmetic word problems", *The Journal of Mathematical Behavior*, 4: 3-21.
- Duval R. (1991), "Interaction des niveaux de représentation dans la compréhension de textes", *Annales de Didactique et de Sciences Cognitives*, 4: 163-196.
- INVALSI (2013), *Rilevazioni nazionali sugli apprendimenti 2012-2013. Rapporto tecnico*, testo disponibile al sito: http://www.invalsi.it/snvpn2013/rapporti/Rapporto_tecnico_SNV2013_12.pdf, data di consultazione: 31 maggio 2017.
- Johnson R.B., Onwuegbuzie A.J. (2004), "Mixed Methods Research: A Research Paradigm whose Time has come", *Educational Researcher*, 33, 7: 14-26.

-
- Kolen M.J., Brennan R.L. (1995), *Test Equating: Methods and Practices*, Springer, New York (NY).
- Laborde C. (1995), “Occorre imparare a leggere e scrivere in matematica?”, *La matematica e la sua didattica*, 2: 121-135.
- Mayer R. (1982), “The psychology of mathematical problem solving”, in F. Lester, L.J. Garofalo (eds.), *Mathematical Problem Solving. Issues in Research*, The Franklin Institute Press, Philadelphia (PA).
- Nesher P. (1982), “Levels of description in the analysis of addition and subtraction word problems”, in J. Moser (ed.), *Addition and Subtraction: A Cognitive Perspective*, Lawrence Erlbaum Associates, Hillsdale (NJ).
- OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris.
- Rasch G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Danmarks Paedagogiske Institut, Copenhagen.
- Thevenot C., Oakhill J. (2005), “The strategic use of alternative representation in arithmetic word problem solving”, *Quarterly Journal of Experimental Psychology*, 58: 1311-1323.

10. LIM e rendimenti scolastici degli studenti italiani: un'analisi di impatto su larga scala

The effects of interactive whiteboards on students' academic performances: a large scale impact analysis

di Gianluca Argentin, Tiziano Gerosa

La popolarità della lavagna interattiva multimediale (LIM) come strumento di supporto alla didattica è notevolmente accresciuta nel corso dell'ultimo decennio, dando il via a onerose politiche di investimento tecnologico in ambito educativo. La letteratura internazionale sul tema è concorde nell'associare la presenza della LIM in classe allo sviluppo di un ambiente di apprendimento più stimolante, con ricadute positive sulla motivazione, l'interesse e la partecipazione degli studenti a lezione. Meno chiaro, invece, rimane il rapporto che intercorre fra l'utilizzo di tale tecnologia e i rendimenti scolastici degli studenti nelle materie curriculari. Le evidenze a oggi disponibili sono ancora scarse, eterogenee e fortemente influenzate dai metodi di analisi utilizzati, dal livello di istruzione indagato e dalle discipline considerate. Inoltre, scarso interesse è stato dedicato all'analisi dell'eterogeneità degli effetti della LIM, quindi, alle sue ricadute su studenti con differenti caratteristiche e in diversi contesti di apprendimento. Il presente studio si concentra sui primi due cicli di istruzione del sistema scolastico italiano (classi seconde e quinte della scuola primaria, classe terza della scuola secondaria di I grado e classe seconda della scuola secondaria di II grado) e sfrutta un approccio contro-fattuale per valutare l'impatto prodotto dalla presenza della LIM in classe sulle performance degli studenti in italiano e matematica. In secondo luogo, lo studio offre un'analisi dell'eterogeneità degli effetti rilevati in funzione di alcune caratteristiche individuali degli studenti (sesso, cittadinanza, titolo di studio dei genitori, bocciatura o iscrizione posticipata) e di composizione della classe (numerosità) tipicamente associate a variazioni nei rendimenti scolastici. La base dati utilizzata è frutto dell'unione tra il Questionario insegnante, il Questionario studente e i test standardizzati INVALSI per l'a.s. 2013-14. Si tratta di un campione nazionale statisticamente rappresentativo che comprende più di 110.000 studenti e 6.000 classi. Per valutare gli effetti prodotti dalla presenza della LIM in classe sono stati realizzati più modelli di stima applicando le tecniche statistiche di Regressione lineare multivariata (RLM) e *Propensity Score Matching* (PSM) a due livelli di analisi (studente e media di classe) e su sottoinsiemi di casi specifici giudicati rilevanti in termini di potenziale eterogeneità degli effetti. I risultati rimandano a un impatto per lo più nullo della LIM in classe sulle performance scolastiche degli studenti, ma moderatamente positivo se limitiamo l'analisi alle classi seconde della scuola secondaria di II grado e alla sola materia di matematica per le classi quinte della primaria. Inoltre, non emergono chiari segnali di eterogeneità degli effetti prodotti dalle LIM sui rendimenti scolastici degli studenti. Nel paragrafo conclusivo si discutono i risultati dello studio, formulando raccomandazioni per valutazioni future più robuste e funzionali alle scelte di investimento.

The popularity of interactive whiteboards (IWB) as teaching support tools has increasingly grown in the last decade, urging educational institutions to invest considerable resources for their supply at school. This choice has been primarily driven by the fact that international literature generally agrees in associating the use of IWBs with a more stimulating learning environment and an increase in students' motivation, interest and involvement. Less clear, however, remains the relationship between the presence of IWBs in classrooms and students' academic performances in the curricular subjects. Evidences in this field are still scarce, conflicting and often affected by the use of analytical methods suffering from clear lacks of robustness. Moreover, relatively little interest has been placed on estimating and comparing the IWBs' effect on different groups of students operating in heterogeneous learning contexts. On these grounds, the present study adopts a counterfactual methodological approach to evaluate the impact of IWBs on students' academic performances (literacy and math) in the first two cycle of the Italian compulsory education system

(grade 2 and grade 5 for primary school, grade 8 for middle school and grade 11 for high school). Secondly, it offers an analysis of the heterogeneity of the IWBs' effects by students' individual traits (groups of students distinguished by sex, ethnicity, cultural background and past schooling experience) and class composition (number of students per class). The data used for the analysis comes from a national statistically representative sample of more than 110,000 students and 6,000 classes, obtained by a merge between the Teacher Questionnaire, the Student Questionnaire and the standardized tests collected by INVALSI during the 2013-14 school year. The impact of IWBs is estimated at the student and at the classroom level of aggregation using the statistical techniques of Multiple Linear Regression (MLR) and Propensity Score Matching (PSM). The results show the existence of moderate positive effects of the IWB only on math performances of students at grade 5 and on both the school subjects of literacy and math for students at grade 11, while there are no clear patterns of heterogeneity in the IWB effects on students' performances at the individual or classroom levels. In the last paragraph the results of the study are discussed, formulating recommendations for future research on ICTs investment policies.

1. Introduzione

La lavagna interattiva multimediale (LIM) è un dispositivo didattico progettato per proiettare su una superficie output provenienti da computer, offrendo agli utenti l'opportunità di manipolarli attraverso vari tipi di software e controlli digitali. Nel corso dell'ultimo decennio, buona parte dei Paesi occidentali ha mostrato crescente interesse per il potenziale educativo offerto da tale tecnologia, adottando onerose politiche d'investimento per la sua integrazione all'interno delle scuole. Uno dei primi esempi di approvvigionamento su larga scala della LIM è rappresentato dall'esperienza del Regno Unito. Tra il 2003 e il 2004, infatti, il Ministero dell'Educazione britannico ha investito più di 25 milioni di sterline nell'acquisto della tecnologia LIM (Schroeder, 2007). A quattro anni dall'attuazione di tale politica, il 98% delle scuole ha riferito di possedere lavagne interattive indirizzate all'*e-learning*, con un aumento del loro numero medio per istituto di 8 unità per le primarie e di 22 unità per le secondarie (Kitchen, Finch e Sinclair, 2007). Più di recente, anche il Ministero dell'Istruzione italiano ha scelto di adottare un piano di lungo periodo per l'introduzione della LIM nelle scuole del Paese. Grazie a un investimento complessivo di oltre 93 milioni di euro, dal 2008 ne sono state assegnate oltre 35.000 agli istituti di ogni ordine e grado, innalzando la quota totale di aule dotate di tale tecnologia dal 10% al 35% soltanto nel corso del quadriennio 2010-2014 (Giusti *et al.*, 2015).

Alla luce dei numeri imponenti di queste politiche di approvvigionamento, sorprende scoprire che la ricerca nazionale e internazionale non è ancora stata in grado di dirci con certezza se e in che misura la LIM contribuisca al miglioramento delle performance scolastiche degli studenti. Buona parte della letteratura sul tema, infatti, si concentra sull'analisi delle percezioni e degli atteggiamenti degli studenti nei confronti della LIM, accennando soltanto indirettamente ai potenziali benefici da essa prodotti sui livelli di apprendimento (Higgins, Beauchamp e Miller, 2007; Di Gregorio e Sobel-Lojeski, 2010). Le LIM sono generalmente percepite come un supporto innovativo alla didattica, particolarmente adatto a soddisfare le esigenze educative dei nativi digitali (Smith, Hardman e Higgins, 2006; Schuck e Kearney, 2007). Viene inoltre enfatizzata l'importanza della loro adattabilità ai bisogni degli insegnanti (Schuck e Kearney, 2007) e ai differenti stili di apprendimento adottati dagli studenti (Wall, Higgins e Smith, 2005; Glover *et al.*, 2005). La maggiore offerta di stimoli visivi, sonori e interattivi di tale tecnologia sarebbe in grado di accrescere l'attenzione, l'interesse e la concentrazione in classe (Glover *et al.*, 2007; Slay, Siebörgher e Hodgkinson-Williams, 2008), producendo ricadute positive anche sulla motivazione personale e la partecipazione attiva (Higgins, Beauchamp e Miller, 2007; Wood e Ashfield, 2008). Ciononostante, a oggi non disponiamo di un bagaglio di evidenze empiriche sufficientemente robusto per individuare proprietà specifiche delle LIM in grado di incidere in modo diretto o indiretto sulle competenze cognitive degli studenti (Armstrong *et al.*, 2005) e sulle loro performance scolastiche (Di Gregorio e Sobel-Lojeski, 2010). Possiamo individuare due cause principali alla base di tale incertezza. Innanzitutto, la ricerca sin qui condotta si basa il più delle volte su studi di caso o indagini quantitative lontane da una logica di stima rigorosa degli effetti. Sono pochi gli studi che si propongono di valutare l'impatto delle LIM sulle performance scolastiche utilizzando un approccio contro-fattuale su larga scala e ancora meno quelli che affrontano il tema dell'eterogeneità dei suoi effetti su individui diversi e a differenti livelli di analisi (Di Gregorio

e Sobel-Lojeski 2010). In secondo luogo, le poche evidenze raccolte attraverso l'adozione di un concetto rigoroso di causalità forniscono risultati tendenzialmente controversi. Higgins e colleghi (2007), per esempio, conducono uno studio sui dati longitudinali del *Secondary Whiteboard Expansion Project* senza rilevare scostamenti degni di nota nel lungo periodo tra le performance degli studenti di scuole dotate di LIM e quelle degli studenti di scuole dove tale dispositivo non è presente. Al contrario, Thompson e Flecknoe (2003) rilevano miglioramenti significativi sui rendimenti dei soggetti coinvolti nel progetto educativo *Easiteach Maths*, indirizzato a incentivare la partecipazione attiva degli alunni durante la lezione attraverso un'interazione diretta con la LIM. Altri studi ancora si collocano in una posizione di maggiore incertezza, evidenziando la presenza di forti eterogeneità negli effetti prodotti dalla LIM imputabili a caratteristiche specifiche dei soggetti trattati. È il caso di Lewin e colleghi (2008), che individuano un impatto positivo della LIM sull'alfabetizzazione e i rendimenti in matematica e scienze degli studenti di scuola primaria e secondaria di I grado, ma solo per quelli di età compresa tra i 7 e 11 anni e già dotati di un più ampio bagaglio di conoscenze pregresse.

In ambito nazionale, la produzione scientifica sulla LIM e, più in generale, sulle tecnologie dell'informazione e della comunicazione (TIC) in ambito educativo si fa ancora più scarna. Il progetto *CI@ssi 2.0*, promosso dal Ministero dell'Istruzione nel 2009, stima gli effetti prodotti dall'introduzione di diverse tecnologie informatiche, tra cui la LIM, sui livelli di apprendimento in italiano e matematica degli studenti di 308 classi di scuola secondaria di I grado (Rettore e Checchi, 2014). I risultati della ricerca rimandano a un effetto positivo delle TIC piuttosto moderato solo per italiano ed esclusivamente per il sottogruppo degli studenti dotati di scarse risorse culturali familiari. Giusti e colleghi (2015), invece, approfondiscono il rapporto esistente tra l'intero stock tecnologico a disposizione delle scuole italiane e le performance degli studenti misurate tramite i test INVALSI. I risultati inerenti le LIM mostrano la presenza di associazioni positive a livello nazionale tra la quota di classi coperte e i punteggi medi di plesso nella prova di italiano. Anche in questo caso, però, si tratta di variazioni moderate ed eterogenee tra le aree del Nord, del Centro e del Sud, frutto di modelli di regressione a effetti fissi che comportano un certo grado di cautela nell'interpretazione causale dei risultati.

Nel complesso, le evidenze emerse dalla ricerca pregressa sulle LIM in Italia, così come in altri Paesi, non sono ancora in grado di offrirci chiare indicazioni sulla loro reale efficacia in ambito educativo. Il presente studio vuole quindi contribuire a produrre evidenza in tal senso, valutando se la massiccia introduzione delle LIM all'interno delle scuole italiane abbia effettivamente prodotto ricadute positive sui livelli di apprendimento degli studenti di ogni ordine e grado. In linea con questo proposito conoscitivo, possiamo identificare due principali obiettivi di ricerca:

- stimare l'effetto della presenza della LIM in classe sui rendimenti scolastici in italiano e matematica degli studenti di scuola primaria, secondaria di I grado e secondaria di II grado;
- esplorare la presenza di eterogeneità negli effetti della LIM a livello individuale e di classe.

I suddetti obiettivi conoscitivi, come si avrà modo di illustrare, sono perseguibili grazie all'utilizzo di una base dati INVALSI rappresentativa a livello nazionale e caratterizzata da elevata numerosità. I paragrafi seguenti offrono una descrizione dettagliata di questa base dati, del metodo di analisi utilizzato e dei principali risultati ottenuti, per poi concludere traendo alcune implicazioni per l'impiego futuro delle LIM nelle scuole.

2. Dati e variabili

La base dati di riferimento per le analisi è frutto dell'unione tra il Questionario insegnante, il Questionario studente e i test standardizzati INVALSI di italiano e matematica per l'anno scolastico 2013-14. Si tratta di un campione rappresentativo a livello nazionale, comprendente più di 110.000 studenti e 6.000 classi. La base dati relativa agli studenti raccoglie informazioni su un campione estratto casualmente da INVALSI, che viene sottoposto a verifica diretta da parte di osservatori esterni durante la compilazione. Il Questionario insegnante viene unito al precedente a livello classe e contiene informazioni dettagliate sugli insegnanti di italiano e matematica. Tra le informazioni disponibili al suo interno vi è anche quella relativa alla presenza o meno della LIM nella classe oggetto di rilevazione. Si tratta di un indicatore dicotomico costruito per le classi in cui l'insegnante di italiano e quello di matematica hanno dato la medesima risposta riguardo l'effettiva disponibilità della LIM in aula, escludendo i casi in cui non si è rilevato consenso tra i due. Quest'analisi preliminare ha permesso di verificare la presenza di una ridotta percentuale di incoerenza nelle risposte degli in-

segnanti, pari a circa l'8% del totale delle classi, accrescendo considerevolmente la fiducia circa la correttezza del dato raccolto relativamente alla variabile di trattamento.

Tab. 1 – Ampiezza dei campioni di studenti, classi e scuole inclusi ed esclusi dall'analisi

		<i>Studenti</i>	<i>Classi</i>	<i>Scuole</i>
Italiano	LIM presente	27.545	1.451	1.134
	LIM assente	60.175	3.357	1.412
	Totale casi inclusi nell'analisi	87.720	4.808	2.546
	(esclusi dall'analisi per mancanza di informazioni concordanti sulla LIM)	(7.172)	(383)	(140)
Matematica	LIM presente	27.577	1.444	1.096
	LIM assente	60.333	3.340	1.454
	Totale casi inclusi nell'analisi	87.910	4.784	2.550
	(esclusi dall'analisi per mancanza di informazioni concordanti sulla LIM)	(7.130)	(377)	(144)

Come mostrato in tab. 1, le dimensioni campionarie emerse sono molto ampie rispetto alla letteratura precedente e con un elevato grado di rappresentatività nazionale, anche soffermandosi soltanto sul sottoinsieme dei casi sottoposti ad analisi. Avere un campione numericamente ricco rappresenta un pre-requisito di fondamentale importanza per il nostro studio, perché più la base numerica è ampia, più piccoli sono gli effetti che possono essere intercettati e riconosciuti come statisticamente significativi. Inoltre, un campione ampio consente di controllare per più variabili senza rischiare di estrapolare risultati in assenza di una base empirica sufficiente, e offre maggiori opportunità di esplorare l'eterogeneità degli effetti stimati in base alle caratteristiche degli studenti giudicate rilevanti.

Grazie all'ampiezza e alla ricchezza informativa della base dati a nostra disposizione, è stato possibile sviluppare un disegno della ricerca che prevede l'utilizzo di un vasto insieme di variabili di controllo. In altri termini, il confronto tra apprendimento degli studenti in una classe con LIM e studenti in una senza LIM viene depurato da un'ampia gamma di caratteristiche potenzialmente in grado di distorcere la stima di impatto delle LIM stesse. Possiamo distinguere quattro macro-categorie di variabili di controllo (per la lista completa, si rimanda alla tab. 2):

- variabili territoriali, costruite al fine di assorbire l'influenza dei fattori contestuali sull'apprendimento degli studenti;
- variabili a livello scuola, che permettono di tenere sotto controllo le caratteristiche istituzionali dell'istituto, quelle del dirigente scolastico e i tratti predominanti del bacino di studenti che serve;
- variabili a livello classe, che tengono conto della composizione di classe, di alcune caratteristiche degli insegnanti che operano al suo interno e dell'effetto prodotto sull'apprendimento dal gruppo dei pari;
- variabili a livello studente, tipicamente utilizzate in letteratura come predittori delle loro performance nei test standardizzati.

L'utilizzo di questo ampio bagaglio di variabili di controllo permette di essere ragionevolmente fiduciosi circa la tenuta dell'assunto di eguaglianza fra il gruppo dei trattati e il gruppo dei controlli, e quindi sulle stime di impatto delle LIM sugli apprendimenti. Ovviamente sarebbe preferibile poter contare su stime di impatto frutto di un disegno sperimentale, ma per raggiungere tale obiettivo sarebbe stato necessario avviare lo studio prima dell'introduzione delle LIM nelle scuole.

Tab. 2 – Descrizione delle variabili impiegate nelle analisi

Variabile	Contenuto
Caratteristiche del territorio	
Area geografica ISTAT	Area geografica in cui la scuola ha sede ricondotta a 5 variabili dicotomiche su indicazioni ISTAT: 1) Nord-Ovest; 2) Nord-Est; 3) Centro; 4) Sud; 5) Isole
Comune capoluogo	Ruolo amministrativo del comune in cui la scuola ha sede in forma dicotomica: 0) la scuola ha sede in un comune non capoluogo; 1) la scuola ha sede in un comune capoluogo di Provincia o Regione
Dimensioni comune	Popolazione residente nel comune in cui la scuola ha sede ricondotta a 3 variabili dicotomiche: 1) paese fino a 10.000 abitanti; 2) città da 10.000-100.000 abitanti; 3) grande città con oltre 100.000 abitanti
Densità abitativa	Densità abitativa del comune in cui la scuola ha sede in residenti per km ² . Variabile in forma continua
Altitudine comune	Altitudine del comune in cui la scuola ha sede in forma dicotomica: 0) comune pianeggiante o collinare al di sotto dei 600 m; 1) comune montano al di sopra dei 600 m
Caratteristiche della scuola	
Dimensioni scuola	Dimensioni della scuola in numero totale di classi presenti ricondotta a 3 variabili dicotomiche: 1) fino a 3 classi; 2) da 4 a 6 classi; 3) oltre 6 classi
Fattore qualità dirigente	Fattore estratto da batteria di 4 item di tipo Likert a 4 modalità di risposta sulle percezioni degli insegnanti circa l'impegno profuso dal dirigente scolastico (metodo di estrazione delle componenti principali; varianza spiegata = 64%; <i>alpha</i> di Cronbach = 0,856). Variabile in forma continua ricondotta alla media per scuola
Fattore qualità relazioni	Fattore estratto da batteria di 6 item a 10 modalità di risposta sulle percezioni degli insegnanti circa la qualità delle relazioni con il personale scolastico (metodo di estrazione delle componenti principali; varianza spiegata = 57%; <i>alpha</i> di Cronbach = 0,835). Variabile in forma continua ricondotta alla media per scuola.
Fattore qualità corpo docente	Fattore estratto da batteria di 3 item a 10 modalità di risposta sulle percezioni degli insegnanti circa la qualità del corpo docente nella scuola (metodo di estrazione delle componenti principali; varianza spiegata = 80%; <i>alpha</i> di Cronbach = 0,869). Variabile in forma continua ricondotta alla media per scuola
Fattore qualità scuola	Fattore estratto da batteria di 4 item a 10 modalità di risposta sulle percezioni degli insegnanti circa la qualità dell'organizzazione scolastica (metodo di estrazione delle componenti principali; varianza spiegata = 65%; <i>alpha</i> di Cronbach = 0,815). Variabile in forma continua ricondotta alla media per scuola
Indirizzo scolastico	(Solo per scuole secondarie di II grado) Indirizzo scolastico della scuola ricondotto a 4 variabili dicotomiche: 1) liceo; 2) istituto tecnico; 3) istituto professionale; 4) istituto professionale di avviamento al lavoro
Caratteristiche della classe	
Presenza LIM	Dichiarazione degli insegnanti intervistati sulla presenza della LIM in classe. Variabile dicotomica costruita a livello classe: 0) LIM non presente nella classe/scuola; 1) presenza stabile di una LIM in classe. La robustezza dell'informazione è garantita attraverso un'analisi preventiva della coerenza nelle risposte offerte dagli insegnanti di italiano e matematica nella medesima classe
Dimensioni classe	Dimensioni della classe in numero totale di studenti iscritti ricondotte a 3 variabili dicotomiche: 1) fino a 15 studenti; 2) da 16 a 25 studenti; 3) oltre 25 studenti
Quota immigrati I	Quota di studenti immigrati di prima generazione sul totale della classe. Variabile in forma continua
Quota immigrati II	Quota di studenti immigrati di seconda generazione sul totale della classe. Variabile in forma continua
Quota studenti bocciati	Quota di studenti bocciati o con ingresso ritardato sul totale della classe. Variabile in forma continua
Quota studenti con retroterra culturale elevato	Quota di studenti con retroterra culturale familiare elevato (laurea o più) sul totale della classe. Variabile in forma continua
Quota studenti con retroterra culturale basso	Quota di studenti con retroterra culturale familiare ridotto (licenza media o meno) sul totale della classe. Variabile in forma continua
Sesso insegnante	Sesso dell'insegnante in forma dicotomica: 0) femmina; 1) maschio
Età insegnante	Età dell'insegnante ricondotta a 4 variabili dicotomiche: 1) fino a 40 anni; 2) da 41 a 50 anni; 3) da 51 a 60 anni; 4) oltre 60 anni
Condizione contrattuale insegnante	Condizione contrattuale dell'insegnante in forma dicotomica: 0) contratto a tempo indeterminato; 1) contratto a tempo determinato annuale, fino al termine dell'attività didattica e di supplenza
Radicamento nella scuola	Anni consecutivi di insegnamento nella scuola oggetto di indagine ricondotta in forma dicotomica: 0) fino a 3 anni; 1) più di 3 anni
Titolo di studio insegnante	(Solo per scuola primaria) Titolo di studio insegnante in forma dicotomica: 0) diploma di maturità; 1) laurea o più
Voto di laurea	(Solo per scuole secondarie di I e II grado) Voto di laurea. Variabile ricondotta in forma dicotomica: 0) voto da 0 a 110; 1) 110 <i>cum laude</i>
Corsi di formazione	Numero di corsi di formazione e aggiornamento svolti nel corso degli ultimi 2 anni. Variabile ricondotta in forma dicotomica: 0) meno di 1 all'anno; 1) almeno 1 all'anno
Altre attività oltre all'insegnamento	Numero di attività svolte dall'insegnante all'interno della scuola oltre all'insegnamento. Variabile ricodificata in forma dicotomica: 0) nessuna; 1) 1 o più

Tab. 2 – Descrizione delle variabili impiegate nelle analisi (continua)

Variabile	Contenuto
Caratteristiche degli studenti	
Punteggi standardizzati dei test SNV italiano e matematica	Punteggi ottenuti dallo studente nei test INVALSI di italiano e matematica pesati per il fattore di correzione del <i>cheating</i> e successivamente standardizzati
Sesso studente	Sesso dello studente in forma dicotomica: 0) femmina; 1) maschio
Regolarità scolastica	Regolarità dello studente rispetto al corso di studi in forma dicotomica: 0) regolare; 1) bocciato o ingresso ritardato
Retroterra culturale familiare	Titolo di studio più elevato posseduto dai genitori dello studente ricondotto a 3 variabili dicotomiche: 1) basso; 2) medio; 3) alto. Un basso background culturale familiare indica la presenza di genitori con licenza elementare o secondaria inferiore, mentre un livello medio corrisponde all'ottenimento di qualifiche professionali triennali, del diploma di maturità o di diplomi post-secondari. Il livello alto, infine, comprende genitori che si sono laureati, che hanno concluso un master o un dottorato di ricerca
Classe occupazionale familiare	Classe occupazionale più elevata dei genitori dello studente ricondotta a 4 variabili dicotomiche: 1) <i>working class</i> ; 2) piccola borghesia; 3) classe media impiegatizia; 4) <i>service class</i> . La <i>working class</i> comprende le figure dell'operaio, dell'addetto ai servizi e del socio di cooperativa. La piccola borghesia è formata dal variegato insieme dei lavoratori in proprio (commerciante, coltivatore diretto, artigiano, meccanico ecc.). La classe media impiegatizia raggruppa insegnanti, impiegati e militari graduati. La <i>service class</i> , infine, comprende i professionisti dipendenti, i liberi professionisti, gli imprenditori/proprietari agricoli, i dirigenti, i docenti universitari, i funzionari e gli ufficiali militari
Status migratorio	Status migratorio dello studente ricondotto a 3 variabili dicotomiche: 1) italiano; 2) immigrato di I generazione; 3) immigrato di II generazione

3. Il metodo

Le stime di impatto della LIM sui rendimenti scolastici sono realizzate facendo ricorso a due tecniche di analisi: la Regressione lineare multivariata (RLM) e il *Propensity Score Matching* (PSM). L'obiettivo alla base dell'utilizzo di entrambe le tecniche è il medesimo. Si vuole verificare la presenza di scarti significativi nei risultati dei test INVALSI di italiano e matematica fra gli studenti che hanno ricevuto il trattamento (che sono quindi in una classe in cui è presente la LIM) e chi, invece, non l'ha ricevuto (coloro che non dispongono di una LIM in classe). Per rendere conto delle differenze nelle caratteristiche dei due gruppi di studenti a livello territoriale, di scuola, di classe e individuale, che potrebbero distorcere la stima di impatto, il confronto viene depurato dalle variabili di controllo sintetizzate nella tab. 2.

La RLM e il PSM sono due tecniche di analisi che si pongono lo stesso fine e si basano sull'assunto che le covariate considerate esauriscono l'insieme di variabili da tenere in considerazione per stimare correttamente l'effetto causale di interesse (Rosenbaum e Rubin, 1983). Il PSM però, a differenza della RLM, si basa su un'idea estremamente semplice, che contribuisce a rendere "visibile" il termine controfattuale ricostruito per effettuare la stima di impatto. Il modello cerca nei dati coppie di gemelli costituite ciascuna da un individuo/classe che ha ricevuto il trattamento (la LIM in classe nel nostro caso) e uno o più individui/classi che invece non l'hanno ricevuto (i cosiddetti controlli)¹. Dal confronto tra i due gruppi viene inferita la stima di impatto del trattamento. Rispetto alla RLM, emergono tre principali differenze: si escludono dall'analisi i casi che non hanno trovato almeno un gemello nei dati perché al di fuori del cosiddetto "supporto comune"; prima di guardare alla stima di impatto si verifica che l'abbinamento sia bilanciato, assicurando che i due gruppi di gemelli (trattati e controlli) siano in media statisticamente equivalenti sulle variabili osservate; infine, ci si limita a un semplice confronto fra medie senza formulare assunti rispetto alla forma funzionale esistente nella relazione tra le variabili oggetto di analisi.

Per stimare gli effetti prodotti dalla presenza di una LIM nella classe vengono impiegate entrambe le tecniche, in più fasi distinte per le scelte operate, realizzando molteplici modelli statistici per i test INVALSI di italiano e matematica e per ognuno dei gradi scolastici considerati (2, 5, 8 e 10). Questa ridondanza di stime derivanti da tecniche diverse ha il fine di testare la robustezza delle nostre conclusioni, verificando il loro grado di stabilità al mutare delle scelte modellistiche.

¹ Esistono molti criteri di appaiamento. Una descrizione più approfondita delle alternative di abbinamento disponibili può essere trovata, in lingua italiana, nel testo di Martini e Sisti (2009).

La prima fase di analisi prevede un semplice confronto tra medie utilizzando la presenza della LIM in classe come unico predittore dei punteggi ottenuti dagli studenti nei test di italiano e matematica (modello 0). Questo confronto risente ovviamente di tutte le variabili non considerate che possono influenzare al contempo la presenza della LIM e l'apprendimento degli studenti; pertanto, si tratta di una stima di effetto poco credibile.

La seconda fase impiega un modello di RLM che vede l'inserimento di tutte le variabili di controllo descritte nella tab. 2 e corregge gli errori standard delle stime per la clusterizzazione dei dati (modello 1). Le stime sono condotte sia a livello individuale, includendo tutte le variabili di controllo, sia a livello classe, controllando soltanto per i predittori territoriali, di scuola e di classe². La stima sui due livelli consente di avere sia l'impatto della LIM sull'apprendimento dei singoli studenti, sia l'impatto sull'apprendimento medio di classe. Queste stime sono molto più credibili delle precedenti perché creano statisticamente condizioni di *ceteris paribus* tra studenti (o classi) che hanno la LIM e quelli che non la hanno, per tutte le caratteristiche considerate nell'analisi.

La terza e ultima fase sfrutta la tecnica di abbinamento statistico PSM (modello 2)³. In questo caso i risultati di classe corrispondono alla stima direttamente ottenuta con tale tecnica, controllando per tutti i predittori a livello territoriale, scolastico e di classe presentati nella tab. 2. Gli abbinamenti realizzati per ognuno dei gradi scolastici e della materia presentano un buon bilanciamento tra classi trattate e di controllo, rendendo il nostro confronto credibile. A differenza del livello classe, le stime del modello 2 a livello studente derivano da una regressione del tutto simile a quella descritta nel modello 1. In questo caso però, il modello è realizzato considerando soltanto il sottoinsieme delle classi abbinate attraverso il processo di abbinamento sopra descritto. La *ratio* di questa procedura consiste nel confronto tra studenti di classi che si distinguono per la presenza/assenza della LIM, ma sono molto simili per tutte le altre caratteristiche territoriali, di scuola e di classe considerate in fase di abbinamento. Quest'ultima è per noi la stima più valida, dal momento che pone a confronto studenti con e senza LIM limitatamente alle classi equivalenti sulle caratteristiche considerate.

Dopo aver completato le stime di impatto della LIM sulle performance degli studenti a livello individuale e di classe, si procede con l'analisi dell'eterogeneità degli effetti. Si va cioè a verificare se gli effetti differiscono in base a caratteristiche dei sottogruppi di studenti di volta in volta considerati. L'analisi degli effetti per sottogruppi è svolta per far fronte alle crescenti esigenze di approfondimento emerse in letteratura (Digregorio e Sobel-Lojesky, 2010), utilizzando la tecnica di regressione con termine di interazione. Nell'esplorare l'eterogeneità degli effetti a livello individuale prendiamo in considerazione il sesso dello studente, il retroterra culturale dei suoi genitori (in formato dicotomico, diploma/laurea vs licenza media), la cittadinanza e la presenza di bocciature o di iscrizione posticipata. Spesso la ricerca sulle TIC ha messo in luce che il sesso e le origini sociali degli studenti possono avere un impatto differente su individui con una diversa dotazione di risorse personali potenzialmente utili per un loro corretto utilizzo (Lewin, Somekh e Steadman, 2008; Rettore e Checchi, 2014) e con diversi livelli di competenza digitale (Gui e Argentin, 2011). A livello aggregato, invece, si prende in considerazione soltanto la numerosità degli studenti presenti in classe in forma continua. La scelta di questo particolare termine di interazione è dipesa dalla volontà di comprendere se la LIM rappresenta uno strumento più efficace in contesti classe ristretti, con un numero ridotto di partecipanti, o in classi particolarmente numerose, dove l'insegnante può avere maggiori difficoltà a mantenere alti i livelli di attenzione e un clima disciplinato (Finne Achilles, 1990; Blatchford, Edmonds e Martins, 2003).

² Le variabili di controllo introdotte nei modelli variano in parte a seconda del grado scolastico e del livello di analisi considerato. Il retroterra culturale dell'insegnante viene approssimato con il possesso di una laurea per gli insegnanti di scuola primaria e il voto di laurea pesato per disciplina per quelli di secondaria di I grado e secondaria di II grado. Questa decisione dipende dal fatto che molti insegnanti di scuola primaria non sono laureati. Se si introducesse tale variabile nel modello di regressione, di conseguenza, un corposo numero di insegnanti verrebbe escluso automaticamente dal modello per mancanza di informazioni. Nei modelli delle scuole secondarie di II grado viene introdotta un'ulteriore variabile di controllo che identifica l'indirizzo della scuola (liceo, istituto tecnico, istituto professionale, istituto professionale di avviamento al lavoro).

³ Dopo aver testato molteplici approcci di abbinamento, attraverso i quali sono stati raggiunti risultati pressoché identici, si è scelto il criterio *nearest neighbour*. Si tratta di un criterio che segue una strategia di tipo 1 a 1, senza effettuare alcun reinserimento e limitando il confronto al solo supporto comune.

4. Risultati

Osservando i risultati riportati in tab. 3 possiamo agilmente accorgerci che in assenza di controlli, limitandoci quindi a confronti tra studenti (o classi) con o senza LIM, l'effetto di tale dispositivo sui punteggi standardizzati dei test INVALSI di italiano e matematica sia positivo per tutti i gradi scolastici considerati. Nelle classi di grado 2 e grado 8 si tratta di un impatto del tutto trascurabile in entrambe le materie, mentre per gli studenti di quinta elementare la LIM sembra rappresentare un chiaro valore aggiunto per l'insegnamento della matematica. Ancora più robusto appare lo scarto positivo prodotto dalla presenza della LIM nella secondaria di II grado, pari a 0,34 deviazioni standard del punteggio nel test di italiano e 0,36 di quello di matematica. Risultati del tutto simili si ritrovano anche nelle stime dei modelli realizzati a livello classe, con un effetto positivo e significativo delle LIM per le quinte elementari di italiano e le seconde superiori di entrambe le materie. Occorre però ricordare che, per i motivi illustrati in precedenza, si tratta di stime poco credibili.

Tab. 3 – Modelli di stima di impatto della LIM sulle performance degli studenti nei test INVALSI standardizzate*

Italiano	Grado 2			Grado 5			Grado 8			Grado 10		
	Effetto LIM	ES	P-val	Effetto LIM	ES	P-val	Effetto LIM	ES	P-val	Effetto LIM	ES	P-val
<i>Stime sull'apprendimento medio di classe</i>												
M0 - Confronto di medie	0,008	0,089	0,925	0,078	0,068	0,252	0,067	0,059	0,256	0,341	0,076	0,000
M1 - MLR tutti i casi	0,105	0,091	0,250	0,129	0,060	0,032	0,067	0,061	0,279	0,124	0,039	0,002
M2 - PSM	0,054	0,104	–	0,087	0,084	–	0,078	0,061	–	0,044	0,079	–
<i>Stime sull'apprendimento degli studenti</i>												
M0 - Confronto di medie	-0,002	0,038	0,952	0,036	0,032	0,257	0,039	0,043	0,355	0,261	0,044	0,000
M1 - MLR tutti i casi	0,036	0,038	0,345	0,067	0,029	0,019	0,042	0,043	0,332	0,096	0,024	0,000
M2 - MLR classi gemelle	0,021	0,040	0,596	0,052	0,036	0,152	0,041	0,043	0,345	0,076	0,028	0,007
Matematica	Grado 2			Grado 5			Grado 8			Grado 10		
	Effetto LIM	ES	P-val	Effetto LIM	ES	P-val	Effetto LIM	ES	P-val	Effetto LIM	ES	P-val
<i>Stime sull'apprendimento medio di classe</i>												
M0 - confronto di medie	0,063	0,083	0,447	0,240	0,078	0,002	0,043	0,059	0,467	0,363	0,080	0,000
M1 - MLR tutti i casi	0,076	0,087	0,385	0,246	0,076	0,001	0,061	0,062	0,323	0,111	0,052	0,033
M2 - PSM	0,057	0,099	–	0,256	0,084	–	0,057	0,060	–	0,110	0,087	–
<i>Stime sull'apprendimento degli studenti</i>												
M0 - confronto di medie	0,028	0,042	0,504	0,156	0,041	0,000	0,022	0,044	0,621	0,275	0,049	0,000
M1 - MLR tutti i casi	0,038	0,044	0,381	0,151	0,040	0,000	0,041	0,045	0,360	0,087	0,035	0,012
M2 - MLR classi gemelle	0,025	0,050	0,621	0,178	0,045	0,000	0,041	0,045	0,364	0,108	0,041	0,008

* Deviazioni standard dei punteggi nei test INVALSI per ognuno dei gruppi di controllo approssimabili a un valore pari a 1.

I passaggi al modello 1 e al modello 2 portano a modifiche sostanziali nelle stime di impatto della LIM, con una netta contrazione dei coefficienti stimati proprio per la classe seconda della scuola secondaria di II grado, dove gli effetti sembravano molto consistenti. Osservando le stime del modello giudicato maggiormente affidabile (il modello 2 a livello studente), si registrano un aumento di 0,18 deviazioni standard in italiano per gli studenti di classe quinta nella scuola primaria e differenze inferiori a 0,1 deviazioni standard alla fine del biennio della scuola secondaria di II grado sia per italiano sia per matematica. Si nota, inoltre, che le stime individuali, oltre a essere tendenzialmente più alte, non mutano nel passaggio dal modello di RLM a quello di PSM, mentre a livello di classe non si trovano riscontri simili. L'incoerenza fra gli andamenti delle due stime a livello individuale e di classe potrebbe essere indice della capacità del PSM di rendere conto di alcuni squilibri nella distribuzione delle caratteristiche individuali degli studenti in grado di influire sulle performance scolastiche che la RLM non è in grado di cogliere.

Nel complesso, le analisi condotte rimandano a un impatto per lo più nullo della LIM in classe sulle performance scolastiche degli studenti in italiano e matematica, ma moderatamente positivo se limitiamo l'analisi alla classe seconda

della scuola superiore e alla sola materia di matematica per le classi quinte della primaria, dove gli scarti sono anche sostanzialmente rilevanti.

Passiamo ora all'approfondimento di eterogeneità degli effetti della LIM su individui e classi con differenti caratteristiche (tab. 4). A dispetto delle ipotesi formulate in sede di pianificazione delle analisi, le variabili inserite nei modelli di regressione non sembrano interagire in modo rilevante con l'efficacia del trattamento. Si registrano soltanto delle lievi tendenze negative per materia a svantaggio degli studenti maschi e di quelli con minori risorse socio-culturali, con alcuni scostamenti significativi per la variabile cittadinanza nelle performance di italiano di grado 5 e per il retroterra culturale nelle performance di matematica nei gradi 2 e 10. In conclusione, possiamo affermare che, stando ai risultati delle analisi, non emergono chiari segnali di eterogeneità degli effetti prodotti dalle LIM sui rendimenti scolastici degli studenti.

Tab. 4 – Eterogeneità degli effetti: modelli di regressione a livello studente e classe con termine di interazione

	Italiano				Matematica			
	Grado 2	Grado 5	Grado 8	Grado 10	Grado 2	Grado 5	Grado 8	Grado 10
<i>Studente</i>								
<i>Sesso (rif. femmina)</i>								
LIM	0,062	0,072*	0,072	0,118**	0,048	0,138**	0,054	0,069
Sesso	0,017	-0,147**	-0,102**	0,014	0,169**	0,146**	0,161**	0,268**
LIM*Sesso	-0,051	-0,009	-0,061*	-0,044	-0,019	0,025	-0,026	0,038
<i>Titolo studio genitori (rif. diploma/laurea)</i>								
LIM	0,033	0,050	0,011	0,097**	0,079	0,136**	0,033**	0,108**
Titolo	-0,275**	-0,346**	-0,263**	-0,084**	-0,277**	-0,315**	-0,224**	0,001
LIM*Titolo	-0,073	-0,011	0,013	-0,025	-0,181**	-0,017	-0,014	-0,113**
<i>Cittadinanza (rif. italiana)</i>								
LIM	0,045	0,083**	0,047	0,099**	0,051	0,160**	0,046	0,092*
Cittad.	-0,339**	-0,285**	-0,258**	-0,326**	-0,188**	-0,156**	0,118**	-0,082**
LIM*Cittad.	-0,078	-0,156*	-0,051	-0,037	-0,107	-0,098	-0,052	-0,049
<i>Bocciatura (rif. no)</i>								
LIM	0,033	0,068*	0,046	0,098	0,039	0,150**	0,048	0,091*
Bocciatura	-0,267	-0,326**	-0,424**	-0,164	-0,096	-0,221**	-0,377	-0,127**
LIM*Bocciatura	0,175	-0,023	-0,046	-0,010	-0,085	0,018	-0,080	-0,023
<i>Classe</i>								
<i>N. studenti per classe</i>								
LIM	0,878**	-0,182	0,395	0,138	-0,258	-0,339	0,536	0,050
N. studenti	0,028*	-0,006	0,009	0,004	0,024	-0,007	0,010	-0,006
LIM*N. studenti	-0,038*	0,015	-0,015	-0,001	0,016	0,029*	-0,022	0,003

* p-value < 0,05; ** p-value < 0,01.

5. Osservazioni conclusive e implicazioni

Le analisi proposte nel corso del capitolo presentano il grosso vantaggio di poter disporre di una base dati ampia e statisticamente rappresentativa a livello nazionale, nonché di un ricco baglio di informazioni per correggere eventuali errori nelle stime degli effetti della LIM. L'insieme di tali risorse ci ha permesso di produrre analisi approfondite dell'impatto della LIM sulle performance scolastiche degli studenti.

Dai risultati dello studio emerge che, a prescindere dal particolare modello di stima utilizzato, le LIM hanno effetti modesti e non uniformi tra i quattro gradi scolastici considerati e, al contempo, impattano in modo tendenzialmente omogeneo su soggetti diversi e classi dotate di differenti caratteristiche. Abbiamo trovato un effetto degno di nota sulle performance in matematica nelle classi quinte della scuola primaria ed effetti più contenuti alla fine del biennio della

scuola secondaria di II grado. Per gli altri gradi scolastici i segni osservati sono positivi, ma molto vicini allo zero e non significativi nonostante l'ampio campione impiegato.

Al contempo, è doveroso rimarcare che le analisi condotte presentano un chiaro elemento di debolezza, che spinge a qualificare il nostro lavoro come esplorativo nonostante la ricchezza di informazioni utilizzate rispetto a molta della letteratura preesistente. La natura dei dati a nostra disposizione, infatti, ci consente di interpretare casualmente i risultati delle analisi assumendo di aver tenuto sotto controllo tutte le variabili antecedenti al trattamento rilevanti anche per il rendimento scolastico degli studenti. Si tratta evidentemente di un assunto forte, che difficilmente può essere ritenuto pienamente soddisfatto anche considerando l'ampio insieme di variabili di controllo a nostra disposizione. Non possiamo escludere, per esempio, che l'assegnazione della LIM agli insegnanti non sia casuale, per quanto i processi di allocazione dei dispositivi tra scuole e nelle scuole siano caratterizzati da forte aleatorietà, e che gli effetti osservati (e non osservati) derivino da caratteristiche degli insegnanti stessi e dal loro modo di utilizzare la LIM invece che dalla sua sola presenza in classe.

La questione cruciale che si pone, a questo punto, è relativa al massiccio investimento che sta avendo luogo nel nostro Paese sul fronte delle LIM. Il nostro lavoro, coerentemente con la letteratura pregressa sul tema, mostra effetti al più moderati derivanti dall'introduzione di questo dispositivo nelle classi e anche non equalizzatori rispetto alle disuguaglianze ascritte degli studenti. Riflettendo sull'allocazione delle (modeste) risorse destinate al sistema di istruzione nazionale, pare quindi discutibile continuare a promuovere la distribuzione a tappeto di un supporto oneroso e dotato di un'efficacia tanto parziale. Sembra più importante invece sperimentare rigorosamente interventi di accompagnamento all'uso della LIM che possano accrescere l'efficacia degli investimenti già fatti. Più precisamente, raccomandiamo di realizzare studi pilota basati su sperimentazioni controllate, attraverso randomizzazioni con assegnazione casuale delle scuole (o classi) al gruppo di trattamento o di controllo. Precedenti lavori hanno mostrato che tale via è percorribile nelle scuole italiane (Abbiati *et al.*, 2013). È tempo che nelle decisioni di investimento relative alla scuola italiana si adotti un approccio pragmatico, in cui l'adozione di strumenti o innovazioni su larga scala (con grande investimento di denaro pubblico) abbia luogo solo dopo aver provato che si tratta della direzione giusta da seguire.

Riferimenti bibliografici

- Abbiati G., Argentin G., Caputo A., Pennisi A., Romano B., Vidoni D. (2013), "Ricomincio da tre. Lezioni da tre esperienze italiane di analisi controfattuale in ambito educativo", *RIV Rassegna italiana di valutazione*, 55, 1: 21-46.
- Armstrong V., Barnes S., Sutherland R., Curran S., Mills S., Thompson I. (2005), "Collaborative research methodology for investigating teaching and learning: the use of interactive whiteboard technology", *Educational Review*, 57, 4: 457-469.
- Blatchford P., Edmonds S., Martin C. (2003), "Class size, pupil attentiveness and peer relations", *British Journal of Educational Psychology*, 73, 1: 15-36.
- Digregorio P., Sobel-Lojeski K. (2010), "The effects of interactive whiteboards (LIMs) on student performance and learning: a literature review", *Journal of Educational Technology Systems*, 38, 3: 255-312.
- Finn J.D., Achilles C.M. (1990), "Answers and questions about class size: a statewide experiment", *American Educational Research Journal*, 27, 3: 557-577.
- Giusti S., Gui M., Micheli M., Parma A. (2015), "Gli effetti degli investimenti in tecnologie digitali nelle scuole del mezzogiorno", *Materiali UVAL Analisi e studi Documenti Metodi*, 33.
- Glover D., Miller D., Averis D., Door V. (2005), "The interactive whiteboard: a literature survey", *Technology, Pedagogy and Education*, 14, 2: 155-170.
- Glover D., Miller D., Averis D., Door V. (2007), "The evolution of an effective pedagogy for teachers using the interactive whiteboard in mathematics and modern languages: an empirical analysis from the secondary sector", *Learning, Media and Technology*, 32, 1: 5-20.
- Gui M., Argentin G. (2011), "Digital skills of internet natives: different forms of digital literacy in a random sample of northern Italian high school students", *New Media & Society*, 13, 6: 963-980.
- Higgins S., Beauchamp G., Miller D. (2007), "Reviewing the literature on interactive whiteboards", *Learning, Media and Technology*, 32, 3: 213-225.
- Kitchen S., Finch S., Sinclair R. (2007), *Harnessing Technology Schools Survey*, Becta, Coventry, testo disponibile al sito: http://dera.ioe.ac.uk/1554/1/-becta_2007_htssfindings_report.pdf, data di consultazione: 12 settembre 2016.

-
- Lewin C., Somekh B., Steadman S. (2008), "Embedding interactive whiteboards in teaching and learning: the process of change in pedagogic practice", *Education and Information Technologies*, 13, 4: 291-303.
- Martini A., Sisti M. (2009), *Valutare il successo delle politiche pubbliche*, il Mulino, Bologna.
- Rettore E., Checchi D. (2014), "La valutazione degli esiti sugli apprendimenti degli alunni", in S. Girardi, V. Pandolfini (a cura di), *Rapporto finale del progetto cl@ssi 2.0*, testo disponibile al sito: https://www.irvap.it/sites/irvap.fbk.eu/fin-les/rapporto_finale_classi_2.0.pdf, data di consultazione: 5 settembre 2016.
- Rosenbaum P.R., Rubin D.B. (1983), "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, 1: 41-55.
- Schroeder R. (2007), "Active learning with interactive whiteboards: a literature review and a case study for college freshmen", *Communications in Information Literacy*, 1, 2: 63-73.
- Schuck S., Kearney M. (2007), *Exploring Pedagogy with Interactive Whiteboards: A Case Study of Six Schools*, University of Technology Sydney, testo disponibile al sito: <http://www.eddev.uts.edu.au/teachered/research/-LIMproject/pdfs/LIMreportweb.pdf>, data di consultazione: 9 settembre 2016.
- Slay H., Siebörgher I., Hodgkinson-Williams C. (2008), "Interactive whiteboards: real beauty or just 'lipstick'?", *Computers & Education*, 51, 3: 1321-1341.
- Smith F., Hardman F., Higgins S. (2006), "The impact of interactive whiteboards on teacher-pupil interaction in the National Literacy and Numeracy Strategies", *British Educational Research Journal*, 32, 3: 443-457.
- Thompson J., Flecknoe M. (2003), "Raising attainment with an interactive whiteboard in Key Stage 2", *Management in Education*, 17, 3: 29-33.
- Wall K., Higgins S., Smith H. (2005), "The visual helps me understand the complicated things: pupil views of teaching and learning with interactive whiteboards", *British Journal of Educational Technology*, 36, 5: 851-867.
- Wood R., Ashfield J. (2008), "The use of interactive whiteboard for creative teaching and learning in literacy and mathematics: a case study", *British Journal of Educational Technology*, 39, 1: 84-96.

Parte seconda
I dati INVALSI per il governo del sistema scolastico

11. *Il tempo pieno e la dispersione dei voti*

Full-time school scheme and inequality in students' learning outcomes

di Giulia Bovini, Marta De Philippis, Paolo Sestito

Il lavoro analizza l'effetto del trascorrere più tempo a scuola sulla media e sulla distribuzione degli apprendimenti degli studenti della scuola primaria. In particolare, si analizza se trascorrere più ore a scuola invece che a casa contribuisca a mitigare l'influenza della famiglia di provenienza e a ridurre i gap di apprendimento tra studenti provenienti da contesti familiari differenti. Si studia il caso dell'Italia, dove coesistono nella scuola primaria due schemi orari, il *tempo pieno* (40 ore settimanali) e il *tempo normale* (da 24 a 30 ore settimanali). L'effetto causale del tempo pieno viene identificato sfruttando variazioni nel tempo della frazione di classi a tempo pieno in una data scuola, che derivano plausibilmente da cambiamenti nell'offerta di tale schema orario e non sono quindi correlate con caratteristiche degli studenti e delle famiglie che scelgono l'uno o l'altro schema orario. I risultati principali mostrano che il tempo pieno ha un effetto quasi nullo e relativamente costante sui decili della distribuzione dei punteggi di italiano, mentre ha un effetto positivo sui punteggi di matematica e concentrato nei decili più bassi della distribuzione. Emerge inoltre una modesta, seppur imprecisamente stimata, riduzione dei gap di apprendimento tra studenti provenienti da contesti familiari differenti. Gli effetti sia sulla media sia sulla dispersione degli esiti scolastici sono maggiori per le classi quinte che per le classi seconde.

This paper analyzes the effect of increasing the amount of time spent at school on the mean and on the distribution of students' performance. In particular, we evaluate whether spending more time at school instead of at home mitigates family and community influences and therefore helps narrowing the observed achievement gaps among pupils with different parental backgrounds. We study the case of Italy, where two instructional schemes – that entail a different amount of time spent at school – coexist in primary schools, namely the “tempo pieno” scheme (TP, 40 hours per week) and the “tempo normale/modulare” scheme (TN, 24 to 30 hours per week). Identification comes from plausibly supply-driven variation in the share of TP classes offered within a given school and grade over subsequent cohorts of second and fifth graders. We find that the effect of time spent at school upon reading is very small and relatively constant among the deciles of the test score distribution within school, while the effect upon math is positive and stronger in the bottom deciles of the grade distribution. Moreover, there is evidence of a reduction, even if imprecisely estimated, of achievement gaps among students having different socio-economic background. The effects are stronger for students in the fifth grade rather than for students in the second grade.

1. Introduzione

Dal momento che il capitale umano costituisce una determinante fondamentale dei redditi individuali e della crescita economica (Barro, 2001; Hanushek e Woessmann, 2008), la stima dei rendimenti medi delle diverse componenti della funzione di produzione di capitale umano è da sempre uno dei temi cruciali affrontati dalla letteratura economica. Tuttavia, l'effetto di tali componenti non solo sulla media ma anche sull'intera distribuzione e dispersione dei risultati scolastici – tramite il modo in cui le risorse sono allocate e i diversi *input* interagiscono tra di loro – è relativamente meno esplorato. Le potenziali implicazioni sulla disegualianza del reddito o sulla mobilità sociale potrebbero invece essere di grande interesse.

Questo lavoro analizza gli effetti dell'espansione dell'offerta di risorse scolastiche, che sono un facile oggetto di *policy*, sulla distribuzione dei risultati scolastici e sui gap di apprendimento tra studenti provenienti da contesti familiari diversi. In particolare, questa ricerca si focalizza sull'impatto dell'incremento del tempo trascorso a scuola, svolgendo attività sia formative sia ricreative. Un aumento nella quantità di tempo trascorso a scuola, in un ambiente relativamente standardizzato e omogeneo, di solito comporta una contestuale riduzione del tempo dedicato allo studio a casa, dove invece i contesti familiari di provenienza assumono una maggiore importanza. I risultati di questo studio verificano se ciò possa mitigare la dispersione nei gap di apprendimento tra studenti con caratteristiche socio-economiche differenti.

Il lavoro si focalizza sulle scuole primarie in Italia, dove coesistono due diversi schemi orari: il *tempo pieno* (TP da ora in avanti, che prevede 40 ore settimanali di permanenza a scuola) e il *tempo normale* (TN da ora in avanti, che prevede dalle 24 alle 30 ore settimanali di permanenza a scuola). Per misurare gli apprendimenti scolastici si utilizzano i dati INVALSI, che contengono una ricca gamma di informazioni sugli studenti e sul loro contesto familiare, oltre ai punteggi ottenuti nei test standardizzati di matematica e italiano al termine delle classi seconda e quinta della scuola primaria.

Il semplice paragone dei risultati degli alunni iscritti alle classi a TP e a TN non restituirebbe l'effetto causale del maggior tempo trascorso a scuola, poiché tali studenti differiscono probabilmente anche in termini di altre caratteristiche non osservabili che concorrono a determinarne gli esiti scolastici. Il lavoro sfrutta pertanto la struttura longitudinale dei dati, utilizzando variazioni nel tempo, all'interno di una data scuola e di un dato grado, della frazione di classi a TP. La strategia identificativa si basa sull'ipotesi che tali variazioni non riflettano cambiamenti significativi dal lato della domanda di TP, ma piuttosto dipendano in gran parte da cambiamenti dal lato dell'offerta, determinati da variazioni nella disponibilità di fondi per finanziare le classi a TP e dalle regole che disciplinano il numero minimo di richieste necessarie per attivare le classi a TP. A corroborare la veridicità dell'ipotesi sottostante alla strategia identificativa, il lavoro presenta evidenza che variazioni nella frazione di classi a TP, all'interno di una stessa scuola e di un dato grado nel tempo, non appaiono associate a cambiamenti nelle caratteristiche osservabili degli studenti iscritti.

I risultati del lavoro conducono a tre conclusioni principali. In primo luogo, un maggiore tempo trascorso a scuola ha un effetto praticamente nullo e non statisticamente significativo sul punteggio medio di italiano, mentre ha un effetto positivo sul punteggio medio di matematica. Se una scuola passasse da non avere alcuna classe a TP ad avere solo classi a TP, il risultato medio nelle prove di matematica aumenterebbe del 2,2%. Tale effetto è statisticamente significativo ancorché quantitativamente piccolo se posto a confronto col maggior costo del TP rispetto al TN: i maggiori apprendimenti equivalgono a circa un decimo della differenza media negli apprendimenti in matematica tra uno studente regolare e uno studente in ritardo nel suo percorso scolastico. D'altro lato, il TP *ceteris paribus* innalza di circa un quarto il fabbisogno di personale docente rispetto alle classi TN. Nel valutare la rilevanza di tali risultati vanno però anche tenuti presenti gli altri due aspetti posti in luce dalle nostre stime. Mentre l'effetto sugli esiti di italiano è costante rispetto ai diversi decili della distribuzione, l'effetto sui punteggi di matematica è concentrato nei decili più bassi della distribuzione. Il TP determina perciò una riduzione nella dispersione degli apprendimenti di matematica. Inoltre, i risultati suggeriscono che il TP comporti una modesta, seppur imprecisamente stimata, riduzione dei gap di apprendimento degli studenti immigrati o provenienti da contesti familiari più difficili. Infine, gli effetti sia sulla media sia sulla dispersione degli esiti scolastici sono maggiori per le classi quinte rispetto alle seconde, suggerendo l'esistenza di effetti cumulativi.

Questo lavoro si inserisce nella letteratura che studia l'effetto di più tempo trascorso a scuola e, in generale, di una maggiore qualità del sistema scolastico, sui risultati degli studenti. Lee e Barro (2001) e Woessman (2003) sfruttano variazioni tra Paesi nella lunghezza del calendario scolastico e riportano effetti modesti sulla performance degli studenti nei test internazionali standardizzati. Tuttavia, a causa di limiti nei dati, entrambi gli studi non affrontano interamente la possibilità che tali variazioni riflettano differenze non osservabili nella popolazione di riferimento, non includendo per esempio effetti fissi a livello di Paese. Lavy (2015) utilizza i dati dell'indagine del *Program of International Student Assessment* (PISA) e analizza l'effetto di differenze nelle ore di insegnamento dedicate alle diverse materie sui rendimenti scolastici. Utilizzando effetti fissi a livello di studente, riporta un effetto positivo e statisticamente significativo sui risultati scolastici, maggiore in scuole di più alta qualità. La letteratura più recente utilizza invece esperimenti naturali. Un primo filone si basa su variazioni del calendario scolastico dovute a chiusure non previste della scuola a causa di condizioni meteorologiche estreme (Marcotte, 2007; Marcotte e Hemelt, 2008; Hansen, 2011; Goodman, 2014). Questi studi trovano generalmente effetti positivi. Tuttavia si basano su variazioni limitate e inaspettate nel numero di giorni, che non implicano una corrispondente ragionata revisione del programma scolastico per adattarlo al minore tempo trascorso a scuola. Un altro filone della letteratura si basa invece su variazioni del calendario dovute a politiche scolastiche. Pisch-

ke (2007) valuta le conseguenze dell'eccezionalmente breve calendario scolastico in vigore in Germania nel 1966-67, mentre Parinduri (2014) analizza le conseguenze dell'eccezionalmente lungo calendario scolastico in vigore in Indonesia nel 1978-79. Entrambi mostrano che più giorni a scuola migliorano i risultati scolastici; il primo non riscontra effetti duraturi sui redditi futuri, il secondo trova invece un effetto positivo anche nel lungo termine. Agüero e Beleche (2013) e Aucejo e Romano (2014) riscontrano similmente un effetto positivo sugli apprendimenti, basandosi invece su variazioni nel numero di giorni di scuola indotte da cambiamenti nella data degli esami o nel calendario scolastico, in Messico e negli Stati Uniti, rispettivamente. Altri lavori (Bellei, 2009; Lavy, 2012) analizzano riforme che incrementano le ore giornaliere trascorse a scuola, invece del numero di giorni, in Cile e in Israele, rispettivamente. Entrambi concludono che i risultati scolastici migliorano all'aumentare del tempo trascorso a scuola. Nel contesto italiano, Battistin e Meroni (2016) analizzano l'effetto di un incremento delle ore di insegnamento per alcune coorti di studenti di scuola secondaria di I grado iscritti a scuole del Sud Italia ove gli esiti scolastici medi sono bassi e riportano, coerentemente con il nostro lavoro, un effetto maggiore sugli apprendimenti di matematica.

Questo lavoro contribuisce alla letteratura sopra citata in svariati modi. In primo luogo, si analizza un contesto dove la variazione nel tempo trascorso a scuola si riferisce alla lunghezza della giornata scolastica e non dei giorni di scuola e dove tale variazione è nota in anticipo – permettendo quindi agli insegnanti di aggiustare il programma scolastico di conseguenza – e quantitativamente considerevole: il TP aumenta infatti le ore giornaliere trascorse a scuola del 30% circa. Specialmente se si ipotizza che il tempo trascorso a scuola abbia effetti non lineari, è importante analizzare l'effetto per variazioni consistenti dell'orario scolastico. In secondo luogo, lo studio si focalizza esplicitamente sull'effetto lungo l'intera distribuzione dei voti e sui gap di apprendimento, mentre la maggior parte dei lavori precedenti si focalizza solo sull'effetto medio (parziali eccezioni sono Lavy, Paserman e Schlosser 2012; Lavy 2015; Marcotte e Hemelt 2008). Una recente eccezione è costituita dal lavoro di Huebener, Kuger e Marcus (2017), che analizza l'effetto di aumentare le ore di insegnamento nelle scuole superiori, in Germania, sulla distribuzione degli apprendimenti e riporta un aumento della dispersione nei punteggi. Il contesto, tuttavia, è notevolmente diverso da quello del presente lavoro, per quanto riguarda sia l'età degli studenti sia il contenuto, puramente formativo, dell'ulteriore tempo trascorso a scuola.

Il lavoro si sviluppa come segue: il paragrafo 2 descrive in dettaglio il contesto istituzionale; il paragrafo 3 descrive i dati; il paragrafo 4 espone la strategia identificativa utilizzata per ottenere i risultati descritti nel paragrafo 5. Il paragrafo 6, infine, conclude.

2. Il contesto istituzionale

Il sistema scolastico italiano è organizzato intorno a tre cicli didattici: le scuole primarie (o elementari, per alunni dall'età di 6 anni all'età di 11 anni), le scuole secondarie di I grado (da 11 a 14 anni) e le scuole secondarie di II grado (da 14 a 19 anni). Il lavoro si focalizza sulle scuole primarie, poiché è il ciclo didattico in cui la differenza tra i diversi schemi orari è più marcata¹.

La scelta dello schema orario viene effettuata all'inizio della scuola primaria e le classi così formate si mantengono nel corso dei cinque anni di scuola primaria. Nel modulo a TN le lezioni durano da 24 a 30 ore settimanali (frequentemente 27) e sono distribuite tipicamente su sei giorni alla settimana, di solito la mattina. Il modulo a TP invece prevede che gli alunni trascorrono a scuola 40 ore alla settimana, divise in 5 giorni, di solito dalle 8.30 alle 16.30. Gli studenti rimangono quindi a scuola anche durante la pausa pranzo e nelle ore pomeridiane, quando sono impegnati sia in attività istruttive (per esempio svolgono i compiti sotto la supervisione degli insegnanti) sia in attività ricreative (con lo scopo di socializzare e aumentare le interazioni tra studenti). Infine, gli schemi del TN e del TP differiscono nel numero di insegnanti a disposizione per alunno: a partire dagli anni Ottanta infatti gli studenti delle classi a TN sono seguiti da tre insegnanti ogni due classi, mentre a quelli delle classi a TP sono assegnati due insegnanti per classe².

¹ Il TP è stato introdotto in centro Italia alla fine degli anni Sessanta (dalla legge 820/1971).

² Prima della cosiddetta riforma Gelmini (legge 133/2008 e 196/2008), era richiesto agli insegnanti a TP di essere presenti in classe contemporaneamente almeno due ore alla settimana (compresenza). Si controlla per differenze nel numero di insegnanti, indotte dalla riforma Gelmini, includendo un controllo per le coorti esposte a tale riforma. Inoltre, sebbene la riforma Gelmini abbia introdotto la possibilità di uno schema a TN di sole 24 ore settimanali, che comporta l'assegnazione alla classe di un solo insegnante, tale schema orario è estremamente poco diffuso (interessa solo lo 0,5% delle classi secondo Battistin e Schuller, 2013).

Al momento dell'iscrizione alla scuola primaria i genitori esprimono le loro preferenze riguardo alla scuola e allo schema orario. Tuttavia l'offerta di TP non sempre soddisfa la domanda e ciò può verificarsi per due ragioni principali. Da un lato per una domanda troppo limitata, poiché la legge prevede un numero minimo di studenti per attivare una classe: se la domanda di TP in una scuola è insufficiente, allora è più difficile attivare la classe a TP. Dall'altro lato per un eccesso di domanda, poiché la legge indica anche un numero massimo di studenti per classe: poiché attivare classi a TP è più costoso, le risorse finanziarie allocate alle scuole potrebbero essere insufficienti in alcuni anni per soddisfare pienamente la richiesta di classi a TP, limitandone così l'offerta. In pratica, in caso di eccesso di domanda, i presidi scelgono quali studenti assegnare alle classi a TP sulla base di diversi criteri, quali la distanza della scuola dal domicilio o la presenza di fratelli nella scuola o la presenza di forme di invalidità.

3. Dati

Il lavoro utilizza i dati sui test standardizzati di italiano e matematica proposti a tutti gli studenti della seconda e della quinta primaria alla fine dell'anno scolastico e raccolti dall'INVALSI. Le scuole, oltre a riportare i punteggi in italiano e in matematica degli studenti, forniscono altre informazioni quali lo schema orario, la numerosità di ciascuna classe e le principali caratteristiche demografiche e socio-economiche degli alunni. Questionari somministrati agli studenti in concomitanza del test per la classe quinta rilevano ulteriori informazioni circa il contesto di provenienza e l'uso del tempo al di fuori della scuola. Il lavoro sfrutta la struttura longitudinale dei dati, collegando i diversi plessi scolastici nel tempo, per gli anni scolastici dal 2011-12 al 2014-15.

Il dataset INVALSI è comparabile a dataset simili disponibili in altri Paesi ma ha il vantaggio di contenere informazioni molto dettagliate sulle caratteristiche delle famiglie, quali il livello di istruzione dei genitori, il loro status occupazionale e indicatori del reddito. Tali informazioni permettono di investigare l'impatto sui gap tra studenti provenienti da contesti familiari diversi. Vi sono tuttavia alcuni limiti: l'assenza di informazioni sugli insegnanti e la presenza di diversi episodi di manipolazione dei risultati (il cosiddetto *cheating*), specialmente per i test somministrati nelle scuole primarie e in alcune regioni del Sud Italia (si vedano Quintano, Castellano e Longobardi, 2009; Falzetti, Longobardi e Sestito, 2015; Pereda-Fernandez, 2016). Per tale motivo il lavoro utilizza i punteggi corretti per il *cheating*, secondo la procedura proposta dall'INVALSI. Analisi basate sui punteggi grezzi restituiscono risultati simili.

4. Strategia empirica

La stima dell'effetto del TP sugli apprendimenti è resa problematica dal meccanismo di allocazione degli alunni alle scuole e ai diversi schemi orari. Le famiglie infatti si auto-selezionano in determinate scuole (e schemi orari) anche sulla base di caratteristiche probabilmente correlate coi risultati scolastici. Questo meccanismo rende difficile interpretare i risultati derivanti da una semplice comparazione della performance degli studenti iscritti al TP e al TN: ogni differenza potrebbe essere legata non allo schema orario di per sé ma a fattori latenti correlati con la scelta di frequentare classi a TP.

Tali problematiche vengono affrontate nel seguente modo. In primo luogo, si includono nella specificazione da stimare effetti fissi a livello di scuola, che controllano per differenze tra le scuole in caratteristiche non osservabili costanti nel tempo quali la qualità degli studenti, degli insegnanti e delle strutture. Inoltre, vengono inclusi anche trend lineari specifici per ogni scuola, che controllano per variazioni lineari nel tempo, all'interno di ogni scuola, di tali caratteristiche non osservabili. In secondo luogo, l'analisi viene svolta aggregando i dati individuali a livello di scuola-grado-anno, eliminando in questo modo i problemi di selezione in classi diverse all'interno della stessa scuola. L'identificazione dell'effetto causale del TP proviene quindi dalla variazione nel tempo, per una data scuola e in un dato grado, della frazione di classi a TP (in deviazione dal trend lineare). Questa strategia di identificazione è stata utilizzata in altri studi (si veda per esempio Lavy, Paserman e Schlosser, 2012).

Si stima quindi la seguente equazione:

$$y_{sgt} = \delta FrazTP_{sgt} + \gamma X_{sgt} + \mu_{sg} + \theta_{sg} t + \varphi_t + \varepsilon_{sgt} \quad (1)$$

dove s indica la scuola; g il grado; t l'anno; y_{sgt} è il momento di interesse della distribuzione del logaritmo dei voti di matematica o di italiano nella scuola s , nel grado g e nell'anno t ; $FrazTP_{sgt}$ indica la frazione di classi a TP nella scuola s , nel grado g e nell'anno t ; X_{sgt} indica un insieme di controlli che variano nel tempo a livello di scuola-grado (la percentuale di studenti maschi; la percentuale di studenti italiani; la percentuale di studenti non ripetenti; la percentuale di madri e di padri con almeno il diploma di scuola superiore; la dimensione della scuola; il numero di classi e una *dummy* per le coorti esposte alla riforma Gelmini); μ_{sg} e φ_t sono rispettivamente effetti fissi a livello di scuola-grado e di anno. θ_{sgt} è il trend lineare a livello di scuola-grado. Infine ε_{sgt} rappresenta un termine di errore statistico.

La variabilità negli anni all'interno della stessa scuola nel numero di classi a TP da un anno scolastico a quello successivo, che è quella utilizzata per la stima degli effetti di interesse, si attesta intorno al 15%.

4.1. Statistiche descrittive

La tab. 1 riporta alcune statistiche descrittive relative agli studenti iscritti a classi a TP e a classi a TN, rispettivamente.

Tab. 1 – Statistiche descrittive – TN e TP

	TN	TP	Diff: TN - TP
I – Caratteristiche dell'area locale			
Popolazione	142.841	375.855	-233.014,42
Frazione residente in capoluoghi di provincia	0,25	0,37	-0,11
Tasso di disoccupazione (2011-14)	0,13	0,10	0,03
Tasso di non partecipazione femminile (2011-13)	0,29	0,20	0,09
II – Caratteristiche della scuola e degli alunni			
Dimensione della classe	17,27	19,00	-1,73
Frazione di alunni maschi*	0,51	0,51	-0,00
Frazione di alunni con la cittadinanza italiana*	0,92	0,85	0,06
Frazione di studenti regolari*	0,96	0,97	-0,01
Frazione di studenti che parlano prevalentemente l'italiano o il dialetto a casa*	0,93	0,90	0,04
Frazione di studenti che hanno frequentato l'asilo*	0,28	0,42	-0,14
Frazione di padri con almeno il diploma di scuola superiore*	0,49	0,53	-0,04
Frazione di madri con almeno il diploma di scuola superiore*	0,56	0,62	-0,05
Frazione di madri che non lavorano*	0,50	0,34	0,16
Indicatore ESCS medio*	0,02	0,11	-0,09
Punteggio medio nella prova di italiano (NC)	0,02	-0,04	0,06
Punteggio medio nella prova di italiano (C)	-0,02	0,03	-0,05
Punteggio medio nella prova di matematica (NC)	0,02	-0,04	0,07
Punteggio medio nella prova di matematica (C)	-0,01	0,02	-0,03
Numero di studenti		4.794.955	

* indica che le frazioni sono condizionate rispetto alle risposte non mancanti.

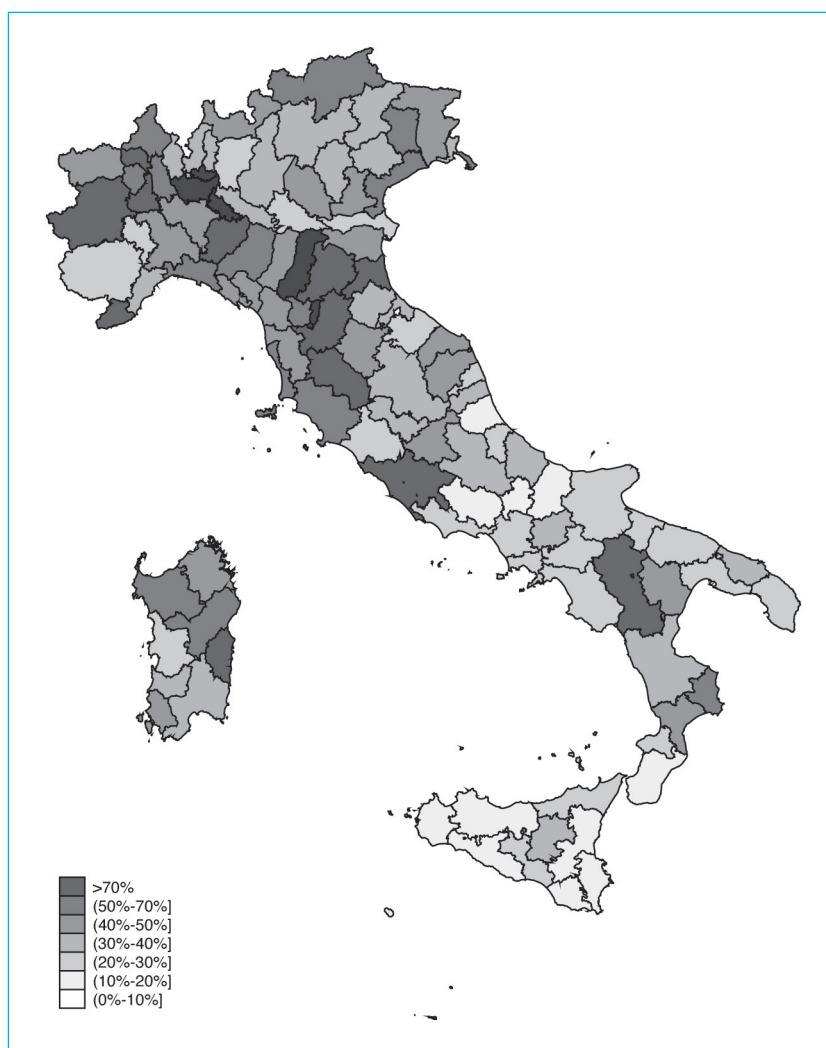
Nota: L'indicatore ESCS misura la condizione socio-economica del nucleo familiare, basata sull'istruzione e l'occupazione dei genitori e sulle risorse disponibili a casa. Le informazioni relative all'ESCS e alla lingua prevalentemente parlata a casa sono disponibili solo per gli alunni della quinta primaria. I punteggi corretti per il *cheating* (C) non sono disponibili per l'anno scolastico 2010-11. L'indicatore ESCS e i punteggi nelle prove di italiano e matematica sono standardizzati per avere media 0 e deviazione standard 1 in un dato grado-anno.

Gli studenti di classi a TP vivono con maggiore probabilità in comuni più densamente popolati e nei capoluoghi di provincia, in aree dove il mercato del lavoro locale è più attivo e dove i tassi di disoccupazione e di inattività femminile sono più bassi. Le classi a TP sono in media di maggiori dimensioni e frequentate in misura maggiore da cittadini non italiani, coerentemente con la maggiore presenza del TP in contesti urbani e nel Nord Italia. Gli studenti a TP hanno frequentato l'asilo nido con maggiore probabilità e hanno più frequentemente una madre lavoratrice. Inoltre provengono generalmente da un contesto familiare più privilegiato: il loro indice di contesto socio-economico

(ESCS) è più alto. Infine, i risultati di matematica e italiano, una volta corretti per il *cheating*, sono in media più alti per gli studenti a TP.

La mappa della fig. 1 mostra la percentuale di studenti iscritti a una classe a TP nelle diverse province italiane. La diffusione del modulo a TP è notevolmente eterogenea ed è significativamente più alta al Nord rispetto al Sud, con alcune notabili eccezioni, come per esempio la regione Basilicata.

Fig. 1 – Percentuale di alunni iscritti al TP a livello provinciale



Nota: I dati sono forniti dall'INVALSI e si riferiscono al numero di studenti iscritti al tempo pieno nella classe I primaria, anno scolastico 2014-15.

4.2. Validità delle ipotesi identificative

L'ipotesi alla base della strategia di identificazione è che variazioni nella frazione di classi a TP nel tempo, in una certa scuola e in un dato grado, non siano sistematicamente associate a cambiamenti nella domanda, quindi nelle caratteristiche (non osservabili) degli studenti e degli insegnanti, ma che derivino invece da fluttuazioni dell'offerta, a causa delle diverse risorse pubbliche disponibili o di movimenti idiosincratichi intorno alla soglia minima di iscrizioni per attivare una classe.

Tab. 2 – Test di bilanciamento

	<i>Studenti iscritti</i>			<i>Studenti assenti</i>		
Fraz. TP	5,905*** (0,315)	-0,728*** (0,226)	-1,000*** (0,289)	1,038*** (0,042)	0,369*** (0,075)	0,071 (0,094)
Oss.	118.662	118.662	118.662	118.662	118.662	118.662
Media		39,86			3,25	
	<i>Fraz. studenti maschi</i>			<i>Fraz. risposte mancanti</i>		
Fraz. TP	0,002*** (0,001)	0,004 (0,002)	0,002 (0,004)	0 (0,000)	0 (0,000)	0 (0,000)
Oss.	118.662	118.662	118.662	118.662	118.662	118.662
Media		0,51			0	
	<i>Fraz. studenti nativi</i>			<i>Fraz. risposte mancanti</i>		
Fraz. TP	-0,054*** (0,002)	-0,006*** (0,002)	-0,003 (0,003)	0,001*** (0,000)	0 (0,000)	0 (0,001)
Oss.	118.633	118.633	118.633	118.662	118.662	118.662
Media		0,9			0	
	<i>Fraz. studenti regolari</i>			<i>Fraz. risposte mancanti</i>		
Fraz. TP	0,005*** (0,001)	0,003*** (0,001)	-0,002 (0,002)	0,000*** (0,000)	0,002** (0,001)	0,003** (0,001)
Oss.	118.641	118.641	118.641	118.662	118.662	118.662
Media		0,96			0	
	<i>Fraz. madri con almeno il diploma</i>			<i>Fraz. risposte mancanti</i>		
Fraz. TP	0,043*** (0,002)	0,018*** (0,003)	-0,002 (0,004)	0,040*** (0,003)	-0,008 (0,005)	-0,013 (0,008)
Oss.	106.403	106.403	106.403	118.662	118.662	118.662
Media		0,58			0,18	
	<i>Fraz. padri con almeno il diploma</i>			<i>Fraz. risposte mancanti</i>		
Fraz. TP	0,037*** (0,003)	0,014*** (0,003)	-0,002 (0,005)	0,043*** (0,003)	-0,008 (0,005)	-0,012 (0,008)
Oss.	106.149	106.149	106.149	118.662	118.662	118.662
Media		0,49			0,2	
Eff. fissi di scuola-grado	No	Si	Si	No	Si	Si
Eff. fissi e trend di scuola-grado	No	No	Si	No	No	Si

* p < 0,1, ** p < 0,05, *** p < 0,01.

Nota: L'unità di osservazione è la scuola-grado, la variabile indipendente è la frazione di classi a TP. Le variabili dipendenti sono: il numero di studenti iscritti; il numero di studenti assenti durante almeno uno dei test; la frazione di studenti maschi; con la cittadinanza italiana; regolari (cioè né anticipatori né posticipatori); di madri e padri con almeno il diploma superiore; di studenti per i quali non sono riportate tali informazioni. Tutte le regressioni includono effetti fissi a livello di anno, le colonne 1 e 4 riportano stime che non includono effetti fissi a livello di scuola-grado, mentre le colonne 2 e 5 riportano stime che includono effetti fissi a livello di scuola-grado. Le colonne 3 e 6 riportano stime che includono sia effetti fissi sia trend lineari a livello di scuola-grado. Gli errori standard sono clusterizzati a livello di scuola.

La tab. 2 mostra una serie di test di bilanciamento. Essi mostrano che le variazioni nel tempo delle caratteristiche osservabili degli studenti in una data scuola-grado non sono correlate con la frazione di classi a TP attivate in tale scuola-grado nei diversi anni presi in considerazione³. Le colonne 1 e 4 riportano i coefficienti delle regressioni che non includono gli effetti fissi di scuola-grado. Le altre colonne invece si riferiscono alle specificazioni finali del lavoro e includono gli effetti fissi di scuola-grado (colonne 2, 3, 5, 6) e il trend specifico di scuola-grado (solo colonne 3 e 6). In generale, l'introduzione di effetti fissi di scuola-grado e del trend specifico a livello di scuola-grado elimina la gran

³ In particolare, le caratteristiche osservabili sono: il numero di studenti iscritti; il numero di studenti presenti ad almeno uno dei test; la percentuale di studenti maschi; la percentuale di studenti non italiani; la percentuale di studenti regolari (né anticipatori né posticipatori); la percentuale di padri e madri con almeno il diploma di scuola superiore; la percentuale di studenti per i quali non sono riportate tali informazioni.

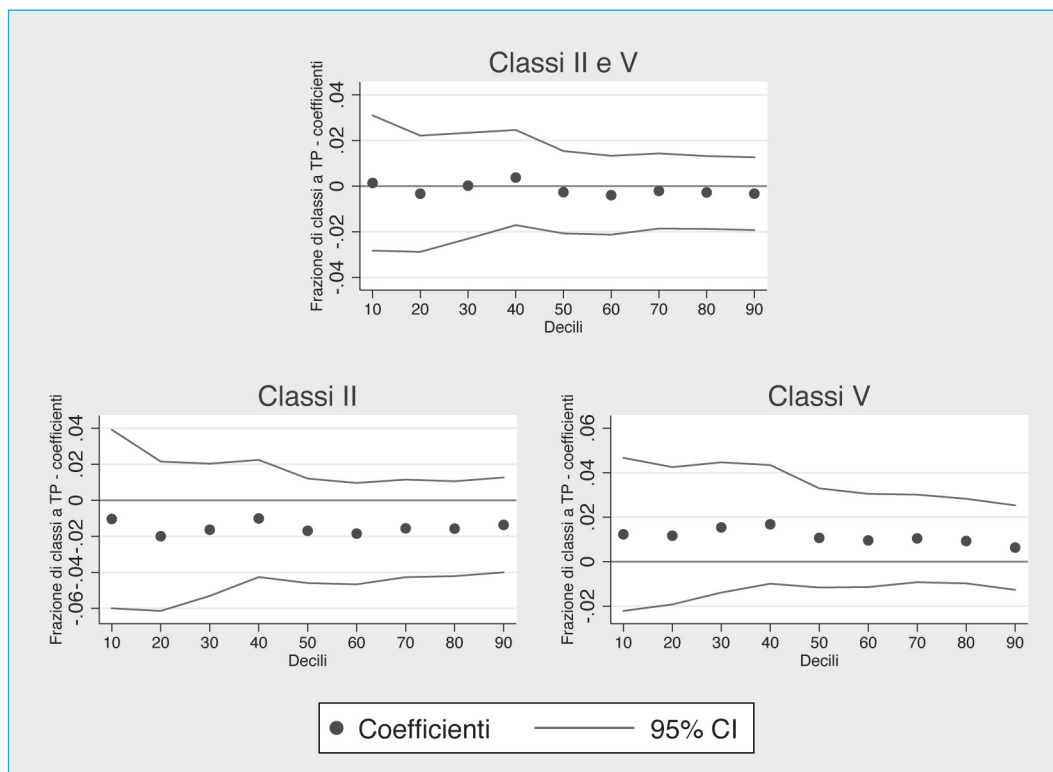
parte della correlazione tra $FrazTP_{sgt}$ e le caratteristiche osservabili degli studenti nella scuola s nel grado g nell'anno t . Se le caratteristiche osservabili e non osservabili sono correlate, questo risultato corrobora la strategia identificativa.

5. Risultati

Le figg. 2 e 3 riassumono i risultati principali, riportando i coefficienti ottenuti dalla stima dell'equazione (1), dove le variabili dipendenti sono i valori dei diversi decili della distribuzione dei punteggi (in logaritmo) nell'anno t nella scuola s e nel grado g , rispettivamente per matematica e italiano. La fig. 2 si riferisce ai punteggi di italiano: l'effetto del TP sugli apprendimenti risulta praticamente nullo e costante lungo i decili della distribuzione. Stimando la specificazione (1) separatamente per grado, l'effetto è leggermente negativo per le classi seconde e leggermente positivo per le classi quinte, sebbene i coefficienti non siano mai statisticamente significativi. Diversamente, l'effetto sui punteggi di matematica (fig. 3) è positivo, statisticamente significativo e maggiore nella coda sinistra della distribuzione: passare dal non avere classi a TP all'avere solo classi a TP innalzerebbe il primo decile della distribuzione dei punteggi in un dato scuola-grado di circa il 3,6%, mentre innalzerebbe il nono decile di circa lo 0,8%. La stima separata per grado mostra che l'effetto eterogeneo sulla distribuzione dei punteggi è più marcato per le classi quinte.

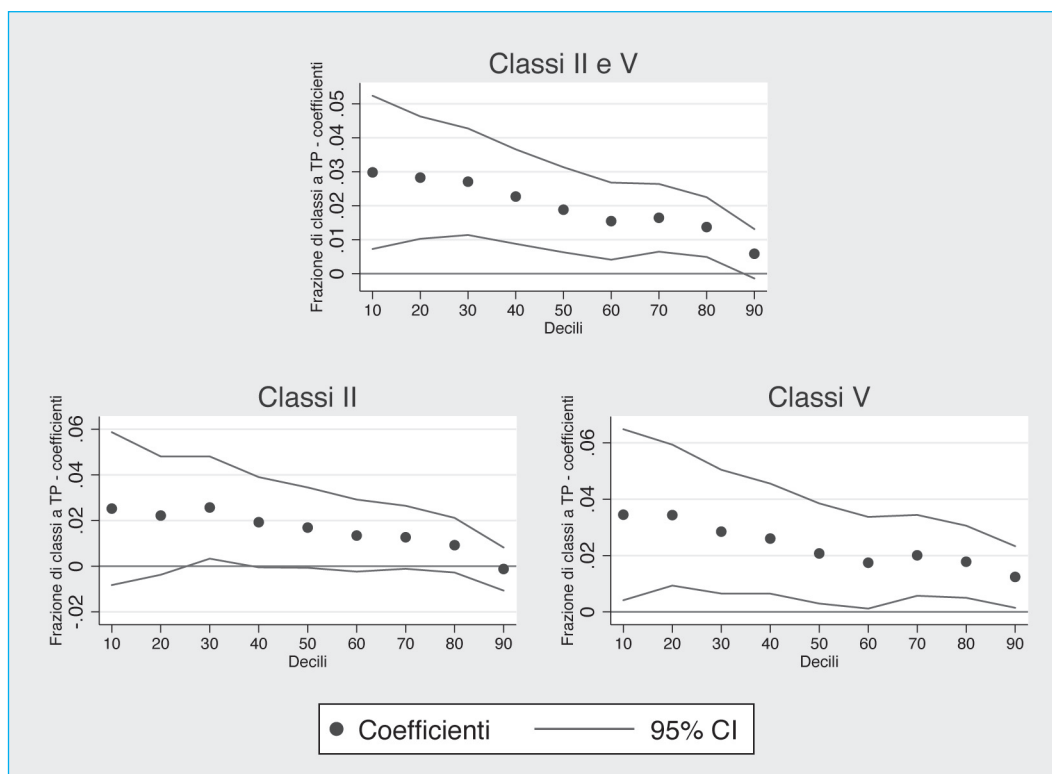
Le tabb. 3 e 4 si concentrano su alcune misure di tendenza centrale e dispersione della distribuzione dei livelli di apprendimento. La tab. 3 mostra la stima degli effetti sulla media e sulla mediana. Le colonne 1-3 mostrano i risultati relativi a italiano, mentre le colonne 4-6 riportano i risultati relativi a matematica. Coerentemente con quanto mostrato nelle figg. 2 e 3, l'effetto sulla media e la mediana dei punteggi di italiano è nullo, mentre è positivo per quanto riguarda la matematica. Passare dal non avere classi a TP all'avere solo classi a TP innalzerebbe il punteggio medio in un dato scuola-grado di circa il 2,2%. Inoltre, l'effetto è maggiore per le classi quinte.

Fig. 2 – Effetto del TP sui decili della distribuzione dei punteggi (in logaritmo) di italiano non corretti per il cheating



Nota: I decili della distribuzione sono calcolati a livello di scuola-grado. Gli errori standard sono clusterizzati a livello di scuola-grado.

Fig. 3 – Effetto del TP sui decili della distribuzione dei punteggi (in logaritmo) di matematica non corretti per il cheating



Nota: I decili della distribuzione sono calcolati a livello di scuola-grado. Gli errori standard sono clusterizzati a livello di scuola-grado.

Tab. 3 – Misure di tendenza centrale

	Italiano			Matematica		
	II e V	II	V	II e V	II	V
<i>Punteggio medio</i>						
Fraz. TP	-0,002 (0,009)	-0,013 (0,014)	0,009 (0,011)	0,022** (0,010)	0,011 (0,015)	0,032** (0,013)
<i>Punteggio mediano</i>						
Fraz. TP	-0,004 (0,009)	-0,016 (0,015)	0,008 (0,012)	0,021* (0,011)	0,013 (0,018)	0,029** (0,014)
Controlli studenti	Sì	Sì	Sì	Sì	Sì	Sì
Controlli scuole	Sì	Sì	Sì	Sì	Sì	Sì
Oss.	81.696	38.442	43.254	81.698	38.442	43.256

* $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$.

Nota: L'unità di osservazione è la scuola-grado. La variabile indipendente è la frazione di classi a TP. Le variabili dipendenti sono i punteggi medi e mediani nelle prove di italiano e matematica, espressi in logaritmi. Tutte le regressioni includono effetti fissi a livello di anno, a livello di scuola-grado e trend lineari a livello di scuola-grado. I controlli relativi alle caratteristiche degli studenti sono i seguenti: la frazione di studenti maschi; con cittadinanza italiana; regolari; di madri e padri con almeno il diploma di scuola superiore; di risposte mancanti per ciascuna di queste informazioni. I controlli relativi alle caratteristiche della scuola includono: il numero di studenti iscritti, il numero di classi, il numero di studenti assenti durante almeno uno dei test, una variabile che indica se un esaminatore esterno ha somministrato il test in almeno una delle classi nella scuola e una variabile che indica se la riforma Gelmini era in vigore. Le colonne 1 e 4 riportano stime basate sia sulle classi seconde sia sulle classi quinte. Le colonne 2 e 5 riportano stime basate solo sulle classi seconde, le colonne 3 e 6 stime basate solo sulle classi quinte. Gli errori standard sono clusterizzati a livello di scuola-grado.

La tab. 4 analizza gli effetti sulla dispersione dei voti e presenta una struttura simile. In particolare si considerano tre misure di dispersione: la deviazione standard, il rapporto tra il 75° e 25° percentile, il rapporto tra il 90° e 10° percentile. Coerentemente con quanto mostrato nelle figg. 2 e 3, emerge che l'effetto di riduzione della dispersione dei livelli di apprendimento è maggiore in matematica e per le classi quinte.

Tab. 4 – Misure di dispersione

	Italiano			Matematica		
	II e V	II	V	II e V	II	V
<i>Std. Dev.</i>						
Fraz. TP	-0,019 (0,018)	-0,013 (0,026)	-0,023 (0,024)	-0,015 (0,015)	-0,004 (0,020)	-0,025 (0,021)
<i>Rapporto 75/25</i>						
Fraz. TP	0,001 (0,009)	0,007 (0,014)	-0,005 (0,012)	-0,016 (0,010)	-0,008 (0,015)	-0,023* (0,013)
<i>Rapporto 90/10</i>						
Fraz. TP	-0,003 (0,014)	-0,001 (0,024)	-0,006 (0,015)	-0,028* (0,014)	-0,014 (0,023)	-0,041** (0,019)
Controlli studente	Si	Si	Si	Si	Si	Si
Controlli scuola	Si	Si	Si	Si	Si	Si
Oss.	81.545	38.353	43.192	81.611	38.396	43.215

* p < 0,1, ** p < 0,05, *** p < 0,01.

Nota: L'unità di osservazione è la scuola-grado. La variabile indipendente è la frazione di classi a TP. Le variabili dipendenti sono la deviazione standard, il rapporto 75/25 e il rapporto 90/10 dei punteggi di italiano e matematica, espressi in logaritmi. Tutte le regressioni includono effetti fissi a livello di anno, effetti fissi a livello di scuola-grado e trend lineari a livello di scuola-grado. I controlli relativi alle caratteristiche degli studenti e delle scuole sono gli stessi elencati nelle note della tab. 3. Le colonne 1 e 4 riportano stime basate sia sulle classi seconde sia sulle classi quinte. Le colonne 2 e 5 stime basate solo sulle classi seconde, mentre le colonne 3 e 6 stime basate solo sulle classi quinte. Gli errori standard sono clusterizzati a livello di scuola-grado.

Tab. 5 – Gap di apprendimento tra studenti

	Italiano			Matematica		
	Edu. padre	Citt.	Lingua a casa	Edu. padre	Citt.	Lingua a casa
<i>Classi II e V</i>						
Fraz. TP	-0,008 (0,007)	-0,003 (0,014)	-0,014 (0,020)	-0,006 (0,008)	-0,006 (0,015)	-0,02 (0,023)
Oss.	77.144	49.782	25.066	77.198	50.004	25.471
<i>Classi II</i>						
Fraz. TP	0,001 (0,011)	-0,001 (0,023)		0,009 (0,012)	0,006 (0,024)	
Oss.	36.384	24.039		36.408	24.145	
<i>Classi V</i>						
Fraz. TP	-0,016* (0,008)	-0,002 (0,017)	-0,014 (0,020)	-0,019* (0,010)	-0,017 (0,020)	-0,02 (0,023)
Oss.	40.760	25.743	25.066	40.790	25.859	25.471
Controlli studente	Si	Si	Si	Si	Si	Si
Controlli scuola	Si	Si	Si	Si	Si	Si

* p < 0,1, ** p < 0,05, *** p < 0,01.

Nota: L'unità di osservazione è la scuola-grado. La variabile indipendente è la frazione di classi a TP. Le variabili dipendenti sono i gap di apprendimento, espressi come logaritmo dei rapporti, tra studenti provenienti da contesti socio-economici differenti. Le colonne 1 e 4 si riferiscono al gap tra alunni con padri con almeno il diploma di scuola superiore e alunni con padri senza tale titolo; le colonne 2 e 5 si riferiscono al gap tra alunni italiani e stranieri; le colonne 3 e 6 al gap tra alunni che a casa parlano prevalentemente italiano o un dialetto e alunni che parlano prevalentemente una lingua straniera. L'informazione circa la lingua parlata a casa è disponibile solo per gli studenti della quinta primaria. Tutte le regressioni includono effetti fissi a livello di anno e di scuola-grado e trend lineari a livello di scuola-grado. I controlli relativi alle caratteristiche degli studenti e delle scuole sono gli stessi elencati nelle note della tab. 3. Gli errori standard sono clusterizzati a livello di scuola-grado.

La tab. 5 studia gli effetti sui gap di apprendimento tra studenti provenienti da contesti familiari diversi. Si analizza l'eterogeneità lungo tre dimensioni: l'istruzione del padre (se avente un titolo di studio pari almeno al diploma di scuola superiore o no), la cittadinanza degli studenti (se immigrati o no) e la lingua parlata maggiormente a casa (se italiano/dialetto o lingua straniera). La variabile dipendente è il logaritmo del rapporto del punteggio medio di questi due gruppi di studenti all'interno di una data scuola-grado. I coefficienti non sono purtroppo quasi mai significativi. Questa minor precisione delle stime potrebbe anche essere legata al minor numero di osservazioni utilizzate, a causa della necessità di includere solo quelle scuole in cui è presente, in un dato grado, almeno uno studente per ciascuno dei gruppi presi in considerazione. Quando statisticamente significativi, i risultati comunque vanno sempre nel senso di contribuire a ridurre i gap di apprendimento.

6. Conclusioni

Comprendere il modo in cui i diversi input della funzione di produzione del capitale umano interagiscono l'uno con gli altri è fondamentale per analizzare le implicazioni di politiche scolastiche alternative e per valutarne l'impatto in termini di equità. Questo lavoro studia l'effetto di trascorrere un maggior numero di ore a scuola sull'apprendimento degli studenti in Italia a 8 e a 11 anni. In particolare, si intende indagare se l'incremento del tempo trascorso a scuola, in un ambiente relativamente omogeneo, e la contestuale riduzione del tempo trascorso a casa, dove la qualità delle opportunità e le risorse familiari variano notevolmente, hanno un effetto sulla distribuzione dei punteggi e sui gap di apprendimento di studenti provenienti da contesti socio-economici differenti. La strategia identificativa si basa sulla coesistenza di due schemi orari nella scuola primaria italiana, il *tempo pieno* (TP, 40 ore settimanali) e il *tempo normale* (TN, da 24 a 30 ore settimanali, tipicamente 27). L'effetto causale del TP viene stimato sfruttando variazioni nel tempo della frazione di classi a TP in una data scuola-grado, che riflettono plausibilmente cambiamenti dal lato dell'offerta dovuti a variazioni nei fondi disponibili e alle leggi che disciplinano la formazione delle classi.

I risultati principali mostrano che aumentare il numero di classi a TP dallo 0% al 100% ha un effetto nullo sui punteggi medi di italiano, mentre ha un effetto statisticamente positivo sui punteggi medi di matematica pari a circa il 2,2%. L'effetto medio più forte in matematica potrebbe dipendere dal fatto che per questa materia – almeno per la maggioranza delle famiglie – “sostituire” la guida di un docente specializzato rispetto a quella di un genitore (che di solito è la madre) è più rilevante. Inoltre, l'effetto del TP è relativamente costante lungo i decili della distribuzione dei punteggi di italiano, mentre è più forte nella coda sinistra della distribuzione dei punteggi di matematica. Ne consegue che il TP riduce la dispersione degli esiti nei test di matematica. Emerge inoltre un effetto di mitigazione dei gap di apprendimento, che tuttavia è stimato con meno precisione, a causa della ridotta numerosità campionaria, e non risulta statisticamente significativo.

A nostra conoscenza, questo lavoro è tra i pochi a proporre un'analisi dell'effetto causale del tempo trascorso a scuola non solo sulla media ma anche sull'intera distribuzione dei punteggi e sui gap di apprendimento tra studenti provenienti da contesti socio-economici differenti. Riteniamo che queste stime siano rilevanti sia per la discussione circa l'esistenza di una significativa percentuale di alunni con bassa performance nei test standardizzati, sia per il dibattito circa le modalità con cui promuovere l'apprendimento degli studenti provenienti da contesti svantaggiati.

Riferimenti bibliografici

- Aguero J.M., Beleche T. (2013), “Test-Mex: Estimating the effects of school year length on student performance in Mexico”, *Journal of Development Economics*, 103: 353-361.
- Angrist J.A., Battistin E., Vuri D. (2017), “In a small moment: Class size and moral hazard in the Mezzogiorno”, *American Economic Journal: Applied Economics*, 9 (4): 216-249.
- Aucejo E., Romano T. (2016), “Assessing the effect of school days and absences on test score performance”, *Economics of Education Review*, 55: 70-87.
- Barro R.J. (2001), “Human capital and growth”, *American Economic Review*, 91, 2: 12-17.
- Battistin E., Meroni E. (2016), “Should we increase instruction time in low achieving schools? Evidence from Southern Italy”, *Economics of Education Review*, 55: 39-56.

-
- Battistin E., Schuller S. (2013), *Evaluation of the “Maestro Unico” Reform, Technical Report*, IRVAPP, Trento.
- Bellei C. (2009), “Does lengthening the school day increase students’ academic achievement? Results from a Natural Experiment in Chile”, *Economics of Education Review*, 28, 5: 629-640.
- Falzetti P., Longobardi S., Sestito P. (2015), “A multistep approach to detect and correct the cheating in Italian students’ data”, paper presented at the *SIS2015 Statistical Conference, Statistics and Demography: the Legacy of Corrado Gini*, 9-11 September, Treviso.
- Goodman J. (2014), “Flaking out: Student absences and snow days as disruptions of instructional time”, *NBER Working Paper Series*, 20221, testo disponibile al sito: <http://www.nber.org/papers/w20221.pdf>, data di consultazione: 3 giugno 2017.
- Hansen B. (2011), *School Year Length and Student Performance: Quasi-Experimental Evidence*, University of California, Santa Barbara (CA), mimeo.
- Hanushek E.A., Woessmann L. (2008), “The role of cognitive skills in economic development”, *Journal of Economic Literature*, 46, 3: 607-668.
- Huebener M., Kuger S., Marcus J. (2017, in press), “Increased instruction hours and the widening gap in student performance”, *Labour Economics*, 47: 15-34.
- Lavy V. (2012), “Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior”, *NBER Working Paper Series*, 18369, testo disponibile al sito: <http://www.nber.org/papers/w18369.pdf>, data di consultazione: 3 giugno 2017.
- Lavy V. (2015), “Do differences in schools’ instruction time explain international achievement gaps? Evidence from developed and developing countries”, *The Economic Journal*, 125, 588: F397-F424.
- Lavy V., Paserman D.M., Schlosser A. (2012), “Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom”, *The Economic Journal*, 122, 559: 208-237.
- Lee J.W., Barro R.J. (2001), “Schooling Quality in a Cross-section of Countries”, *Economica*, 68, 272: 465-488.
- Marcotte D.E. (2007), “Schooling and test scores: a mother-natural experiment”, *Economics of Education Review*, 26, 5: 629-640.
- Marcotte D.E., Hemelt S.W. (2008), “Unscheduled school closings and student performance”, *Education Finance and Policy*, 3, 3: 316-338.
- Mariani V., Recchia P., Sestito P., Vacca V. (2012), *Tempo pieno, Test Scores and Behaviour*, mimeo.
- Parinduri R.A. (2014), “Do children spend too much time in schools? Evidence from a longer school year in Indonesia”, *Economics of Education Review*, 41: 89-104.
- Pereda-Fernandez S. (2016), “A new method for the correction of test scores manipulation”, *Temi di Discussione (Working Papers)*, 1047.
- Pischke J.S. (2007), “The impact of length of the school year on student performance and earnings: Evidence from the German short school years”, *The Economic Journal*, 117, 523: 1216-1242.
- Quintano C., Castellano R., Longobardi S. (2009), “A fuzzy clustering approach to improve the accuracy of Italian student data. An experimental procedure to correct the impact of outliers on assessment test scores”, *Statistica & Applicazioni*, 7, 2: 149-171.
- Woessmann L. (2003), “Schooling resources, educational institutions and student performance: the international evidence”, *Oxford Bulletin of Economics and Statistics*, 65, 2: 117-170.

12. Predittori della comprensione del testo nei primi anni di scuola primaria: un'analisi multilivello

Predictors of reading comprehension in early primary school grades: a multilevel analysis

di Marta Desimoni, Antonella Mastrogiovanni, Alessia Mattei

La maggior parte degli studi sulla comprensione della lettura nei bambini nei primi anni di scuola primaria si è focalizzata sulle associazioni tra tale abilità e variabili a livello individuale (per es. la fluenza nella lettura, il genere ecc.). Il presente studio si propone, in una prospettiva multilivello, di indagare se la variabilità nella comprensione nella lettura nei bambini di seconda primaria è, almeno in parte, legata a differenze tra le classi frequentate dai bambini e di esplorare gli effetti di variabili individuali (a livello studente) e contestuali (a livello classe) sulla comprensione. Al fine di esplorare tali effetti, sono stati esaminati i dati raccolti dall'Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione (INVALSI). INVALSI conduce regolarmente rilevazioni standardizzate delle abilità e delle conoscenze degli studenti ai fini della valutazione della qualità del sistema di educazione. Nel presente lavoro saranno presi in esame i dati relativi a un campione di studenti che frequentavano la seconda primaria alla fine dell'a.s. 2014-15. I dati sono stati esaminati attraverso una serie di modelli multilivello a due livelli: studenti (livello 1) e classi (livello 2). In tutti i modelli, la variabile dipendente è la stima dell'abilità di comprensione del testo degli allievi. Le variabili indipendenti, a livello dello studente, includono un indice di fluenza nella lettura, il genere, la cittadinanza, il livello di istruzione dei genitori, l'aver frequentato la scuola dell'infanzia e l'anno di entrata nella scuola primaria. Le variabili contestuali, a livello classe, includono il livello di fluenza nella lettura, la percentuale di laureati tra i genitori, la percentuale di studenti stranieri, il numero di studenti per classe e le ore di scuola a settimana. I risultati indicano che una porzione significativa della variabilità nella comprensione della lettura è dovuta a differenze tra le classi. Si osserva l'effetto significativo sia di variabili a livello del singolo studente sia di variabili a livello della classe scolastica, confermando l'importanza dell'adottare un approccio multilivello nello studio della relazione tra comprensione del testo, fluenza nella lettura e altre variabili rilevanti.

Most studies on reading comprehension in children attending early primary school have been focused on the associations between this ability and individual – level variables (e.g. reading accuracy, reading fluency, vocabulary, gender etc.). Turning into a multi-level perspective, the present study aims at investigating whether variability in reading comprehension in Grade 2 Italian children is at least in part due to differences between classes and at examining the effects of individual (student-level) and contextual (classroom-level) variables on reading comprehension. In order to explore the effects of individual and contextual variables on reading comprehension, we analyzed data from the Italian National Institute for the Educational Evaluation of Instruction and Training (INVALSI). INVALSI regularly carries out standardized tests to assess students' knowledge and skills and to evaluate the overall quality of the educational offering of schools. In the present work, we considered data collected on a sample of students attending Grade 2 at the end of the 2014-15 school year. A series of multilevel models were fitted to INVALSI data, considering two nested level: pupils (level-1) and classes (level-2). The outcome in all models was the IRT person ability estimate of reading comprehension ability, based on the whole set of questions of the INVALSI Italian language test. Individual-level independent variables included an index of children reading fluency, obtained from children performance at INVALSI reading test and other students characteristics like gender, citizenship, parental education levels, preschool attendance and primary school entry age. Contextual variable included class reading fluency level, parental educational level, percentage of immigrants, number of students per class, hours per week. Results suggest that a significant portion of variance in reading comprehension is

due to differences between classes. Significant fixed effects at individual level and at class level emerge. Overall, these results highlight the importance of adopting a multilevel perspective when investigating the relationship between reading comprehension, reading fluency and other relevant variables.

1. Introduzione

Imparare a leggere correttamente, fluentemente e con un'adeguata comprensione del testo è riconosciuto in ambito nazionale e internazionale come uno dei più importanti obiettivi delle fasi iniziali del percorso di apprendimento dei bambini in età scolare, dato il ruolo fondamentale della competenza di lettura per gli apprendimenti scolastici e per la futura vita lavorativa, sociale e relazionale degli studenti (per es. Pontecorvo e Pontecorvo, 1986; Duncan *et al.*, 2007; Hulme e Snowling, 2011). Negli ultimi decenni, numerosi studi nella ricerca psicologica hanno indagato le interrelazioni tra abilità di comprensione del testo e i processi di lettura e scrittura strumentali (per es. Mehta *et al.*, 2005; Bashir e Hook, 2009; Campodifiori *et al.*, 2011; Desimoni, Scalisi e Orsolini, 2012), nonché le associazioni tra comprensione del testo e altri fattori, esterni al dominio delle competenze di lettura, ma rilevanti per l'apprendimento della lingua scritta (per una rassegna, vedi Kirby *et al.*, 2008; Pufpaff, 2009). Come sottolineato da Kim, Petscher e Foorman (2015), le numerose ricerche condotte si sono focalizzate prevalentemente su variabili a livello di singolo studente, mentre un numero minore di studi ha approfondito l'effetto di caratteristiche dei contesti di apprendimento, quali la classe e la scuola.

Alla luce di tale riflessione, il presente lavoro si propone di contribuire alla ricerca sulla relazione tra comprensione del testo, lettura strumentale e altre variabili rilevanti nelle prime fasi di apprendimento della lingua scritta, adottando una prospettiva multilivello che tenga conto di variabili sia a livello studente sia a livello del micro-sistema in cui gli allievi sono inseriti, ossia le classi scolastiche frequentate, nonché l'interazione tra variabili individuali e contestuali. In particolare, saranno esaminati in tale prospettiva i dati della rilevazione nazionale degli apprendimenti condotta da INVALSI alla fine del secondo anno di scuola primaria, focalizzandosi in particolare sulla prova INVALSI di italiano.

Gli studi sulle variabili che a livello di singolo individuo si configurano come buoni predittori e correlati cognitivi della comprensione del testo hanno evidenziato la relazione tra tale abilità e altre competenze di lettura di base. Seppure sia ampiamente riconosciuto che la comprensione del testo è un'abilità complessa, cui sono associate numerose abilità linguistiche, metalinguistiche e cognitive (per es. Goff, Pratt e Ong, 2005; Jenkins *et al.*, 2003; Klaua e Guthrie, 2008; Pazzaglia, Cornoldi e Tressoldi, 1993; Verhoeven, Reitsma e Siegel, 2011), vi è un generale accordo nel considerare l'abilità di lettura come uno dei correlati più importanti della capacità di comprensione del testo, soprattutto nelle prime fasi del percorso di apprendimento (per es. Hulme e Snowling, 2011). Gli studi che hanno evidenziato le relazioni tra dimensioni sottese all'abilità di lettura strumentale e la comprensione del testo sono molto numerosi (per es. Storch e Whitehurst, 2002, Paris, 2005; Desimoni *et al.*, 2012; Kim *et al.*, 2015) e si sono focalizzati su diversi aspetti della decodifica, per esempio, sull'accuratezza della decodifica delle parole, sulla rapidità di decodifica o su indici di fluency nella lettura, che può essere definita nell'accezione più estesa, come abilità di leggere il testo scritto accuratamente, rapidamente e con un'adeguata prosodia (per una rassegna, vedi Bashir e Hook, 2009).

Come sottolineato da Hulme e Snowling (2011), la comprensione del testo dipende, almeno in parte, dalla capacità di riconoscere le parole, definibile come capacità di decifrare le stringhe ortografiche correttamente e automaticamente, attribuendo un contenuto semantico alle stringhe decifrate. Nelle prime fasi di apprendimento, la relazione tra le prove di comprensione e le prove di riconoscimento delle parole è molto elevata: per esempio, Juel, Griffith e Gough (1986) in uno studio su bambini del primo e secondo anno di scuola primaria hanno riscontrato una correlazione pari a 0,70. Come sottolineato da Paris (2005), i bambini che riescono a leggere correttamente poche parole nel testo avranno difficoltà nella comprensione del testo stesso, con una relazione specifica tra accuratezza e comprensione per i bassi livelli di abilità di decodifica, relazione che andrebbe diminuendo man mano che il bambino diviene un lettore più accurato. La relazione tra comprensione e abilità di riconoscimento delle parole, tuttavia, non si limita alla componente dell'accuratezza, ma anche all'automaticità, nel senso di grado di coinvolgimento di processi attentivi con cui le parole sono riconosciute.

La capacità di riconoscimento delle parole è infatti considerata come competenza alla base della fluency nella lettura, considerata da alcuni autori come l'anello di congiunzione tra riconoscimento delle parole e comprensione del testo (per

una rassegna, vedi Bashir e Hook, 2009). Come sottolineato da Wolf e Katzir-Cohen (2001), la fluenza nella lettura, che sottende sia l'accuratezza sia la rapidità nella lettura, è basata sulla capacità di riconoscere le parole accuratamente e in modo automatizzato.

Secondo una delle ipotesi più diffuse sulla relazione tra fluenza nella lettura e comprensione del testo (LaBerge e Samuels, 1974; Perfetti, 1985), man mano che nel percorso di apprendimento il bambino diviene un lettore più fluente, la maggiore automatizzazione del processo di lettura consente di "liberare" le risorse attentive precedentemente dedicate alla decodifica del testo scritto e di dedicarle maggiormente ai processi di comprensione del testo. È dunque possibile ipotizzare che nella fase del percorso scolastico in cui l'abilità di lettura strumentale è in corso di stabilizzazione possano emergere delle differenze individuali nella comprensione del testo basate sui differenti livelli di automatizzazione del processo di riconoscimento delle parole e, più in generale, della decodifica del testo. Per gli allievi poco fluenti, infatti, le risorse potrebbero essere ancora dedicate in gran parte al processo di decifrazione del testo, a discapito degli altri processi sottesi alla comprensione piena del testo; per altri allievi, invece, il processo di decodifica è più automatizzato e più risorse possono essere dedicate alla comprensione. Inoltre, un processo di decodifica delle parole poco automatizzato potrebbe portare gli allievi poco fluenti a difficoltà nella memorizzazione di elementi utili alla ricostruzione del significato del testo, con conseguenti difficoltà nella comprensione (Carlisle e Rice, 2002).

Quale, tra le dimensioni dell'accuratezza e della rapidità nella decodifica, sia maggiormente associata alla comprensione del testo è a oggi oggetto di dibattito. Gli studi sull'argomento sottolineano il ruolo di moderatore, nel *pattern* di relazioni tra accuratezza, rapidità e comprensione, del grado di trasparenza del contesto ortografico di riferimento (Seymour, Aro e Erskine, 2003). L'italiano, così come altre lingue (per es. il finlandese, il tedesco, il ceco) è una lingua caratterizzata dall'alta consistenza del rapporto tra grafemi e fonemi, dunque, è un'ortografia "trasparente".

La letteratura scientifica sull'argomento sottolinea che, a differenza delle lingue a ortografia opaca (per es. l'inglese) in cui emerge l'importanza della correttezza nella lettura, nelle lingue a ortografia trasparente la correttezza nella lettura raggiunge un effetto "soffitto" alla fine del primo anno di scuola primaria (per es. Landerl e Wimmer, 2008; Orsolini *et al.*, 2006) e la fluenza nella lettura, con particolare riferimento alla rapidità, è la dimensione che caratterizza meglio e più a lungo i cattivi lettori rispetto ai buoni lettori (Landerl e Wimmer, 2008), con una maggiore associazione della comprensione del testo alla fluenza/rapidità di lettura, rispetto all'accuratezza (De Jong e van der Leij, 2002; Leppänen *et al.*, 2008; Müller e Brady, 2001; Verhoeven e van Leeuwe, 2008). La relazione tra fluenza/rapidità di lettura e comprensione del testo sembra decrescere, tuttavia, anche nelle lingue a ortografia trasparente all'aumentare del livello scolastico (per es. Müller e Brady, 2001; Verhoeven e van Leeuwe, 2008), con il rafforzarsi, invece, del ruolo della conoscenza pregressa e dell'abilità nel compiere inferenze (Perfetti, 1985), nonché di altre variabili cognitive e metacognitive, quali, per esempio, la capacità del lettore di pianificare, mettere in atto e integrare le diverse strategie cognitive e linguistiche necessarie per un'accurata comprensione (Gaskins, Satlow e Pressley, 2007; Bashir e Hook, 2009).

Gli studi condotti in ambito psicologico sulle caratteristiche degli studenti importanti per lo sviluppo delle comprensione del testo si sono ampiamente soffermate su variabili cognitive, come le sopra descritte abilità di lettura strumentale, la comprensione del linguaggio orale, il livello cognitivo generale e abilità cognitive specifiche (per esempio, la memoria di lavoro, la consapevolezza fonologica, la denominazione rapida automatizzata), il vocabolario, le abilità metacognitive, solo per citarne alcune. Al di là di tali variabili, un certo numero di studi ha anche esaminato il ruolo di fattori socio-demografici per lo sviluppo delle competenze di lettura e, più in generale, per gli apprendimenti scolastici. Variabili ampiamente considerate in letteratura sono quelle relative allo status socio-economico culturale. Associazioni positive tra apprendimenti scolastici e status socio-economico sono state riscontrate nelle rilevazioni INVALSI, condotte annualmente sul territorio italiano negli ambiti disciplinari dell'italiano e della matematica (per es. *Rapporto nazionale prove INVALSI 2016*: INVALSI, 2016). In particolare, la relazione tra tali variabili è indagata da INVALSI in V primaria e II secondaria di II grado, attraverso il calcolo di un indice ESCS che include, in linea con l'indagine internazionale PISA, il grado d'istruzione dei genitori, il prestigio della professione da essi esercitata e le dotazioni presenti in casa (quantità di libri, una scrivania per studiare ecc.).

Un lavoro di Campodifiori e collaboratori (2011), condotto sulla base dei dati INVALSI dell'anno scolastico 2009-2010 e in cui è stato utilizzato il titolo di studio dei genitori come indice del livello culturale della famiglia dell'allievo, ha riscontrato che tale variabile ha un impatto unico significativo sulla comprensione del testo, al netto di altre variabili rilevanti, tra cui l'abilità di lettura strumentale degli studenti. Il contributo del livello di istruzione dei genitori e del

loro status socio-economico è stato evidenziato in numerosi studi e in diversi contesti ortografici (per es. Adams, 1990; Bowey, 1995; Leppänen *et al.*, 2004; Leppänen *et al.*, 2008).

Per esempio, nel contesto di un'ortografia trasparente (il finlandese) Leppänen e colleghi (2008) hanno riscontrato che il titolo di studio della madre è un buon predittore della comprensione del testo dei bambini in quarta primaria, anche dopo aver controllato l'abilità di lettura strumentale nel primo anno di scuola primaria. Quali siano i meccanismi che legano il livello di istruzione dei genitori, il loro status socio-economico e gli apprendimenti scolastici è oggi ancora poco chiaro e diverse ipotesi sono state formulate sull'argomento (per es. Adams, 1990; Davis-Kean, 2005; Bergen *et al.*, 2016).

Adams (1990) ha ipotizzato che la relazione tra status socio-economico culturale dei genitori e apprendimento della lingua scritta è mediato dal grado in cui i genitori coinvolgono i figli in attività che potenziano abilità alla base dello sviluppo della letto-scrittura, per esempio, la consapevolezza fonologica.

Davis-Kean (2005) ha ipotizzato e verificato su un campione di bambini americani che la relazione tra livello di istruzione dei genitori e competenze di lettura dei bambini è mediato dalle aspettative dei genitori sul successo in ambito scolastico/accademico dei figli, che a loro volta influenzano la costruzione di un ambiente che sia di supporto all'apprendimento della lingua scritta e alla messa in atto di attività mirate a promuovere la lettura.

Più recentemente, Bergen e collaboratori (2016) hanno proposto un modello in cui fattori genetici e fattori ambientali interagiscono, sottolineando la relazione tra livello di fluenza nella lettura raggiunto dai genitori e fluenza nella lettura dei figli, in associazione a fattori legati all'ambiente di apprendimento e alle risorse disponibili (per es. numero di libri posseduti).

Oltre allo status socio-economico culturale della famiglia di provenienza, differenze sulle competenze di lettura sono state evidenziate sulla base di altre variabili socio-demografiche, per esempio il genere, l'età di ingresso nella scuola primaria e l'aver frequentato la scuola dell'infanzia. Nelle rilevazioni degli apprendimenti a cura di INVALSI, emerge in generale un gap di genere a favore delle allieve per l'italiano (per es. per tutti i livelli nell'anno scolastico 2014-15; per la quinta primaria, terza secondaria di I grado e seconda secondaria di II grado nell'anno scolastico 2015-16; INVALSI, 2015 e 2016). Tale gap di genere trova riscontro anche nelle indagini comparative internazionali.

Per esempio, facendo riferimento all'indagine internazionale PIRLS, condotta su bambini di età più vicina a quella del campione preso in esame nel presente studio (livello 4), emergono prestazioni migliori delle ragazze nelle prove di lettura rispetto ai ragazzi nella gran parte dei Paesi che partecipano alla rilevazione (per es. INVALSI, 2011). Tale differenza è stata confermata anche da altri lavori (per es. Elley, 1991; Leppänen *et al.*, 2008), seppure sia stato osservato che è modulata da fattori quali il livello scolastico e il tipo di materiale proposto (per es. testi continui vs testi non continui, Elley, 1991).

Altre variabili indagate dalla letteratura sull'argomento sono relative al percorso scolastico degli studenti, come aver frequentato o meno la scuola pre-primaria (per es. Campodifiori, Falzetti e Papini, 2016) e l'età di ingresso nella scuola primaria, ossia l'essere o meno anticipatorio (per una rassegna, vedi Thoren, Heinig e Brunner, 2016). In letteratura è riportato uno svantaggio per i bambini anticipatori rispetto ai bambini che iniziano regolarmente il loro percorso di istruzione formale, soprattutto per quanto riguarda gli apprendimenti nei primi anni di scolarità (Thoren *et al.*, 2016).

Nel contesto italiano, in cui si osserva una notevole variabilità, data la possibilità per i genitori di iscrivere i bambini anticipatamente, Desimoni, Pelagaggi e Scalisi (2006) hanno osservato che, all'inizio del percorso di apprendimento, l'essere anticipatorio nel percorso di studi non ha un impatto sulla lettura e scrittura strumentali, ma ha un impatto forte negativo sulla comprensione del testo.

Se la letteratura psicologica sui fattori individuali alla base dello sviluppo delle competenze di lettura è molto ampia, un minor numero di studi ha approfondito il ruolo di caratteristiche delle classi scolastiche o della scuola sullo sviluppo dell'abilità di comprensione, nonché del possibile ruolo di moderatore delle caratteristiche della classe e della scuola frequentata dai bambini sul *pattern* di relazioni tra fattori individuali e sviluppo degli apprendimenti.

In un recente lavoro condotto su bambini americani dal livello 3 al livello 8, Kim e collaboratori (2015) hanno evidenziano che nei livelli scolastici più avanzati (dal 6° all'8° grado), la proporzione della variabilità nella comprensione del testo legata a differenze tra le classi è elevata (range = 41-46%) e di quota simile a quella spiegata dalle differenze individuali degli studenti (range = 48-51%), con una proporzione invece più bassa di variabilità tra le scuole/distretti (circa il 4%). Nei livelli inferiori, la gran parte della variabilità osservata nella comprensione del testo è legata a differenze tra gli studenti (circa il 67%). Emerge, tuttavia, una quota significativa e non trascurabile di variabilità legata a

differenze tra le classi scolastiche (tra il 21 e il 23%) e una proporzione minore di variabilità a livello di scuola/distretto (3-4%). Tale dato è confermato anche da altri lavori in cui emerge che le differenze nelle *literacy skills* attribuibili alle caratteristiche delle classi variano, nei primi anni di scuola, tra 0,08 e 0,26 (Al Otaiba *et al.*, 2012; Mehta *et al.*, 2005). Tali risultati suggeriscono che le differenze nella comprensione del testo non debbano essere esaminate esclusivamente a livello dei singoli studenti, ma anche del contesto di apprendimento.

In particolare, come recentemente evidenziato da Martínez (2012) un ruolo particolarmente rilevante sembra essere ascrivibile al micro-sistema in cui lo studente è inserito, ossia la classe scolastica. Alcuni studi hanno approfondito il ruolo di alcune variabili come possibili fattori alla base di tali differenze, quali, per esempio, la composizione della classe in base alle caratteristiche cognitive o al livello di apprendimento in ingresso degli studenti, il livello di istruzione e il livello socio-economico dei genitori degli allievi della classe, il numero di studenti immigrati o fattori inerenti il “clima” di classe e variabili relative agli insegnanti (per es. De Fraine *et al.*, 2003; Mehta *et al.*, 2005; Suchodoletz *et al.*, 2015). Il ruolo di tali variabili, tuttavia, è ancora controverso e ulteriori studi sono necessari per approfondire in che modo fattori individuali e caratteristiche del contesto di apprendimento dello studente interagiscono nel favorire lo sviluppo dell’abilità di comprensione del testo.

2. Il presente studio

Sulla base di quanto emerso dall’esame della letteratura sull’argomento, il presente lavoro si propone di esplorare e approfondire la relazione tra caratteristiche individuali degli studenti, caratteristiche del contesto di apprendimento e abilità di comprensione del testo alla fine del secondo anno di scuola primaria. Lo studio è stato condotto a partire dai dati raccolti dall’Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione (INVALSI) nell’ambito delle rilevazioni standardizzate degli apprendimenti degli studenti in italiano e matematica, rilevazioni condotte ai fini della valutazione della qualità del sistema di educazione in Italia.

Nel presente lavoro saranno presi in esame i dati relativi a un ampio campione di studenti che frequentavano la seconda primaria alla fine dell’a.s. 2014-15. Lo studio, dunque, fa riferimento a studenti che si trovano in una fase relativamente iniziale del percorso di apprendimento della lingua scritta, in cui si ipotizza che i bambini stiano stabilizzando le loro abilità di lettura e scrittura strumentali (per es. AID, Associazione italiana dislessia, 2007), e nel contesto ortografico della lingua italiana, ortograficamente trasparente. Il contesto di apprendimento oggetto di studio è la classe scolastica. La scelta della classe scolastica come unità sovraordinata (o di secondo livello) agli studenti, invece della scuola, è stata effettuata accogliendo il suggerimento di alcuni autori (per es. Rowe e Holmes-Smith, 1995; Goldstain, 1997; De Fraine, 2003; Martínez, 2012) secondo i quali, soprattutto per i livelli scolastici in cui gli studenti sono inseriti in gruppi classe relativamente stabili, è proprio la classe scolastica, e non la scuola, l’unità da privilegiare nelle analisi con approccio multilivello. Ovviamente, ciò non implica che classe e scuola non possano, in teoria, essere prese in considerazione contemporaneamente nella stessa analisi; tuttavia, dato che nel caso del campione esaminato (classi in cui la rilevazione INVALSI è stata condotta alla presenza di un osservatore esterno) il numero di classi incluse nel campione per ogni scuola è molto basso (massimo due classi), si è posto il problema di dover scegliere tra uno dei due livelli di analisi, scegliendo la classe.

Il primo obiettivo prefissato è quello di verificare se, già dalla seconda primaria, una quota significativa di variabilità nella comprensione del testo, valutata attraverso la prova INVALSI di italiano, è legata a differenze tra le classi scolastiche o se, in tale livello di scolarità, la variabilità nella prestazione alla prova di italiano è legata solo a differenze individuali tra gli allievi. L’esame della letteratura sull’argomento suggerisce che le differenze tra le classi scolastiche giocano un ruolo importante nello spiegare la variabilità negli apprendimenti scolastici (per es. Hill e Rowe, 1996; Kim *et al.*, 2015). In molti casi, tuttavia, si tratta di studi che hanno preso in esame livelli scolastici superiori alla seconda primaria (Kim *et al.*, 2015). Dato il risultato, riportato da alcuni autori (per es. Kim *et al.*, 2015), di un decrescere della porzione di variabilità legata a differenze tra le classi al diminuire del livello di scolarità interessato, è dunque importante investigare se, nel contesto italiano, la porzione di variabilità tra le classi è significativa già in seconda primaria.

Il secondo obiettivo è quello di verificare quali variabili, a livello individuale e a livello della classe scolastica, hanno un effetto significativo sulla prestazione alla prova INVALSI di italiano e contribuiscono a spiegare la variabi-

lità tra gli studenti entro le classi e tra le classi. In particolare, sulla base delle variabili presenti nel database fornito da INVALSI e alla luce dei principali risultati della letteratura sull'argomento, saranno considerate le variabili socio-demografiche relative al genere, al livello di istruzione dei genitori, alla cittadinanza (allievi di origine italiana vs allievi stranieri di I e II generazione), all'essere anticipatari rispetto al percorso scolastico e all'aver frequentato la scuola dell'infanzia.

Sarà inoltre presa in esame la prestazione alla prova di lettura strumentale INVALSI (per una descrizione dettagliata della prova, vedi Campodifiori *et al.*, 2011). Tale strumento, pensato per una valutazione dell'abilità di lettura strumentale dei bambini su larga scala e attraverso una somministrazione collettiva all'intero gruppo classe, richiede al bambino di leggere silenziosamente una lista di parole e di individuare, per ogni parola, la figura corrispondente al significato della parola stessa, tra le quattro alternative presentate. Il compito deve essere svolto dai bambini in un massimo di due minuti, con un punteggio relativo al numero di scelte correttamente operate. La prestazione a tale prova nel presente lavoro è considerata come un indice della correttezza e rapidità con cui il bambino è in grado di codificare le stringhe ortografiche con accesso al significato delle parole stesse, dunque, come indicatore dell'abilità di riconoscimento delle parole e *proxy* della fluenza della lettura. Considerata l'importanza della fluenza nella lettura per la comprensione del testo nei primi anni di scolarità e nel contesto delle lingue a ortografia trasparente, si ipotizza di riscontrare un'associazione positiva tra fluenza e comprensione nella lettura, anche al netto delle altre variabili considerate. In un'ottica esplorativa, si vuole inoltre indagare se tale relazione è stabile tra le classi o se l'intensità dell'associazione tra componenti dell'abilità di lettura varia tra le classi.

A livello della classe, saranno presi in esame il ruolo di alcune variabili inerenti l'organizzazione della classe, quali l'organizzazione oraria della classe e la numerosità della classe; e variabili relative alla composizione della classe rispetto alla percentuale di bambini stranieri, alla percentuale di figli di genitori con livello di istruzione elevato, alla presenza di bambini poco fluenti nella lettura. Sempre considerata la centralità della fluenza della lettura per la comprensione del testo, sarà inoltre indagato se la composizione della classe rispetto all'abilità di lettura strumentale degli allievi modera la relazione tra fluenza nella lettura del singolo allievo e la comprensione del testo. In particolare, si vuole verificare se l'essere un allievo di una classe con un numero relativamente alto di bambini che a fine seconda primaria sono ancora lettori poco fluenti ha un effetto sulla comprensione del testo del bambino e per quale livello di abilità tale effetto si verifica. Il ruolo di tali fattori sarà infine esaminato anche al netto delle differenze legate alla collocazione geografica delle classi.

3. Metodo

3.1. Disegno e procedura

I dati esaminati sono stati raccolti da INVALSI nella rilevazione nazionale degli apprendimenti di italiano e matematica nell'anno scolastico 2014-15. La somministrazione delle prove, di tipo carta e matita, è avvenuta alla fine dell'anno scolastico, nel mese di maggio. La somministrazione delle prove di italiano e della prova di lettura strumentale (prova preliminare di lettura) è collettiva, dunque, è svolta contemporaneamente dall'intero gruppo classe. Per la prova di comprensione del testo, in particolare, è importante sottolineare che il bambino ha sempre a disposizione il testo cui le domande fanno riferimento, evitando un successivo carico della memoria di lavoro. Per lo svolgimento della prova di italiano è previsto un tempo massimo di 45 minuti; come verificato in fase di pre-test tale arco temporale è sufficiente allo svolgimento della prova, che non deve, dunque, essere considerata una prova a tempo. La rilevazione INVALSI degli apprendimenti è di tipo censuario, ossia la rilevazione è condotta su tutti gli studenti della popolazione per i livelli interessati. In generale, la somministrazione delle prove, la codifica e la registrazione delle risposte è effettuata dalle insegnanti. Tuttavia, dalla popolazione complessiva, è estratto un campione rappresentativo di scuole in cui le prove sono somministrate alla presenza di un osservatore esterno, al fine di assicurare la conformità con il protocollo di somministrazione.

3.2. Partecipanti

Sono stati considerati esclusivamente i dati delle classi campione, ossia delle classi in cui le rilevazioni sono state effettuate alla presenza di un osservatore esterno. Dal campione complessivo, sono stati esclusi i casi con dati mancanti; sono inoltre state escluse le classi con numero di studenti inferiore a 5 e le classi con percentuale di studenti con dati mancanti superiore al 50%. Il campione finale è costituito da 17.392 studenti in 1.055 classi scolastiche. Per quanto riguarda la distribuzione degli studenti per classe, la moda è pari a 22 studenti per classe e la mediana a 20 studenti per classe. Meno del 10% delle classi è composta da 12 studenti o meno e meno del 10% delle classi è composto da più di 25 studenti. Nel 74% delle classi esaminate, la percentuale di studenti di cui sono a disposizione i dati su tutte le variabili considerate è pari almeno all'80%.

3.3. Materiali

Per la seconda primaria la rilevazione degli apprendimenti a cura di INVALSI prevede due prove, una prova di italiano e una prova di lettura strumentale.

Nell'anno scolastico 2014-15, la prova INVALSI di italiano per la seconda primaria si compone di un testo continuo narrativo, corredato da ventuno domande e da due esercizi linguistici.

Il testo, di 53 righe, è preceduto da tre domande a risposta multipla semplice. Agli studenti è richiesto di rispondere a queste tre domande prima di leggere l'intero testo basandosi esclusivamente sul titolo del racconto che viene proposto successivamente. Lo scopo è quello di sollecitare ambiti di significato e mobilitare previsioni-anticipazioni-aspettative rispetto al testo che dovrà essere letto. Dopo la lettura del testo gli studenti devono rispondere ad altre 18 domande; di cui 17 a scelta multipla e una a risposta aperta.

Per richiamare l'attenzione sul testo e su alcuni punti specifici del testo ed evitare che lo studente sia costretto a tornare indietro e si perda nella ricerca del particolare passo del racconto su cui alcune domande vertono, accanto a queste domande è stato riportato la parte di testo in questione.

Nello specifico è stato proposto il testo "Sua Maestà si annoia" (tratto e adattato da M. Sabas, *Sua Maestà si annoia*, Ape Junior, Milano, 2002). Il testo proposto risponde al criterio di adeguatezza rispetto al livello scolastico, ma allo stesso tempo consente di formulare domande di diverso grado di difficoltà, relative ai diversi aspetti della lettura indagati dalle prove INVALSI; dall'individuazione di informazioni fino alla ricostruzione del significato di singole parti e del testo nel suo insieme. Le domande, incentrate su punti nodali per la ricostruzione del significato del testo, si propongono di indagare la comprensione della lettura focalizzandosi su specifici aspetti a essa sottesi. La seconda parte della prova è costituita da due esercizi volti a valutare lo sviluppo linguistico degli alunni dal punto di vista lessicale e sintattico-semantic. Il primo esercizio chiede di indicare, in una lista di 12 coppie di parole, se le parole di ogni coppia hanno significato eguale o contrario; si tratta di una domanda a scelta multipla complessa. Il secondo esercizio chiede invece di mettere in relazione in modo appropriato la prima parte (gruppo nominale) con la seconda parte (gruppo verbale) di cinque frasi, in questo caso il formato della domanda è chiamato corrispondenze (*matching*). Indipendentemente dal formato della domanda, il tipo di codifica finale per ogni domanda è di tipo dicotomico (1 = risposta corretta; 0 = risposta errata).

La prova preliminare di lettura strumentale comprende 40 item, ciascuno dei quali formato da una parola seguita da una serie di quattro figure, tra le quali l'alunno deve indicare quella corrispondente alla parola letta. Il tempo previsto per leggere le 40 parole e scegliere la figura corrispondente è stato di due minuti. Nel presente lavoro è stato considerato come punteggio alla prova il numero di risposte corrette nel tempo massimo previsto; tale è considerato come un indice della correttezza e rapidità con cui il bambino è in grado di codificare le stringhe ortografiche con accesso al significato delle parole stesse, dunque, come indicatore dell'abilità di riconoscimento delle parole e *proxy* della fluenza della lettura.

4. Analisi dei dati e risultati

La variabile dipendente considerata nelle analisi è la stima del livello di padronanza linguistica ottenuta attraverso il modello di Rasch (1960) sulla base della prestazione degli allievi alla prova di italiano di seconda primaria. Nella cali-

brazione degli item e nella stima del livello di abilità degli studenti, lo zero corrisponde al livello medio di abilità degli studenti italiani della coorte considerata (anno scolastico 2014-15). Nel campione sottoposto ad analisi, risultante dalle operazioni di *data-cleaning* (come descritto nel paragrafo relativa ai partecipanti), la media osservata è pari a -0,01 (deviazione standard = 1,17), media che non si discosta significativamente da quella del campione complessivo ($p = 0,22$).

In tutte le analisi, saranno condotte analisi multilivello a due livelli: studenti e classi. Nella strategia di analisi, saranno testati una serie di modelli: dapprima un modello a intercetta *random* vuoto, che sarà utilizzato al fine della valutazione dei modelli successivi, e poi modelli con variabili a livello studente e a livello classe.

Le variabili a livello studente considerate nelle analisi sono:

- il punteggio alla prova di lettura strumentale. Tale punteggio, qui considerato come un *proxy* della fluenza nella lettura, indica quante parole lo studente è riuscito a decodificare correttamente in due minuti, recuperando dal lessico mentale il significato della parola stessa e selezionando la figura corretta tra le quattro presentate. Ai fini delle analisi il punteggio è stato standardizzato in riferimento alla media e alla deviazione standard del campione sottoposto ad analisi, trasformando, dunque, la distribuzione originaria in una distribuzione con media 0 e deviazione standard 1;
- il genere, codificato considerando come categoria di riferimento il genere femminile ($F = 0$; $M = 1$);
- la cittadinanza, trasformata in variabile dicotomica (cittadinanza italiana vs straniero) e codificata considerando come categoria di riferimento i bambini italiani (italiano = 0; straniero di prima o seconda generazione = 1);
- l'aver frequentato la scuola dell'infanzia, variabile dicotomica codificata considerando 0 se il bambino ha frequentato la scuola dell'infanzia; 1 se il bambino non ha frequentato la scuola dell'infanzia;
- l'essere o meno anticipatorio rispetto all'ingresso nella scuola primaria, variabile codificata considerando come categoria di riferimento il non essere anticipatori (0);
- il titolo di studio dei genitori. Poiché non sono disponibili, in seconda primaria, i dati del questionario studenti per il calcolo dell'indicatore di status socio-economico-culturale, qui è considerato un *proxy* di tale indice, calcolato come titolo di studio più elevato conseguito dal padre e/o dalla madre. Sono state calcolate due *dummy*: in tutte e due le variabili la categoria di riferimento è relativa agli allievi i cui genitori hanno acquisito come più alto titolo di studio il diploma superiore o una qualifica professionale triennale; nella variabile "licenza media", la categoria di riferimento (diploma o qualifica triennale, codificata come 0) è contrastata a quella degli studenti i cui genitori hanno acquisito come titolo più elevato la licenza elementare o media (codificata come 1); nella variabile "laurea", la categoria di riferimento è contrastata a quella degli studenti i cui genitori hanno acquisito come titolo più elevato un titolo superiore al diploma ("laurea o titolo superiore", "altro titolo superiore al diploma", categoria codificata come 1).

Le variabili a livello classe considerate sono:

- l'organizzazione oraria della classe (0 = fino a 30 ore settimanali; 1 = 40 ore settimanali, tempo pieno);
- la numerosità della classe (variabile centrata rispetto al valore mediano);
- la presenza, nella classe, di una percentuale di bambini stranieri superiore al 15%;
- la presenza, nella classe, di una percentuale di bambini con genitori con titolo di studio superiore al diploma elevata rispetto alla distribuzione complessiva (più del 41% laureati o con altro titolo superiore al diploma, pari al 75° percentile nella distribuzione della percentuale di laureati nelle classi del campione complessivo);
- la composizione della classe rispetto alla capacità di lettura strumentale degli allievi, operazionalizzata considerando la percentuale, sul totale della classe, di allievi con basse prestazioni al test di lettura strumentale (< 1 ds). La variabile identifica con 1 le classi che si collocano al di sopra del 75° percentile per numero di studenti con basse prestazioni e con 0 tutte le altre classi.

È infine considerata la collocazione geografica delle classi, considerando come categoria di riferimento il centro e ottenendo 4 variabili *dummy*: Nord-Ovest, Nord-Est, Sud, Sud-Isole.

Il primo interrogativo sottoposto a indagine è se, considerando gli allievi di seconda primaria, una porzione significativa della variabilità osservata nella comprensione del testo scritto è legata a differenze tra le classi scolastiche, ossia se l'essere allievo di una determinata classe scolastica ha un effetto sulla comprensione del testo. Per rispondere a tale interrogativo, è stata condotta una regressione multilivello a 2 livelli (allievi, livello 1, entro le classi, livello 2) a intercetta *random* "vuoto", ossia in cui non sono presenti predittori. Tale modello consente di ottenere una partizione della varianza in componente tra le classi (le unità di livello 2) e tra gli allievi entro le classi (le unità di livello 1). Il coefficiente di partizione della varianza è pari a 0,123; dunque, circa il 12% della variabilità nella comprensione del

testo è legata a differenze tra le classi scolastiche frequentate dagli allievi. Tale porzione di variabilità è significativa, come è possibile osservare dal test Z di Wald (15,45; $p < 0,001$) e dal *Likelihood ratio* (LR) *test*, in cui sono confrontate le devianze (“-2*Log Likelihood”, 22LL) del modello a intercetta *random* e del modello a intercetta fissa. La statistica test è pari a $LR = 54.929,407 - 53.952,991 = 976,416$, significativamente diversa da 0 (1 gdl; $p < 0,001$).

Il secondo punto indagato è quali variabili rilevate a livello del singolo studente hanno un effetto unico significativo sulla comprensione della lettura e se l'introduzione di tali variabili nel modello porta a una riduzione della varianza non spiegata tra e entro le classi. L'effetto delle variabili a livello studente considerate è globalmente significativo, come indicato dal *Likelihood ratio* (LR) *test*, calcolato come differenza tra le devianze del modello a intercetta *random* vuoto e del modello a intercetta *random* con effetti fissi a livello studente ($LR = 3511,176$; gdl = 7; $p < 0,001$). Le variabili indipendenti considerate nelle analisi hanno un effetto significativo sulla comprensione nella lettura ad eccezione della frequentazione della scuola dell'infanzia ($p = 0,086$), il cui effetto sulla comprensione del testo non è significativo. Tale variabile è stata dunque esclusa dal modello (la differenza tra tale modello e il precedente, calcolata attraverso il *Likelihood ratio test*, non è significativa).

La fluenza nella lettura ha un impatto unico positivo, significativo e di intensità elevata (coefficiente non standardizzato = 0,45; coefficiente standardizzato = 0,38). Ha un effetto unico di entità non trascurabile sulla comprensione della lettura anche il titolo di studio dei genitori, considerato al netto degli altri fattori, inclusa l'abilità di lettura strumentale. Considerando infatti un'allieva con fluenza nella lettura nella media, italiana e che non è entrata in anticipo nella scuola primaria, l'essere figlia di genitori laureati porta a un vantaggio pari a 0,29 punti sulla scala di comprensione del testo, mentre l'essere figlia di genitori con licenza elementare o media a uno svantaggio di circa 0,31 punti sulla stessa scala (coefficienti standardizzati, 0,25 per la variabile che identifica i bambini figli di laureati e -0,27 per la variabile che identifica gli studenti figli di genitori con licenza elementare o media). Inoltre, a parità degli altri fattori, inclusa l'abilità di lettura strumentale, si osserva un'associazione negativa significativa tra l'essere straniero e il punteggio alla prova di comprensione del testo (parametro non standardizzato, -0,22; parametro standardizzato = -0,26). Un'associazione di direzione negativa significativa, seppure di intensità meno elevata, si osserva tra comprensione del testo, il genere e l'essere entrati in anticipo nella scuola primaria. A parità di altri fattori, infatti, i maschi hanno uno svantaggio di 0,06 punti circa rispetto alle femmine e gli allievi anticipatori uno svantaggio di circa 0,13 punti. Con l'introduzione delle variabili a livello studente, la varianza totale non spiegata varia da 1,38 a 1,12. La varianza non spiegata tra gli allievi entro le classi diminuisce da 1,21 a 1,00, dunque, di circa il 18% rispetto alla varianza tra gli allievi del modello nullo; la varianza tra le classi diminuisce da 0,17 a 0,12, dunque, di circa il 27% rispetto alla varianza tra le classi del modello nullo.

Nel modello con le variabili a livello studente, è emerso in particolare il ruolo del punteggio alla prova di lettura strumentale come correlato della comprensione del testo. Tale dato è coerente con la letteratura sull'argomento, considerando l'età dei bambini esaminati. Rispetto ai modelli a un livello, attraverso modelli multilivello è possibile verificare se la forza di tale relazione varia nelle diverse classi scolastiche. A tal fine, è stato aggiunto al modello precedente la componente casuale relativa alla *slope* della fluenza nella lettura. In altre parole, si ipotizza che la variabilità tra le classi non sia solo rispetto all'intercetta, ma anche rispetto alla forza dell'associazione tra fluenza nella lettura e comprensione del testo. Il *Likelihood ratio* (LR) *test*, calcolato come differenza tra i valori pari a “-2*massima verosimiglianza” del modello a intercetta *random* con soli effetti fissi a livello studenti e del modello a intercetta *random* e *random slope* per la fluenza nella lettura è significativo ($LR = 32,190$; gdl = 2; $p < 0,01$). Tale risultato indica una variabilità significativa tra le classi rispetto all'impatto della fluenza nella lettura sulla comprensione del testo. La covarianza tra componenti *random* è significativa e positiva ($\hat{\rho} = 0,18$, $p < 0,05$). Tale risultato indica che le classi con residui più grandi rispetto all'intercetta tendono ad avere residui più grandi anche rispetto alla *slope*, dunque, considerando il segno dell'intercetta (0,012) e della *slope* (0,45), positivo, questo implica che nelle classi in cui l'intercetta è più alta della media la pendenza è maggiore della media mentre la pendenza è più bassa per le classi con intercetta bassa. In altre parole, le differenze tra le classi nella comprensione del testo si accentuano per gli allievi con i livelli più alti di fluenza nella lettura. In alcune classi, in cui a livelli medi di fluenza nella lettura corrisponde un'alta abilità di comprensione del testo, si osservano differenze più ampie nella comprensione del testo tra allievi poco fluenti e allievi con alto livello di fluenza; in altre classi, in cui a livelli medi di fluenza corrispondono prestazioni più basse alla prova di comprensione, le differenze nella comprensione del testo tra allievi fluenti e allievi poco fluenti risultano più ridotte.

Il passo successivo è quello di verificare se e quali variabili a livello della classe hanno un impatto significativo sulla comprensione nella lettura e l'eventuale impatto di tali variabili nel ridurre la variabilità tra e entro le classi. L'organizzazione oraria della classe, la numerosità della classe e il frequentare una classe con percentuale di immigrati superiore al 15% non hanno un effetto significativo e sono dunque rimosse dal modello finale. Due variabili, tra quelle considerate, hanno un effetto significativo e sono mantenute nel modello finale.

La prima è la variabile relativa al titolo di studio dei genitori della classe, che ha un effetto positivo significativo ($p < 0,05$). A parità di altre condizioni, il frequentare una classe con percentuale di allievi con genitori laureati superiore al 41% (75° percentile della distribuzione della percentuale di laureati per classe) porta a un vantaggio di circa 0,08 punti nell'abilità di comprensione del testo.

La seconda variabile di secondo livello (classe) che ha un effetto sull'abilità di comprensione del testo è la composizione della classe rispetto alla presenza di allievi con bassa fluenza nella lettura. Tale variabile è stata ottenuta individuando i bambini con prestazione alla prova di lettura inferiore a una deviazione standard dalla media generale distinguendo le classi in cui la percentuale di tali allievi è elevata rispetto alla distribuzione generale (75° percentile, pari a una percentuale di studenti poco fluenti pari al 28% circa) dalle altre classi. È inoltre considerata l'interazione *cross-level* tra fluenza nella lettura (a livello 1) e l'essere parte di una classe con alta percentuale di studenti con bassa fluenza nella lettura (livello 2). Sia l'effetto principale sia l'interazione sono significativi.

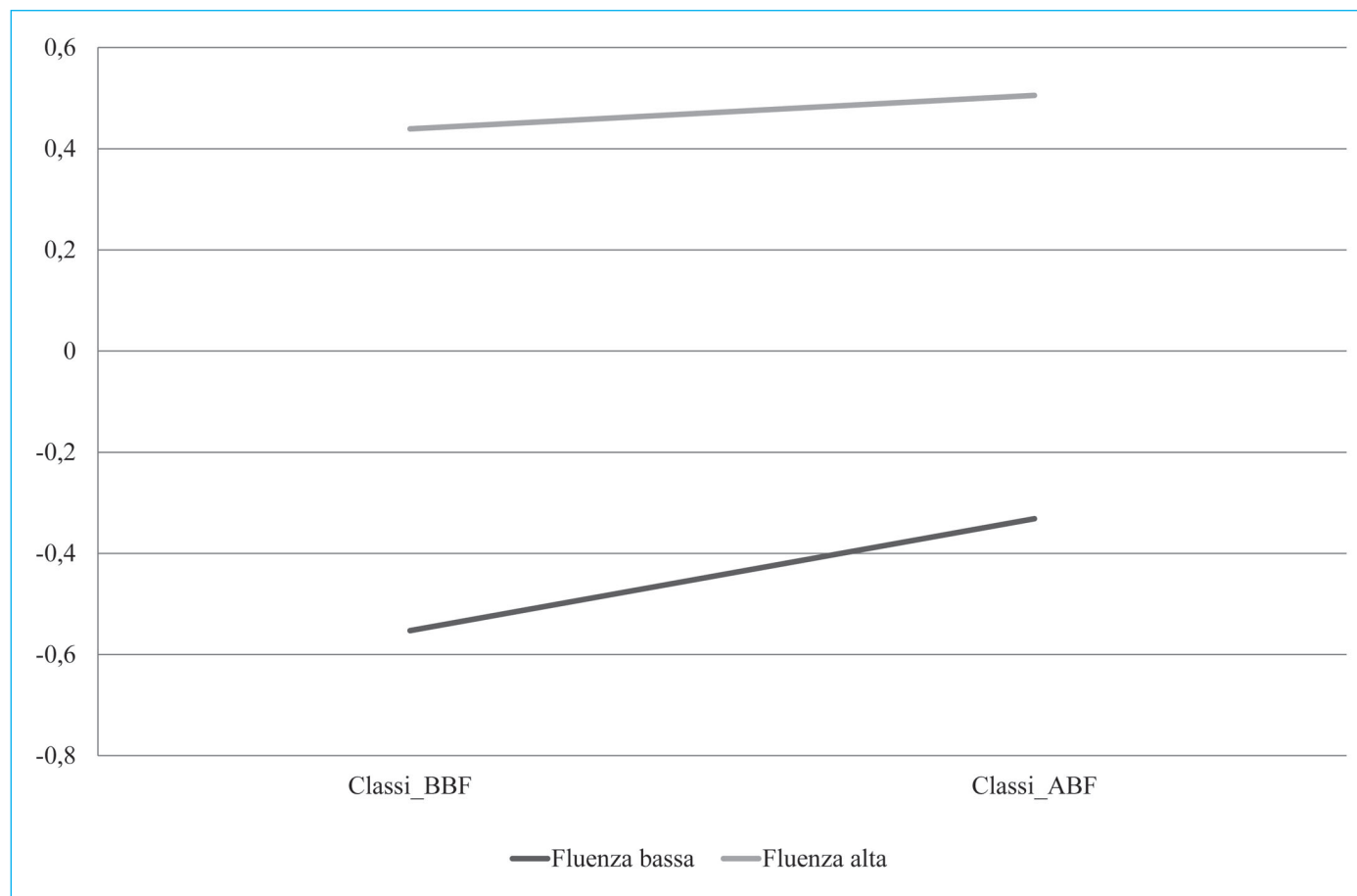
Interpretiamo, dunque, l'effetto della composizione della classe rispetto alla fluenza nella lettura alla luce dell'interazione, attraverso l'esame delle *simple intercepts* e *simple slopes* ai valori condizionati del moderatore, ossia gli effetti di una variabile indipendente sulla variabile dipendente ai diversi livelli del moderatore. Prendendo i valori di fluenza di lettura pari a -1, per le classi con bassa percentuale di studenti con bassa fluenza nella lettura il punteggio di comprensione stimato è pari a -0,55 (*simple intercept*) e la differenza tra frequentare una classe con alta percentuale di studenti con bassa fluenza nella lettura rispetto alle altre classi è pari a 0,22 punti (*simple slope*), differenza significativa ($z = 6,6775$, $p < 0,01$). Per valori di fluenza nella lettura pari a +1, per le classi con bassa percentuale di studenti con bassa fluenza nella lettura il punteggio di comprensione stimato è pari a 0,4391 (*simple intercept*); la differenza tra frequentare una classe con alta percentuale di studenti con bassa fluenza nella lettura rispetto alle altre classi è trascurabile e non significativa (*simple slope* = 0,066; $z = 1,5152$, $p = 0,1298$).

L'introduzione di tali variabili a livello classe porta a una riduzione della variabilità non spiegata per l'intercetta di circa il 5% rispetto al modello precedente. Nell'ultimo modello sottoposto a verifica si vuole indagare se gli effetti fissi precedentemente considerati risultano significativi anche dopo aver inserito nel modello la collocazione geografica delle classi, e se tale variabile di livello 2 porta a una sostanziale riduzione della variabilità tra le classi.

L'introduzione delle variabili relative alla collocazione geografica delle classi porta a una riduzione della variabilità non spiegata a livello dell'intercetta: la varianza tra le classi non spiegata è pari 0,100346; con una riduzione della variabilità tra le classi pari al 9% rispetto al modello precedente e una riduzione della variabilità tra le classi pari al 13% rispetto al modello con i soli predittori di livello 1. Il test LR rispetto al modello *random intercept* e *random slope* con effetti fissi solo a livello 1 è statisticamente significativo (LR = 101,688; $df = 7$, $p < 0,001$), dunque, il modello in cui sono incluse le variabili di livello 2 sopra descritte e la collocazione geografica della scuola si adatta meglio ai dati osservati rispetto al modello con variabili esplicative collocate solo a livello del singolo soggetto. In particolare, osservando i coefficienti degli effetti fissi, osserviamo che, considerando come categoria di riferimento l'essere un allievo di una scuola del Centro Italia, a parità di altre condizioni l'essere un allievo del Nord-Ovest porta a un vantaggio significativo ($p < 0,05$) di circa 0,08 punti, l'essere un allievo di scuole delle regioni del Sud e Sud e Isole porta invece a uno svantaggio di 0,15 e 0,17 punti, rispettivamente ($p < 0,001$).

Anche dopo l'introduzione delle variabili relative alla collocazione geografica delle classi scolastiche, tuttavia, gli altri effetti di livello 2 e 1 introdotti nel modello precedente rimangono significativi, ad eccezione dell'età di ingresso nella scuola primaria, che approssima soltanto la significatività ($p = 0,12$). In particolare, per quanto riguarda l'interazione *cross-level* tra prestazione alla prova, si conferma il vantaggio, rispetto alla comprensione del testo, di frequentare classi con più alta percentuale di bambini con bassa prestazione alla prova di lettura strumentale per i bambini poco fluenti, vantaggio che non è invece significativo per i livelli più alti di abilità (per fluenza nella lettura = -1, *simple intercept* = -0,5098, *errore standard* = 0,0378, $z = -13,4884$, $p = 0,01$; *simple slope* = 0,2174, *errore standard* = 0,0325, $z = 6,6791$, $p < 0,01$; per fluenza nella lettura pari a +1, *simple intercept* = 0,4947, *errore standard* = 0,0362, $z = 13,6764$, $p < 0,01$; *simple slope* = 0,046, *errore standard* = 0,0431, $z = 1,0661$, $p = 0,2864$).

Fig. 1 – Interazione cross-level tra livello di fluenza di lettura dello studente (rappresentata per i valori +1 e -1) e il frequentare una classe con alta (ABF) o medio-bassa (BBF) percentuale di studenti poco fluenti. In ordinata, il punteggio alla prova di comprensione del testo



5. Discussione e conclusioni

Come sottolineato nell'introduzione teorica, gli studi psicologici sulle differenze individuali nella comprensione del testo nelle prime fasi di apprendimento formale della lingua scritta si sono soffermate soprattutto su variabili a livello del singolo studente, quali l'effetto delle altre componenti delle *literacy skills* (per es. la correttezza nella decodifica, la rapidità di decodifica, la fluenza nella lettura), l'effetto del vocabolario, delle abilità di comprensione del linguaggio orale, delle abilità metacognitive, del livello intellettivo e di altri fattori relativi a variabili cognitive associate alla lettura, ma al di fuori del suo dominio (per es. la consapevolezza fonologica, la memoria a breve termine e di lavoro, l'abilità di denominazione rapida automatizzata), nonché l'effetto di alcune variabili relative allo status socio-economico della famiglia di provenienza degli allievi e di altre caratteristiche demografiche, quali per esempio il genere. Se esiste, dunque, una solida letteratura scientifica su un'ampia pletora di variabili a livello studente che possono avere un effetto sulla comprensione del testo, risulta meno approfondito, almeno nelle prime classi della scuola primaria, il ruolo di variabili a livello sovraordinato rispetto all'allievo, quali la classe e la scuola (Kim, Petscher e Foorman, 2015).

Nel presente lavoro, basandosi sulla ricchezza dei dati raccolti da INVALSI nella rilevazione censuaria sulla qualità del Sistema educativo di istruzione e formazione, è stato esplorato in un'ottica multilivello l'effetto di alcune variabili a livello di singolo studente e a livello di classe scolastica sulla prestazione alla prova INVALSI di italiano. Il livello di scolarità considerato è la fine del secondo anno di scuola primaria, dunque, una fase relativamente iniziale del percorso di apprendimento, in cui ci si attende che i bambini stiano stabilizzando le loro abilità di lettura (e scrittura) di base che diventeranno sempre di più, negli anni successivi, "strumentali" agli altri apprendimenti. In particolare, lo studio si è

focalizzato sull'anno scolastico 2014-15, in cui la prova di italiano consta di un brano narrativo seguito da 21 quesiti, volti a indagare i diversi aspetti della competenza di lettura previsti dal Quadro di riferimento in linea con le Indicazioni nazionali per il curricolo per la scuola del primo ciclo di istruzione, e due esercizi linguistici afferenti agli ambiti lessico e semantica/semantica e sintassi.

Il primo interrogativo posto è relativo all'opportunità di adottare un approccio multilivello, ampiamente utilizzato in studi che hanno preso in esame dati relativi agli apprendimenti di studenti in gradi superiori di scolarità (per es. Kim *et al.*, 2015) già alla fine del secondo anno di scuola primaria. I dati ottenuti hanno indicato che una porzione significativa di variabilità nel punteggio alla prova INVALSI di italiano è legata a differenze tra le classi scolastiche. Dunque, già in una fase relativamente iniziale del percorso di apprendimento e nel contesto italiano, in cui non sono previste differenze nei curricula, si osservano differenze significative tra le classi scolastiche, oltre che tra gli allievi entro le classi.

Il secondo interrogativo riguarda quali variabili, a livello studente e a livello della classe scolastica, hanno un effetto significativo sulla prestazione alla prova INVALSI di italiano e contribuiscono a spiegare la variabilità tra le classi e tra gli studenti entro le classi.

In generale, i risultati indicano che le variabili considerate a livello individuale, sia di tipo socio-demografico sia di tipo cognitivo, hanno un impatto unico significativo sulla comprensione del testo, ad eccezione dell'aver frequentato la scuola dell'infanzia (3-6 anni). Quest'ultimo dato, tuttavia, deve essere considerato con cautela. La percentuale di studenti di seconda primaria che non ha frequentato la scuola dell'infanzia è molto bassa (meno del 5%). È possibile che con un maggior dettaglio relativo alla frequentazione della pre-primaria, per esempio con una variabile che tenga conto del numero di anni di scuola dell'infanzia frequentati, si ottengano risultati diversi, sostenendo i dati riscontrati a livello internazionale di un vantaggio per i bambini che hanno frequentato la scuola pre-primaria (per es. OECD, 2013). Il tipo di relazione tra variabili socio-demografiche e comprensione del testo è coerente con i dati presenti in letteratura: a parità di altri fattori è emerso, infatti, l'effetto significativo del genere, con prestazione migliori delle femmine rispetto ai maschi (per es. Elley, 1991; Leppänen *et al.*, 2008), uno svantaggio degli studenti stranieri rispetto agli studenti italiani, e un effetto della regolarità rispetto al corso di studi, con livelli più bassi di comprensione negli studenti anticipatori (per es. Desimoni *et al.*, 2006; Thoren *et al.*, 2016). Si conferma, inoltre, l'impatto del livello di istruzione dei genitori, con un forte vantaggio, a parità di altri fattori, dell'essere figli di genitori con titolo superiore al diploma di secondaria superiore e uno svantaggio di essere figli di genitori con titolo pari o inferiore alla licenza media (per es. Adams, 1990; Davis-Kean, 2005; Bergen, Bishop e de Jong, 2016). Ovviamente tale dato richiede ulteriori approfondimenti, anche condotti alla luce dei modelli proposti da alcuni autori (per es. Davis-Kean, 2005), al fine di indagare quali variabili mediano la relazione tra titolo di studio dei genitori e abilità di lettura dei figli.

Tra le variabili considerate a livello di singolo studente, emerge l'impatto unico significativo della prestazione alla prova di lettura strumentale, al netto di tutte le variabili socio-demografiche inserite nel modello, incluso il livello di istruzione dei genitori. La fine del secondo anno di scuola primaria è considerata un'età particolarmente "critica" per il consolidarsi dei processi alla base del riconoscimento automatizzato delle parole. In particolare, dati presenti in letteratura (per es. Burani, Marcolini e Stella, 2002; Martini *et al.*, 2002; Orsolini *et al.*, 2006) suggeriscono che molti bambini hanno già automatizzato il riconoscimento delle parole, integrando la lenta decodifica attraverso la conversione grafema-fonema (via fonologica) con il più rapido processo di lettura lessicale, con riconoscimento della forma ortografica della parola e accesso diretto al significato. Per altri bambini, invece, potrebbe ancora prevalere la più lenta e laboriosa via fonologica di lettura. In linea con la teoria più diffusa sulla relazione tra fluenza nella lettura e comprensione del testo (LaBerge e Samuels, 1974; Perfetti, 1985), una decodifica più laboriosa, con un carico delle risorse attentive e della memoria di lavoro, può portare a una minore disponibilità di risorse per gli altri processi alla base della piena comprensione del testo e questo potrebbe spiegare la forte relazione, anche al netto di altre caratteristiche degli studenti, tra comprensione e prestazione alla prova di lettura strumentale a fine seconda primaria.

Un dato interessante emerso nel presente lavoro è la variabilità nella forza dell'associazione tra fluenza nella lettura e comprensione del testo tra le classi scolastiche. In particolare, l'effetto della classe frequentata sulla comprensione del testo sembra accentuarsi per i bambini che hanno buone prestazioni alla prova di lettura strumentale e che quindi ipotizziamo avere buone capacità di riconoscimento automatizzato delle parole scritte. È possibile ipotizzare che, per tali bambini, variabili legate alle pratiche didattiche messe in atto dalle insegnanti possano potenziare in modo più o meno efficace i processi cognitivi e metacognitivi alla base della comprensione del testo, portando a un divario tra allievi con buona padronanza della lettura strumentale frequentanti classi diverse. Tale ipotesi, tuttavia, richiede suc-

cessive verifiche e può essere di stimolo a successivi approfondimenti rispetto alla relazione tra lettura strumentale e comprensione.

I dati relativi alle variabili individuali, dunque, confermano quanto ipotizzato in letteratura sulla relazione tra caratteristiche individuali e comprensione nella lettura. Facendo riferimento alla partizione della variabilità in variabilità tra le classi e variabilità tra gli allievi entro le classi, possibile grazie all'approccio multilivello utilizzato nel presente lavoro, è interessante osservare che le variabili considerate a livello individuale non solo contribuiscono a spiegare la variabilità tra le gli allievi, dato atteso considerando che tali caratteristiche fanno proprio riferimento a caratteristiche degli studenti, ma anche a una riduzione sostanziale della variabilità tra le classi. Questo risultato suggerisce che le classi si differenziano sostanzialmente per composizione rispetto alle variabili considerate a livello individuale, e che tali differenze nella composizione della classe hanno un effetto sulle differenze tra le classi rispetto all'abilità di comprensione del testo.

L'effetto di variabili individuali nel ridurre la variabilità tra le classi non è nuovo nella letteratura sull'argomento: per esempio Kim e collaboratori (2015) hanno evidenziato che la fluenza nella lettura, misurata a livello individuale, spiega una porzione significativa di variabilità tra le classi (e tra le scuole) nella comprensione del testo, misurata a fine anno scolastico, anche dopo aver considerato il livello di comprensione della lettura degli allievi all'inizio dell'anno scolastico per tutti i livelli considerati (livello 3 -10). Il ruolo delle caratteristiche delle classi scolastiche, in particolare delle caratteristiche degli allievi che le compongono, per lo sviluppo degli apprendimenti è stata sottolineata da De Fraine e collaboratori (2003). Secondo gli autori, la composizione della classe ha un ruolo sugli apprendimenti che può essere, in alcuni casi, anche più forte di quello ascrivibile alla qualità dell'insegnamento. Mehta e colleghi (2005), in uno studio longitudinale su bambini americani dalla prima alla quarta classe di scuola primaria, hanno riscontrato che la composizione della classe rispetto alle abilità linguistiche e al livello di apprendimento in ingresso spiega una porzione significativa della variabilità nella competenza di lettura, con un ruolo indipendente (e relativamente più forte) da quello esercitato da altre variabili a livello della classe, quale il punteggio ottenuto dalle insegnanti a una scala sulla qualità dell'insegnamento. Nel presente lavoro è stato esplorato il ruolo di più caratteristiche delle classi scolastiche, sia rispetto alla composizione sia rispetto ad alcuni fattori organizzativi.

I risultati indicano che, una volta considerato l'effetto di variabili a livello di singolo studente, l'organizzazione oraria della classe e la numerosità della classe non hanno un effetto significativo sulla comprensione del testo; analogo risultato è stato ottenuto rispetto alla percentuale di studenti stranieri nella classe. Dunque, a parità di altre condizioni, gli studenti italiani hanno prestazioni migliori alla prova INVALSI di italiano rispetto agli studenti stranieri, ma il frequentare una classe con un elevato numero di stranieri non sembra portare a prestazioni peggiori nella comprensione delle letture. Hanno invece un effetto significativo sulla comprensione della lettura, anche al netto della collocazione geografica della scuola di appartenenza, il livello di istruzione dei genitori dei bambini della classe e la composizione della classe rispetto alla fluenza nella lettura degli allievi.

I dati indicano che, a parità di altre condizioni, essere inserito in una classe con alta percentuale di compagni figli di laureati porta a un vantaggio sostanziale nella comprensione della lettura, anche dopo aver tenuto sotto controllo l'effetto della variabile individuale relativa al titolo di studio. Per esempio, a parità di altre condizioni, un bambino figlio di diplomati avrà prestazioni più elevate nella prova di comprensione del testo se inserito in una classe con un'alta percentuale di bambini figli di laureati rispetto all'essere inserito in una classe con compagni i cui genitori hanno come titolo di studio il diploma o un livello inferiore di istruzione. L'importanza dello status socio-economico-culturale sugli esiti scolastici trova ampio riscontro in letteratura e l'effetto sugli apprendimenti della composizione della classe rispetto a tale variabile trova conferma in altre ricerche sull'argomento (per es. Martinez, 2012). Come suggerito da Davis-Kean (2005), tuttavia, ulteriori ricerche sono necessarie per spiegare i meccanismi che possono mediare la relazione tra livello di istruzione dei genitori e competenze di lettura. Alla luce dei risultati ottenuti nel presente lavoro, una possibile strada di ricerca è quella di approfondire non solo come tali meccanismi possono agire entro la famiglia del bambino, ma come le eventuali opportunità, credenze e atteggiamenti che possono derivare da un livello superiore di istruzione possano influenzare, anche indirettamente, il micro-sistema in cui il bambino si trova a interagire, la classe scolastica, provvedendo a colmare il possibile svantaggio derivante per alcuni bambini dal basso status socio-economico culturale della famiglia di provenienza.

Risulta, infine, interessante il dato relativo alla composizione della classe rispetto alla percentuale di studenti poco fluenti nella lettura. Considerato il livello scolastico preso in esame, è possibile che in alcune classi si osservi una per-

centuale rilevante di bambini che sono ancora poco fluenti nella lettura, con una bassa automatizzazione del processo di riconoscimento delle parole con accesso al significato delle stesse. Ci si è dunque chiesti se essere parte di tali classi porti a un vantaggio o a uno svantaggio rispetto alla comprensione del testo e per quale livello di abilità tale effetto si verifichi. Se facciamo riferimento a studenti con buone capacità di lettura strumentale, per esempio, è possibile che il frequentare una classe con compagni che presentano una decodifica lenta e laboriosa porti a uno svantaggio, in quanto è possibile che gli insegnanti siano costretti, sulla base delle difficoltà incontrate dai bambini, a soffermarsi su attività che potenziano la lettura strumentale, a discapito della comprensione del testo; oppure potrebbero essere gli studenti in difficoltà nella lettura a beneficiare dell'essere inseriti in una classe in cui i compagni hanno già sviluppato adeguate capacità di lettura strumentale, oppure dall'essere inseriti in classi in cui altri compagni condividono le stesse difficoltà nell'automatizzazione dei processi di lettura. I dati ottenuti nel presente lavoro sembrano confermare quest'ultima ipotesi. Infatti, i risultati indicano che a parità di altri fattori a prestazioni più basse nella prova INVALSI di lettura strumentale corrispondono punteggi più bassi alla prova di italiano sia nelle classi con alta percentuale di studenti poco fluenti sia nelle altre classi; per i bassi livelli di abilità di lettura strumentale, tuttavia, emerge una differenza tra le classi in funzione della percentuale di studenti con bassa fluenza nella lettura.

A parità di altri fattori, gli studenti poco fluenti hanno prestazioni migliori nella prova di comprensione del testo se inseriti in una classe in cui la percentuale di studenti con livello basso di fluenza è più elevata; la differenza tra frequentare una classe con alta percentuale di studenti poco fluenti e le altre classi si riduce all'aumentare del livello di fluenza di lettura dello studente, con differenze non significative per i livelli alti di abilità. Tale risultato è aperto a numerose interpretazioni, che potranno essere oggetto di indagine in future ricerche: per esempio, è possibile che i bambini che a fine seconda primaria sono ancora lettori poco fluenti, se inseriti in classi in cui gli altri studenti hanno ormai adeguate abilità di lettura strumentale possano sviluppare maggiori sentimenti di inadeguatezza, con atteggiamenti negativi nei confronti di attività associate alla lettura e poca motivazione a leggere anche al di là del contesto scolastico, con conseguenze negative per lo sviluppo della comprensione del testo. È altresì possibile che l'osservare che nella classe ci sono numerosi bambini con difficoltà nella lettura strumentale porti l'insegnante a svolgere una maggiore attività di potenziamento delle attività di lettura, con effetti positivi sulla comprensione del testo.

In conclusione, nel presente studio, attraverso l'esame dei dati di una rilevazione su larga scala nel contesto italiano, è stato possibile estendere le analisi delle associazioni tra comprensione del testo, fluenza nella lettura e altre variabili rilevanti nei primi anni di scolarità obbligatoria considerando non solo fattori a livello del singolo studente, ma anche caratteristiche delle classi scolastiche, con particolare riferimento alla composizione delle classi. Ulteriori ricerche sono necessarie per confermare quanto emerso e per individuare i meccanismi alla base degli effetti riscontrati, prendendo in esame variabili che possono agire da possibili mediatori nel *pattern* di effetti e interazioni riscontrati, nonché per individuare ulteriori caratteristiche delle classi scolastiche che possono avere un effetto sulla comprensione della lettura.

Riferimenti bibliografici

- Adams M.J. (1990), *Beginning to Read: Thinking and Learning about Print*, MIT Press, Cambridge (MA).
- AID, Associazione italiana dislessia (2007), *Consensus Conference, Disturbi evolutivi specifici di apprendimento. Raccomandazioni per la pratica clinica definite con il metodo della Consensus Conference*, Montecatini Terme, 22-23 settembre 2006, Milano 26 gennaio 2007, testo disponibile al sito: <http://hubmiur.pubblica.istruzione.it/alfresco/d/d/workspace/SpacesStore/a9cf25e3-2cd0-471a-9b44-f2d12d17e5a3/raccomandazionidsa2007.pdf>, data di consultazione: 3 aprile 2017.
- Al Otaiba S., Kosanovich M.L., Torgesen J.K., Kamhi A.G., Catts H.W. (2012), "Assessment and instruction for phonemic awareness and word recognition skills", *Language and Reading Disabilities*, 3: 112-140.
- Arcolini I., Zardini G. (a cura di) (2002), *I disturbi di apprendimento della lettura e della scrittura*, FrancoAngeli, Milano.
- Bashir A.S., Hook P.E. (2009), "Fluency: a key link between word identification and comprehension", *Language, Speech and Hearing Services in Schools*, 40, 2: 196-200.
- Bergen E., Zuijzen T., Bishop D., Jong, P.F. (2016), "Why are home literacy environment and children's reading skills associated? What parental skills reveal", *Reading Research Quarterly*, 52, 2: 147-160.
- Bowey J.A. (1995), "Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement", *Journal of Educational Psychology*, 87, 3: 476-487.

-
- Burani, C., Marcolini, S., Stella, G. (2002), "How early does morpholexical reading develop in readers of a shallow orthography?", *Brain and Language*, 81, 1: 568-586.
- Campodifiori E., Falzetti P., Papini M. (2016), "La frequenza della scuola pre-primaria e gli apprendimenti scolastici", *INVALSI Working Papers*, 27, testo disponibile al sito: http://www.invalsi.it/download/wp/wp27_Campodifiori_Falzetti_Papini.pdf, data di consultazione: 3 aprile 2017.
- Campodifiori E., Figura E., Martini A., Papini M. (2011), "La prova di lettura strumentale di II Primaria e la relazione con la comprensione del testo", *INVALSI Working Papers*, 15, testo disponibile al sito: http://www.invalsi.it/download/wp/wp15_Campodifiori.pdf, data di consultazione: 3 aprile 2017.
- Carlisle J., Rice M.S. (2002), *Improving Reading Comprehension: Research-Based Principles and Practices*, York Press, Baltimora.
- Davis-Kean P.E. (2005), "The influence of parent education and family income on child achievement: the indirect role of parental expectations and the home environment", *Journal of Family Psychology*, 19, 2: 294-304.
- De Fraine B., Van Damme, J., Van Landeghem G., Opdenakker M.C., Onghena P. (2003), "The effect of schools and classes on language achievement", *British Educational Research Journal*, 29, 6: 841-859.
- De Jong P.F., van der Leij A. (2002), "Effects of phonological abilities and linguistic comprehension on the development of reading", *Scientific Studies of Reading*, 6, 1: 51-77.
- Desimoni M., Pelagaggi D., Scalisi T.G. (2006), "Differenze di età nelle prestazioni di lettura e scrittura in un campione di bambini alla fine del primo anno della scuola primaria", *Atti del XV Congresso Nazionale AIRIPA*, Roma, 20-21 ottobre 2006: 23.
- Desimoni M., Scalisi T.G., Orsolini M. (2012), "Predictive and concurrent relations between literacy skills in grades 1 and 3: a longitudinal study of Italian children", *Learning and Instruction*, 22, 5: 340-353.
- Duncan G.J., Dowsett C.J., Claessens A., Magnuson K., Huston A.C., Klebanov P. (2007), "School readiness and later achievement", *Developmental Psychology*, 43, 1428-1446.
- Elley W.B. (1991), "Acquiring literacy in a second language: the effect of book-based programs", *Language Learning*, 41, 3: 375-411.
- Gaskins I.W., Satlow E., Pressley M. (2007), "Executive control of reading comprehension in the elementary school", in L. Metzler (ed.), *Executive Function in Education: From Theory to Practice*, Guilford Press, New York: 194-215.
- Goff D.A., Pratt C., Ong B. (2005), "The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample", *Reading and Writing*, 18, 7-9: 583-616.
- Goldstein H. (1997), "Methods in school effectiveness research", *School Effectiveness and School Improvement*, 8: 369-395.
- Hill P.W., Rowe K.J. (1996), "Multilevel modeling in school effectiveness research", *School Effectiveness and School Improvement*, 7, 1: 1-34.
- Hulme C., Snowling M.J. (2011), "Children's reading comprehension difficulties nature, causes, and treatments", *Current Directions in Psychological Science*, 20, 1: 139-142.
- INVALSI (2011), *Rapporto indagini IEA 2011 PIRLS e TIMSS: i risultati degli studenti italiani in lettura, matematica e scienze*, testo disponibile al sito: http://www.invalsi.it/invalsi/ri/timss2011/documenti/Rapporto_PIRLS_TIMSS.pdf, data di consultazione: 20 marzo 2017.
- INVALSI (2015), *Rilevazioni nazionali degli apprendimenti 2014-15, Rapporto risultati*, testo disponibile al sito: http://www.invalsi.it/invalsi/doc_evidenza/2015/034_Rapporto_Prove_INVALSI_2015.pdf, data di consultazione: 20 marzo 2017.
- INVALSI (2016), *Rilevazioni nazionali degli apprendimenti 2015-16, Rapporto risultati*, testo disponibile al sito: http://www.invalsi.it/invalsi/doc_evidenza/2016/07_Rapporto_Prove_INVALSI_2016.pdf, data di consultazione: 20 marzo 2017.
- Jenkins J.R., Fuchs L.S., Van Den Broek P., Espin C., Deno S.L. (2003), "Sources of individual differences in reading comprehension and reading fluency", *Journal of Educational Psychology*, 95, 4: 719-729.
- Juel C., Griffith P.L., Gough P.B. (1986), "Acquisition of literacy: a longitudinal study of children in first and second grade", *Journal of Educational Psychology*, 78, 4: 243-255.
- Kim Y.S., Petscher Y., Foorman B. (2015), "The unique relation of silent reading fluency to end-of-year reading comprehension: understanding individual differences at the student, classroom, school, and district levels", *Reading and Writing*, 28, 1: 131-150.
- Kirby J.R., Desrochers A., Roth L., Lai S.S. (2008), "Longitudinal predictors of word reading development", *Canadian Psychology/Psychologie canadienne*, 49, 2: 103-110.
- Klauda S.L., Guthrie J.T. (2008), "Relationships of three components of reading fluency to reading comprehension", *Journal of Educational Psychology*, 100: 310-321.
- LaBerge D., Samuels S.J. (1974), "Toward a theory of automatic information processing in reading", *Cognitive Psychology*, 6, 2: 293-323.
- Landerl K., Wimmer H. (2008), "Development of word reading fluency and spelling in a consistent orthography: an 8-year follow-up", *Journal of Educational Psychology*, 100, 1: 150-161.
- Leppänen U., Aunola K., Niemi P., Nurmi J.E. (2008), "Letter knowledge predicts Grade 4 reading fluency and reading comprehension", *Learning and Instruction*, 18, 6: 548-564.
-

-
- Leppänen U., Niemi P., Aunola K., Nurmi J.E. (2004), "Development of reading skills among preschool and primary school pupils", *Reading Research Quarterly*, 39, 1: 72-93.
- Martínez J.F. (2012), "Consequences of omitting the classroom in multilevel models of schooling: an illustration using opportunity to learn and reading achievement", *School Effectiveness and School Improvement*, 23, 3: 305-326.
- Martini A., Brizzolara D., Pecini C., Dinetti D. e Negrin E. (2002), "Aspetti diagnostici e riabilitativi in bambini dislessici italiani", in I. Arcolini, G. Zardini (a cura di), *I disturbi di apprendimento della lettura e della scrittura*, FrancoAngeli, Milano: 223-235.
- Mehta P.D., Foorman B.R., Branum-Martin L., Taylor W.P. (2005), "Literacy as a unidimensional multilevel construct: validation, sources of influence, and implications in a longitudinal study in grades 1 to 4", *Scientific Studies of Reading*, 9, 2: 85-116.
- Metzler L. (ed.) (2007), *Executive Function in Education: From Theory to Practice*, Guilford Press, New York.
- Müller K., Brady S. (2001), "Correlates of early reading performance in a transparent orthography", *Reading and Writing*, 14, 7-8: 757-799.
- OECD (2013), *PISA 2012 Results: What makes Schools Successful? Resources, Policies and Practices* (vol. IV), PISA, OECD Publishing, testo disponibile al sito: <http://dx.doi.org/10.1787/9789264201156-en>, data di consultazione: 20 marzo 2017.
- Orsolini M., Fanari R., Tosi V., De Nigris B., Carrieri R. (2006), "From phonological recoding to lexical reading: a longitudinal study on reading development in Italian", *Language and Cognitive Processes*, 21, 5: 576-607.
- Paris S.G. (2005), "Reinterpreting the development of reading skills", *Reading Research Quarterly*, 40, 2: 184-202.
- Pazzaglia F., Cornoldi C., Tressoldi P.E. (1993), "Learning to read: Evidence on the distinction between decoding and comprehension skills", *European Journal of Psychology of Education*, 8, 3: 247-258.
- Perfetti C.A. (1985), *Reading Ability*, Oxford University Press.
- Pontecorvo C., Pontecorvo M. (1986), *Psicologia dell'educazione. Conoscere a scuola*, il Mulino, Bologna.
- Pufpaff L.A. (2009), "A developmental continuum of phonological sensitivity skills", *Psychology in the Schools*, 46, 7: 679-691.
- Rasch, G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Nielsen & Lydiche, Copenhagen.
- Rowe K.J., Hill P.W., Holmes-Smith P. (1995), "Methodological issues in educational performance and school effectiveness research: A discussion with worked examples", *Australian Journal of Education*, 39, 3: 217-248.
- Seymour P.H., Aro M., Erskine J.M. (2003), "Foundation literacy acquisition in European orthographies", *British Journal of Psychology*, 94, 2: 143-174.
- Storch S.A., Whitehurst G.J. (2002), "Oral language and code-related precursors to reading: evidence from a longitudinal structural model", *Developmental Psychology*, 38, 6: 934-947.
- Suchodoletz A., Larsen R.A., Gunzenhauser C., Fäsche A. (2015), "Reading and spelling skills in German third graders: Examining the role of student and context characteristics", *British Journal of Educational Psychology*, 85, 4: 533-550.
- Thoren K., Heinig E., Brunner M. (2016), "Relative age effects in mathematics and reading: investigating the generalizability across students, time and classes", *Frontiers in Psychology*, 7: 1-12.
- Verhoeven L., Reitsma P., Siegel L.S. (2011), "Cognitive and linguistic factors in reading acquisition", *Reading and Writing*, 24, 4: 387-394.
- Verhoeven L., Van Leeuwe J. (2008), "Prediction of the development of reading comprehension: a longitudinal study", *Applied Cognitive Psychology*, 22, 3: 407-423.
- Wolf M., Katzir-Cohen T. (2001), "Reading fluency and its intervention", *Scientific Studies of Reading*, 5, 3: 211-239.

13. La Prova nazionale INVALSI e l'esame conclusivo del primo ciclo d'istruzione

The role of INVALSI national test in the leaving exam of lower secondary education

di Angela Martini

Il paper affronta il ruolo giocato dalla Prova nazionale INVALSI (PN) nell'esame di licenza media. Basandosi sui dati dell'anno 2014-15, esso analizza le distribuzioni dei voti di tutte le prove d'esame, tra cui quello della Prova nazionale, per l'Italia nel suo insieme e per ciascuna delle tre grandi aree geografiche del Paese (Nord, Centro e Sud), nonché le correlazioni fra di essi. Il peso esercitato dalla Prova nazionale sul voto finale è esaminato calcolando la media dei voti delle prove sia comprendendo la Prova nazionale sia escludendola. Sono anche analizzate le relazioni tra i punteggi, non corretti e corretti per il *cheating*, nelle prove INVALSI di italiano e di matematica (che insieme confluiscono nel voto della Prova nazionale) da una parte, e il voto della Prova nazionale e delle prove interne alle scuole nelle stesse materie dall'altra parte. Dalle analisi emerge che:

- le distribuzioni dei voti differiscono in parte tra le diverse aree del Paese;
- l'incidenza del voto della Prova nazionale sul voto finale è limitata: il 91% degli alunni è promosso con un voto pari al voto di ammissione, nel 6% dei casi il voto della Prova nazionale abbassa il voto finale e nel 3% lo alza;
- la relazione tra i punteggi, corretti per il *cheating*, nelle prove INVALSI di italiano e matematica e i voti della Prova nazionale e delle prove interne differisce nelle varie aree dell'Italia.

The paper deals with the role played by the INVALSI national test (PN) in the leaving exam at the end of lower secondary education. Based on the 2014-15 data, it analyzes the distributions of all the examination marks (including the mark of the national test) for Italy as a whole and for each of three geographical areas: North Italy, Centre Italy and South Italy. The correlations of each examination mark with the others are also analyzed. The weight the national test has on the final evaluation of the students is estimated by averaging the marks calculated with and without the national test mark.

Further analyses have also been carried out to examine, on one side, the relation between the scores, biased and not biased by cheating, of the Italian and of the Mathematical test (that contribute together to the unique mark of the national test) and on the other side the relation between the marks of the national test and of the school examinations in the same subjects. The paper conclusions are:

- the marks distributions are partly different in the three Italian geographical areas;
- the weight of the national test mark on the student final evaluation is small: 91% of all successful students get the same mark they got as admitted candidates to the examinations, 6% get a lower mark, 3% get a higher mark;
- the relation between the scores, not biased by cheating, of the INVALSI Italian and Mathematical tests and the marks both of the national test as a whole and of school examinations differs from one Italian area to another.

1. Premessa

Nel 2008 è stata per la prima volta introdotta fra le prove dell'esame di conclusione del primo ciclo d'istruzione una prova standardizzata eguale per tutti gli alunni composta di due parti: una prova di italiano, a sua volta costituita da domande di comprensione della lettura di almeno due testi (narrativo ed espositivo) e da una decina di quesiti gram-

ticali, e una prova di matematica con domande in ognuno dei quattro ambiti dell'aritmetica (Numeri), della geometria (Spazio e figure), dell'algebra (Relazioni e funzioni) e della statistica (Dati e previsioni). Le due prove che sono alla base della Prova nazionale – come viene ufficialmente denominata la prova standardizzata di cui si sta discutendo – hanno la stessa struttura e impostazione delle prove di italiano e matematica a cui l'INVALSI sottopone ogni anno gli alunni della seconda e quinta classe di scuola primaria e della seconda classe di scuola secondaria di II grado, ma mentre queste ultime hanno l'obiettivo di monitorare i risultati raggiunti, in termini di livelli d'apprendimento obiettivamente misurati, dal sistema d'istruzione nel suo insieme e nelle sue articolazioni territoriali (macro-aree e regioni) e l'efficacia delle singole scuole, nel caso delle prove della terza classe della scuola secondaria inferiore, alla finalità di monitoraggio che esse condividono con quelle degli altri livelli scolari, si aggiunge la finalità di contribuire alla valutazione dei singoli studenti. Il voto della Prova nazionale entra infatti – insieme al voto d'ammissione, alle prove scritte interne a ogni singola scuola (italiano, matematica, lingua straniera)¹ e al colloquio d'esame – nel calcolo della media dei voti che dà luogo, dopo arrotondamento all'intero, alla valutazione finale d'uscita assegnata a ogni studente.

È opportuno sottolineare, prima di proseguire, due punti:

- 1) il voto della Prova nazionale è unico, benché si basi, come già detto, su due prove distinte, una di italiano e una di matematica;
- 2) il passaggio dai punteggi nelle prove d'italiano e matematica all'unico voto della Prova nazionale non è un passaggio meccanico – come accadrebbe se, per esempio, si traducesse direttamente la media dei punteggi percentuali delle due prove in un voto in decimi – ma implica una procedura più complessa, caratterizzata dalla suddivisione degli item delle due prove in tre blocchi, ai quali viene dato un peso diverso a seconda del grado di difficoltà dei quesiti che li compongono misurato dalla “prova sul campo” a cui, nella fase di costruzione, vengono sottoposti².

Da notare, però, che la procedura di conversione dai punteggi delle due prove al voto unitario in decimi, uniformemente applicata in tutte le scuole grazie anche a un software apposito fornito dall'INVALSI, è stata introdotta solo nel 2010, cosicché è solo da questa data che i voti della Prova nazionale sono confrontabili tra loro³; nei due anni precedenti (2008 e 2009), la valutazione della Prova nazionale era affidata alle commissioni d'esame costituite presso le singole scuole e dunque i voti a essa assegnati, decisi individualmente da ogni commissione, non erano comparabili. Rimane però un problema: i punteggi di italiano e matematica delle due parti della Prova nazionale, che sono il punto di partenza della procedura di conversione nell'unico voto decimale, non sono corretti per il *cheating*⁴ ed è dunque presumibile un certo “gonfiamento” sia dei voti della Prova nazionale sia, di conseguenza, dei voti finali d'esame, specie in quelle aree dove comportamenti opportunistici da parte di studenti e insegnanti sono più diffusi.

Come ultima considerazione, prima di chiudere questo paragrafo introduttivo, osserviamo che la Prova nazionale, eguale per tutti gli alunni e corretta e valutata con le medesime modalità in tutte le scuole italiane a differenza di quanto accade per le altre prove, è la sola che conferisca all'esame di licenza media un certo grado di esternalità, riducendo, sebbene in misura assai limitata, la “soggettività” della valutazione finale degli studenti, che rimarrebbe altrimenti del tutto autoreferenziale. Dicendo questo non si vuol dire, come si è avuto occasione di osservare altrove (Martini, 2009), che gli insegnanti che fanno parte delle commissioni d'esame valutino male i loro alunni ma solo che, data la diversità delle prove interne da una scuola all'altra, dei criteri di valutazione e soprattutto delle popolazioni di studenti che frequentano i vari istituti, i voti degli insegnanti, come la ricerca dimostra, non sono confrontabili fra loro (Dardanoni, Modica e Pennisi, 2007). Nel prosieguo di questo lavoro si cercherà di vedere, basandosi sui risultati delle prove dell'esame di conclusione del primo ciclo dell'anno 2015, quale ruolo giochi la Prova nazionale, quale sia la correlazione tra il voto di questa prova e quelli delle altre prove e quale ne sia l'impatto sul voto finale dei singoli studenti.

¹ Le prove scritte di lingua straniera possono essere due o una qualora la seconda lingua straniera venga valutata solo all'interno del colloquio. Nel primo caso, il più frequente, i voti che contribuiscono al voto finale d'esame sono 7, mentre nel secondo sono 6 e di conseguenza la Prova nazionale INVALSI viene a pesare su di esso per 1/7 o per 1/6.

² La procedura di passaggio dai punteggi delle due prove di italiano e matematica all'unico voto della Prova nazionale è stabilita ogni anno dal comitato di esperti che presiede alla costruzione delle prove INVALSI ed è descritta sul sito web dell'Istituto nei documenti di accompagnamento alla Prova nazionale.

³ Bisogna precisare che il confronto è al momento possibile solo sincronicamente per i voti relativi alla prova di ciascun anno, ma non è ancora possibile in senso diacronico, da un anno all'altro, non essendo le prove ancorate fra loro.

⁴ Il termine – che in inglese significa “imbrogliare” – designa quei casi in cui i risultati di un test riflettono non tanto il grado in cui i soggetti che vi sono sottoposti posseggono l'attributo che il test intende misurare ma piuttosto l'effetto del ricorso a mezzi direttamente o indirettamente truffaldini per alterare verso l'alto le misure (come, per esempio, permettere agli alunni di copiare gli uni dagli altri o suggerire le risposte).

2. I dati

Le analisi di cui si espongono qui gli esiti sono state effettuate sui dati delle prove d'esame 2015 del campione di circa 28.500 studenti estratto su tutti gli studenti italiani che frequentavano il terzo anno della scuola secondaria inferiore. Dal campione originale sono stati eliminati poco più di 2.500 studenti per i quali non è stato possibile raccogliere i voti di tutte le prove d'esame. Il campione di studenti su cui sono state condotte le analisi si compone, dopo le eliminazioni, di 25.946 studenti, così ripartiti fra le tre grandi aree geografiche del Paese: 9.790 nell'Italia settentrionale (area Nord), 5.532 nell'Italia centrale (area Centro), 10.624 nell'Italia meridionale e insulare (area Sud). Le elaborazioni sono state fatte sui dati non pesati relativi, oltre che a tutti i voti d'esame, ai punteggi percentuali ottenuti dagli studenti in ciascuna delle due prove INVALSI di italiano e matematica che fanno parte congiuntamente della Prova nazionale. I pochissimi dati mancanti di alcune delle variabili prese in esame (meno dello 0,5%) sono stati imputati con la media della serie o col valore mediano.

3. I voti d'esame nelle tre aree dell'Italia

Le tabelle che seguono mostrano la distribuzione percentuale, la media e la mediana dei voti di ammissione all'esame, delle prove scritte interne di italiano, lingua straniera e matematica, della Prova nazionale INVALSI, del colloquio orale e della valutazione finale per ognuna delle tre principali aree geografiche e per l'Italia nel suo insieme.

Tab. 1 – Distribuzione percentuale, media e mediana dei voti di ammissione all'esame di licenza per area

Area	6	7	8	9	10	Media	Mediana
Nord	21,9	31,7	26,1	15,2	5,1	7,5	7
Centro	21,2	31,7	25,8	15,7	5,7	7,5	7
Sud	23,4	29,2	22,7	16,5	8,2	7,6	7
Italia	22,4	30,7	24,6	15,8	6,5	7,5	7

Tab. 2 – Distribuzione percentuale, media e mediana dei voti della prova interna di italiano per area

Area	≤ 4	5	6	7	8	9	10	Media	Mediana
Nord	0,2	2,3	18,9	30,8	24,0	15,0	8,8	7,6	7
Centro	0,1	1,6	17,4	28,5	24,6	16,1	11,7	7,7	8
Sud	0,1	1,3	19,7	26,4	21,6	15,3	15,6	7,8	8
Italia	0,1	1,8	18,9	28,5	23,1	15,4	12,2	7,7	8

Tab. 3 – Distribuzione percentuale, media e mediana dei voti della prova interna di lingua straniera per area

Area	≤ 4	5	6	7	8	9	10	Media	Mediana
Nord	1,2	8,0	22,4	24,5	20,0	14,9	9,0	7,4	7
Centro	1,2	6,7	21,6	25,4	20,0	14,4	10,7	7,4	7
Sud	0,3	3,7	25,4	23,5	18,8	15,5	12,9	7,6	7
Italia	0,8	6,0	23,4	24,3	19,5	15,0	10,9	7,4	7

Tab. 4 – Distribuzione percentuale, media e mediana dei voti della prova interna di matematica per area

Area	≤ 4	5	6	7	8	9	10	Media	Mediana
Nord	6,9	13,7	17,5	16,0	17,0	15,2	13,7	7,2	7
Centro	5,9	11,7	17,3	17,4	15,9	15,8	16,0	7,4	7
Sud	2,0	8,0	23,2	18,9	15,7	14,3	17,9	7,5	7
Italia	4,7	10,9	19,8	17,5	16,2	15,0	15,9	7,4	7

Tab. 5 – Distribuzione percentuale, media e mediana dei voti della Prova nazionale INVALSI per area

Area	4	5	6	7	8	9	10	Media	Mediana
Nord	4,1	13,8	15,7	16,0	20,1	17,8	12,4	7,4	8
Centro	3,6	14,2	17,6	17,3	20,6	16,8	9,9	7,3	7
Sud	4,4	15,1	17,7	18,6	18,9	15,0	10,3	7,2	7
Italia	4,1	14,4	16,9	17,4	19,7	16,5	11,0	7,3	7

Tab. 6 – Distribuzione percentuale, media e mediana dei voti del colloquio d'esame per area

Area	≤ 4	5	6	7	8	9	10	Media	Mediana
Nord	1,1	6,5	19,8	22,0	19,6	14,9	16,1	7,6	8
Centro	0,9	5,0	17,9	21,1	20,2	14,9	20,1	7,8	8
Sud	0,3	3,6	19,4	19,4	19,1	15,2	23,1	7,9	8
Italia	0,7	5,0	19,2	20,7	19,5	15,0	19,8	7,8	8

Tab. 7 – Distribuzione percentuale, media e mediana dei voti finali d'esame per area

Area	5	6	7	8	9	10	Media	Mediana
Nord	0,2	25,8	26,4	23,1	17,3	7,2	7,5	7
Centro	0,1	23,3	26,0	24,3	17,5	8,8	7,6	8
Sud	0,0	23,8	25,1	21,8	17,7	11,5	7,7	8
Italia	0,1	24,5	25,8	22,8	17,5	9,3	7,6	7

Scorrendo le tabelle, si può notare che, in generale, i maggiori scostamenti fra le tre aree dell'Italia nelle distribuzioni percentuali dei voti assegnati dalle scuole (esclusa, dunque, la Prova nazionale) si osservano nei voti estremi: in corrispondenza del voto 10, la differenza tra la percentuale registrata al Nord, più bassa, e quella del Sud, più alta, va dai 4 punti circa nella prova di lingua straniera ai 7 punti nel colloquio orale; all'inverso, in corrispondenza dei due voti più bassi, 4 e 5, il Sud fa registrare percentuali che sono la metà di quelle del Nord. Il Centro si colloca in una posizione intermedia fra le due aree settentrionale e meridionale, ma ora più vicino al Nord, ora al Sud, a seconda della distribuzione di voti considerata. Un andamento diverso e opposto si osserva invece per la Prova nazionale, dove la percentuale di voti 9 e 10 è più alta al Nord rispetto al Centro e più ancora al Sud, mentre la percentuale dei voti 4 e 5 è maggiore al Sud rispetto sia al Nord sia al Centro, che hanno valori abbastanza simili fra loro. In definitiva, dunque, le distribuzioni dei voti, tranne quella della Prova nazionale, appaiono più spostate verso i valori più alti nel Sud in confronto al Centro ma soprattutto al Nord.

Ulteriori spunti di riflessione emergono se si considerano le medie e le mediane delle distribuzioni. Si ricorda, innanzitutto, che quando la media è superiore alla mediana, ciò significa che la maggioranza dei dati registra valori più bassi della media, mentre quando la media è inferiore alla mediana prevalgono i valori più alti della media. In genere, in tutte le aree e in tutte le prove la media tende a essere più alta della mediana, ma sono constatabili alcune significative eccezioni: una prima eccezione riguarda proprio la Prova nazionale, dove la media dei voti è nel Nord più bassa della mediana, al contrario di quanto accade nelle altre due aree, confermando così i migliori risultati del Nord in questa prova. La seconda eccezione riguarda la prova interna di italiano, dove si assiste a un fenomeno opposto al precedente: mentre al Nord la media è maggiore della mediana, al Centro e al Sud succede il contrario, denotando come la propensione delle scuole di queste due aree a valutazioni più generose rispetto alle scuole del Nord sia particolarmente visibile in questa prova. Un ultimo dato d'interesse riguarda i voti del colloquio orale: qui in tutta Italia la media è più bassa della mediana e quest'ultima risulta, in ognuna delle tre aree, pari al voto di 8. Nella valutazione del colloquio sembra dunque profilarsi un'uniformità di comportamenti da parte degli insegnanti i quali, a prescindere dall'area geografica dove operano, condividono una comune tendenza a largheggiare nel voto, probabilmente anche perché questa prova, di cui non esiste un'autonoma documentazione (il verbale è redatto dagli stessi docenti della commissione), si presta particolarmente a compensare eventuali carenze degli studenti in altre prove. Per quanto riguarda, infine, il voto finale, la relazione tra media e mediana nelle tre aree è analoga a quella già osservata nella prova interna di italiano: ancora una volta nel Nord prevalgono i voti più bassi della media, mentre nel Centro e nel Sud accade l'inverso.

Da ultimo, notiamo che la frequenza delle lodi, coerentemente al quadro d'insieme fin qui osservato, è più alta al Sud e al Centro rispetto al Nord ma ciò è soprattutto frutto della maggiore frequenza di 10 nell'assegnazione delle valutazioni finali nelle prime due aree, come si può vedere dalla tabella seguente.

Tab. 8 – Alunni licenziati con lode su totale alunni promossi e su totale promossi con 10 per area

	N. alunni promossi con 10 (tra parentesi %)	% licenziati con lode su totale alunni promossi con 10	% licenziati con lode su totale alunni promossi
Nord	702 (7,2)	34,5	2,5
Centro	485 (8,8)	39,0	3,4
Sud	1.217 (11,5)	33,9	3,9
Italia	2.404 (9,3)	35,1	3,2

4. La relazione tra i voti d'esame e il ruolo della Prova nazionale

La tab. 9 riporta le correlazioni (*rho* di Spearman) tra tutti i voti che concorrono alla valutazione finale dello studente al termine del primo ciclo. Come si può vedere, le correlazioni sono per la quasi totalità piuttosto elevate e comunque significative con un margine di errore minore dell'1%. Particolarmente alta (0,92) è la correlazione tra il voto d'ammissione all'esame e il voto finale e quasi altrettanto elevate sono la correlazione tra il voto d'ammissione e il voto del colloquio orale (0,81) e tra questo e il voto finale (0,86), a riprova di quanto in precedenza osservato sulla funzione che il colloquio orale riveste. Il voto della Prova nazionale ha con i voti delle altre prove correlazioni più basse di quanto non abbiano questi ultimi tra di loro.

Tab. 9 – Correlazioni (*rho*) tra i voti d'esame – Italia

	Voto amm.	Italiano	Lingua str.	Matem.	PN INVALSI	Colloquio	Voto finale
Voto amm.	1,00	0,77**	0,76**	0,77**	0,64**	0,81**	0,92**
Italiano	0,77**	1,00	0,67**	0,62**	0,56**	0,70**	0,81**
Lingua str.	0,76**	0,67	1,00	0,63**	0,58**	0,66**	0,81**
Matem.	0,77**	0,62**	0,63**	1,00	0,62**	0,67**	0,82**
PN INVALSI	0,64**	0,56**	0,58**	0,62**	1,00	0,54**	0,73**
Colloquio	0,81**	0,70**	0,66**	0,67**	0,54**	1,00	0,86**
Voto finale	0,92**	0,81**	0,81**	0,82**	0,73**	0,86**	1,00

** La correlazione è significativa a livello 0,01 (a due code).

In particolare, la correlazione tra il voto della Prova nazionale e i voti delle prove interne di italiano e di matematica, le due materie su cui essa verte, è pari nel primo caso a 0,56 e nel secondo a 0,62, mentre la correlazione con il voto d'ammissione e con il voto finale è rispettivamente 0,64 e 0,73. Non a caso la correlazione più bassa in assoluto (0,54) si registra tra il voto della Prova nazionale e il voto del colloquio orale. In conclusione, sembrerebbe che la Prova nazionale spezzi in qualche misura la compattezza del quadro che emerge dalle valutazioni dei docenti ma senza incidere più di tanto sul voto finale, che ricalca nella stragrande maggioranza dei casi quello con cui lo studente è stato ammesso all'esame, al di là dell'esito della prova INVALSI e delle differenze di risultati che questa ha a seconda dell'area geografica. Una conferma di quanto si è ora detto si ha confrontando a due a due il voto della Prova nazionale, il voto di ammissione e il voto finale. Le tabelle che seguono mostrano il risultato di tali confronti per l'Italia nel suo insieme e per le tre aree geografiche del Paese.

Tab. 10 – Confronto tra voto della Prova nazionale e voto di ammissione per area (%)

Area	Voto PN = Voto a.	Voto PN > Voto a.	Voto PN < Voto a.
Nord	28,0	32,9	39,1
Centro	29,3	27,9	42,8
Sud	29,0	24,8	46,2
Italia	28,7	28,5	42,8

Tab. 11 – Confronto tra voto della Prova nazionale e voto finale per area (%)

Area	Voto PN = Voto f.	Voto PN > Voto f.	Voto PN < Voto f.
Nord	33,0	28,6	38,4
Centro	33,7	21,3	45,0
Sud	32,1	18,6	49,4
Italia	32,8	22,9	44,3

Tab. 12 – Confronto tra voto finale e voto di ammissione per area (%)

Area	Voto f. = Voto a.	Voto f. > Voto a.	Voto f. < Voto a.
Nord	74,0	14,5	11,5
Centro	73,7	17,7	8,6
Sud	78,7	16,1	5,2
Italia	75,8	15,9	8,3

Dall'esame delle tre tabelle scaturiscono alcune interessanti osservazioni. Innanzitutto, in tutta Italia la frequenza percentuale dei casi in cui il voto della Prova nazionale è più basso del voto di ammissione e del voto finale è maggiore rispetto a quella dei casi in cui accade il contrario, ma lo scarto tra i due valori, come si vede dalle tab. 10 e 11, si allarga progressivamente in modo consistente passando da Nord a Sud (da 6 fino a 21 punti nel primo caso e da 10 fino a 31 nel secondo). Nello stesso tempo, come si vede dalla tab. 12, la frequenza percentuale dei casi in cui il voto finale risulta più basso del voto di ammissione si riduce di alcuni punti procedendo da Nord a Sud, mentre aumenta proporzionalmente quella dei casi in cui il voto finale è uguale o superiore al voto di ammissione, che passa dall'88,5% nel Nord al 91,4% nel Centro e al 94,8% nel Sud. Per riassumere, non solo la Prova nazionale non influisce in modo determinante sul voto finale, ma, benché i suoi risultati siano migliori al Nord, ciò non ha un riscontro nella valutazione conclusiva, dove i casi in cui il voto finale risulta più basso sia di quello della Prova nazionale sia del voto di ammissione sono più frequenti al Nord rispetto al Centro e al Sud: a questo esito paradossale concorre la maggior severità nella valutazione delle prove scritte interne nel Nord, come si è visto nel paragrafo precedente, senza dire che, a causa di una più diffusa propensione a comportamenti opportunistici, i voti della Prova nazionale nel Sud sono più alti di quanto sarebbero in assenza di tali comportamenti. Tuttavia, un limitatissimo effetto equilibratore la Prova nazionale sembra averlo, come si può vedere dalla tabella che segue, dove si pone a confronto il voto finale con la media dei voti calcolata senza di essa.

Tab. 13 – Confronto tra il voto finale e la media dei voti d'esame esclusa la Prova nazionale per area (%)

Area	Voto finale = Media dei voti esclusa PN	Voto finale > Media dei voti esclusa PN	Voto finale < Media dei voti esclusa PN
Nord	91,2	4,0	4,8
Centro	90,8	2,9	6,3
Sud	90,9	1,9	7,2
Italia	91,0	2,9	6,1

Poiché il voto finale attribuito allo studente è dato dalla media, arrotondata all'intero, del voto d'ammissione e dei voti di tutte le prove, quando esso è superiore alla media dei voti computata senza tener conto del voto della Prova nazionale ciò significa che quest'ultimo ha inciso positivamente sulla valutazione conclusiva, mentre quando è più basso ciò significa che la sua incidenza è stata negativa. In primo luogo è da rilevare che, a livello nazionale e in ciascuna

delle grandi aree geografiche del Paese, i casi in cui la Prova nazionale altera in un senso o nell'altro il voto finale non raggiungono il 10%. In tal caso, però, tale effetto si esercita nella direzione attesa: infatti, come si può constatare dalla tab. 13, i casi in cui il voto finale è inferiore a quello che si ottiene dalla media dei voti calcolata escludendo la Prova nazionale sono percentualmente meno numerosi al Nord e più frequenti al Centro e più ancora al Sud, e parallelamente sono più frequenti al Nord rispetto al Centro e al Sud in casi in cui la Prova nazionale aumenta il voto finale.

5. I punteggi in italiano e matematica sottostanti al voto della Prova nazionale

Nel paragrafo introduttivo si è accennato al fatto che i punteggi delle prove standardizzate di italiano e matematica che costituiscono le due parti della Prova nazionale e il cui risultato confluisce nell'unico voto attribuito non vengono corretti per il *cheating*, a differenza di quanto accade per gli stessi punteggi quando divengono oggetto di pubblicazione nel rapporto annuale stilato dall'INVALSI sugli esiti delle rilevazioni e all'atto della loro restituzione alle scuole. La correzione per il *cheating* – che porta a un abbassamento dei punteggi osservati – viene operata, mediante una complessa procedura statistica (Quintano, Castellano e Longobardi, 2009), sui dati delle prove di rilevazione di tutti i livelli scolari interessati, sia sui dati campionari⁵ sia su quelli di popolazione prima della restituzione alle singole scuole. Dalle rilevazioni effettuate sulla scuola primaria e sulla secondaria superiore non emergono, normalmente, indizi di comportamenti opportunistici nelle classi campione, dove la somministrazione delle prove INVALSI e la correzione e tabulazione delle risposte degli studenti avviene sotto il controllo di un osservatore esterno, mentre il fenomeno è presente nelle classi non campione ma in misura differente nelle varie parti d'Italia, con una maggiore incidenza nel Sud del Paese in confronto al Centro e al Nord (Bertoni, Brunello e Rocco, 2013). Le prove della terza classe della scuola secondaria di I grado, rispetto alla questione ora discussa, rappresentano in qualche modo un'eccezione: poiché esse fanno parte dell'esame di Stato di conclusione del primo ciclo, durante il loro svolgimento nelle classi campione non è presente l'osservatore esterno, il cui compito è demandato al presidente della commissione d'esame.

Purtroppo, ciò non è sufficiente a evitare il prodursi di fenomeni di *cheating*, cosa che rende necessaria la correzione dei punteggi anche per le classi campione di questo livello scolare. La correzione, tuttavia, riguarda solo i punteggi dell'Italia e delle sue articolazioni territoriali pubblicati nel rapporto annuale e i punteggi delle classi e delle scuole restituiti agli istituti scolastici dopo ogni rilevazione, ma non i punteggi dei singoli studenti utilizzati come base per formulare il voto della Prova nazionale. Questo perché la procedura di individuazione del *cheating*, come ogni procedura statistica, è soggetta a un margine ineliminabile di errore, errore che – per ovvi motivi – non può andare a pesare sulla valutazione individuale degli alunni. È interessante, però, a questo punto, vedere quale sia l'andamento dei punteggi percentuali delle prove INVALSI di italiano e matematica, corretti e non corretti per il *cheating*, in funzione del voto unico della Prova nazionale. I grafici seguenti mostrano tale andamento per l'Italia e le tre aree geografiche.

Dai grafici emerge che, mentre le linee spezzate che si ottengono mettendo in relazione il punteggio medio non corretto con il voto della Prova nazionale in ciascuna delle tre aree e in Italia si sovrappongono pressoché completamente per quanto riguarda sia la prova INVALSI di italiano sia la prova di matematica, quando si considerano i punteggi corretti per il *cheating* le spezzate cominciano a distinguersi una dall'altra, nel caso dell'italiano a partire più o meno dal voto 7, e nel caso della matematica già a partire dal voto 6. Inoltre, in italiano non solo il divario dei punteggi corretti tra il Nord e il Centro, da una parte, e il Sud, dall'altra, aumenta progressivamente man mano che il voto della Prova nazionale aumenta, ma nelle prime due aree essi continuano a crescere, nella terza, invece, dal voto 8 in poi la crescita si azzerava. In matematica, gli andamenti dei punteggi corretti sono un po' diversi: in questo caso essi crescono sistematicamente in tutte e tre le aree con l'aumentare del voto della Prova nazionale, mentre il divario del Nord rispetto sia al Centro sia – in maggior misura – al Sud, si amplia progressivamente. In definitiva, comunque, quel che sembra si possa affermare è che la correzione per il *cheating* interessa soprattutto i punteggi più alti ma in misura diversa nelle tre aree.

⁵ Sebbene le prove INVALSI siano censuarie, dalle popolazioni di studenti dei vari livelli scolari interessati viene estratto annualmente un campione di classi, dove è inviato un osservatore esterno con il compito di garantire l'attendibilità dei dati raccolti e la loro rapida trasmissione all'Istituto di Valutazione, che calcola su di essi le stime dei parametri pubblicati nel rapporto sull'esito delle rilevazioni.

Fig. 1 – Punteggi percentuali non corretti (grafico a sinistra) e corretti per il cheating (grafico a destra) della prova INVALSI di italiano in funzione del voto della Prova nazionale

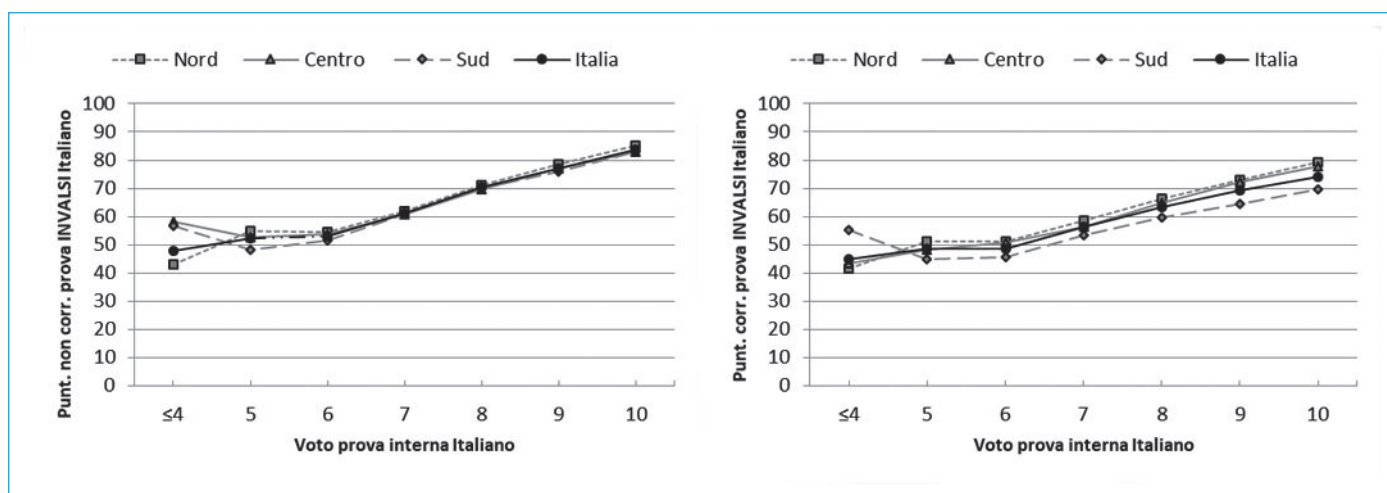
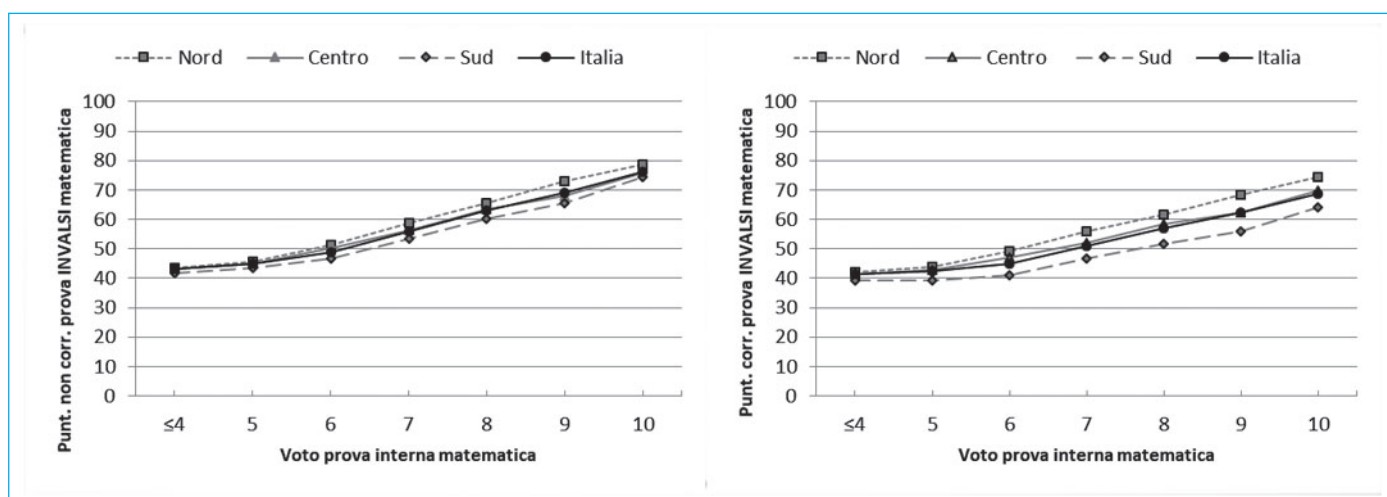


Fig. 2 – Punteggi percentuali non corretti (grafico a sinistra) e corretti per il cheating (grafico a destra) della prova INVALSI di matematica in funzione del voto della Prova nazionale



Se si ripete l’esercizio mettendo in relazione i punteggi percentuali, corretti e non corretti, delle prove INVALSI di italiano e matematica con il voto della prova scritta interna dell’una e dell’altra materia, il quadro che si ottiene è illustrato dai grafici che seguono.

In italiano è da rilevare che, in corrispondenza dei voti più bassi (4 o meno), il punteggio medio non corretto e corretto per il *cheating* è nel Nord inferiore rispetto al Centro e al Sud, dato che forse potrebbe essere spiegato con la maggiore presenza di alunni stranieri in quest’area. In corrispondenza dei voti successivi al 4, in particolare dal 6 in poi, i punteggi non corretti delle tre aree sono molto simili tra loro, mentre i punteggi corretti sono più alti al Nord e al Centro rispetto al Sud. In matematica le cose vanno un po’ diversamente: i punteggi non corretti sono quasi sempre, a parità di voto assegnato alla prova interna, più alti al Nord in confronto al Centro e al Sud, e il divario fra le tre aree si amplia quando si considerano i punteggi corretti, specie in corrispondenza dei voti più alti.

Fig. 3 – Punteggi percentuali non corretti (grafico a sinistra) e corretti per il cheating (grafico a destra) della prova INVALSI di italiano in funzione del voto della prova interna di italiano

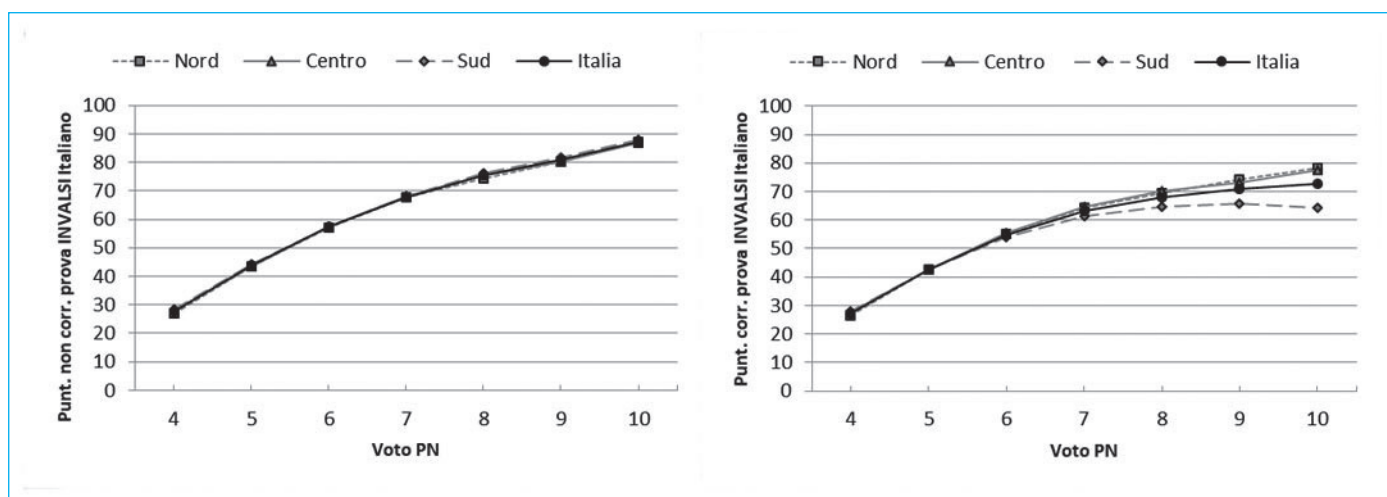
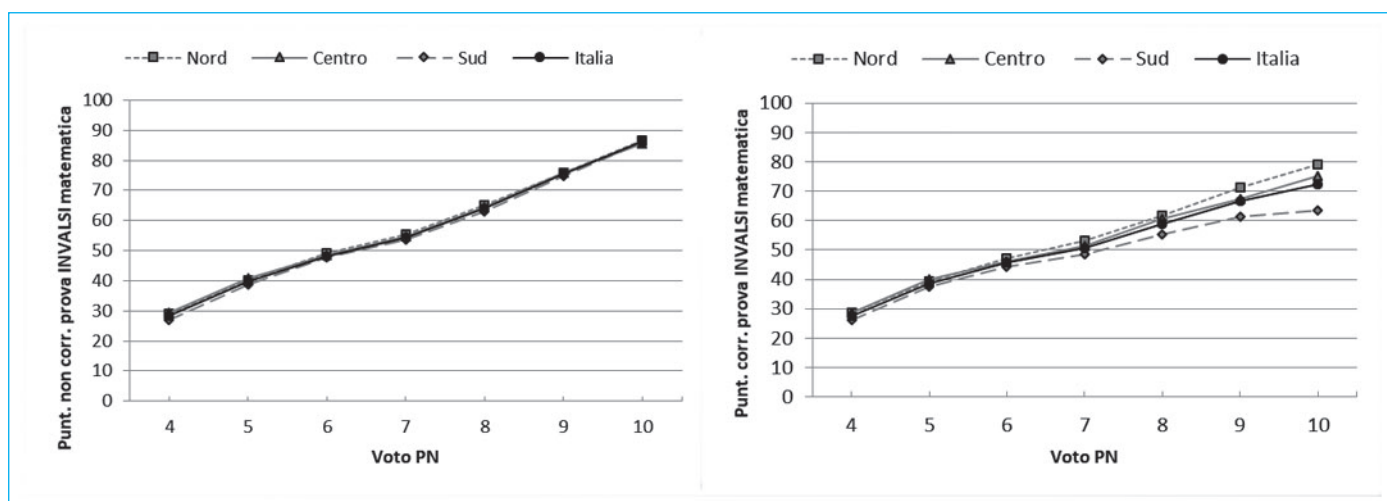


Fig. 4 – Punteggi percentuali non corretti (grafico a sinistra) e corretti per il cheating (grafico a destra) della prova INVALSI di matematica in funzione del voto della prova interna di matematica



Come ultima notazione, aggiungiamo che la correlazione (r di Pearson) tra il risultato della prova di italiano e quello della prova di matematica, sul piano nazionale, è pari a 0,64 quando si considerano i punteggi non corretti, e a 0,55 quando si considerano i punteggi corretti.

6. Conclusioni

L'esame di Stato è stato introdotto nel nostro ordinamento dalla riforma Gentile del 1923 con il duplice scopo, da un lato, di assicurare il rigore degli studi e la sostanziale equivalenza su tutto il territorio nazionale – in termini di preparazione raggiunta – dei diplomi ottenuti con il suo superamento, dall'altro di realizzare la parità di trattamento fra gli studenti delle scuole statali e quelli delle scuole non statali, gli uni e gli altri ugualmente tenuti all'obbligo di sottoporsi al giudizio di commissioni formate da insegnanti esterni alle scuole frequentate. Il principio dell'esame di Stato "per l'ammissione ai vari ordini e gradi di scuole o per la conclusione di essi", è stato poi accolto, successivamente alla caduta del fascismo, nel testo dell'articolo 33 della Costituzione repubblicana entrata in vigore in Italia il primo gennaio del 1948. Da allora tale principio, anche a seguito dei processi di massificazione dell'istruzione che hanno segnato, in Italia

come in tutti i Paesi sviluppati, la seconda metà del secolo scorso, è andato incontro a una continua erosione sul piano della realtà effettuale se non su quello giuridico-formale. Il carattere di “esternalità”⁶ dell’esame di Stato, che dovrebbe costituirne l’essenza e la ragion d’essere (appaltatore e collaudatore non possono coincidere nel medesimo soggetto, diceva Salvemini), è infatti quasi completamente venuto meno per l’esame di licenza media ed è stato fortemente compromesso per l’esame di conclusione degli studi superiori. Se da una parte ciò è stato anche il frutto di politiche volte alla democratizzazione dell’istruzione attraverso l’eliminazione delle barriere all’accesso, dall’altra parte ha portato alla trasformazione degli esami di Stato in un esercizio ritualistico stravolgendone gli obiettivi originari. Nel 2015, i promossi all’esame di conclusione del I ciclo sono stati, secondo i dati pubblicati dal MIUR, il 99,8% (sul totale degli ammessi, a sua volta pari al 97,2% degli alunni di classe terza secondaria di I grado) e una percentuale di poco inferiore, il 99,4%, sono stati i promossi all’esame di conclusione del II ciclo (su un totale di 95,6% di ammessi a sostenerlo). La selezione, che pure è un fenomeno presente nella scuola italiana come in quella di altri Paesi, avviene dunque durante il percorso degli studi nel passaggio da una classe alla successiva, in particolare nel biennio della secondaria superiore, ma certamente non attraverso il vaglio degli esami.

Non intendiamo approfondire qui ulteriormente questo punto né tanto meno cercare di rispondere all’ovvio interrogativo che si pone sull’opportunità di mantenere nell’ordinamento un sistema di esami di cui non si comprende quale sia la funzione, rinviando per questo ad altri contributi (Martini, 1998 e 2006). Per tornare al nostro tema principale – il peso esercitato dalla Prova nazionale sulla valutazione finale degli studenti che sostengono l’esame di Stato – la prima cosa da rilevare, sulla base dell’analisi dei dati di cui si è dato conto nei paragrafi precedenti, è che esso è alquanto modesto anche se non completamente trascurabile. Questa prima constatazione risponde a una delle critiche più frequenti rivolte dagli insegnanti all’inserimento delle prove INVALSI tra le prove d’esame di conclusione del primo ciclo (Martini e Papini, 2015), e cioè che ciò determinerebbe un abbassamento della valutazione finale. Come si è visto, questo è vero solo per il 6% degli studenti che hanno sostenuto nel 2015 l’esame, mentre per un altro 3% circa il voto della Prova nazionale fa aumentare il voto finale. Nella stragrande maggioranza dei casi, tuttavia, la Prova nazionale non modifica né in un senso né nell’altro la valutazione finale degli studenti, che risulta per lo più eguale al voto con cui sono stati ammessi all’esame (vedi tab. 12). C’è da chiedersi, naturalmente, se il ruolo molto marginale che la Prova nazionale esercita sia dovuto al fatto che il suo esito, tutto sommato, collima con gli altri elementi di valutazione che concorrono al voto finale, oppure sia da attribuire a interventi di compensazione operati dagli insegnanti in presenza di un risultato nella prova non in sintonia con le loro aspettative. Probabilmente è vera sia l’una sia l’altra cosa, senza dimenticare che, in ogni caso, il voto della Prova nazionale contribuisce solo per un settimo al voto finale.

Quali sono, a questo punto, le lezioni che si possono trarre dall’esperienza ormai quasi decennale di inclusione della Prova nazionale INVALSI tra le prove dell’esame di conclusione del I ciclo?

La prima è che, evidentemente, ciò non è sufficiente a rendere l’esame più oggettivo e imparziale – ammesso che questa fosse l’intenzione – in assenza di un ripensamento complessivo della sua architettura e delle sue finalità: la prova standardizzata non può esser semplicemente giustapposta alle altre – come di fatto è accaduto – senza prima stabilire quale funzione essa debba avere e senza una visione d’insieme unitaria e coerente. La seconda e più importante lezione da trarre è che, quando ci si accinga a intervenire con misure di riforma nella materia degli esami, è necessario chiarire preliminarmente gli obiettivi che si intendono raggiungere, soppesando vantaggi e svantaggi, rischi e opportunità che possono derivare dal cambiamento, e solo in seguito vagliare attentamente quali procedure siano meglio in grado di assicurare il conseguimento degli obiettivi che ci si è proposti.

⁶ L’esternalità di un esame è data dall’essere i candidati sottoposti alle stesse prove, sostenute nelle stesse condizioni e corrette e valutate secondo criteri uniformi da esaminatori terzi. Essa, secondo Ludger Wöessmann (2003), costituisce la sola giustificazione di un sistema centrale d’esami.

Riferimenti bibliografici

- Bertoni M., Brunello G., Rocco L. (2013), “When the cat is near, the mice won’t play”, *Journal of Public Economics*, 104: 75-77.
- Dardanoni V., Modica S., Pennisi A. (2007), “Grading in heterogeneous schools”, testo disponibile al sito: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.491.5366>, data di consultazione: 21 aprile 2016.
- Martini A. (1998), “La riforma dell’esame di maturità, il valore legale del titolo di studio e la formazione delle classi dirigenti”, *Scuola e città*, 11: 489-498.
- Martini A. (2006), “Rifare l’esame. A proposito della nuova maturità”, *Il Mulino*, 5: 894-903.
- Martini A. (2009), “Prove INVALSI: limiti e vantaggi”, *La vita scolastica*, 6: 16-19.
- Martini A. e Papini M. (2015), “Che cosa ne pensano gli insegnanti delle prove INVALSI”, *INVALSI Working Papers*, 24, testo disponibile al sito: www.invalsi.it/invalsi/istituto.php?page=working_papers, data di consultazione: 21 aprile 2016.
- Quintano C., Castellano R., Longobardi S. (2009), “A fuzzy clustering approach to improve the accuracy of Italian student data. An experimental procedure to correct the impact of outliers on assessment test scores”, *Statistica & applicazioni*, 7, 2: 149-171.
- Woessmann L. (2003), “Central exit exams and student achievement: international evidence”, in P.E. Peterson, M.R. West (eds.), *No Child Left Behind? The Politics and Practice of School Accountability*, Brookings Institution Press, Washington (DC): 292-323.

14. Stima pesata delle abilità degli studenti nei test standardizzati di profitto attraverso modelli IRT multidimensionali

Weighted estimate of students' abilities in standardized assessment tests through multidimensional IRT models

di Simone Del Sarto, Michela Gnaldi

In questo capitolo introduciamo una nuova metodologia endogena per attribuire i pesi alle variabili, o item, per costruire una misura composita di abilità degli studenti nei test standardizzati di profitto. Tale procedura è sviluppata attraverso un'estensione dei modelli *Item Response Theory* (IRT) e tiene conto della dimensionalità del dataset. In particolare, la procedura si sviluppa attraverso due fasi consecutive. Nella prima si applica un algoritmo gerarchico di *clustering*, con lo scopo di determinare il numero di dimensioni misurate dai dati. Nella seconda invece vengono stimati i parametri di discriminazione, tenendo conto della struttura di dimensionalità dei dati accertata nella prima fase. Questa tecnica di attribuzione dei pesi viene illustrata attraverso un'applicazione a un test standardizzato di profitto, sviluppato e raccolto dall'Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione (INVALSI). Questa procedura può essere molto utile anche in termini di valutazione degli item più informativi all'interno di ciascuna dimensione, consentendo di ridurre la lunghezza del test (in fase di pre-test), oppure di semplificare la fase di restituzione dei risultati alle scuole (in fase post-somministrazione).

In this paper we introduce a new endogenous methodology for weighting the variables, or items, in order to build a composite measure of the students' abilities in standardized assessment test. Such procedure is developed within an extension of the *Item Response Theory* (IRT) models and takes into account the dimensionality of the dataset. Specifically, the procedure consists of two consecutive steps. In the first one a hierarchical clustering algorithm is applied with the aim of obtaining the number of dimensions measured by the data. In the second one the discrimination parameters are estimated, considering the dimensionality structure ascertained in the previous phase. Such weighting technique is illustrated through an application to a standardized assessment test, developed and collected by the Italian National Institute for the Evaluation of the Education System (INVALSI). This procedure can be very useful in terms of evaluation of the most informative items within each dimension, allowing to reduce the test length (in the pre-test phase), or to simplify the phase of results restitution to schools (during the post-administration phase).

1. Introduzione

L'abilità complessiva a un test di valutazione degli apprendimenti è assimilabile a un indicatore composito, ottenuto come sintesi o aggregazione, normalmente non pesata, di singole componenti (variabili o item) che concorrono a misurare i diversi aspetti del fenomeno oggetto di interesse. Generalmente la costruzione di un indicatore composito avviene attraverso specifiche procedure di aggregazione, standardizzazione ecc., nelle quali una questione molto discussa riguarda le modalità con cui vengono assegnati i pesi a singoli indici. Esistono due principali criteri per assegnare i pesi nel processo di costruzione di un indicatore composito: criteri soggettivi, basati sulle opinioni dei ricercatori o sui punti di vista della società, oppure criteri oggettivi che fanno ricorso a tecniche statistiche. Diversamente dai primi, anche detti approcci "normativi", che dipendono da un giudizio di valore, i secondi approcci, cosiddetti *data-driven*, assegnano i pesi in modo endogeno, ovvero a partire dai dati e in funzione di essi.

In questo capitolo proponiamo un nuovo approccio statistico per l'assegnazione dei pesi, sviluppato grazie a un'estensione dei modelli *Item Response Theory* (IRT). Il criterio proposto consente di ottenere una misura complessiva

pesata dell'abilità dello studente a un test di valutazione degli apprendimenti, che tiene conto sia del diverso peso informativo degli item, sia della multidimensionalità del test, ovvero del fatto che gli item di un test di profitto sono spesso fortemente correlati. Nello specifico, suggeriamo di assegnare i pesi sulla base dei parametri di discriminazione, stimati tramite un modello IRT multidimensionale con una parametrizzazione logistica a due parametri (2-PL). Questo modello è stato inizialmente proposto da Bartolucci (2007) e successivamente applicato per varie ricerche, in primo luogo in campo educativo e sanitario (si vedano, per esempio, Bacci e Gnaldi, 2015; Bartolucci, Bacci e Gnaldi, 2015; Gnaldi, Bacci e Bartolucci, 2015; Gnaldi e Bacci, 2015). In un recente lavoro di Bartolucci, Montanari e Pandolfi (2012), gli autori hanno applicato questo modello per ridurre il numero di item di un questionario in ambito sanitario. Nel presente lavoro proponiamo di usare, allo stesso tempo, il modello introdotto da Bartolucci (2007) e le procedure descritte in Bartolucci, Montanari e Pandolfi (2012) per valutare le abilità degli studenti, stimate assegnando pesi diversi agli item del test. Tali pesi esprimono il diverso potenziale informativo degli item nelle diverse dimensioni del test.

Poiché il sistema di attribuzione dei pesi è molto più accurato se si tiene conto delle differenti dimensioni che contribuiscono a caratterizzare un test (ossia gruppi di item che misurano lo stesso concetto latente), suggeriamo di assegnare i pesi in base ai parametri di discriminazione stimati considerando tale dimensionalità, mediante un modello IRT 2-PL multidimensionale. In particolare, la procedura proposta viene implementata attraverso due fasi consecutive. Nella prima, l'obiettivo è verificare il numero effettivo di dimensioni misurate dai dati. Nella seconda fase vengono stimati i parametri di discriminazione mediante il modello multidimensionale selezionato nella prima fase.

La procedura proposta è illustrata con un'applicazione a dati educativi, riferiti a un test nazionale di valutazione delle competenze, sviluppati e raccolti dall'Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione (INVALSI). Nell'applicazione verrà mostrato come costruire un indicatore composito pesato attraverso la procedura proposta, così da valutare le differenze in termini di classificazioni ricevute dagli studenti se il peso degli item entro dimensione viene o meno considerato.

Questo capitolo è organizzato nel modo seguente. Nel paragrafo successivo viene descritta in dettaglio la procedura proposta, mentre le sue potenzialità vengono illustrate nel paragrafo 3 attraverso un'applicazione ai dati relativi a un test INVALSI. Infine, nel paragrafo 4 vengono fornite alcune conclusioni finali.

2. La procedura per attribuire i pesi a singole variabili

In questo paragrafo descriviamo la procedura finalizzata a ottenere i pesi da attribuire alle singole variabili durante la procedura di costruzione di un indicatore di sintesi delle abilità degli studenti a un test di verifica degli apprendimenti. In particolare, nel paragrafo 2.1, specificando la natura latente e multidimensionale dell'abilità degli studenti, si chiariscono "a monte" i motivi dell'utilità del modello proposto. Nel paragrafo 2.2 vengono forniti alcuni dettagli sul modello utilizzato, mentre nel paragrafo 2.3 vengono illustrate le due fasi della procedura per ottenere i pesi.

2.1. La multidimensionalità latente dell'abilità nei test di verifica degli apprendimenti

Come è noto, le abilità degli studenti non possono essere direttamente osservate, ma possono essere studiate solo considerando le risposte degli studenti ai singoli item di un test (Gnaldi, 2016). L'abilità di uno studente è quindi una caratteristica latente, in quanto può essere derivata soltanto indirettamente misurando la performance nei singoli item che compongono un test di apprendimento. Un altro aspetto tipico dell'abilità è la multidimensionalità, specchio della sua complessità. Per esempio, l'abilità di matematica è un'abilità complessa, che si sostanzia in una serie di sottodimensioni più specifiche, quali la capacità di risolvere problemi, la capacità di interpretare dati e fare previsioni sulla base di essi, e così via. Sebbene l'ipotesi di unidimensionalità sia molto comune nelle analisi in ambito educativo e sociale in generale, generalmente si tratta di un'assunzione restrittiva (Gnaldi, Del Sarto e Maggino, 2017). L'applicazione di misure unidimensionali a dati multidimensionali causa una incompatibilità tra modello e dati (Bonifay *et al.*, 2015) e ciò accade ogni volta che cerchiamo di sintetizzare fenomeni complessi attraverso un singolo valore, come accade quando collochiamo le unità di analisi (studenti, classi, scuole ecc.) in un'unica posizione entro una singola classificazione, o ranking.

I modelli IRT costituiscono una metodologia di analisi particolarmente adatta in contesti di ricerca in cui gli oggetti dell'analisi sono costrutti latenti e multidimensionali. Grande parte della letteratura corrente sui modelli IRT e loro applicazioni fa ricorso a tali strumenti per derivare scale di misura. Diversamente, in questo lavoro proponiamo di utilizzare un'estensione dei modelli IRT per pesare le singole componenti (ossia gli item o le variabili) di una scala già sviluppata e validata (un test INVALSI). La procedura che descriveremo, basata su modelli IRT che tengono conto della dimensionalità latente dei dati, può essere particolarmente utile nel processo di costruzione di un indicatore composito di abilità degli studenti. Nello specifico, la procedura proposta consiste in due fasi. La prima mira a studiare la dimensionalità latente presente nei dati in questione, mentre la seconda assegna i pesi alle singole variabili, raggruppate nella medesima dimensione durante la prima fase.

In particolare, la fase 1 è basata su una metodologia IRT sviluppata da Bartolucci (2007), che consiste in un algoritmo di *clustering* per il raggruppamento, nello stesso cluster, di variabili che contribuiscono a misurare lo stesso costrutto latente. La fase 2 prosegue con la stima del parametro di discriminazione relativo a ogni variabile, utilizzando il modello IRT multidimensionale selezionato nella fase precedente. Successivamente, al fine di assegnare i pesi agli item del test, la distribuzione delle abilità latente viene standardizzata, mentre i parametri di discriminazione vengono riparametrizzati e trasformati per renderli comparabili tra dimensioni.

2.2. I modelli IRT multidimensionali

I modelli IRT tradizionali assumono che le risposte a un insieme di item (o variabili di un questionario) dipendano da un singolo tratto latente (per esempio, l'abilità degli studenti). Tuttavia, un questionario spesso è composto da alcuni sottoinsiemi di item che misurano costrutti differenti, ma potenzialmente correlati. In questi casi l'assunzione IRT tradizionale di una singola variabile latente sottostante il modello risulta restrittiva (Bartolucci, Bacci e Gnaldi, 2015).

Per superare i limiti dei modelli IRT unidimensionali, Bartolucci (2007) ha proposto un modello IRT multidimensionale, in cui si assume che il tratto latente misurato dal questionario abbia una struttura multidimensionale, composta da s dimensioni. Inoltre, un'altra differenza del modello in questione rispetto ai classici modelli IRT riguarda l'ipotesi che il campione osservato sia estratto da una popolazione di unità (per esempio studenti) diviso in k sotto-popolazioni, o classi latente. Le unità che appartengono alla stessa classe latente condividono caratteristiche molto simili in termini di abilità latente sottostante.

Consideriamo un test composto da J item dicotomici, in cui la risposta può essere corretta o non corretta. Il modello utilizza i parametri λ_{jc} (probabilità condizionata di successo), che denota la probabilità di corretta risposta all'item j per soggetti appartenenti alla classe latente c , e π_c , detta probabilità a priori della classe c , $c = 1, \dots, k$. Nello specifico, utilizziamo la cosiddetta parametrizzazione 2-PL (due parametri logistici) di λ_{jc} , così da includere anche un parametro che misuri il potere discriminatorio dell'item.

Quindi, il modello IRT 2-PL multidimensionale è basato sulla seguente equazione:

$$\text{logit}(\lambda_{jc}) = \gamma_j \left(\sum_{d=1}^s \delta_{jd} \theta_{cd} - \beta_j \right), \quad j = 1, \dots, J, \quad c = 1, \dots, k \quad (1)$$

dove γ_j è l'indice di discriminazione dell'item j e β_j è il parametro relativo alla difficoltà dell'item j , ossia la tendenza globale a rispondere in maniera errata all'item j . Inoltre, δ_{jd} è una variabile indicatrice, che assume valore 1 se l'item j contribuisce a misurare la dimensione d , e 0 altrimenti. Infine, θ_{cd} è la misura del livello di tratto latente (abilità) per soggetti appartenenti alla classe latente c , rispetto alla dimensione d . Per ulteriori dettagli sul modello in questione, si rimanda ai lavori di Bartolucci (2007) e Bartolucci, Montanari e Pandolfi (2012).

2.3. Le due fasi della procedura per attribuire i pesi alle variabili

Allo scopo di ottenere i pesi da attribuire alle variabili, nell'ambito del processo di costruzione di una misura di sintesi dell'abilità degli studenti, proponiamo una procedura a due fasi.

Una volta ottenuto il numero di classi latenti k , la procedura inizia con la valutazione della dimensionalità della struttura latente dei dati. A tale scopo, viene adottato un algoritmo gerarchico per raggruppare gli item in un numero ridotto di gruppi, supponendo che gli item dello stesso gruppo misurino la stessa dimensione latente (ossia un'abilità) e ogni gruppo misuri una dimensione diversa. L'algoritmo inizia stimando il modello (1) e considerando il caso in cui ogni item misuri una diversa dimensione, quindi $s = J$. Esso termina con la stima di un modello unidimensionale ($s = 1$). A ogni passo, l'algoritmo stima un modello con una dimensione in meno rispetto al modello stimato nel passo precedente. Alla fine avremo $J - 1$ modelli, tra i quali deve essere scelto il modello "migliore", che porterà quindi alla selezione del numero di dimensioni più adeguato per i dati in questione. Tale scelta può essere effettuata seguendo vari criteri, tra cui il p-value, il BIC, l'AIC ecc.

Dopo aver selezionato il numero di dimensioni s , nella seconda fase della procedura proposta vengono stimati i parametri di discriminazione mediante un modello IRT 2-PL s -dimensionale: tali parametri saranno poi utilizzati per costruire un indicatore pesato delle abilità degli studenti al test. È fondamentale sottolineare che, se il fenomeno in studio può essere considerato multidimensionale (quindi $s > 1$), i parametri di discriminazione degli item non sono direttamente comparabili tra dimensioni. Il valore massimo che può essere ottenuto per questo indice è 1, che corrisponde all'item che presenta il più alto potere di discriminazione in ogni dimensione (Bartolucci, Montanari e Pandolfi, 2012). Quindi, per rendere tali parametri direttamente comparabili tra le varie dimensioni, la distribuzione del tratto latente viene prima standardizzata per ogni dimensione $d = 1, \dots, s$, calcolando le seguenti quantità:

$$\hat{\mu}_d = \sum_c \hat{\theta}_{cd} \hat{\pi}_c, \hat{\sigma}_d = \sqrt{\sum_c (\hat{\theta}_{cd} - \hat{\mu}_d)^2 \hat{\pi}_c} \quad (2)$$

dove $\hat{\theta}_{cd}$ denota la stima della misura del tratto latente per soggetti appartenenti alla classe latente c rispetto alla dimensione d , mentre $\hat{\pi}_c$ è la stima della probabilità a priori di appartenere alla classe latente c . Successivamente, i parametri di discriminazione stimati possono essere trasformati entro ogni dimensione, utilizzando la seguente formula (Bartolucci, Montanari e Pandolfi, 2012):

$$\hat{\gamma}_j^* = \hat{\sigma}_d \hat{\gamma}_j, j \in I_d \quad (3)$$

dove I_d è l'insieme degli item che contribuiscono a misurare la dimensione d , con $d = 1, \dots, s$. Tali parametri trasformati possono essere quindi utilizzati per assegnare i pesi alle singole variabili, in ottica di costruzione di un indicatore composto di sintesi delle abilità degli studenti.

3. Applicazione a dati reali

La procedura descritta nel paragrafo precedente viene ora applicata ai dati relativi al test INVALSI di matematica, somministrato nel mese di giugno 2014 a un campione di 25.348 studenti delle scuole medie italiane. Il test prevede quesiti a scelta multipla e domande aperte, per un totale di 33 item dicotomici.

Per quanto riguarda la definizione del numero di classi latenti k , poiché in ricerche precedenti (Gnaldi, Bacci e Bartolucci, 2015) si è trovata evidenza di $k = 3$ classi latenti per questo dataset, nella presente applicazione adotteremo questo numero di classi.

La procedura proposta inizia con la stima della dimensionalità del test. A tale scopo, utilizziamo l'algoritmo di *clustering* introdotto nel paragrafo 2.3. Come già evidenziato in Gnaldi e Del Sarto (2016), troviamo evidenza per una struttura a tre dimensioni (quindi $s = 3$), poiché in corrispondenza di tale numero di dimensioni possiamo osservare l'ultimo valore negativo della differenza tra il BIC del modello in questione e quello del modello iniziale (differenza pari a -110.67). La prima dimensione include sei item (con etichetta 2, 10, 13, 18, 22 e 26): guardando ai contenuti degli item secondo le classificazioni riportate nei documenti INVALSI (2012a e 2012b), possiamo considerare questa dimensione come espressiva della dimensione relativa alla cultura statistica e conoscenza e uso di relazioni e funzioni. Analogamente, la seconda dimensione è misurata da dieci item ed è possibile ricondurla al riconoscimento e utilizzo di grandezze

di misura in campo geometrico. Infine, diciassette item contribuiscono a misurare la terza dimensione, riguardante la risoluzione di problemi in ambito numerico e di relazioni e funzioni.

Il secondo passo della procedura proposta consiste nell'ottenere i parametri di discriminazione mediante un modello IRT 2-PL multidimensionale, considerando le tre dimensioni determinate nella fase precedente. In seguito alla standardizzazione dei tratti latenti, tali parametri di discriminazione sono trasformati secondo le formule riportate nel paragrafo 2.3. In tab. 1 si riportano le stime dei parametri di discriminazione $\hat{\gamma}_j$ e il loro valore trasformato $\hat{\gamma}_j^*$, per ognuna delle tre dimensioni. Osservando le stime originali dei parametri di discriminazione per ogni dimensione, $\hat{\gamma}_j$, possiamo affermare che, per quanto riguarda la prima dimensione, l'item 18 è il più discriminante, mentre gli altri item della stessa dimensione hanno simili stime di $\hat{\gamma}_j$. D'altra parte, tali stime sono più variabili nella seconda dimensione. Globalmente, gli item 18, 17 e 4 possono essere visti come i più informativi delle tre dimensioni considerate.

Tab. 1 – Parametri di discriminazione per i 33 item del test INVALSI 2014 di matematica, relativi a tre dimensioni

a) Dimensione 1			b) Dimensione 2			c) Dimensione 3		
j	$\hat{\gamma}_j$	$\hat{\gamma}_j^*$	j	$\hat{\gamma}_j$	$\hat{\gamma}_j^*$	j	$\hat{\gamma}_j$	$\hat{\gamma}_j^*$
2	0,662	0,488	1	0,748	0,856	4	1,000	1,300
10	0,622	0,459	3	0,698	0,799	5	0,808	1,051
13	0,700	0,516	6	0,633	0,725	7	0,892	1,160
18	1,000	0,738	8	0,401	0,459	9	0,655	0,851
22	0,603	0,445	14	0,803	0,919	11	0,681	0,885
26	0,653	0,481	17	1,000	1,144	12	0,779	1,013
			27	0,598	0,684	15	0,878	1,141
			28	0,736	0,842	16	0,870	1,132
			31	0,550	0,629	19	0,758	0,985
			32	0,776	0,888	20	0,627	0,815
						21	0,524	0,681
						23	0,702	0,913
						24	0,450	0,586
						25	0,437	0,569
						29	0,867	1,127
						30	0,557	0,724
						33	0,830	1,079

Nota: Si riportano i $\hat{\gamma}_j$ originali e la loro trasformazione $\hat{\gamma}_j^*$, così da renderli comparabili tra dimensioni.

Qualora fossimo interessati a rendere comparabile la discriminazione degli item tra dimensioni, è necessario utilizzare i valori trasformati $\hat{\gamma}_j^*$. Osservando di nuovo la tab. 1, possiamo notare come l'item più discriminante su tutte e tre le dimensioni sia l'item 4, che è risultato essere pure il più discriminante all'interno della terza dimensione. È inoltre possibile osservare che i pesi associati a tutti gli item della prima dimensione tendono a essere i più bassi, segno del fatto che la prima dimensione può essere considerata meno importante rispetto alle altre nel bilancio globale del test.

Passiamo ora a illustrare le modalità con cui i pesi ottenuti nel paragrafo precedente possono essere utilizzati per costruire indicatori compositi che ci diano una misura dell'abilità degli studenti al test in questione. Nello specifico, vengono calcolati due indicatori per ogni studente, combinando le risposte ottenute agli item del test. Il primo è un semplice punteggio non pesato per ogni dimensione, ottenuto come media delle risposte fornite dallo studente agli item che contribuiscono a misurare ogni dimensione:

$$A_d = \frac{1}{n_d} \sum_{j \in I_d} y_j, \quad d = 1, 2, 3 \quad (4)$$

dove n_d è il numero di item presenti nella dimensione d e y_j è la risposta fornita dallo studente all'item j , che può essere corretta (codificata con 1), oppure sbagliata (codificata con 0). Il secondo indicatore è invece una combinazione pesata delle risposte, dove i pesi riflettono la discriminatività di ogni item:

$$A_d^p = \frac{1}{\sum_{j \in I_d} \hat{y}_j^*} \sum_{j \in I_d} \hat{y}_j^* y_j, \quad d = 1, 2, 3 \quad (5)$$

Tab. 2 – Ranking degli studenti di due classi selezionate, per ognuna delle tre dimensioni, utilizzando l'indicatore non pesato e pesato, A_d e A_d^p , rispettivamente

a) Classe 1 – 22 studenti									
Dimensione 1 (6 item)									
Stud.	A_1	A_1^p	$r(A_1)$	$r(A_1^p)$	Stud.	A_1	A_1^p	$r(A_1)$	$r(A_1^p)$
1	0,667	0,679	9,0	10,0	12	0,500	0,537	16,0	16,0
2	0,333	0,390	20,5	20,0	13	0,500	0,546	16,0	14,0
3	0,833	0,844	3,5	3,5	14	0,833	0,853	3,5	2,0
4	0,667	0,699	9,0	6,0	15	0,667	0,610	9,0	11,0
5	0,500	0,537	16,0	16,0	16	1,000	1,000	1,0	1,0
6	0,500	0,466	16,0	18,0	17	0,500	0,537	16,0	16,0
7	0,833	0,835	3,5	5,0	18	0,833	0,844	3,5	3,5
8	0,333	0,301	20,5	21,0	19	0,500	0,457	16,0	19,0
9	0,667	0,688	9,0	8,5	20	0,667	0,688	9,0	8,5
10	0,167	0,147	22,0	22,0	21	0,667	0,599	9,0	12,0
11	0,500	0,555	16,0	13,0	22	0,667	0,697	9,0	7,0
Dimensione 2 (10 item)									
Stud.	A_2	A_2^p	$r(A_2)$	$r(A_2^p)$	Stud.	A_2	$r(A_2^p)$	$r(A_2)$	A_2^p
1	0,900	0,856	8,0	10,0	12	0,600	0,612	22,0	22,0
2	0,800	0,835	17,0	15,0	13	0,800	0,777	17,0	18,0
3	0,700	0,659	20,5	20,5	14	0,900	0,856	8,0	10,0
4	1,000	1,000	1,0	1,0	15	0,800	0,798	17,0	16,0
5	0,900	0,942	8,0	2,0	16	0,900	0,914	8,0	4,0
6	0,900	0,856	8,0	10,0	17	0,900	0,921	8,0	3,0
7	0,900	0,899	8,0	5,0	18	0,900	0,856	8,0	10,0
8	0,900	0,856	8,0	10,0	19	0,900	0,856	8,0	10,0
9	0,900	0,856	8,0	10,0	20	0,900	0,856	8,0	10,0
10	0,900	0,856	8,0	10,0	21	0,800	0,777	17,0	18,0
11	0,700	0,659	20,5	20,5	22	0,800	0,777	17,0	18,0
Dimensione 3 (17 item)									
Stud.	A_3	A_3^p	$r(A_3)$	$r(A_3^p)$	Stud.	A_3	A_3^p	$r(A_3)$	$r(A_3^p)$
1	0,647	0,662	11,0	10,0	12	0,353	0,390	22,0	22,0
2	0,647	0,647	11,0	11,0	13	0,529	0,540	16,5	16,0
3	0,471	0,460	20,0	20,0	14	1,000	1,000	1,0	1,0
4	0,765	0,797	6,5	6,0	15	0,882	0,848	3,0	4,0
5	0,824	0,857	4,5	3,0	16	0,471	0,481	20,0	19,0
6	0,471	0,451	20,0	21,0	17	0,588	0,645	13,5	12,0
7	0,765	0,788	6,5	7,0	18	0,941	0,930	2,0	2,0
8	0,529	0,535	16,5	17,5	19	0,647	0,640	11,0	13,0
9	0,529	0,535	16,5	17,5	20	0,588	0,624	13,5	14,0
10	0,824	0,843	4,5	5,0	21	0,706	0,739	8,5	8,0
11	0,529	0,556	16,5	15,0	22	0,706	0,719	8,5	9,0

Tab. 2 – Ranking degli studenti di due classi selezionate, per ognuna delle tre dimensioni, utilizzando l'indicatore non pesato e pesato, A_d e A_d^p , rispettivamente (continua)

b) Classe 2 – 20 studenti									
Dimensione 1 (6 item)									
Stud.	A_1	A_1^p	$r(A_1)$	$r(A_1^p)$	Stud.	A_1	A_1^p	$r(A_1)$	$r(A_1^p)$
1	0,333	0,392	10,5	9,0	11	0,000	0,000	19,0	19,0
2	0,167	0,154	15,5	16,5	12	0,500	0,537	4,5	5,0
3	0,333	0,312	10,5	11,0	13	0,667	0,688	1,0	1,0
4	0,333	0,310	10,5	12,0	14	0,500	0,532	4,5	6,0
5	0,500	0,475	4,5	7,0	15	0,167	0,156	15,5	15,0
6	0,333	0,392	10,5	9,0	16	0,000	0,000	19,0	19,0
7	0,333	0,303	10,5	13,0	17	0,500	0,548	4,5	3,0
8	0,000	0,000	19,0	19,0	18	0,500	0,539	4,5	4,0
9	0,167	0,236	15,5	14,0	19	0,500	0,557	4,5	2,0
10	0,333	0,392	10,5	9,0	20	0,167	0,154	15,5	16,5
Dimensione 2 (10 item)									
Stud.	A_2	A_2^p	$r(A_2)$	$r(A_2^p)$	Stud.	A_2	A_2^p	$r(A_2)$	$r(A_2^p)$
1	0,700	0,707	11,0	11,0	11	0,600	0,552	16,0	16,0
2	0,700	0,750	11,0	6,0	12	0,500	0,499	19,0	18,5
3	0,800	0,744	6,0	7,0	13	0,500	0,499	19,0	18,5
4	0,700	0,686	11,0	13,0	14	1,000	1,000	1,5	1,5
5	0,900	0,856	3,5	3,5	15	1,000	1,000	1,5	1,5
6	0,800	0,740	6,0	9,0	16	0,600	0,573	16,0	15
7	0,700	0,691	11,0	12,0	17	0,700	0,743	11,0	8,0
8	0,600	0,537	16,0	17,0	18	0,500	0,488	19,0	20,0
9	0,900	0,856	3,5	3,5	19	0,800	0,777	6,0	5,0
10	0,700	0,671	11,0	14,0	20	0,700	0,718	11,0	10,0
Dimensione 3 (17 item)									
Stud.	A_3	A_3^p	$r(A_3)$	$r(A_3^p)$	Stud.	A_3	A_3^p	$r(A_3)$	$r(A_3^p)$
1	0,294	0,262	18,5	19,0	11	0,353	0,317	15,0	16,5
2	0,412	0,431	11,5	10,0	12	0,294	0,280	18,5	18,0
3	0,412	0,380	11,5	12,0	13	0,353	0,317	15,0	16,5
4	0,353	0,318	15,0	15,0	14	0,588	0,602	4,0	2,0
5	0,647	0,591	2,0	3,0	15	0,588	0,585	4,0	4,0
6	0,353	0,323	15,0	14,0	16	0,529	0,502	6,5	7,0
7	0,471	0,442	9,0	8,0	17	0,353	0,372	15	13,0
8	0,529	0,537	6,5	6,0	18	0,471	0,403	9,0	11,0
9	0,706	0,681	1,0	1,0	19	0,588	0,566	4,0	5,0
10	0,235	0,188	20,0	20,0	20	0,471	0,435	9,0	9,0

Nota: La colonna etichettata con $r(\cdot)$ riporta il ranking in base all'indicatore entro parentesi.

Per valutare il funzionamento dei due indicatori, essi sono stati applicati alle risposte di ogni studente presente nel campione analizzato. Successivamente sono state confrontate le classificazioni degli studenti all'interno della classe cui appartengono, utilizzando entrambi gli indicatori. I risultati mostrano che, utilizzando gli indicatori non pesati o viceversa quelli pesati, le classificazioni degli studenti rimangono invariate soltanto in una piccola proporzione di classi. Considerando il totale di 1.465 classi analizzate, soltanto in cinque classi non osserviamo variazioni nella classificazione degli studenti per quanto riguarda la prima dimensione, mentre per le altre due dimensioni le classi "invariate" sono solo sei e sette, rispettivamente. Inoltre, per ogni classe è stato calcolato il numero di studenti che subisce una variazione

nella loro classificazione, considerando o meno i pesi nel calcolare l'indicatore di abilità. In media circa il 70% degli studenti subisce una variazione nel loro ranking all'interno della classe considerando la prima dimensione, mentre nelle altre due dimensioni questo dato è pari a circa il 65%.

Nella tab. 2 è possibile osservare i risultati ottenuti su due classi selezionate come esempio illustrativo, composte rispettivamente da 22 e 20 studenti. Utilizzando gli indicatori riportati in (4) e (5), sono state calcolate misure composite di abilità complessiva dello studente nel test INVALSI di matematica. In seguito, ogni studente è stato classificato all'interno della classe di appartenenza in base a questi due indicatori: qualora vi fossero studenti con la stessa posizione nella classificazione, a essi viene assegnata la posizione "media" in base ai possibili ranking. Per esempio, se i primi quattro studenti di una classe condividono la prima posizione, essi riceveranno la posizione 2,5, ottenuta come media delle posizioni 1, 2, 3 e 4.

Osservando la tab. 2 è possibile notare i diversi ranking ricevuti dagli studenti in base ai due indicatori composti. Per esempio, nella classe 1 (parte a), gli studenti 4 e 21 condividono la stessa posizione nella prima dimensione se consideriamo l'indicatore non pesato, ma diverse posizioni se consideriamo la versione pesata. Pesare le loro risposte in base al potere discriminante degli item consente infatti di differenziare meglio le loro posizioni. Un discorso simile è valido per gli studenti 2 e 6 nella classe 2 (parte b), per quanto riguarda la seconda dimensione. In questo caso, è possibile osservare addirittura un'inversione delle posizioni in classifica di questi due studenti. Infatti, se consideriamo l'indicatore non pesato, lo studente 2 ottiene un punteggio di 0,7 e l'undicesima posizione all'interno della classe, mentre 0,8 è il punteggio dello studente 6, che occupa così la sesta posizione. Considerando invece l'indicatore pesato, tenendo quindi conto della discriminatività degli item, lo studente 2 risale fino alla sesta posizione, scalando ben cinque posizioni, mentre lo studente 6 retrocede di tre posizioni fino alla nona. Ciò accade perché lo studente 2 ha risposto correttamente agli item più discriminanti, viceversa lo studente 6 ha risposto correttamente agli item meno discriminanti.

4. Conclusioni

Nel presente lavoro è stata illustrata una procedura a due fasi per costruire indicatori di sintesi delle abilità degli studenti mediante i modelli IRT multidimensionali. Questi modelli costituiscono una metodologia particolarmente adatta quando l'oggetto di studio non è direttamente osservabile ed è caratterizzato da una struttura multidimensionale (come l'abilità di uno studente).

Il primo passo di tale procedura consente di ottenere il numero di dimensioni misurate dai dati, mediante un algoritmo gerarchico di *clustering*. Nel secondo passo viene stimato un modello IRT 2-PL multidimensionale, considerando il numero di dimensioni determinato nella prima fase. Dalla stima di tale modello, è possibile ottenere i parametri di discriminazione relativi a ogni item del test e utilizzarli per la costruzione di una misura composita pesata, che sintetizzi le abilità degli studenti a un test di verifica degli apprendimenti.

Questa procedura è stata applicata, a scopo esemplificativo, ai dati relativi al test INVALSI di matematica, somministrato nell'anno 2014 a studenti delle scuole medie. In questo esempio abbiamo mostrato come costruire un indicatore pesato di sintesi delle abilità in matematica degli studenti in questione, considerando la diversa discriminazione degli item del test. È stata poi confrontata la performance di tale indicatore "pesato" rispetto al suo corrispondente non pesato, ottenuto come semplice somma dei punteggi ai singoli item del test. La valutazione delle performance è stata effettuata confrontando le classificazioni degli studenti entro la classe di appartenenza, considerando l'indicatore pesato piuttosto che quello non pesato. È stato osservato che circa 7 studenti su 10 subiscono una variazione nella classificazione, e dunque nella valutazione, all'interno della classe, se teniamo conto dell'indicatore pesato, anziché della sua versione non pesata. È inoltre possibile apprezzare come gli studenti che hanno risposto correttamente agli item più discriminanti (quindi con pesi maggiori) occupano le posizioni più alte. Di conseguenza, riteniamo che, quando vi è interesse nel valutare l'abilità complessiva a un test, il diverso peso informativo degli item deve essere preso in considerazione; in caso contrario, presupponendo il medesimo potere informativo per tutti gli item, potremmo incorrere in una valutazione distorta e parziale dell'abilità dello studente.

È inoltre importante ricordare che questa procedura consente di individuare gli item più informativi, cioè quelli che ricevono un peso maggiore, tenendo conto, allo stesso tempo, della multidimensionalità dei dati. Individuare gli item più informativi può essere rilevante per l'INVALSI, in ottica di restituzione dei risultati alle scuole, in quanto la

selezione degli item più informativi (due o tre) entro ogni dimensione può agevolare in modo considerevole la fase di restituzione dei risultati alle scuole, sia in termini di costi sia di tempi sia, infine, di semplificazione della lettura dei risultati da parte delle scuole. Un altro aspetto a favore della procedura proposta riguarda la possibilità di impiegarla in fase di pre-test con l'obiettivo di ridurre la lunghezza di un test, rimuovendo gli item meno discriminanti e dunque meno informativi in ciascuna dimensione.

Infine, la procedura proposta può essere estesa in vari modi, per esempio, per classificare unità di livello maggiore (ossia le classi, nel nostro esempio, o le scuole) in gruppi omogenei e tenendo conto della struttura multilivello dei dati.

Riferimenti bibliografici

- Bacci S., Gnaldi M. (2015), "A classification of university courses based on students' satisfaction: An application of a two-level mixture item response model", *Quality & Quantity*, 49, 3: 927-940.
- Bartolucci F. (2007), "A class of multidimensional IRT models for testing unidimensionality and clustering items", *Psychometrika*, 72, 2: 141-157.
- Bartolucci F., Bacci S., Gnaldi M. (2015), *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*, CRC Press, Boca Raton.
- Bartolucci F., Montanari G.E., Pandolfi S. (2012), "Dimensionality of the latent structure and item selection via latent class multidimensional IRT models", *Psychometrika*, 77, 4: 782-802.
- Bonifay W.E., Reise S.P., Scheines R., Meijer R.R. (2015), "When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index", *Structural Equation Modeling: A Multidisciplinary Journal*, 22, 4: 504-516.
- Gnaldi M. (2017), "A multidimensional IRT model approach for dimensionality assessment of standardised students' tests in mathematics", *Quality & Quantity*, 51, 3: 1167-1182.
- Gnaldi M., Bacci S. (2016), "Joint assessment of the latent trait dimensionality and observed differential item functioning of students' national tests", *Quality & Quantity*, 50, 4: 1429-1447.
- Gnaldi M., Bacci S., Bartolucci F. (2016), "A multilevel finite mixture item response model to cluster examinees and schools", *Advances in Data Analysis and Classification*, 10, 1: 53-70.
- Gnaldi M., Del Sarto S. (2016), "Variable weighting via multidimensional IRT models in Composite Indicators construction", *Social Indicators Research*, doi: 10.1007/s11205-016-1500-5.
- Gnaldi M., Del Sarto S., Maggino F. (2017), "The role of extended IRT models for composite indicators construction", in F. Maggino (ed.), *Complexity in Society – From Indicators Construction to Their Synthesis*, Social Indicator Research Series, 70: 213-227, doi: 10.1007/978-3-319-60595-1_9.
- INVALSI (2012a), *Quadro di riferimento per il primo ciclo di istruzione*, rapporto tecnico INVALSI, testo disponibile al sito: http://www.invalsi.it/snv2012/documenti/QDR/QdR_Mat_I_ciclo.pdf, data di consultazione: 6 marzo 2017.
- INVALSI (2012b), *Quadro di riferimento per il secondo ciclo di istruzione*, rapporto tecnico INVALSI, testo disponibile al sito: http://www.invalsi.it/snv2012/documenti/QDR/QdR_Mat_II_ciclo.pdf, data di consultazione: 6 marzo 2017.

15. *Tecniche statistiche avanzate per la previsione della performance scolastica* *Advanced statistical models for predicting student performances*

di Anna Siri, Nicola Luigi Bragazzi, Luca Oneto

Offrire uno strumento predittivo dei fattori che influenzano la performance scolastica è un obiettivo ambizioso nella ricerca educativa, tanto più se realizzato in modo da consentire di estrarre rapidamente informazioni operative dall'enorme massa di dati riguardanti ogni singolo studente.

La presente ricerca ha come obiettivi lo studio, lo sviluppo e l'implementazione delle metodologie avanzate di *educational data mining* più opportune per predire le performance degli studenti della scuola superiore.

Per il nostro studio abbiamo utilizzato i dati della rilevazione INVALSI sugli apprendimenti in italiano e matematica per la seconda classe della scuola superiore (a.s. 2013-14). Si tratta di un test standard che intende misurare negli studenti alla fine dell'obbligo scolastico, indipendentemente dal tipo di scuola e dall'indirizzo di studio frequentato, le conoscenze e le competenze necessarie sia per proseguire gli studi all'università sia per inserirsi nel mondo del lavoro.

Le tecniche proposte sono risultate efficaci sia nel fornire un modello in grado di predire le performance scolastiche, sia nel selezionare le variabili che più ne influenzano il risultato.

A crucial issue in the educational field is represented by the ability to provide different stakeholders, including educators and policymakers, with an advanced tool able to predict the variables that influence school performance at the student level by efficiently extracting actionable information from massive amounts of data.

The research project is focused on the study, the development and the implementation of the most appropriate data mining methodologies for predicting school student performance. We used data from the INVALSI standardised national test assessing the reading and mathematical skills of Italian pupils carried out in secondary schools in the scholastic year 2013-14. The test aims to measure the essential skills and knowledge required to pursue university degree courses or to get a job.

We have proved the effectiveness of our proposal both in predicting the school performance of the student and in detecting the most influencing variables of the phenomenon.

1. Introduzione

Le tecniche di *machine learning* e di *data mining*, nonché i più recenti sistemi di raccolta, gestione e fruizione dei dati stanno rivoluzionando ogni aspetto della vita umana (Bennato, 2015).

Le nuove tecnologie consentono di raccogliere, memorizzare ed elaborare grandi quantità di dati provenienti da fonti diverse ed eterogenee. Software per la memorizzazione e l'analisi di dati come Hadoop (Shvachko *et al.*, 2010) e Spark (Zaharia *et al.*, 2012) e loro ecosistemi consentono di accedere ed elaborare facilmente dati di diverso formato anche provenienti da fonti diverse (per esempio, banche dati istituzionali, social network). A ciò si aggiungono i recenti progressi nella costruzione di algoritmi di apprendimento (Evgeniou e Pontil, 2004; Chapelle, Scho'lkopf e Zien, 2006; Hinton, Osindero e Teh, 2006; Pan e Yang, 2010) e di rigorose procedure di inferenza statistica (Bartlett, Boucheron e Lugosi, 2002; Anguita *et al.*, 2012) che hanno notevolmente migliorato la capacità di estrarre informazioni significative da enormi quantità di dati, frammentati e difficilmente interpretabili se presi singolarmente, consentendo quindi una maggiore fruibilità in termini di facilità di accesso, di interpretazione e di condivisione delle informazioni. Anche il

problema legato alla riservatezza dei dati, un tempo apparentemente insormontabile, è stato superato grazie ai recenti progressi nel campo (Dwork, 2008).

Il recente rapporto OECD 2016 *Science, Technology and Innovation Outlook – Tenkey Technology Trends for the Future* delinea la grande potenzialità e insieme l'impatto di ordine sociale e culturale delle nuove tecnologie che stanno emergendo e fa comprendere la portata delle sfide che ci troveremo ad affrontare.

In questo contesto, anche la ricerca sociale, da anni legata ad approcci statistici basati sul modello classico, si sta avvicinando sempre più a tali tecniche per meglio comprendere i fenomeni studiati.

L'applicazione del *data mining* nella ricerca educativa (Mason, Vaughane e Wallach, 2014; Koedinger *et al.*, 2015) ha infatti di recente attirato l'interesse crescente di studiosi appartenenti ad aree scientifiche diverse, consentendo così la creazione di gruppi di ricerca interdisciplinari.

Tra gli obiettivi principali di applicazione del *data mining* in educazione, così come evidenziati da Baker e Yacef (2009), vi sono la possibilità di prevedere il comportamento futuro degli studenti, di scoprire nuovi modelli o migliorare gli esistenti, nonché di studiare gli effetti di tutorato e counselling in ambito educativo e di consentire il progresso delle conoscenze scientifiche sul comportamento degli studenti.

Offrire uno strumento di previsione dei fattori che influenzano la performance scolastica costruito in modo da consentire una rapida estrazione delle informazioni operative dall'enorme numero di dati riguardanti ogni studente è un obiettivo molto ambizioso nella ricerca educativa.

Lo sviluppo di un modello predittivo basato sull'utilizzo di metodi classici di statistica monovariata o lineare è infatti limitato dall'elevata quantità di variabili in gioco e dalla complessità di interazioni tra di esse, il cui effetto sulla variabile dipendente (la performance scolastica) risulta essere altamente non-lineare.

Le tecniche più avanzate di *educational data mining* (Papamitsiou ed Economides, 2014), per esempio le tecniche di apprendimento automatico basate su *kernel* (*Support Vector Machines*) o le tecniche *ensemble* (*Random Forest*), consentono di superare queste limitazioni e di realizzare modelli predittivi.

Il campo dell'*educational data mining* è recente e non è stato ancora sviluppato un protocollo che indirizzi il ricercatore a usare una tecnica piuttosto che un'altra in una data situazione. Perciò intendiamo comparare diversi approcci computazionali per ottenere risultati robusti e solidi, in quanto confermati da varie tecniche di *data mining*.

2. Scopo della ricerca

Lo scopo della ricerca è quello di valutare se i dati raccolti attraverso rilevazioni nazionali sugli apprendimenti in italiano e matematica possono fornire informazioni utili alla previsione del successo scolastico e, subordinatamente, di selezionare le variabili che più ne influenzano i risultati.

3. I dati

L'Istituto nazionale per la valutazione del sistema educativo di istruzione e formazione (INVALSI) ha il compito di "attuare verifiche periodiche e sistematiche sulle conoscenze e abilità degli studenti" (d.lgs. n. 286/2004), con l'obiettivo di valutare l'efficacia e l'efficienza del Sistema educativo di istruzione e formazione. A tal fine, ogni anno, elabora prove standardizzate con l'obiettivo di misurare e monitorare i livelli di apprendimento degli studenti in italiano e matematica. I livelli scolastici interessati sono le classi seconda e quinta della scuola primaria, la classe terza della scuola secondaria di I grado e la classe seconda della scuola secondaria di II grado.

Nell'anno scolastico considerato, l'a.s. 2013-14, la rilevazione degli apprendimenti ha coinvolto tutte le scuole statali e paritarie (nell'anno scolastico di riferimento circa 13.200) e tutti gli studenti dei quattro livelli scolari interessati, ossia 2.287.745 alunni (INVALSI, 2014).

Più specificamente, il dataset utilizzato si riferisce esclusivamente agli esiti della rilevazione sugli apprendimenti in italiano e matematica somministrata agli allievi della scuola secondaria di II grado nell'a.s. 2013-14, pari a 560.672 alunni, appartenenti a 26.540 classi.

Il test INVALSI per la scuola secondaria di II grado è un test standardizzato che intende misurare, indipendentemente dal tipo di scuola e dall'indirizzo di studio frequentato dagli studenti alla fine dell'obbligo scolastico, le conoscenze e le competenze necessarie sia per proseguire gli studi all'università sia per inserirsi nel mondo del lavoro.

Al fine di approfondire l'analisi, alle variabili del dataset sono state associate anche le informazioni fornite dal Questionario studente, sempre relative all'a.s. 2013-14.

Sono state utilizzate tutte le variabili disponibili nei due dataset INVALSI (prove nazionali di italiano e matematica e Questionario studente), dopo l'eliminazione di un certo numero di variabili, quali per esempio le variabili "voto orale di italiano" e "voto orale di matematica", per la presenza di dati mancanti (*missing data*). I valori mancanti non sono stati sostituiti, in quanto si riferiscono a informazioni (talune soggettive) ritenute non omologabili ad altro valore. Il database utilizzato è composto da 261 variabili.

4. Analisi

La soluzione considerata propone un approccio modellistico innovativo con le tecniche computazionali più recenti, il cui scopo è la formulazione e validazione di metodi e modelli in grado di inferire informazione da insiemi di dati complessi, di riconoscere in essi la presenza di correlazioni non note a priori e, in ultima analisi, di fornire previsioni attendibili sul comportamento futuro di alcune variabili di interesse.

Per andare incontro a questa difficoltà abbiamo utilizzato una varietà di approcci, che vanno dai modelli generalizzati alle tecniche di apprendimento automatico quali le tecniche basate su *kernel* (*Support Vector Machines* o SVM) o le tecniche *ensemble* (*Random Forest*) per la realizzazione di modelli predittivi.

Grande attenzione è stata, in particolare, rivolta al problema della selezione del modello tramite l'ottimizzazione degli iperparametri delle tecniche di apprendimento automatico attraverso sia i classici metodi di ricampionamento sia i metodi più avanzati basati sulla teoria dell'apprendimento statistico ("imparare dai dati").

Il nostro approccio metodologico ha permesso inoltre di acquisire, in fase pre-sperimentale, ulteriori informazioni circa l'incidenza e il peso relativo dei fattori di rischio considerati. A tale scopo i modelli predittivi sono stati utilizzati congiuntamente a tecniche avanzate di selezione delle variabili di maggiore rilevanza, come il test di permutazione, e tecniche di regolarizzazione per aumentare la sparsità di rappresentazione.

In particolare, le *Random Forest* (RF) sono uno strumento di classificazione introdotto per la prima volta nel 2001 da Leo Breiman. Oltre alla migliorata accuratezza, sono interessanti per la loro efficienza, dovuta al parallelismo sia in fase di costruzione sia in fase di classificazione.

In un classico problema di regressione (Vapnik, 1998, Shawe-Taylor e Cristianini, 2004) si ha a disposizione un insieme di dati $\mathcal{D}_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ con $x_i \in \mathcal{X} \in \mathbb{R}^d$ e $y_i \in \mathcal{Y} \in \mathbb{R}$. In questo contesto lo scopo è quello di identificare il sistema incognito $\mathfrak{S}: \mathcal{X} \rightarrow \mathcal{Y}$ attraverso il modello $\mathfrak{M}: \mathcal{X} \rightarrow \mathcal{Y}$ scelto da un algoritmo $\mathcal{A}_{\mathcal{H}}$ caratterizzato dal suo insieme di iperparametri \mathcal{H} .

L'accuratezza del modello \mathfrak{M} nel rappresentare \mathfrak{S} può essere valutata attraverso diverse misure di accuratezza (Ghelardoni, Ghio e Anguita, 2013, Elattar, Goulermas e Wu, 2010). In particolare, dato un insieme di nuovi dati $\mathcal{T}_m = \{(x_1, y_1), \dots, (x_m, y_m)\}$, il modello creato produrrà un insieme di uscite $\{\hat{y}_1, \dots, \hat{y}_m\}$ dati gli ingressi $\{x_1, \dots, x_m\}$.

In base a questi dati è possibile calcolare le seguenti misure di accuratezza [19]:

- Errore assoluto medio: $MAE = 1/m \sum_{i=1}^m |y_i - \hat{y}_i|$
 - Correlazione: $PPMCC = \sum_{i=1}^m (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) / \left(\sqrt{\sum_{i=1}^m (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^m (\hat{y}_i - \bar{\hat{y}})^2} \right)$
- con $\bar{\hat{y}} = 1/m \sum_{i=1}^m \hat{y}_i$

Una volta che il modello è stato costruito e validato in modo da garantire la sua accuratezza nel rappresentare il fenomeno, è importante capire come lo stesso possa essere influenzato dalle varie variabili che sono state utilizzate durante la sua costruzione.

Nel campo dell'analisi dati questa procedura prende il nome di "selezione e ordinamento delle variabili" (Hong, 1997; Friedman, Hastie e Tibshirani, 2001; Guyon ed Elisseeff, 2003; Yoon, Yang e Shahabi, 2005; Chang e Lin, 2008). Questo processo permette di capire se alcune variabili, che a priori sono note per influenzare il problema in esame, sono tenute in considerazione dal modello \mathfrak{M} . L'incapacità del modello statistico nel tenere in considerazione queste variabili potrebbe indicare una scarsa qualità del modello o l'identificazione di correlazione spurie.

È noto in letteratura che combinare la predizione di molti modelli produce un risultato nettamente migliore rispetto all'utilizzo di uno solo di essi (Breiman, 2001; Germain *et al.*, 2015). Per questo motivo molti algoritmi costruiscono una combinazione opportunamente pesata di modelli semplici (Lever, Laviolette e Shawe-Taylor, 2013): *bagging* (Breiman, 2001), *boosting* (Schapire *et al.*, 1998), tecniche Bayesiane (Gelman *et al.*, 2014), così anche le reti neurali (NN) (Bishop, 1995) o le tecniche basate su *kernel* (SVM) (Vapnik, 1998, Anguita *et al.*, 2011a). Ciononostante, ottimizzare l'accuratezza del modello finale rappresenta un problema largamente irrisolto.

Come vanno costruiti i modelli semplici? Quanti modelli semplici occorre costruire? Come si combinano questi modelli semplici? Esiste una teoria che ci può aiutare a prendere queste decisioni?

Breiman (2001) ha provato a dare una risposta a queste domande proponendo le *Random Forest* (RF) di alberi di regressione e classificazione, che rappresentano probabilmente uno degli algoritmi più performanti presenti in letteratura (Fernandez-Delgado *et al.*, 2014). Le *Random Forest* combinano il *bagging* con un campionamento casuale delle variabili presenti nell'insieme di dati originali. Nella tecnica del *bagging* ogni albero è costruito in modo indipendente utilizzando la procedura di *bootstrap* sull'insieme di dati a disposizione (Efron, 1979). RF aggiunge un altro livello di casualità al *bagging*. Oltre a costruire ogni albero sfruttando la tecnica di *bootstrap* sull'insieme di dati a disposizione, RF cambia il modo in cui ogni albero è costruito. Classicamente ogni nodo di un albero è costruito usando la variabile più discriminante nell'insieme di dati a disposizione. In RF invece ogni nodo di un albero è costruito usando la variabile più discriminante in un sottoinsieme casuale delle variabili dell'insieme di dati a disposizione. Alla fine gli alberi sono combinati con una votazione, nel caso della classificazione, o con la media nel caso della regressione.

L'accuratezza del modello finale dipende principalmente da tre fattori: il numero di alberi costruiti, l'accuratezza di ogni albero e la correlazione tra di essi. Essa cresce man mano che si aumenta il numero di alberi nella foresta, oppure aumentando l'accuratezza di ogni singolo albero e infine diminuendo la correlazione tra gli stessi. La strategia delle RF si è dimostrata molto efficace in quanto molto robusta al rumore rispetto ad altre tecniche (Breiman, 2001; Fernandez-Delgado *et al.*, 2014).

La fase di apprendimento di ogni singolo albero degli n_t alberi che compongono la RF è molto semplice. Da \mathcal{D}_n , $[bn]$ campioni sono estratti con reinserimento creando l'insieme di dati $\mathcal{D}'_{[bn]}$. L'albero è costruito usando $\mathcal{D}'_{[bn]}$, ma la variabile più discriminante è scelta in un sottoinsieme casuale di grandezza n_v delle variabili dell'insieme di dati a disposizione. L'albero cresce fino a quando il nodo contiene un massimo di n_l campioni. Durante la fase di predizione dell'output associato a un input x , ogni albero produce un output $y_i \in \{1, \dots, n_t\}$, e il risultato finale non è altro che la $\{p_1, \dots, p_{n_t}\}$ -pesata combinazione delle risposte dei vari alberi che compongono la foresta (si noti come occorra che $\sum_{i=1}^{n_t} p_i = 1$).

Imponendo

$$b = 1, n_v = \sqrt{n}, n_l = 1 \text{ e } P_{i \in \{1, \dots, n_t\}} = 1/n_t$$

si ottiene la formulazione originale delle RF (Breiman, 2001), dove n_t è scelto solitamente come compromesso tra accuratezza e efficienza (Hernández-Lobato, Martínez-Muñoz e Suárez, 2013) oppure basandosi sul cosiddetto errore out-of-bag (Breiman, 2001), o infine basandosi su un principio di consistenza (Hernández-Lobato, Martínez-Muñoz e Suárez, 2013).

Un comune errore è quello di considerare le RF come un algoritmo privo di iperparametri (Biau, 2012). Esistono infatti molti iperparametri che influenzano l'accuratezza del modello finale: il numero di alberi n_t , il numero di campioni da campionare durante la procedura di *bootstrap* b , la profondità di ogni albero n_l e la cardinalità n_v del sottoinsieme di variabili per la selezione del miglior discriminante. Oltre a questo i pesi $\{p_1, \dots, p_{n_t}\}$ sono di grande importanza per l'accuratezza del modello finale (Anguita *et al.*, 2012; Orlandi, Oneto e Anguita, 2016). Per questa ragione in questo lavoro useremo la strategia proposta da Catoni (2007) e recentemente sviluppata in Lever, Laviolette e Shawe-Taylor (2013) e Oneto, Ridella e Anguita (2016) per pesare ogni albero T_i in base al suo errore di out-of-bag $\hat{L}(T_i)$ (Catoni, 2007, Lever, Laviolette e Shawe-Taylor, 2013; Orlandi, Oneto e Anguita, 2016; Oneto, Ridella e Anguita, 2016):

$$p_i = \frac{e^{-\gamma \widehat{L}(T_i)}}{\sum_{j=1}^{n_t} e^{-\gamma \widehat{L}(T_j)}}$$

dove γ è un altro iperparametro. Al fine di selezionare il miglior valore dei vari iperparametri n_p , b , n_v , e n_t per ottenere la più alta accuratezza possibile, occorre utilizzare quindi una tecnica di selezione del modello (Anguita *et al.*, 2012).

RF, oltre a essere un efficace metodo di modellazione di problemi di regressione, può essere utilizzato al fine di effettuare una robusta e accurata selezione delle variabili.

Esistono due metodi fondamentali in questo ambito. Il primo può essere descritto come segue: a ogni accrescimento di un albero l'errore out-of-bag deve essere conservato. Quindi viene effettuata una permutazione casuale della variabile nell'insieme di dati out-of-bag e l'errore out-of-bag viene ricalcolato. La differenza mediata sugli alberi della foresta di questi due errori rappresenta l'importanza della variabile j . Questo approccio nasce dal test di permutazione (Good, 2013) usato in maniera estensiva in letteratura, molto efficiente ed efficace nel caso delle RF (White e Liu, 1994; Deng, Runger e Tuv, 2011). Il secondo approccio, meno efficace, consiste nel calcolare la diminuzione dell'impurità dovuta alla variabile mediata sugli alberi della foresta. In regressione l'impurità è misurata con i residui ai minimi quadrati.

La selezione del modello (MS) affronta il problema di settare correttamente gli iperparametri di un algoritmo di apprendimento (Anguita *et al.*, 2012).

Esistono molti metodi di MS ma i metodi di ricampionamento quali la *K-fold Cross Validation* (KCV) (Kohavi, 1995) e il *bootstrap* non parametrico (BTS) (Anguita *et al.*, 2011b) rappresentano lo stato dell'arte nella soluzione di molti problemi pratici (Anguita *et al.*, 2012).

I metodi di ricampionamento nascono da una semplice idea. L'insieme di dati \mathcal{D}_n è ricampionato una e più (n_r) volte, con o senza reinserimento, in modo da creare due insiemi indipendenti chiamati, rispettivamente, insieme di apprendimento \mathcal{L}_1^r e insieme di validazione \mathcal{V}_v^r con $r \in \{1, \dots, n_r\}$. Si noti come $\mathcal{L}_1^r \cap \mathcal{V}_v^r = \emptyset$ e $\mathcal{L}_1^r \cup \mathcal{V}_v^r = \mathcal{D}_n$. Di conseguenza, al fine di trovare il miglior insieme di iperparametri \mathcal{H} in un insieme di configurazioni $\mathfrak{S} = \{\mathcal{H}_1, \mathcal{H}_2, \dots\}$ per l'algoritmo $\mathcal{A}_{\mathcal{H}}$, la seguente procedura deve essere applicata:

$$\mathcal{H}^*: \min \frac{1}{n_r} \sum_{r=1}^{n_r} \frac{1}{v} \sum_{(x_i, y_i) \in \mathcal{V}_v^r} |\mathcal{A}_{\mathcal{H}, \mathcal{L}_1^r}(x_i) - y_i|$$

dove $\mathcal{A}_{\mathcal{H}, \mathcal{L}_1^r}$ è il modello costruito con l'algoritmo \mathcal{A} , gli insiemi di iperparametri \mathcal{H} e l'insieme di dati \mathcal{L}_1^r . Visto che \mathcal{L}_1^r è indipendente da \mathcal{V}_v^r l'idea è che \mathcal{H}^* dovrebbe essere l'insieme di iperparametri che garantisce di ottenere un errore piccolo su un insieme di dati indipendenti da quelli usati per costruire il modello.

Si noti come imponendo $r = 1$, con l e v decisi a priori tali per cui $n = l + v$, e con il ricampionamento effettuato senza reinserimento, si ottenga la tecnica dell'insieme di test.

Al fine di implementare la KCV occorre invece imporre:

$$r \leq \binom{n}{k} \binom{n - \frac{n}{k}}{k}, \quad l = (k - 1) \frac{n}{k} \quad \text{e} \quad v = \frac{n}{k}$$

utilizzando un ricampionamento senza reinserimento. Infine per il BTS $r \leq n^n, l = n$, il ricampionamento di \mathcal{L}_1^r viene effettuato con reinserimento dove \mathcal{V}_v^r sono i dati in \mathcal{D}_n non presenti in \mathcal{L}_1^r .

In questo lavoro abbiamo usato il BTS in quanto rappresenta il metodo più efficace (Efron e Tibshirani, 1993; Anguita *et al.*, 2011b; Anguita *et al.*, 2012).

Per il modello generalizzato è stato utilizzato il software SPSS versione 23 (IBM, NY, USA), mentre le tecniche di SVM e RF sono state realizzate mediante script *ad hoc* in ambiente R.

5. Risultati

Nel modello generalizzato relativo alla prova di matematica emergono quali variabili più importanti l'indice di propensione al *cheating*¹, il voto orale nella materia e il genere, come evidenziato nella tab. 1 sotto riportata.

Tab. 1 – Modello generalizzato relativo alla prova di matematica

Sorgente	F	Sig.	Eta quadrato parziale	Potenza osservata
Genere	6.990,196	0,000	0,034	1,000
Orario_scolastico	680,207	0,000	0,013	1,000
Regolarità	5,168	0,023	0,000	0,623
Cittadinanza	97,198	0,000	0,001	1,000
Cod_provincia_ISTAT	28,307	0,000	0,011	1,000
ESCS	857,639	0,000	0,004	1,000
Voto_orale_mat	15.307,115	0,000	0,071	1,000
Cheating_Matematica	23.947,695	0,000	0,106	1,000
Q12_E	4,711	0,003	0,000	0,899
Q12_S	3,134	0,024	0,000	0,732
Q13_F	3,412	0,017	0,000	0,772
Q13_N	2,954	0,031	0,000	0,704
Q14_D	2,847	0,036	0,000	0,686
Q14_F	2,632	0,048	0,000	0,648

Relativamente alla prova di italiano, il modello generalizzato evidenzia come variabile con maggiore peso il *cheating* e a seguire il luogo di nascita e la tipologia di scuola.

Tab. 2 – Modello generalizzato relativo alla prova di italiano

Sorgente	F	Sig.	Eta quadrato parziale	Potenza osservata
Tipologia_scuola	164,959	0,000	0,002	1,000
Luogo_nascita	2.436,459	0,000	0,035	1,000
Orario_scolastico	24,051	0,000	0,000	1,000
Cod_provincia_ISTAT	26,924	0,000	0,010	1,000
Cod_reg	32,543	0,000	0,000	1,000
Q22	6,276	0,000	0,000	0,997
ESCS	17,802	0,000	0,000	0,988
Q5_ma	3,172	0,042	0,000	0,610
Voto_orale_ita	5,115	0,024	0,000	0,619
Cheating_italiano	17.586,545	0,000	0,079	1,000

L'approccio modellistico mediante l'uso di tecniche di apprendimento automatico ha visto l'esecuzione di 48 esperimenti, in cui sono stati fatti variare deliberatamente dei fattori (controllabili) di input, per osservare la risposta in uscita e quindi, grazie a opportune elaborazioni statistiche inferenziali, determinare quali di essi abbiano indotto una variazione significativa nella risposta.

I risultati e le analisi mostrano che i migliori modelli hanno un'accuratezza di predizione superiore all'89% nel caso dei punteggi di matematica, del 97% nel caso dei punteggi di italiano.

¹ Per stabilire l'attendibilità dei dati raccolti, l'INVALSI ha deciso di rilevare indirettamente "comportamenti impropri che hanno consentito, secondo modi e forme differenti, agli allievi di fornire le risposte corrette non in virtù delle loro competenze, ma perché copiate dagli altri studenti o da libri e altre fonti (*student cheating*) o, persino, suggerite più o meno esplicitamente dai docenti (*teacher cheating*)" (INVALSI, 2012, p. 9). L'effetto del *cheating* è misurato mediante un indicatore percentuale che esprime quale parte del punteggio osservato è mediamente da attribuire alle predette anomalie.

Tab. 3 – Modelli Random Forest relativi alle prove di matematica e italiano

N_Sample Train	Esperimento n. 34 Output WLE_MAT_200		Esperimento n. 46 Output WLE_ITA_200	
	Err	Cor	Err	Cor
100	16.932.944	0,855518	15.069.278	0,879898
200	16.065.773	0,856856	12.181.772	0,943116
500	15.109.419	0,869808	10.738.474	0,955042
1000	14.886.417	0,871373	10.244.150	0,954379
2000	14.142.728	0,884859	9.629.210	0,961071
5000	13.625.343	0,890078	8.825.712	0,965912
10000	13.449.459	0,891273	8.478.030	0,969501
20000	13.101.209	0,896504	8.152.886	0,971940

Le variabili più importanti relativamente ai livelli di apprendimento della matematica sono risultate essere l'indice di propensione al *cheating*, la tipologia di scuola, il voto orale di matematica e l'orario scolastico, mentre per la prova di italiano il *cheating*, il lavoro dei genitori, l'indice di status socio-economico-culturale e il luogo di nascita.

I risultati confermano l'importanza delle politiche di contrasto alla pratica del *cheating* che risulta avere un peso determinante negli esiti di entrambe le prove. Nella scuola secondaria di II grado emergono differenze di genere soprattutto in matematica, così come risulta pesare sui punteggi anche la tipologia di scuola, in maniera più significativa rispetto a quanto succeda per l'italiano.

6. Conclusioni

Le possibilità di raccolta e memorizzazione di quantità sempre maggiori di dati, offerte dagli attuali strumenti informatici, hanno consentito di aumentare notevolmente il numero di dati da sottoporre ad analisi statistiche.

Di fronte a una tale esplosione di informazioni, l'uso esclusivo di tecniche tradizionali di analisi statistica risulta essere spesso incapace di gestire pienamente l'intero patrimonio informativo. In questa situazione, gli algoritmi del *machine learning* vengono utilizzati per individuare particolari relazioni fra i dati, anche quelle apparentemente meno significative o sconosciute a priori.

In particolare, le tecniche di *machine learning* e *data mining* applicate nel nostro studio, per la prima volta, a dati di tipo educativo provenienti da dataset pubblici nazionali, rappresentano i più innovativi e attuali approcci di modellazione predittiva. Essi presentano alcune importanti differenze rispetto alle tradizionali analisi statistiche, in particolare nella fase di trasformazione dei dati, dal momento che non richiedono che gli stessi siano normalmente distribuiti o lineari. Offrono inoltre il vantaggio di non richiedere ipotesi a priori da parte del ricercatore, neppure sulla forma distributiva delle variabili, di consentire l'esplorazione di un numero elevato di osservazioni, di minimizzare i tempi di elaborazione, nonché di ottenere risultati di semplice interpretazione.

I risultati della presente ricerca ne dimostrano l'efficacia anche con l'utilizzo di informazioni che mai prima erano state utilizzate per tale scopo. Si tratta di risultati preliminari in quanto il dataset utilizzato raccoglie le informazioni scaricabili pubblicamente e non contempla altre variabili individuali e di contesto che possono incidere sui rendimenti che sono disponibili in altri dataset non pubblici o pubblici ma non collegabili senza specifici codici riservati. L'applicazione di tali tecniche con informazioni più esaustive, così come è stata svolta dagli autori con dati relativi ad altri contesti educativi (Siri, 2014; Bragazzi, Briki e Siri, 2016), potrebbe consentire di validare *policies* che impiegano ingenti risorse economiche, specificamente negli ambiti scolastici che si rivelano significativamente correlati agli apprendimenti degli studenti, depurandole da fattori che non presentano un impatto significativo.

Risulta evidente come le tecniche di *data mining* possano rappresentare uno strumento di supporto a indagini mirate sul campo, uno strumento utile alla pianificazione delle attività di monitoraggio, in grado inoltre di favorire l'ottimizzazione delle risorse economiche, organizzative e umane per migliorare i risultati educativi.

I nuovi modelli di *data mining* possono aprire infatti nuove frontiere di analisi rispetto a quelle convenzionalmente impiegate, in termini di capacità di trattamento della varietà di dati, del volume e della velocità di elaborazione. Mentre

i tradizionali modelli di *data mining* sono funzionali a estrarre conoscenza secondo un approccio statico e retrospettivo, le tecniche di ultima generazione permettono invece di incrementare le capacità elaborative di straordinari volumi di dati, strutturati e non strutturati, costantemente aggiornati. Esse rappresentano una grande opportunità per aggregare, analizzare e sfruttare le diverse combinazioni informative provenienti da fonti nazionali e internazionali. L'incremento di tali processi permette di generare nuove informazioni e nuova conoscenza.

Riferimenti bibliografici

- Anguita D., Ghio A., Oneto L., Ridella S. (2011a), "Selecting the hypothesis space for improving the generalization ability of Support Vector Machines", in *International Joint Conference on Neural Networks*, 31 July-5 August 2011, San Jose (CA): 1169-1176, testo disponibile al sito: <http://ieeexplore.ieee.org/document/6033356/>, data di consultazione: 23 maggio 2017.
- Anguita D., Ghio A., Oneto L., Ridella S. (2011b), "In-sample model selection for Support Vector Machines", in *The 2011 International Joint Conference on Neural Networks*, 31 July-5 August 2011, San Jose, CA: 1154-1161, testo disponibile al sito: <http://ieeexplore.ieee.org/document/6033354/>, data di consultazione: 23 maggio 2017.
- Anguita D., Ghio A., Oneto L., Ridella S. (2012), "In-sample and out-of-sample model selection and error estimation for support vector machines", *IEEE Transactions on Neural Networks and Learning Systems*, 23, 9: 1390-1406.
- Arlot S., Celisse A. (2010), "A survey of cross-validation procedures for model selection", *Statistics Surveys*, 4: 40-79.
- Baker R., Yacef K. (2009), "The state of educational data mining in 2009: A review and future visions", *JEDM – Journal of Educational Data Mining*, 1, 1: 3-17.
- Bartlett P.L., Boucheron S., Lugosi G. (2002), "Model selection and error estimation", *Machine Learning*, 48, 1: 85-113.
- Bennato D. (2015), *Il computer come macroscopio. Big Data e approccio computazionale per comprendere i cambiamenti sociali e culturali*, FrancoAngeli, Milano.
- Bernard S., Heutte L., Adam S. (2009), "Influence of hyperparameters on Random Forest accuracy", in J.A. Benediktsson, J. Kittler, F. Roli (eds.), *Multiple Classifier Systems. MCS 2009*, Springer, Berlin, Heidelberg: 171-180.
- Biau G. (2012), "Analysis of a random forests model", *Journal of Machine Learning Research*, 13: 1063-1095.
- Bishop C.M. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press, Oxford, New York.
- Bragazzi N.L., Briki W., Siri A. (2016), "Predicting academic success in health professions training and education: Insights from a data mining analysis", in C.H. Gallagher (ed.), *Academic Performance: Student Expectations, Environmental Factors and Impacts on Health*, Nova Science Publishers, New York: 91-106.
- Breiman L. (2001), "Random Forests", *Machine Learning*, 45, 1: 5-32.
- Catoni O. (2007), *Pac-Bayesian Supervised Classification: The Thermodynamics of Statistical Learning*, Institute of Mathematical Statistics, Beachwood (Ohio).
- Chang Y.W., Lin C.J. (2008), "Feature ranking using linear SVM", in I. Guyon, C. Aliferis, G. Cooper, A. Elisseeff, J.-P. Pellet, P. Spirtes, A. Statnikov (eds.), *Causation and Prediction Challenge. Challenges in Machine Learning*, vol. 2, Microtome Publishing, Brookline (Mass.): 47-58.
- Chapelle O., Scho'kopf B., Zien A. (eds.) (2006), *Semi-Supervised Learning*, The MIT Press, Cambridge (Mass.).
- Deng H., Runger G., Tuv E. (2011), "Bias of importance measures for multi-valued attributes and solutions", in T. Honkela, W. Duch, M. Girolami, S. Kaski (eds.), *Artificial Neural Networks and machine learning – ICANN 2011*, Springer, Berlin, Heidelberg: 293-300.
- Dwork C. (2008), "Differential privacy: a survey of results", in M. Agrawal, D. Du, Z. Duan, A. Li (eds.), *Theory and Applications of Models of Computation. TAMC 2008*, Springer, Berlin, Heidelberg: 1-19.
- Efron B. (1979), "Bootstrap methods: another look at the jackknife", *The Annals of Statistics*, 7, 1: 1-26.
- Efron B., Tibshirani R.J. (1993), *An Introduction to the bootstrap*, Chapman & Hall, New York.
- Elattar E.E., Goulermas J., Wu Q.H. (2010), "Electric load forecasting based on locally weighted support vector regression", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40, 4: 438-447.
- Evgeniou T., Pontil M. (2004), "Regularized multi-task learning", in R. Kohavi, J. Gehrke, W. DuMouchel, J. Ghosh (eds.), *KDD 2004 – Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 22-25, Seattle (WA): 109-117.
- Fernandez-Delgado M., Cernadas E., Barro S., Amorim D. (2014), "Do we need hundreds of classifiers to solve real world classification problems?", *Journal of Machine Learning Research*, 15, 1: 3133-3181.
- Friedman J., Hastie T., Tibshirani R. (2001), *The Elements of Statistical Learning. data mining, Inference, and Prediction*, Springer, New York.

-
- Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., Rubin D.B. (2014), *Bayesian Data Analysis*, Chapman & Hall/CRC Press, Boca Raton (FL).
- Germain P., Lacasse A., Laviolette F., Marchand M., Roy J.F. (2015), “Risk bounds for the majority vote: From a PAC-Bayesian analysis to a learning algorithm”, *Journal of Machine Learning Research*, 16, 1: 787-860.
- Ghelardoni L., Ghio A., Anguita D. (2013), “Energy load forecasting using empirical mode decomposition and support vector regression”, *IEEE Transactions on Smart Grid*, 4, 1: 549-556.
- Good P. (2013), *Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer, New York.
- Guyon I., Elisseeff A. (2003), “An introduction to variable and feature selection”, *Journal of Machine Learning Research*, 3: 1157-1182.
- Hernández-Lobato D., Martínez-Muñoz G., Suárez A. (2013), “How large should ensembles of classifiers be?”, *Pattern Recognition*, 46, 5: 1323-1336.
- Hinton G.E., Osindero S., Teh Y.W. (2006), “A fast learning algorithm for deep belief nets”, *Neural Computation*, 18, 7: 1527-1554.
- Hong S.J. (1997), “Use of contextual information for feature ranking and discretization”, *IEEE Transactions on Knowledge and Data Engineering*, 9, 5: 718-730.
- INVALSI (2012), *Rilevazioni nazionali sugli apprendimenti 2011-12. Il quadro di sistema*, testo disponibile al sito: http://www.invalsi.it/snv2012/documenti/Rapporti/Rapporto_rilevazione_apprendimenti_2012.pdf, data di consultazione: 10 maggio 2017.
- INVALSI (2014), *Rilevazioni nazionali degli apprendimenti 2013-14. Rapporto risultati*, testo disponibile al sito: http://www.invalsi.it/areaprove/rapporti/Rapporto_SNV_PN_2014_10.pdf, data di consultazione: 20 luglio 2016.
- Koedinger K.R., D’Mello S., McLaughlin E.A., Pardos Z.A., Rosé C.P. (2015), “Data mining and education”, *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 4: 333-353.
- Kohavi R. (1995), “A study of cross-validation and bootstrap for accuracy estimation and model selection”, in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (II)*, August 20-25, Montreal, Quebec, Canada: 1137-1143.
- Lever G., Laviolette F., Shawe-Taylor J. (2013), “Tighter PAC-Bayes bounds through distribution-dependent priors”, *Theoretical Computer Science*, 473: 4-28.
- Mason W., Vaughan J.W., Wallach H. (2014), “Computational social science and social computing”, *Machine Learning*, 95, 3: 257-260.
- OECD (2016), *OECD Science, Technology and Innovation Outlook 2016*, OECD Publishing, Paris.
- Oneto L., Ridella S., Anguita D. (2016), “Tuning the distribution dependent prior in the PAC-Bayes framework based on empirical data”, in *ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 27-29 April, Bruges (Belgium), testo disponibile al sito: <https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2016-46.pdf>, data di consultazione: 24 maggio 2017.
- Orlandi I., Oneto L., Anguita D. (2016), “Random Forests model selection”, in *ESANN 2016 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 27-29 April, Bruges (Belgium), testo disponibile al sito: <https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2016-48.pdf>, data di consultazione: 24 maggio 2017.
- Pan S.J., Yang Q. (2010), “A survey on transfer learning”, *IEEE Transactions on Knowledge and Data Engineering*, 22, 10: 1345-1359.
- Papamitsiou Z.K., Economides A.A. (2014), “Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence”, *Journal of Educational Technology & Society*, 17, 4: 49-64.
- Schapire R.E., Freund Y., Bartlett P., Lee W.S. (1998), “Boosting the margin: a new explanation for the effectiveness of voting methods”, *The Annals of Statistics*, 26, 5: 1651-1686.
- Shawe-Taylor J., Cristianini N. (2004), *Kernel Methods for Pattern Analysis*, Cambridge University Press, Cambridge.
- Shvachko K., Kuang H., Radia S., Chansler R. (2010), “The hadoop distributed file system”, in *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, May 3-7, Incline Village (NV): 1-10.
- Siri A. (2014), *Predicting Students’ Academic Dropout Using Artificial Neural Networks*, Nova Science Publishers, New York.
- Vapnik V.N. (1998), *Statistical Learning Theory*, Wiley, New York.
- White A.P., Liu W.Z. (1994), “Technical note: bias in information-based measures in decision tree induction”, *Machine Learning*, 15, 3: 321-329.
- Yoon H., Yang K., Shahabi C. (2005), “Feature subset selection and feature ranking for multivariate time series”, *IEEE Transactions on Knowledge and Data Engineering*, 17, 9: 1186-1198.
- Zaharia M., Chowdhury M., Das T., Dave A., Ma J., McCauley M., Franklin M.J., Shenker S., Stoica I. (2012), “Resilient distributed datasets: a fault-tolerant abstraction for in-memory cluster computing”, in *Proceedings of NSDI ’12: 9th USENIX Symposium on Networked Systems Design and Implementation*, April 25-27, San Jose (CA): 15-28.
-

16. Oltre l'effetto "in media": uno studio sulle prestazioni degli studenti nei test INVALSI utilizzando l'approccio quantile

Thinking beyond the "average case": exploring students' performance in INVALSI test through a quantile regression perspective

di Antonella Costanzo, Marta Desimoni

Nella ricerca educativa la questione legata all'eterogeneità degli effetti di fattori individuali e contestuali sugli apprendimenti rispetto ai livelli di abilità risulta di particolare interesse. In effetti, il background socio-economico, il genere, le differenze territoriali possono incidere in modo diverso sulle prestazioni degli studenti più bravi o con maggiori difficoltà di apprendimento. In tale contesto, un approccio basato sulla regressione quantile rappresenta un utile complemento alle analisi tradizionali in quanto offre la possibilità di una più completa caratterizzazione del ruolo di tali fattori sui risultati scolastici rispetto alle differenze nei livelli di performance degli studenti. Tale approccio consente di ricavare una maggiore quantità di informazioni utili non solo a individuare se i fattori potenzialmente correlati agli apprendimenti hanno mediamente un effetto, ma anche se essi pesano in modo diverso, per quali categorie di studenti e in che misura.

Questo lavoro si propone di illustrare i punti di forza del metodo quantile per esplorare le determinanti dei risultati scolastici degli studenti partecipanti alle Rilevazioni nazionali.

Over the last decades, in educational studies, there is a growing interest in exploring the heterogeneous effects of educational predictors affecting students' performance. For instance, the impact of gender stereotypes, regional disparities and socio-economic background could be different for different levels of students' ability, e.g. between low performing and high performing students. In this framework, quantile regression is a useful complement to standard analysis, as it offers a different perspective to investigate educational data particularly interesting for researchers and policymakers. As a matter of fact, it allows to explore in a more effective way potential inequalities in the access to education as well as peculiar patterns of the relationship between predictors affecting performance at different levels of students' attainment. Through an analysis of data collected in the National Annual Survey carried out by INVALSI, this paper illustrates the added value of quantile regression in order to obtain additional information concerning the impact of educational factors in driving students' performance.

1. Introduzione

Nella ricerca educativa nazionale e internazionale, l'eterogeneità degli effetti di fattori potenzialmente correlati agli apprendimenti e agli esiti scolastici rispetto ai livelli di competenza degli studenti riveste notevole importanza sia dal punto di vista concettuale sia da un punto di vista misuratorio. L'evidenza empirica suggerisce, infatti, che fattori individuali e contestuali, per esempio il background socio-economico e culturale, il genere, le differenze territoriali possono incidere in modo diverso sulle prestazioni degli studenti più bravi o per quelli con maggiori difficoltà di apprendimento (Hanushek e Woessmann, 2008).

Nella letteratura prevalente, gli approcci tradizionalmente impiegati per studiare la relazione tra diversi fattori e i risultati scolastici si basano sulla stima degli effetti in media (OECD, 2005 e 2007; Marks, Cresswell e Ainley, 2006) mediante l'utilizzo di un modello di regressione lineare classico (OLS, abbreviazione di *Ordinary Least Squares*). Tuttavia, la natura complessa degli apprendimenti e il legame articolato degli esiti scolastici con altre variabili rilevanti sono tali da rendere il ragionamento "in media" un'eccessiva semplificazione della realtà. Il rischio è quello di trascurare effetti

che possono essere differenti, sia in segno sia in forza, qualora l'interesse si sposti dalla media a posizioni più estreme della distribuzione della performance, dunque verso gli allievi con maggiori difficoltà di apprendimento o, al contrario, quando l'attenzione si sposta sulle eccellenze.

Di conseguenza, negli ultimi anni si riscontra la tendenza al completamento del quadro informativo risultante dall'utilizzo degli approcci tradizionali. Questo avviene mediante l'applicazione di modelli più flessibili che permettono di esplorare l'intera distribuzione delle prestazioni (Eide e Showalter, 1998; Tian, 2006; Ohinata e van Ours, 2013; Di Tommaso, Mendolia e Contini, 2016) tenendo conto, nella stima, della variabilità degli effetti in funzione del livello di abilità degli studenti. Tra questi, si annovera il modello di regressione quantile (Koenker e Basset, 1978) che offre la possibilità di una più completa caratterizzazione del ruolo di alcune variabili rispetto alle differenze nei livelli di performance degli studenti. In altre parole, tale approccio consente di ricavare una maggiore quantità di informazioni da sfruttare non solo per individuare se i fattori potenzialmente correlati agli apprendimenti hanno mediamente un effetto, ma anche per comprendere se essi pesano in modo diverso, per quali categorie di studenti e in che misura.

La possibilità di utilizzare approcci più puntuali alla valutazione del *pattern* di relazioni tra esiti scolastici e altri fattori risulta di estrema utilità per le possibili applicazioni dal punto di vista sociale, istituzionale ed educativo. In particolare, la scelta di ragionare in un'ottica quantile consente alle Istituzioni scolastiche di individuare in modo approfondito i fattori che sottostanno alla presenza di eccellenze o di carenze particolarmente pronunciate, e quindi di scegliere opportunamente quali provvedimenti eventualmente adottare non soltanto per raggiungere risultati mediamente migliori ma anche per conseguire obiettivi di maggiore equità nello sviluppo degli apprendimenti.

Alla luce di tali riflessioni, questo lavoro si propone un duplice obiettivo: da un lato, illustrare i principali punti di forza e le criticità legati all'utilizzo dell'approccio quantile per la valutazione delle relazioni tra fattori rilevanti e apprendimenti scolastici; dall'altro, sfruttando la strategia di analisi dei dati funzionale all'applicazione dell'approccio quantile, mettere in luce la ricchezza delle informazioni potenzialmente ricavabili dalle Rilevazioni nazionali che costituiscono una risorsa fondamentale per la ricerca in campo educativo.

Nel seguente paragrafo è descritto, negli aspetti metodologici essenziali, l'approccio basato sulla regressione quantile (QR). Successivamente tale metodo di analisi sarà applicato nell'ambito di uno studio basato sui dati relativi agli apprendimenti di un campione di studenti delle classi quinte di scuola primaria che hanno partecipato alle Rilevazioni nazionali dell'anno scolastico 2014-15 nell'ambito del Sistema nazionale di valutazione (INVALSI, 2015).

1.1. Il modello di regressione quantile nella valutazione degli apprendimenti: aspetti metodologici essenziali

Il modello di regressione quantile (QR) rappresenta un'estensione del modello classico di regressione (OLS) poiché consente di analizzare l'intera distribuzione condizionata della variabile di risposta attraverso la stima di tanti modelli separati per i differenti quantili di interesse (Hao e Naiman, 2007).

La scelta di utilizzare un modello quantile permette di fornire una più completa descrizione di come la distribuzione condizionata degli apprendimenti varia al variare delle caratteristiche individuali e contestuali non solo rispetto alla posizione o alla scala ma anche alla forma della distribuzione. In effetti, da un punto di vista strettamente metodologico, le stime ottenute con il metodo di regressione classico esprimono l'effetto in media delle variabili esplicative sulla performance e, pertanto, consentono di ragionare con riferimento al centro della distribuzione condizionata dei punteggi. D'altro canto, le stime quantili permettono di valutare l'impatto dei regressori per ogni quantile della distribuzione condizionata degli apprendimenti.

Da un punto di vista operativo, i livelli di abilità degli studenti per entrambe le discipline sono rappresentati dai quantili della distribuzione delle competenze. A tale proposito, si definiscono quantili di una variabile quantitativa degli indici di posizione che descrivono la distribuzione della variabile. Tali indici dividono la variabile in parti caratterizzate da pari numerosità e possono essere più o meno numerosi a seconda del livello di dettaglio desiderato (Piccolo, 2010).

Tra le altre proprietà (Hao e Naiman, 2007; Davino, Furno e Vistocco, 2013), la regressione quantile consente di ottenere stime robuste, non sensibili a *outliers* nei valori della variabile dipendente. In effetti, la proprietà della robustezza assicura che la stima quantile sia influenzata solamente dal comportamento locale della distribuzione condizionata della variabile di risposta in corrispondenza del quantile considerato (Koenker, 2005). Inoltre non sono necessarie l'ipotesi di normalità degli errori, che invece è tipica del modello classico di regressione, e l'ipotesi di omoschedasticità.

Il modello di regressione quantile per l' i -esimo quantile condizionato può essere espresso come segue:

$$Q_{y_i(\theta)|x_i} = \mathbf{x}_i^T \beta_\theta$$

dove y è la variabile dipendente, \mathbf{x}_i^T è il vettore $k \times 1$ delle variabili esplicative, β è il vettore dei coefficienti di regressione, θ è il quantile condizionato di interesse. Il vettore dei coefficienti β varia a seconda dei quantili scelti per la stima. È utile ribadire che le stime ottenute nell'ambito del modello QR sono semi-parametriche in quanto non è necessaria alcuna specificazione circa la forma della distribuzione degli errori, ma la relazione funzionale tra i punteggi e i predittori è di tipo lineare. Pertanto, in questo lavoro si fa riferimento a un approccio semi-parametrico.

In analogia con quanto accade nel caso del modello di regressione classica per la media condizionata, che può essere definita come soluzione al problema di minimizzazione della somma degli scarti al quadrato, è possibile definire ogni singolo quantile come la soluzione del seguente problema di minimo (Koenker, 2005):

$$\min_{\beta} \sum_{y_i \geq \mathbf{x}_i^T \beta} \theta |y_i - \mathbf{x}_i \beta| + (1 - \theta) \sum_{y_i < \mathbf{x}_i^T \beta} \theta |y_i - \mathbf{x}_i \beta|$$

Specificando $\theta = 0,5$ è possibile ottenere la regressione mediana. Per quanto riguarda l'interpretazione dei coefficienti associati ai diversi regressori, considerando un modello lineare, essa avviene allo stesso modo della regressione classica, per cui il coefficiente angolare β_θ rappresenta la misura dell'effetto che il cambiamento unitario del particolare regressore ha sulla variabile di risposta quando i valori di tutti gli altri regressori rimangono fissi. A differenza della regressione classica, in cui tale effetto riguarda la media, nel caso della regressione quantile è possibile misurare l'effetto di ciascun regressore su tutti i punti che definiscono la distribuzione della variabile di risposta.

2. Metodo

2.1. Partecipanti

In questo lavoro sono stati analizzati i dati raccolti dall'INVALSI nella Rilevazione nazionale degli apprendimenti dell'italiano e della matematica che si è svolta nell'anno scolastico 2014-15. In particolare, sono stati esaminati i dati degli studenti frequentanti il quinto anno di scuola primaria nelle classi delle cosiddette "scuole campione", in cui la somministrazione avviene alla presenza di un osservatore esterno (INVALSI, 2015). Dopo aver effettuato le operazioni preliminari di *data cleaning*, il campione oggetto di analisi è costituito da 13.744 studenti (7.033 maschi e 6.711 femmine) per l'italiano e da 14.282 studenti (7.328 maschi e 6.954 femmine) per la matematica.

3. Materiali e procedura

La prova INVALSI di italiano è composta da un testo narrativo e uno espositivo accompagnati, ciascuno, da quesiti di vario formato caratterizzati da diversi livelli di difficoltà e da una sezione di grammatica i cui quesiti sono ugualmente di vario formato e con diversi livelli di difficoltà.

La prova di matematica è composta da quesiti di vario formato, in particolare si trovano domande a risposta chiusa (scelta multipla semplice e complessa), domande a risposta aperta e *cloze* che richiedono il completamento di frasi, calcoli o espressioni mediante l'utilizzo di elementi forniti dal testo della prova o dall'alunno stesso. I quesiti afferiscono sia alla dimensione dei contenuti inerenti ai diversi ambiti matematici cui le domande fanno riferimento, sia alla dimensione cognitiva riguardante i processi che gli allievi attivano quando rispondono ai quesiti. In generale, in entrambe le prove di italiano e di matematica, i quesiti sono costruiti in base a dei Quadri di riferimento (INVALSI, 2012 e 2013) in cui si esplicitano, per ciascuna disciplina, i punti di riferimento concettuali e i criteri operativi da utilizzare nella costruzione delle prove.

La procedura di somministrazione prevede che, nella scuola primaria, vista l'età degli alunni coinvolti, le prove si svolgano in due giornate distinte al fine di evitare l'effetto affaticamento. Sia per l'italiano sia per la matematica, il tempo massimo per lo svolgimento delle prove è di 75 minuti.

4. Risultati

4.1. Le caratteristiche del campione

Nella tab. 1 e nella tab. 2 sono riportate le principali statistiche di sintesi relative alle prestazioni degli studenti nelle prove INVALSI di italiano e matematica, sia nel campione generale sia rispetto alle variabili considerate nel lavoro. I punteggi degli allievi sono a media 200 e deviazione standard 40.

Si è scelto di utilizzare, per ogni disciplina, la stima del livello di abilità degli studenti secondo il modello di Rasch (1960 e 1980). In accordo con la letteratura prevalente, sono stati considerati il genere, la condizione socio-economica e culturale espressa dall'indice ESCS¹ (per maggiori dettagli, si veda Campodifiori *et al.*, 2010), la cittadinanza (italiano, straniero), la regolarità negli studi (studente regolare, anticipatorio o posticipatorio) e le informazioni connesse alla frequenza della scuola dell'infanzia (sì, no) come variabili rappresentative delle caratteristiche individuali degli studenti. Inoltre, l'area geografica di riferimento (Nord, Centro e Sud) e il bacino di utenza socio-economico delle scuole (Oliveira e Santos, 2005) rappresentato dall'indicatore di ESCS a livello di scuola sono stati individuati come fattori di contesto potenzialmente in grado di influenzare i risultati scolastici.

I dati relativi agli apprendimenti rispetto alle variabili analizzate si trovano nel *Rapporto risultati* a cura dell'INVALSI per l'anno di riferimento (INVALSI, 2015) e, quindi, ulteriori approfondimenti esulano dallo scopo del presente lavoro. Tuttavia, alcuni aspetti specifici relativi al campione in esame sono stati considerati per poter apprezzare i punti di forza derivanti dall'applicazione del modello di regressione quantile.

Tab.1 – Statistiche di sintesi dei punteggi in italiano nel campione complessivo e rispetto alle categorie delle variabili considerate

	μ	σ	Min	$\theta_{.25}$	$\theta_{.5}$	$\theta_{.75}$	Max
Maschio	199,13	40,60	26,52	171,11	199,40	224,42	343,19
Femmina	201,79	39,07	26,52	175,86	204,20	229,87	343,19
Nativo	201,60	39,62	50,75	175,86	204,20	229,87	343,19
Straniero	184,32	40,05	26,52	156,45	185,26	209,08	343,19
Regolare	200,74	39,73	26,52	175,86	199,40	224,42	343,19
Anticipatorio	204,70	38,88	100,10	180,60	204,20	229,90	343,20
Posticipatorio	175,37	43,02	50,75	146,07	173,49	204,20	289,89
Sc_materna:sì	200,98	39,78	26,52	175,86	199,40	224,42	343,19
Sc_materna:no	194,98	40,47	50,75	166,31	194,66	224,42	319,13
Nord	203,66	37,77	50,75	180,57	204,20	229,87	343,19
Centro	203,97	39,43	26,52	180,57	204,20	229,87	343,19
Sud	196,38	41,14	50,75	166,31	194,66	224,42	343,19
Totale campione	200,40	39,88	26,52	175,90	199,40	224,40	343,20

¹ L'indice ESCS è basato su informazioni relative ai beni strumentali e culturali presenti in casa dello studente, al titolo di studio conseguito dai genitori e alla professione da loro svolta.

Tab. 2 – Statistiche di sintesi dei punteggi in matematica nel campione complessivo e rispetto alle categorie delle variabili considerate

	μ	σ	Min	$\theta_{.25}$	$\theta_{.5}$	$\theta_{.75}$	Max
Maschio	203,55	41,01	56,45	176,48	205,31	232,22	323,16
Femmina	196,52	37,90	56,45	172,23	197,05	222,69	323,16
Nativo	201,09	39,60	56,45	172,23	201,16	227,36	323,16
Straniero	187,80	38,72	76,90	161,20	184,80	213,80	323,20
Regolare	200,51	39,55	56,45	172,23	201,16	227,36	323,16
Anticipatario	201,30	35,50	119,1	180,70	197,00	222,70	323,20
Posticipatario	175,67	43,28	56,45	143,98	176,48	205,31	323,16
Sc_materna:si	200,38	39,55	56,45	172,23	201,16	227,36	323,16
Sc_materna:no	197,71	40,86	56,45	167,91	197,71	227,36	323,16
Nord	203,48	38,87	56,45	176,48	205,31	222,36	323,16
Centro	203,27	39,47	56,45	176,48	205,31	227,36	323,16
Sud	196,04	39,99	56,45	167,91	197,05	222,69	323,16
Totale campione	200,10	39,67	56,45	172,20	201,20	227,40	323,20

Il primo di essi è la caratterizzazione delle distribuzioni dei punteggi in termini di asimmetria e di curtosi. In particolare, mentre gli indici di asimmetria indicano un sostanziale adattamento alla distribuzione normale (indice di asimmetria pari a 0,01 per italiano e 0,07 per matematica), la curtosi indica un certo spessore nelle code (curtosi = 3,22 per italiano e 3,02 per matematica) evidenziando l'opportunità di applicare anche un approccio più flessibile alla valutazione della relazione tra i fattori e gli esiti nelle prove INVALSI che tenga conto dell'intera distribuzione degli apprendimenti.

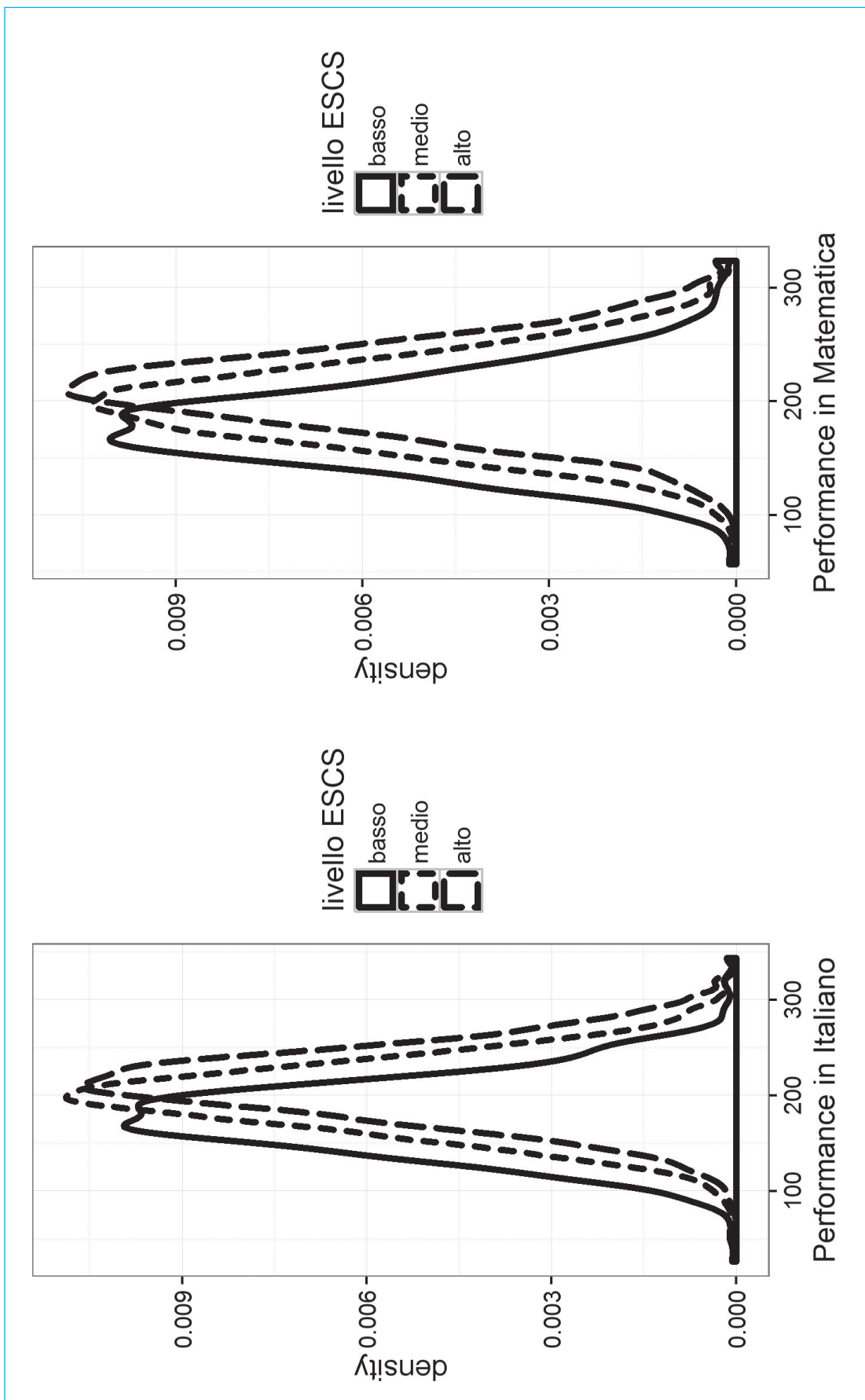
Ulteriori spunti di riflessione circa la forma della distribuzione derivano dal *pattern* della performance in funzione di alcune delle variabili considerate. Per esempio, nella fig. 1 sono riportate le distribuzioni dei punteggi nelle due discipline rispetto ai livelli dello status socio-economico e culturale (ESCS) degli studenti.

Il background familiare rappresenta una variabile interessante nella valutazione della performance degli allievi (Tian, 2006; OECD, 2007; Giambona e Porcu, 2015). A tale proposito, la fig. 1, da un punto di vista descrittivo, mostra che l'indicatore ESCS è in generale positivamente associato con i risultati conseguiti in entrambe le discipline: gli allievi con un background familiare più elevato ottengono, per tutti i livelli di abilità considerati, punteggi più alti rispetto a coloro che si trovano in condizioni di maggiore difficoltà. Se si presta attenzione alla forma della distribuzione, il *pattern* dei punteggi conseguiti dagli studenti cambia per ogni livello di ESCS.

Nelle ricerche sulle differenze individuali negli apprendimenti, una delle variabili su cui si sono focalizzati numerosi studi è il genere degli allievi (Di Tommaso, Mendolia e Contini, 2016; McCullough, 2016). A scopo descrittivo, la fig. 2 riporta i punteggi conseguiti dai maschi e dalle femmine del campione considerato nelle due discipline ed è interessante notare le differenze distributive ascrivibili all'essere femmina soprattutto in matematica e in corrispondenza di valori più estremi (coda a destra) del *pattern* della performance.

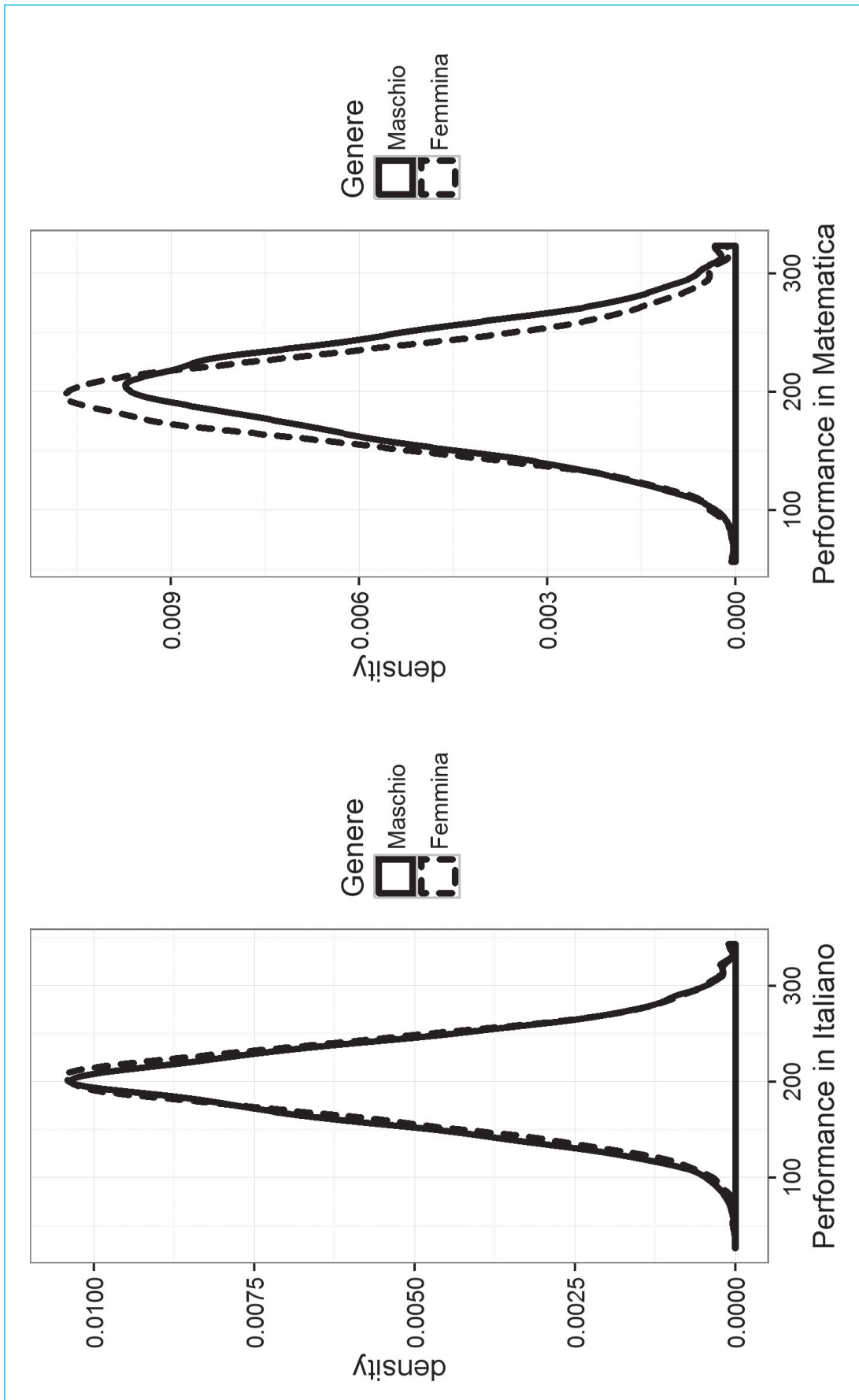
L'approccio statistico per la rappresentazione dei dati fin qui utilizzato consente di tracciare un quadro preliminare di natura descrittiva dei diversi fattori e dei risultati scolastici degli allievi. Il passo successivo sarà quello di stimare l'incidenza e la significatività statistica del contributo di tali fattori alla performance osservata sia in termini di effetto in media sia tenendo conto dei diversi livelli di abilità degli studenti.

Fig. 1 – Distribuzione dei punteggi in italiano e in matematica rispetto alla condizione socio-economica e culturale



Nota: La distribuzione dell'ESCS è stata suddivisa in tre classi di uguale ampiezza indicative del background familiare degli studenti, tali che $(-2,79, -1,1]$ rappresenta il livello basso, $(-1,1, 0,583]$ il livello medio, $(0,583, 2,27]$ rappresenta il livello alto.

Fig. 2 – Distribuzione dei punteggi in italiano e in matematica rispetto al genere



4.2. Le determinanti delle prestazioni degli studenti in italiano e in matematica: effetti rispetto ai livelli di abilità

Gli effetti² del genere, della cittadinanza, del background socio-economico individuale e di scuola, della regolarità, della frequenza della scuola materna e dell'area geografica di riferimento sulle prestazioni degli studenti per entrambe le materie vengono stimati considerando sia il modello OLS (Gujarati, 2003) sia il modello QR per il quale si è ritenuto utile selezionare cinque differenti quantili rappresentativi dei livelli di apprendimento. Tale approccio permette di tener conto dell'eterogeneità dell'impatto dei predittori prescelti sugli esiti in funzione dei diversi livelli di competenze. Le tabb. 3 e 4 riportano le stime OLS e QR risultanti dai due approcci; le figg. 3 e 4 illustrano le stime dell'impatto di ciascuna variabile sulle prestazioni in italiano e in matematica, rispettivamente ottenute utilizzando entrambi gli approcci. Per ciascun predittore, la linea orizzontale tratteggiata rappresenta la stima dell'effetto in media (OLS) mentre la linea spezzata corrisponde all'effetto stimato mediante il metodo QR sui quantili condizionati della variabile di risposta.

Tab. 3 – Effetti dei fattori individuali e contestuali sugli apprendimenti in italiano: stime in media (OLS) e per i cinque quantili di interesse (QR) rappresentativi dei livelli di abilità

	OLS	QR				
		$\theta_{.1}$	$\theta_{.25}$	$\theta_{.5}$	$\theta_{.75}$	$\theta_{.9}$
Intercept	198,95	152,55	172,92	200,91	224,09	246,78
Femmina	2,75	4,72	4,62	2,06	1,74	1,90
Anticipatario	0,34	1,40	2,29	-0,50	-2,09	-2,88
Posticipatario	-13,93	-19,72	-16,19	-18,23	-9,40	-4,52
ESCS	10,92	11,51	11,14	10,93	10,83	11,13
Sc_materna:si	2,71	3,05	4,93	1,34	0,62	0,26
Straniero	-10,90	-11,52	-11,89	-11,83	-8,19	-10,61
Centro	-0,30	-3,59	-1,74	-0,25	1,11	1,43
Sud	-6,00	-10,97	-9,42	-6,86	-2,92	-1,76
ESCS scuola	1,15	-0,45	0,76	2,00	2,93	0,37

Nota: Risultati statisticamente significativi (p -value < 0,05) in grassetto. Errori standard delle stime QR calcolati con metodo *bootstrap* (package: quantreg, www.r-project.org).

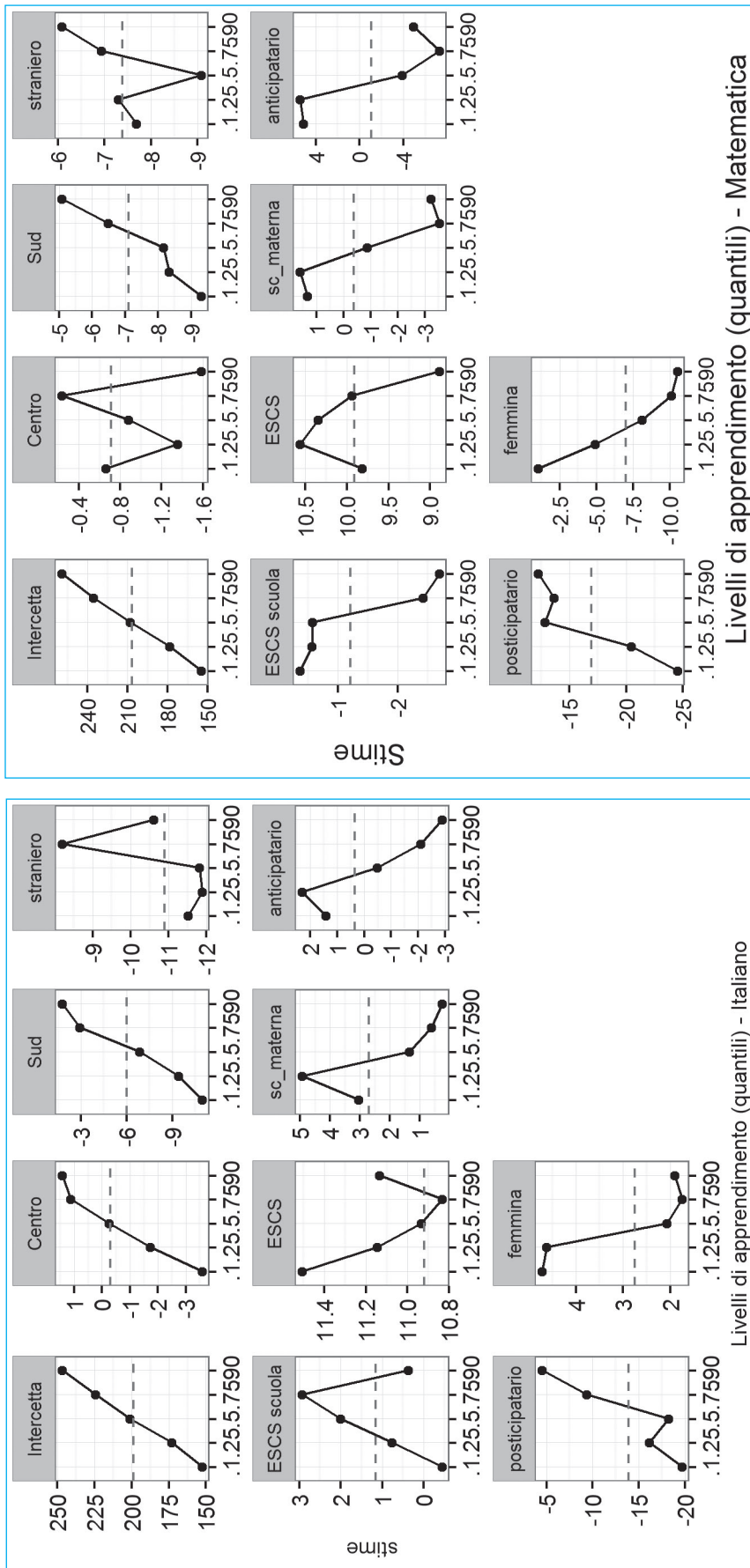
Tab. 4 – Effetti dei fattori individuali e contestuali sugli apprendimenti in matematica: stime in media (OLS) e per i cinque quantili di interesse (QR) rappresentativi dei livelli di abilità

	OLS	QR				
		$\theta_{.1}$	$\theta_{.25}$	$\theta_{.5}$	$\theta_{.75}$	$\theta_{.9}$
Intercept	206,57	154,80	178,48	207,87	235,56	259,23
Femmina	-7,02	-1,04	-4,935	-8,11	-10,13	-10,55
Anticipatario	-1,06	5,11	5,44	-3,93	-7,31	-4,94
Posticipatario	-16,95	-24,55	-20,49	-12,86	-13,65	-12,25
ESCS	9,91	9,82	10,57	10,34	9,94	8,89
Sc_materna:si	-0,37	1,33	1,60	-0,87	-3,53	-3,23
Straniero	-7,38	-7,69	-7,30	-9,08	-6,93	-6,08
Centro	-0,71	-0,66	-1,35	-0,88	-0,23	-1,59
Sud	-7,08	-9,31	-8,34	-8,17	-6,50	-5,08
ESCS scuola	0,88	-0,36	-0,56	-0,57	-2,42	-2,701

Nota: Risultati statisticamente significativi (p -value < 0,05) in grassetto. Errori standard delle stime QR calcolati con metodo *bootstrap* (package: quantreg, www.r-project.org).

² È importante sottolineare la natura associazionale e non causale dei modelli statistici utilizzati. In particolare, gli effetti delle variabili esplicative sugli esiti devono essere interpretati come la forza con la quale tali predittori si associano ai risultati scolastici e non in termini di relazione causa-effetto sugli esiti medesimi (Ricci, 2008).

Fig. 3-4 – Stime OLS e QR degli effetti dei fattori individuali e contestuali sugli apprendimenti in italiano e in matematica



Nota: Sull'asse delle ascisse sono riportati i quantili mentre sull'asse delle ordinate i coefficienti stimati della regressione secondo i due approcci.

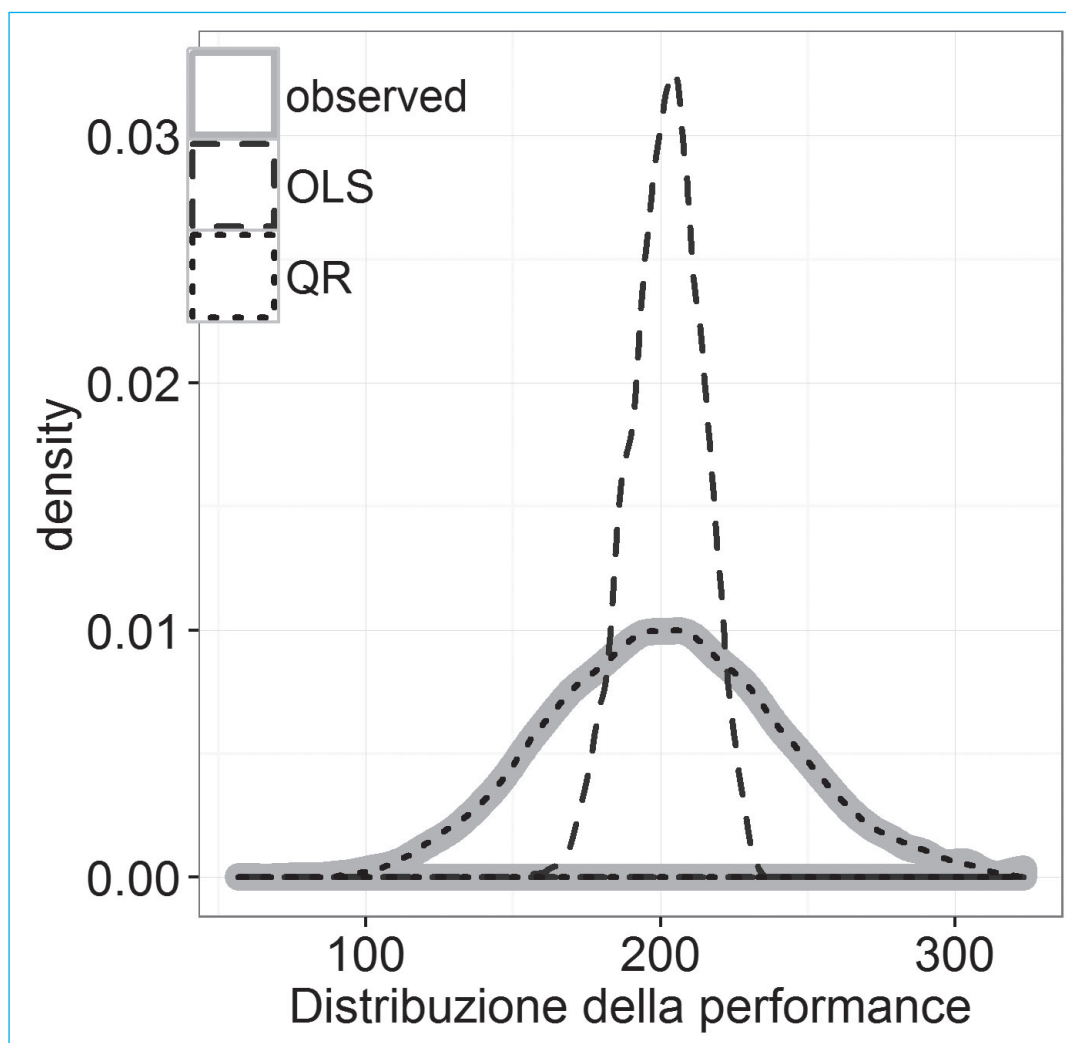
Le stime ottenute dall'applicazione dell'approccio OLS sono in linea con quanto emerge dagli studi sugli apprendimenti realizzati con i dati delle Rilevazioni nazionali INVALSI (Cipollone, Montanaro e Sestito, 2009; Montanaro e Sestito, 2014; INVALSI, 2015) relativamente al livello scolastico considerato e controllando per le variabili di contesto utilizzate.

I risultati ottenuti mostrano che, sebbene particolarmente utile a una prima disamina circa l'impatto delle variabili sui risultati scolastici degli allievi, la scelta di ragionare "in media", mediante l'utilizzo di un approccio basato sulla regressione lineare classica, può comportare un'eccessiva semplificazione della realtà con il rischio di trascurare effetti che possono essere differenti (sia in segno sia in forza) qualora l'interesse si sposti dalla media a posizioni più estreme della variabile di risposta.

Queste considerazioni sono in linea con quanto discusso in letteratura a proposito dell'importanza di un approccio più flessibile alla valutazione degli apprendimenti (Tian, 2006; Schnepf, 2008; Fryer e Levitt, 2010; Robinson e Lubien-sky, 2011; European Commission, 2012; Cobb-Clark e Moschion, 2015; McCullough, 2016; Giambona e Porcu, 2015; Gursakal, Murat e Gursakal, 2016).

Al fine di apprezzare il vantaggio informativo derivante dalle stime QR nello studio degli apprendimenti, la fig. 5 riporta la distribuzione osservata delle prestazioni in matematica, la distribuzione stimata utilizzando il modello quantile e quella ottenuta a partire dalle stime OLS del modello classico.

Fig. 5 – Distribuzione della performance in matematica osservata degli studenti (linea grigia continua) e della performance stimata utilizzando l'approccio classico di regressione (OLS, linea tratteggiata) e l'approccio basato sulla regressione quantile (QR, linea puntinata)



Nota: Il grafico si riferisce agli apprendimenti in matematica, tuttavia è possibile osservare un simile andamento anche per la distribuzione delle prestazioni in italiano.

Come è possibile notare, la distribuzione stimata della variabile di risposta ottenuta utilizzando l'approccio QR risulta praticamente equivalente alla distribuzione osservata delle prestazioni; d'altro canto la distribuzione dei punteggi stimata a partire dal modello OLS costituisce un'approssimazione poiché tiene conto solo degli effetti sulla media condizionata.

5. Discussione

La lettura congiunta dei risultati derivanti dall'applicazione del modello classico di regressione OLS e dell'approccio quantile consentono di ottenere un quadro maggiormente informativo delle relazioni tra i diversi predittori e gli apprendimenti degli studenti.

Dalle stime OLS si evince che, in media, l'impatto del genere è significativo in entrambe le discipline. Tuttavia, il segno e l'entità della relazione con il punteggio varia a seconda della disciplina: le studentesse mostrano migliori prestazioni in italiano con uno svantaggio invece evidente nel caso delle prestazioni in matematica.

I risultati della QR mostrano che le differenze di apprendimento legate al genere si caratterizzano per una maggiore eterogeneità rispetto a quanto stimato in media. Per ogni livello di abilità, le bambine ottengono risultati migliori dei loro coetanei in italiano e viceversa in matematica; tuttavia il vantaggio delle femmine sulle prestazioni in italiano tende a essere maggiore in corrispondenza di quantili più bassi della distribuzione della performance, quindi tra gli allievi con maggiori difficoltà di apprendimento. Per gruppi di studenti con livelli di prestazione più elevati l'impatto del genere non risulta statisticamente significativo. In matematica, lo svantaggio legato all'essere femmina appare maggiormente evidente tra gli allievi più bravi. Le informazioni relative alla prevalenza delle differenze di genere in corrispondenza di taluni livelli di apprendimento a partire dalla scuola primaria consentono ai docenti, alla scuola e agli studiosi di approfondire le motivazioni alla base del gap tra maschi e femmine, monitorare in che misura esso è particolarmente pronunciato rispetto ai livelli di competenza allo scopo di trovare soluzioni adeguate a una maggiore armonizzazione nello sviluppo degli apprendimenti.

La frequenza della scuola dell'infanzia è un fattore che ha un impatto positivo e statisticamente significativo sulle prestazioni. Dalle stime QR emerge che, a parità degli altri fattori, aver frequentato la scuola materna incide positivamente sulle prestazioni in italiano e che l'intensità della relazione è maggiore per gli allievi più deboli; d'altro canto l'effetto non è significativo per gli apprendimenti in matematica per tutti i livelli di abilità considerati.

Il gap nei risultati scolastici tra allievi stranieri e italiani risulta statisticamente significativo in media e per ogni livello di abilità sia in italiano sia in matematica. Le stime QR suggeriscono che, a differenza dell'effetto in media, lo svantaggio degli stranieri rispetto ai nativi è maggiore per gli allievi con più elevate difficoltà di apprendimento. Tale effetto diviene meno incisivo quando si considerano i quantili rappresentativi degli studenti con livelli di abilità più elevati.

Il background socio-economico e culturale ha un peso rilevante sui risultati scolastici: gli studenti con uno status familiare più elevato hanno, in media, esiti di apprendimento migliori. Le corrispondenti stime quantili indicano, tuttavia, che il coefficiente associato all'indicatore ESCS degli studenti è positivo ma la sua intensità è variabile in funzione dei livelli di abilità e varia a seconda della materia. In particolare, il background socio-economico assume maggiore rilevanza in corrispondenza delle code della distribuzione delle prestazioni in italiano: per gli studenti più deboli, che sono spesso quelli con problemi di accesso alle risorse educative, e per gli studenti molto bravi che beneficiano di migliori standard di vita e, quindi, hanno maggiori possibilità di sviluppo del loro capitale umano (Sen, 2005). Sulle prestazioni in matematica, l'effetto dell'ESCS è significativo e più ampio solo per gli studenti con maggiori difficoltà di apprendimento.

L'analisi della relazione tra la condizione socio-economica degli studenti e i risultati scolastici tenendo conto dei diversi livelli di competenze può essere di particolare utilità per le Istituzioni scolastiche poiché permette loro di individuare eventuali fenomeni di disuguaglianza, specie con riferimento agli allievi con maggiori difficoltà di apprendimento e di predisporre percorsi funzionali al recupero di situazioni complesse sul piano economico sociale e delle competenze (Coleman *et al.*, 1966; OECD, 2012; Giambona e Porcu, 2015). Infine, si osserva che, nel campione analizzato, l'indicatore ESCS a livello di scuola non incide sulla performance degli allievi in media e per tutti i livelli di abilità considerati.

Per quanto riguarda la regolarità negli studi, essere studenti anticipatori nella scuola primaria, in media, non sembra incidere significativamente sulle prestazioni in entrambe le discipline; d'altro canto, come è ragionevole

attendersi, il ritardo scolastico pesa negativamente sui risultati ottenuti. L'approccio quantile offre la possibilità di una maggiore caratterizzazione dell'impatto della scolarizzazione anticipata sulla performance scolastica. A tale proposito, mentre in italiano permane un effetto non significativo per tutti i livelli di abilità, in matematica le stime QR suggeriscono che per gli allievi con livelli di competenza più elevati, essere anticipatorio comporta uno svantaggio significativo in termini di risultati scolastici. Sarebbe utile indagare se la scelta di un accesso anticipato alla scuola primaria dei bambini che spesso mostrano elevati livelli di performance sia effettivamente funzionale allo sviluppo delle competenze in questa materia. D'altra parte, nella riflessione sull'impatto della scolarizzazione anticipata sugli apprendimenti, è importante tener conto dell'interazione con gli aspetti territoriali dal momento che, negli ultimi anni, il fenomeno di accesso anticipato alla scuola primaria è in aumento soprattutto nelle regioni del Sud Italia. In particolare, nel campione analizzato circa il 65% degli studenti anticipatori è del Sud, il 27% e il 6,5% sono del Centro e del Nord, rispettivamente. È verosimile che nelle regioni del Sud Italia la scuola primaria venga talvolta considerata come un'opportunità per accogliere bambini provenienti soprattutto da realtà familiari e contestuali molto difficili.

Infine, le differenze territoriali hanno un impatto sulle prestazioni in entrambe le discipline a parità degli altri fattori: gli studenti del Mezzogiorno ottengono risultati peggiori degli studenti del Nord e l'area geografica ha un peso maggiore sulle prestazioni in italiano. Tuttavia, l'effetto negativo del contesto territoriale è meno incisivo all'aumentare dei livelli di abilità, e questo *pattern* riguarda sia gli apprendimenti in italiano sia quelli in matematica.

6. Conclusioni

Sulla base dei dati analizzati in questo studio è possibile concludere che il metodo della regressione quantile per l'analisi della performance degli studenti rappresenta un utile complemento agli approcci tradizionalmente impiegati per la valutazione degli apprendimenti che, ragionando sugli effetti delle variabili in media, prescindono dalle differenze nei differenti livelli di abilità. Naturalmente tale approccio non deve essere considerato in antitesi alle tecniche tradizionalmente usate in tali contesti quanto piuttosto complementare a esse, poiché consente di esplorare caratteristiche del fenomeno che potrebbero, altrimenti, essere trascurate.

La lettura congiunta dei risultati alle prove INVALSI derivanti dall'applicazione di entrambi i modelli OLS e QR ha permesso di ottenere un quadro maggiormente informativo delle relazioni tra i diversi predittori e gli apprendimenti degli studenti evidenziando la ricchezza dei dati delle Rilevazioni nazionali INVALSI come strumento per la ricerca in campo educativo.

Riferimenti bibliografici

- Agasisti T. (2011), "Does competition affect schools' performance? Evidence from Italy through OECD-PISA data", *European Journal of Education*, 46, 4: 549-565.
- Cade B.S., Noon B.R. (2003), "A gentle introduction to quantile regression for ecologists", *Frontiers in Ecology and the Environment*, 1, 8: 412-420.
- Campodifiori E., Figura E., Papini M., Ricci R. (2010), "Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia", *INVALSI Working Papers*, 2, testo disponibile al sito: http://www.invalsi.it/download/wp/wp02_Ricci.pdf, data di consultazione: 3 aprile 2017.
- Cipollone P., Montanaro P., Sestito P. (2009), *I divari territoriali nel capitale umano e nella qualità dell'istruzione*, Banca d'Italia, mimeo.
- Cipollone P., Sestito P. (2007), *Quanto imparano gli studenti italiani: i divari Nord-Sud*, Banca d'Italia, mimeo.
- Cobb-Clark D., Moschion J. (2015), "Gender gaps in early educational achievement", *IZA Discussion Paper*, 9535, testo disponibile al sito: <http://ftp.iza.org/dp9535.pdf>, data di consultazione: 3 aprile 2017.
- Coleman J.S., Campbell E.Q., Hobson C.J., McPartland J., Mood A.M., Weinfeld F.D., York R.L. (1966), *Equality of Educational Opportunity*, US Department of Health, Education, and Welfare, Office of Education, Washington (DC).
- Davino C., Furno M., Vistocco D. (2013), *Quantile Regression: Theory and Applications*, Wiley, Chichester.
- Davino C., Vistocco D. (2008), "Quantile regression for the evaluation of student satisfaction", *Statistica applicata*, 20, 3-4: 179-196.

-
- Di Tommaso M.L., Mendolia S., Contini D. (2016), "The gender gap in mathematics achievement: evidence from Italian data", *IZA Discussion Paper*, 10053, testo disponibile al sito: <http://ssrn.com/abstract=2810464>, data di consultazione: 5 aprile 2017.
- Eide E., Showalter J. (1998), "The effect of school quality on student performance: a quantile regression approach", *Economics Letters*, 5: 345-350.
- European Commission (2012), *Structural Change in Research Institutions: Enhancing Excellence, Gender Equality and Efficiency in Research and Innovation*, Publications Office of the European Union, Luxembourg.
- Fryer R.G., Levitt S.D. (2010), "An empirical analysis of the gender gap in mathematics", *American Economic Journal: Applied Economics*, 2, 2: 210-240.
- Giambona F., Porcu M. (2015), "Student background determinants of reading achievement in Italy. A quantile regression analysis", *International Journal of Educational Development*, 44: 95-107.
- Gujarati D.N. (2003), *Basic Econometrics*, McGraw Hill, Boston.
- Gursakal S., Murat D., Gursakal N. (2016), "Assessment of PISA 2012 results with quantile regression analysis within the context of inequality in educational opportunity", *Alphanumeric Journal*, 4, 2: 41-54.
- Halle T.G., Kurtz-Costes B., Mahoney J.L. (1997), "Family influences on school achievement in low-income, African American children", *Journal of Educational Psychology*, 89, 3: 527-537.
- Hanushek E.A., Woessmann L. (2008), "The role of cognitive skills in economic development", *Journal of Economic Literature*, 46, 3: 607-668.
- Hao L., Naiman D.Q. (2007), *Quantile Regression*, Sage, Thousand Oaks.
- INVALSI (2012), *Quadro di riferimento primo ciclo di istruzione prova di matematica*, testo disponibile al sito: http://www.invalsi.it/snv2012/documenti/QDR/QdR_Mat_I_ciclo.pdf, data di consultazione: 5 aprile 2017.
- INVALSI (2013), *Quadro di riferimento della prova di italiano*, testo disponibile al sito: https://invalsi-areaprove.cineca.it/docs/file/QdR_Italiano_Obligo_Istruzione.pdf, data di consultazione 5 aprile 2017.
- INVALSI (2015), *Rilevazioni nazionali degli apprendimenti 2014-15*, testo disponibile al sito: https://invalsi-areaprove.cineca.it/docs/attach/035_Rapporto_Prove_INVALSI_2015.pdf, data di consultazione 5 aprile 2017.
- ISTAT (2012), "La scuola e le attività educative", *Statistiche Report annuali*, testo disponibile al sito: <http://www.istat.it/it/archivio/71706>, data di consultazione: 3 aprile 2017.
- Koenker R. (2005), *Quantile Regression*, Cambridge University Press, Cambridge.
- Koenker R., Basset G. (1978), "Regression quantiles", *Econometrica*, 46, 1: 33-50.
- Maragos E.K., Despotis D.K. (2004), "Evaluating school performance over time in the frame of regional socio-economic specificities", *WSEAS Transactions on Mathematics*, 3, 3: 664-670.
- Marks G.N., Cresswell J., Ainley J. (2006), "Explaining socioeconomic inequalities in student achievement: the role of home and school factors", *Educational Research and Evaluation*, 12, 2: 105-128.
- McCullough S. (2016), "Gender lessons: patriarchy, sextyping & schools", *Gender and Education*, 28, 5:703-705.
- Montanaro P., Sestito P. (2014), "La qualità dell'istruzione italiana: un confronto tra PISA e le Rilevazioni nazionali dell'INVALSI", *Bank of Italy Occasional Paper*, 218, testo disponibile al sito: <https://ssrn.com/abstract=2489887>, data di consultazione: 2 maggio 2017.
- OECD (2005), *School Factors Related to Quality and Equity. Results from PISA 2000*, OECD Publishing, Paris.
- OECD (2007), *Education at a Glance 2007: OECD Indicators*, OECD Publishing, Paris.
- OECD (2012), *Low-Performing Students: Why They Fall Behind and How to Help Them Succeed*, OECD Publishing, Paris.
- Ohinata A., van Ours J.C. (2013), "Spillover effects of studying with immigrant students: a quantile regression approach", *IZA Discussion Paper*, 7720, testo disponibile al sito: <http://ftp.iza.org/dp7720.pdf>, data di consultazione: 2 maggio 2017.
- Oliveira M.A., Santos C. (2005), "Assessing school efficiency in Portugal using FDH and bootstrapping", *Applied Economics*, 37, 8: 957-968.
- Piccolo D. (2010), *Statistica*, il Mulino, Bologna.
- Rasch G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Danmarks Paedagogiske Institut, Copenhagen.
- Rasch G. (1980), *Probabilistic Models for Some Intelligence and Attainment Tests*, University of Chicago Press, Chicago, expanded ed.
- Ricci R. (2008), "La misurazione del valore aggiunto nella scuola", *Programma Education, FGA Working Paper*, 9.
- Robinson J.P., Lubiensky S.T. (2011), "The development of gender achievement gaps in mathematics and reading during elementary and middle school: examining direct cognitive assessments and teacher ratings", *American Educational Research Journal*, 48, 2: 268-302.
- Schnepf S.V. (2008), "Inequality of learning amongst immigrant children in industrialized countries", *IZA Discussion Paper*, 3337, testo disponibile al sito: <http://ssrn.com/abstract=1135913>, data di consultazione: 2 maggio 2017.
- Sen A. (2005), "Human rights and capabilities", *Journal of Human Development*, 6, 2: 151-166.
- Tian M. (2006), "A quantile regression analysis of family background factor effects on mathematical achievement", *Journal of Data Science*, 4: 461-478.
-

17. Non proprio la stessa scuola.

Segregazione degli insegnanti tra scuole e abbinamento insegnanti-studenti come meccanismi nascosti di disuguaglianza nel sistema scolastico italiano

Not really the same school.

Teachers inter-school segregation and teachers-students matching as hidden inequality mechanisms in the Italian school system

di Gianluca Argentin, Giovanni Abbiati, Tiziano Gerosa

Questo contributo si propone di verificare se esista, in Italia, una relazione sistematica tra gli insegnanti migliori e gli studenti con più elevate origini sociali. Nonostante il sistema scolastico italiano conservi un forte impianto centralizzato, volto in principio a garantire le stesse opportunità educative per tutti gli alunni del Paese, esso mostra forti fenomeni di segregazione tra studenti di diversa origine sociale. Questi fenomeni coinvolgono in primo luogo le famiglie, ma nemmeno gli insegnanti sono esenti da processi di auto-segregazione, come mostrano le ricerche condotte recentemente su dati PISA e INVALSI. Si tratta di tendenze che di per sé non dimostrano l'emergere di fenomeni di disuguaglianza nell'abbinamento tra insegnanti e studenti e, fino a oggi, questa pista di ricerca in Italia non è stata percorribile per la mancanza di basi-dati adeguate. A partire dal 2012, grazie alla contemporaneità delle rilevazioni SNV sugli studenti con la somministrazione del Questionario insegnante condotte dall'INVALSI, è stato possibile esplorare più in profondità tale fenomeno. Questo lavoro sfrutta le basi-dati INVALSI raccolte su insegnanti e studenti (campione SNV e Prova nazionale) nell'anno scolastico 2013-14, per un totale di circa 10.000 insegnanti e 100.000 studenti. Sono state individuate alcune caratteristiche collegate alla qualità dell'insegnamento: la precarietà del contratto; il radicamento all'interno della scuola; il possesso di una laurea (per le scuole elementari); il possesso di una laurea nella materia di insegnamento; il voto di laurea. Come caratteristica di stratificazione degli studenti è stato scelto il livello di istruzione dei genitori. L'associazione tra i vari indicatori di qualità dell'insegnante e le origini dello studente è stimata tramite modelli di regressione di probabilità lineare. Per verificare se le associazioni esistenti tra variabili non siano il riflesso di vincoli per gli insegnanti nella scelta delle scuole o della struttura a indirizzi del sistema scolastico, nei modelli utilizzati vengono impiegate variabili di controllo di tipo geografico e, nelle scuole superiori, l'indirizzo scolastico. Le analisi mostrano l'esistenza di un livello di segregazione crescente passando dalle scuole elementari alle scuole superiori, particolarmente evidente per gli insegnanti di italiano. Nelle scuole superiori una parte consistente del fenomeno è imputabile alla divisione delle scuole per indirizzi, che accentua fortemente i livelli segregazione studentesca già presenti nel nostro Paese e, tramite i meccanismi di reclutamento, acuisce l'auto-segregazione degli insegnanti. Questi risultati segnalano, una volta di più, come l'ordinamento del nostro sistema scolastico agisca come un meccanismo istituzionale di rafforzamento delle disuguaglianze di origine sociale.

The present chapter explores the phenomenon of teacher-students matching in Italy, in order to evaluate whether teachers' quality predictors are unevenly associated with the social background of their students. Despite the Italian school system has a centralized structure which guarantees, at least in principle, the equality of opportunities for all the students, recent research showed that it seems far from being egalitarian: huge gaps in the quality of education have been found between different areas of the country and also between schools. Unfortunately, due to a general lack of data, between-school differences attributable to teaching quality have been seldom inspected in previous research. The few available evidences show that schools characterized by higher shares of failing and/or lower background students are staffed with teachers normally trying to go away, and this is likely to influence the quality of their professional performances. To give a complete account of the phenomenon, then, it becomes crucial to understand how "high quality teachers" are distributed into the school system.

This paper explores the issue of teacher-students matching by exploiting a large-scale dataset that links students' standardized tests and questionnaires to the results of a web survey administered to their teachers.

The teacher questionnaire was experimentally launched in June 2012 by INVALSI to collect information on every teacher of the classes involved in the sampling activities of the National Evaluation Service. In this contribution we use the data extracted from its second wave (2013-14), which got higher response rates than the first one (about 10.000 teachers and 100.000 students). Linear probability models are used to estimate the associations between students' social background and their teachers' characteristics. In absence of value-added measures of teacher quality, we will use indicators such as the presence of a tertiary degree (for primary schools only), in-field teaching experience, university graduation mark, and years spent in the school (indicator of teaching continuity).

Results show the existence of an increasing level of segregation passing from primary to upper secondary schools. The uneven relationship between “good” teachers and students with higher social background is particularly evident for language teachers. In upper secondary schools, where this phenomenon is more intense, a high share of the differences registered between social groups is explained by tracking. Tracking acts in two ways: on the one hand, it fosters the segregation of students and, on the other hand, it promotes teachers' segregation. Teaching-recruiting mechanisms in upper secondary schools, in fact, tend to select those with better academic credentials for the academic schools, normally attended by high-status students. These results confirm how the structure of the Italian school system contributes to perpetuate social inequalities.

1. Introduzione

Nel nostro Paese una consistente parte della ricerca educativa, soprattutto in ambito sociologico, ha posto attenzione alle disuguaglianze di istruzione imputabili al background familiare degli studenti (per una rassegna si veda Argentin e Barone, 2016). La cosa non è sorprendente, soprattutto considerando che l'Italia si caratterizza per una forte associazione tra le origini sociali degli studenti e i loro esiti scolastici in termini di conseguimento dei titoli di studio (Barone e Ruggera, 2015). Questo filone di ricerca si è focalizzato sul fatto che le risorse familiari e le scelte di studenti e famiglie sfociano in percorsi e performance differenziali, evidenziando che gli individui di origini sociali più agiate finiscono per ottenere titoli di studio più alti, accedono più spesso a indirizzi liceali e, in generale, conseguono migliori risultati scolastici. In anni recenti, alcuni studi hanno iniziato a gettare luce sui meccanismi che agiscono sulla riproduzione delle disuguaglianze sociali dentro la scuola stessa. In tale direzione, si è messo a fuoco il tema del *tracking* scolastico nella scuola secondaria di II grado (Gasperoni, 1997; Checchi e Flabbi, 2013), una caratteristica istituzionale del sistema italiano che tende a segregare gli studenti tra diversi indirizzi rafforzando vantaggi e svantaggi legati al loro background parentale. Recentemente, a tale proposito, è stata evidenziata l'importanza del ruolo attivo che gli insegnanti della scuola secondaria di I grado hanno nell'orientare studenti di diverso background verso gli indirizzi della secondaria di II grado, rafforzando le disuguaglianze preesistenti (Romito, 2014). Un ulteriore aspetto rilevante messo in luce dalla ricerca pregressa è rappresentato dalle distorsioni rilevate nei voti dati dagli insegnanti agli studenti dal background svantaggiato, a parità di performance standardizzata nei test INVALSI (Argentin e Triventi, 2015).

Minore attenzione è stata posta a un altro processo di segregazione in grado di rafforzare le disuguaglianze educative, ovvero la distribuzione degli studenti dal differente background tra scuole con gradi diversi di qualità del servizio erogato. Questa carenza di attenzione è probabilmente imputabile al fatto che il sistema scolastico italiano è stato per molto tempo fortemente centralizzato e, nonostante la recente riforma introdotta dalla L. 107/2015 (la cosiddetta riforma della “buona scuola”) abbia rafforzato alcuni elementi di autonomia scolastica (Argentin e Barone, 2016), molti processi gestionali e organizzativi sono tuttora governati da meccanismi burocratici che lasciano poco spazio di manovra a insegnanti e dirigenti. L'uniformità del sistema scolastico italiano sembra però più apparente che reale, soprattutto guardando ai recenti studi che hanno messo in luce importanti differenziazioni tra scuole nella qualità del servizio erogato in termini di sviluppo delle competenze degli studenti. In particolare, grazie all'ultimo *Rapporto risultati* (INVALSI, 2016) si è evidenziata la presenza di scarti considerevoli tra i contributi dati dai singoli istituti scolastici all'apprendimento degli studenti (il cosiddetto “valore aggiunto”), soprattutto per ciò che riguarda le regioni del Mezzogiorno (Pavolini *et al.*, 2015). Tali differenze nell'efficacia scolastica si intrecciano con i noti fenomeni di distribuzione differenziata tra istituti degli studenti con diverso indice di background socio-economico-culturale (ESCS) e diverso background migra-

torio (INVALSI, 2010), nonché con le differenti capacità di attrazione o repulsione degli insegnanti tra scuole, in grado di condurre a concentrazioni molto diversificate di figure precarie all'interno dei singoli istituti (Barbieri, Cipollone e Sestito, 2007; Barbieri, Rossetti e Sestito, 2011).

Il presente lavoro intende approfondire proprio il tema della mancata uniformità tra scuole e di come quest'ultima possa contribuire all'affermazione di disuguaglianze educative. Più precisamente, ci si concentra sulla distribuzione tra scuole della forza lavoro insegnante, fattore la cui qualità è determinante per l'efficacia del sistema di istruzione (Hanushek, 1992; Sanders e Rivers, 1996). Ci si interroga sulla misura in cui a studenti socialmente avvantaggiati o svantaggiati "capita" di incontrare buoni insegnanti e ci si chiede se tale abbinamento sia casuale oppure se generi sistematicamente ulteriori svantaggi per chi ha origini sociali più basse. Si può ipotizzare la presenza di due tipi di abbinamento differenziale tra studenti e insegnanti: da un lato, l'istituirsi di scuole (o classi) con studenti di basso profilo nelle quali si concentrano insegnanti di minore qualità; dall'altro, la presenza di scuole (o classi) con studenti d'élite che incontrano insegnanti di qualità elevata. Cercheremo quindi di identificare se in Italia abbia luogo un abbinamento differenziale insegnanti-studenti e in che misura esso sia riconducibile a ognuno dei due tipi sopra descritti.

Nel corso del contributo passeremo brevemente in rassegna i principali studi sul fenomeno, per descrivere poi i dati INVALSI da noi impiegati nelle analisi, le scelte metodologiche assunte e passare alla descrizione dei risultati e alla discussione delle loro implicazioni.

2. Rassegna della letteratura

Gli studi sull'abbinamento tra insegnanti e studenti rispondono principalmente a interrogativi sull'equità del sistema scolastico. Le prime ricerche in materia sono state condotte negli Stati Uniti, un contesto caratterizzato da una forte segregazione abitativa e da una certa libertà di reclutamento dei docenti da parte delle scuole. L'abbinamento tra insegnanti e studenti è stato studiato osservando l'associazione tra le caratteristiche di background socio-economico degli studenti (etnia, condizione di povertà) e alcuni tratti dei docenti utilizzati come *proxy* dell'efficacia della loro azione educativa (Wayne e Youngs, 2003): l'anzianità di servizio, il possesso di una laurea nella materia insegnata (in gergo: insegnare *in-field*), la partecipazione a corsi di formazione che rilasciano qualifiche specifiche per l'insegnamento, il *rating* dell'università di provenienza, il voto di laurea e il punteggio ottenuto in test abilitanti all'insegnamento. Grazie all'ampia disponibilità di dati longitudinali che caratterizza il contesto statunitense, recentemente sono state anche utilizzate misure di valore aggiunto, che consentono di attribuire a ogni singolo docente una stima della sua efficacia.

I risultati di questo filone di letteratura convergono ampiamente: gli insegnanti efficaci, o dotati dei tratti "desiderabili" sopra elencati, sono concentrati in scuole con alunni in prevalenza bianchi, di elevata estrazione sociale e con punteggi nei test standardizzati più elevati già prima di incontrarli (Hanushek e Luque, 2000; Clotfelter, Ladd e Vigdor, 2002; De Angelis *et al.*, 2005; Hanushek *et al.*, 2005; Goldhaber, DeArmond e DeBurgomaster, 2011; Kalogrides e Loeb, 2013; Sass *et al.*, 2010; Goldhaber, Lavery e Theobald, 2015), anche se l'intensità di questa relazione non è uniforme, ma varia tra differenti distretti (Glazerman e Max, 2011).

Le spiegazioni alla base di questo fenomeno sperequativo, che vede nel sistema scolastico stesso un motore di riproduzione delle disuguaglianze, sono molteplici: la segregazione residenziale, i meccanismi che regolano i mercati del lavoro locali degli insegnanti e le preferenze intrinseche di questi ultimi. Studiando le domande di trasferimento degli insegnanti, Clotfelter, Ladd e Vigdor (2002) osservano che il personale docente tende ad abbandonare sedi scolastiche caratterizzate dall'elevata presenza di minoranze in favore di scuole situate in distretti più ricchi. Studi successivi confermano questo quadro (Hanushek *et al.*, 2005; Boyd *et al.*, 2005; Horng, 2009; Scafidi, Sjoquist e Stinebricker, 2007; Allensworth, Ponisciak e Mazzeo, 2009; Jackson, 2009), mostrando come gli insegnanti tendano a preferire scuole più vicine a casa e con alunni benestanti, in cui l'insegnamento è reputato un'attività più facile. Le scuole dei quartieri più poveri finiscono quindi per sperimentare livelli di turnover molto elevati, in cui si alternano leve successive di insegnanti novizi meno preparati (Hanushek *et al.*, 2005). Provenendo raramente dai distretti più poveri, le preferenze degli insegnanti si intrecciano poi con i fenomeni di segregazione abitativa che caratterizzano le città americane (Boyd *et al.*, 2005).

Queste spiegazioni prendono principalmente in considerazione i movimenti di docenti tra scuole o tra i distretti. Alcuni contributi più recenti mettono anche in evidenza l'esistenza di un ulteriore livello di abbinamento interno alle

scuole (e quindi meno visibile), dovuto alla costituzione delle diverse classi e all'attribuzione a queste ultime di diversi insegnanti (Isenberg *et al.*, 2013; Kalogrides, Loeb e Beteille, 2013; Goldhaber, Lavery e Theobald, 2015). Le stesse dinamiche che operano a livello di scuola e di distretto opererebbero quindi anche all'interno delle istituzioni scolastiche stesse: gli insegnanti con più anzianità di servizio hanno più probabilità di insegnare in classi con livelli di apprendimento pregressi maggiori, così come insegnanti bianchi hanno più probabilità di insegnare in classi con maggioranza di alunni bianchi. In questo caso, si ipotizza che l'abbinamento sia conseguente a processi di negoziazione tra il dirigente scolastico, gli insegnanti e i genitori, ma tale ipotesi rimane a oggi scarsamente supportata dall'evidenza empirica (Kalogrides, Loeb e Beteille, 2013).

Diversamente dagli Stati Uniti, sul piano formale le scuole italiane non dovrebbero giocare alcun ruolo nei processi di selezione e gestione del personale. Sino alla riforma introdotta dalla L. 107/2015, la gestione del personale era realizzata a livello centrale in base a graduatorie in cui l'elemento dirimente era l'anzianità di servizio. Tale scenario è leggermente mutato, ma resta valido ancora oggi in larga misura. Un sistema così centralizzato dovrebbe, in teoria, garantire una maggiore equità, governando più da vicino i trasferimenti di risorse e di personale alle scuole. Una crescente mole di ricerche mostra, però, l'esistenza di profonde differenze nella qualità dell'istruzione offerta dai diversi istituti scolastici (INVALSI, 2016), con differenze particolarmente marcate anche entro specifiche aree territoriali e tenendo conto del contesto socio-economico in cui le scuole operano (Pavolini *et al.*, 2015). Ciò lascia supporre che i fattori che generano apprendimento differiscano profondamente tra scuole e che, tra questi fattori, spicchi la qualità della forza docente. Gli studi che riguardano gli insegnanti e la loro distribuzione tra scuole sono relativamente pochi, data la scarsità di dati disponibili. Barbieri, Cipollone e Sestito (2007) sfruttano l'unione di archivi amministrativi relativi ai trasferimenti degli insegnanti con i risultati delle prove di matematica OCSE-PISA 2003 per studiare la relazione tra turnover, richieste di trasferimento e performance degli studenti. Lo studio mostra l'esistenza di una correlazione negativa tra il livello di turnover dei docenti e i risultati degli studenti alla prova PISA. Ricerche più recenti basate sui test INVALSI confermano questo risultato anche in relazione alle performance in lingua italiana (Ferrer-Esteban, 2011). I docenti preferiscono andarsene dalle scuole in cui è più difficile insegnare, le quali spesso si trovano in contesti socio-economici più svantaggiati. Risultati simili sono stati ottenuti anche attraverso l'analisi dei *patterns* delle richieste di trasferimento tra scuole dei docenti, che confermano come le sedi disagiate siano considerate meno appetibili (Barbieri, Rossetti e Sestito, 2011 e 2013). Una volta maturata l'anzianità di servizio necessaria per poter ottenere il trasferimento, gli insegnanti italiani si spostano tendenzialmente verso sedi non solo più facilmente raggiungibili dalla propria abitazione, ma anche caratterizzate da un corpo studentesco mediamente più abile e di estrazione sociale superiore. Nel complesso, dunque, l'evidenza empirica che va accumulandosi suggerisce che anche il sistema scolastico italiano, nonostante la sua formale uniformità allocativa, presenti meccanismi di abbinamento insegnanti-studenti che rafforzano le disuguaglianze esistenti.

3. Dati, variabili e metodo

La base dati utilizzata in sede di analisi è frutto dell'unione tra il campione delle Rilevazioni nazionali INVALSI (SNV e Prova nazionale) per l'anno scolastico 2013-14 e il Questionario insegnante per l'a. s. 2013-14. Quest'ultimo strumento, indirizzato agli insegnanti di matematica e italiano delle classi campione, è stato avviato in via sperimentale nell'a.s. 2011-12 per poi entrare a regime nell'a.s. 2013-14. La rilevazione mira a raccogliere non solo informazioni di tipo anagrafico o professionale, ma anche aspetti attitudinali e relativi alle pratiche di insegnamento.

La scelta dell'anno di riferimento è dipesa principalmente da due considerazioni sostantive circa la qualità delle risorse a disposizione. Innanzitutto, nel 2014 il Questionario insegnante era già al secondo anno di somministrazione e risultava quindi sufficientemente consolidato sia nella forma sia nelle procedure di somministrazione. Per ciò che riguarda i dati INVALSI, invece, l'approvazione della L. 107 nel luglio del 2015 ha dato il via a fenomeni di boicottaggio delle prove, che hanno condotto a una sensibile riduzione dei tassi di risposta degli studenti, inficiando la qualità dei dati raccolti. L'anno scolastico 2013-14, di conseguenza, rappresenta a oggi la migliore finestra temporale per lo studio dell'associazione tra indicatori di qualità degli insegnanti e delle origini sociali degli studenti attraverso le risorse offerte da INVALSI.

Tab. 1 – Tassi di copertura e numerosità del campione per materia, grado scolastico e livello

	Tasso di risposta insegnanti	N. insegnanti	N. studenti abbinati	N. scuole	Insegnanti per scuola
<i>Italiano</i>					
Primaria	81%	2.376	42.862	700	3,4
Secondaria I grado	89%	1.289	25.021	1.289	1,0
Secondaria II grado	73%	1.593	27.555	914	1,7
Totale	80%	5.258	95.438	2.903	
<i>Matematica</i>					
Primaria	80%	2.353	42.510	698	3,4
Secondaria I grado	89%	1.295	25.223	1.295	1,0
Secondaria II grado	74%	1.628	28.096	935	1,7
Totale	81%	5.276	95.829	2.928	–
Totale generale	80%	10.534	104.937	5.831	–

I risultati dell’abbinamento fra i dati INVALSI e il Questionario insegnante, riportati in tab. 1, evidenziano la presenza di un tasso medio di copertura del campione studenti pari all’80%. La perdita di informazione varia dai 27 ai 10 punti percentuali a seconda della materia e del livello considerato. Si tratta di tassi di caduta modesti per le indagini campionarie con questionari. Complessivamente disponiamo di 10.534 insegnanti dislocati in 5.831 scuole di vario ordine e grado sparse sull’intero territorio nazionale. All’interno di ogni singolo istituto è garantita la presenza di almeno un insegnante per materia e, nell’80% delle classi, disponiamo di dati su entrambi gli insegnanti. Dei 104.937 studenti facenti parte del campione, infatti, 86.330 hanno un insegnante di riferimento per entrambe le materie, mentre 9.108 dispongono solamente di quello di italiano e 9.499 di quello di matematica. Per questo motivo il numero complessivo di studenti parte del campione di riferimento risulta nettamente inferiore alla somma dei totali per materia¹.

Nei dataset si sono identificate la variabile indipendente relativa al background sociale degli studenti e le variabili dipendenti, che intendono invece rilevare diverse dimensioni di qualità dell’insegnante abbinato ai differenti profili sociali degli studenti.

Per quanto riguarda la variabile indipendente di interesse, le origini sociali degli studenti, si è deciso di impiegare il titolo di istruzione più alto in famiglia e di non considerare invece l’occupazione dei genitori. Si tratta di una scelta compiuta sulla base di ragioni di carattere sostantivo e tecnico. In primo luogo, l’influenza dell’istruzione dei genitori è particolarmente rilevante sulle scelte educative degli studenti, e lo è in misura maggiore rispetto alla posizione occupazionale (Bukodi e Goldthorpe, 2013). Inoltre, le scuole riescono a fornire a INVALSI informazioni qualitativamente migliori per l’istruzione dei genitori (stabile nel tempo e piuttosto semplice da raccogliere e codificare) rispetto a quelle sull’occupazione. In particolare, si sono definiti tre gruppi di studenti sulla base del loro livello di istruzione familiare (costruito adottando il principio di dominanza, quindi considerando il titolo più alto in famiglia): i figli dei laureati; i figli dei diplomati; i figli di chi possiede al massimo una licenza media.

Per quanto attiene alla variabile dipendente, la “qualità degli insegnanti”, si è ovviamente nella difficile situazione di non avere una misura predefinita e di avere a che fare con un concetto non privo di potenziali controversie. Infatti, la letteratura relativa a quali caratteristiche degli insegnanti contino nel determinare la loro qualità si basa su evidenze circoscritte di tipo correlazionale ed è tutt’altro che risolutiva (Goldhaber *et al.*, 2010). Si sono identificate cinque caratteristiche che si reputano poco desiderabili per conseguire un buon grado di apprendimento degli studenti:

- l’instabilità contrattuale, che si associa alla mobilità tra scuole (Ferrer-Esteban, 2011);
- lo scarso radicamento dell’insegnante nell’istituto scolastico (la sua presenza nella scuola da 3 anni o meno), che comporta una minore conoscenza del contesto organizzativo formale e informale in cui si opera (Creemers e Kyriakides, 2012) e può influire negativamente sul grado di collaborazione con i colleghi e sull’efficacia della propria azione educativa;
- una performance formativa non eccellente degli insegnanti, che assume due declinazioni: l’assenza della laurea nella scuola primaria oppure, nella scuola secondaria, il fatto di avere un voto di laurea appartenente ai due terzi

¹ Pari a $(95.438 + 95.829) = (86.330 * 2 + 9.108 + 9.499)$.

più bassi della sua distribuzione pesata per ambito disciplinare (per una rassegna dei risultati si veda Wayne e Youngs, 2003);

- l’insegnamento di una materia diversa dalla disciplina di laurea (il cosiddetto *out of field teaching*), come per esempio insegnare matematica con una laurea in biologia oppure insegnare italiano con una laurea in filosofia. Tale elemento potrebbe rendere le conoscenze disciplinari degli insegnanti meno solide e quindi rendere la loro azione didattica più incerta (Wayne e Youngs, 2003).

Per stimare il segno e la forza delle associazioni tra background educativo degli studenti e caratteristiche degli insegnanti, si è deciso di lavorare con i dati a livello studente, clusterizzando per scuola la stima degli errori standard nei modelli di probabilità lineare adottati². Si è preferita questa soluzione a elaborazioni con dati medi di classe per non comprimere fortemente la varianza della variabile indipendente e rischiare così di nascondere eventuali associazioni preesistenti.

I modelli impiegati per stimare le associazioni sfruttano la tecnica di regressione di probabilità lineare:

- modello 0: calcola la semplice associazione statistica tra ciascun indicatore di presunta qualità dell’insegnante (la nostra variabile dipendente) e le origini degli studenti (la variabile indipendente);
- modello 1: corregge l’associazione stimata con il modello precedente, controllandola per la provincia in cui è sita la scuola; tramite questo accorgimento è possibile verificare se e quanto la presenza di un’associazione statistica tra le due variabili sia in realtà il riflesso delle caratteristiche dei bacini scolastici, che limitano la libertà di movimento degli insegnanti tra scuole e che si caratterizzano per distribuzioni differenziate degli studenti stessi³;
- modello 2: stima per le sole scuole secondarie di II grado ulteriori modelli controllando anche per l’indirizzo di scuola superiore (liceo, tecnico, professionale). Lo scopo è verificare in che misura l’eventuale abbinamento osservato sia dovuto alla diversificazione dei percorsi scolastici, la quale opera sia nel senso di segregare gli studenti in diversi tipi di scuola, sia in quello di separare le carriere degli insegnanti in base al sistema dell’assunzione tramite classi di concorso.

I modelli sono replicati separatamente per italiano e matematica, dal momento che l’abbinamento potrebbe variare in funzione della materia e del livello scolastico. Basti pensare al diverso grado di difficoltà di reclutamento degli insegnanti di italiano e di matematica, così come al fatto che sono diversi gli insegnanti che scelgono di insegnare nei tre livelli scolastici e negli indirizzi della secondaria di II grado.

4. Risultati

Iniziamo osservando la distribuzione degli indicatori di presunta qualità degli insegnanti tra livelli, materie e, nelle scuole secondarie di II grado, tra tipi di scuola (tab. 2).

Gli insegnanti di italiano e matematica delle scuole primarie mostrano di essere assai simili rispetto a tutti gli indicatori utilizzati: sono per la maggior parte stabili, radicati nella loro scuola e privi di una laurea. È con le scuole medie che si osserva una prima differenziazione sostanziale: gli insegnanti di italiano hanno ottenuto un voto di laurea migliore e sono quasi tutti laureati in lettere, mentre solo una minoranza degli insegnanti di matematica è laureato nella disciplina insegnata. Alle scuole secondarie di II grado, gli insegnanti di matematica e italiano esibiscono notevoli differenze anche nel livello di radicamento, ma le differenze maggiori si riscontrano tra insegnanti della stessa materia all’interno dei diversi indirizzi. Gli insegnanti dei licei sono più radicati all’interno dell’istituto, meno precari, più spesso laureati nella materia di insegnamento e con voti migliori. In virtù di queste differenze, che segnalano ancora una volta la frammentazione del nostro sistema scolastico, le analisi di questo capitolo saranno condotte separatamente per livello e per materia.

² Inoltre, per evitare di dare eccessivo peso alle classi in cui la raccolta dati è stata deficitaria, le analisi sono state limitate alle classi con almeno 8 alunni in cui almeno il 75% degli studenti possiede l’informazione sulle proprie origini sociali (pari all’85% del campione iniziale).

³ Le procedure di trasferimento degli insegnanti tra scuole sono guidate da un meccanismo basato su graduatorie che, tra i vari criteri adottati, premia gli insegnanti che operano in una scuola situata nello stesso comune di quella in cui si chiede il trasferimento e, a seguire, gli insegnanti della stessa provincia. La provincia della scuola è quindi una variabile *proxy* sub-ottimale per studiare il mercato del lavoro degli insegnanti, ma anche la più precisa utilizzabile con i dati in nostro possesso. Soluzioni alternative (che utilizzano come *proxy* il numero di scuole dello stesso grado scolastico nel comune e l’ampiezza dello stesso) conducono a risultati analoghi.

Tab. 2 – Caratteristiche degli insegnanti per grado e disciplina (%)

Grado	Materia	Tempo det.	Non radicati	Out-of field	Voto di laurea*	Privi di laurea
Primarie	ITA	5,3	12,5	–	–	73,4
Primarie	MAT	6,1	15,7	–	–	75,0
Secondarie I grado	ITA	8,9	27,7	22,9	57,8	–
Secondarie I grado	MAT	12,1	29,2	82,3	63,5	–
Secondarie II grado	ITA	12,2	36,1	12,0	52,4	–
Licei	ITA	9,5	24,6	5,8	42,5	–
Tecnici	ITA	10,6	41,3	15,2	60,9	–
Professionali	ITA	18,4	47,9	18,0	62,2	–
Sec. II grado	MAT	12,1	27,6	28,2	65,3	–
Licei	MAT	10,2	26,5	22,4	61,8	–
Tecnici	MAT	10,7	25,1	30,5	66,6	–
Professionali	MAT	16,7	32,5	34,0	71,0	–

* Terzili medio e basso della distribuzione del voto di laurea pesato per ambito disciplinare.

Si stima ora la forza dell'associazione tra ciascun indicatore di qualità dell'insegnante e le origini sociali dello studente mediante i modelli di probabilità lineare descritti in precedenza. La tab. 3 mostra i risultati delle analisi per gli insegnanti di italiano, mentre la tab. 4 si concentra su quelli di matematica. Per facilitare la lettura dei risultati, la categoria di riferimento delle analisi è costituita dai figli di laureati e, come per le analisi precedenti, il possesso di ciascuna caratteristica poco desiderabile è indicato da una variabile binaria che assume valore 1 se il soggetto la possiede e 0 altrimenti. Inoltre i coefficienti sono stati percentualizzati, così da accrescerne la leggibilità. Coefficienti positivi indicano quindi un rischio maggiore, per i figli di diplomati (o di genitori con la sola licenza media) rispetto ai figli dei laureati, di avere insegnanti che presentano la caratteristica di volta in volta considerata.

Partiamo quindi dagli insegnanti di italiano. Osservando la tab. 3 il primo elemento che emerge è la concentrazione delle differenze più rilevanti nell'abbinamento di studenti e insegnanti nella scuola secondaria di II grado. Più precisamente, nella scuola primaria si osserva solo una più frequente allocazione degli studenti di origini sociali più modeste in classi con insegnanti privi della laurea. Nel caso delle scuole secondarie di I grado esistono invece differenze nel voto di laurea dell'insegnante, con i figli di genitori non laureati più spesso associati a insegnanti con voti nei terzili medio-bassi della distribuzione. Queste differenze segnalano che l'esposizione a insegnanti di minore qualità aumenta al decrescere del livello culturale della famiglia degli studenti, mostrando l'esistenza di entrambi i meccanismi di concentrazione in scuole/classi di élite e segregazione in scuole/classi svantaggiate. Le associazioni in questione si riducono notevolmente una volta che si controlla per i vincoli di mobilità degli insegnanti (modello 1), ma rimangono comunque sostanziali oltre che statisticamente significative.

Come si anticipava, la situazione muta radicalmente nel passaggio alle scuole secondarie di II grado. I divari tra studenti di diversa estrazione sono presenti con maggiore intensità su tutti gli indicatori a eccezione (parziale) dei contratti temporanei. La probabilità di abbinamento a un insegnante non radicato con un titolo di laurea non nella materia insegnata cresce di ben 13 punti percentuali per i figli di chi ha una licenza media, mentre la differenza nell'avere un insegnante laureato nella materia che insegna è pari a 7 punti percentuali. Divari notevoli, ancorché più contenuti, si osservano tra figli di laureati e i figli di diplomati, replicando lo schema "a tre gradini" osservato anche in precedenza: accanto alla creazione di classi e scuole altamente segregate per i figli dei meno abbienti, osserviamo parimenti una fuga dei figli dei laureati verso contesti privilegiati. Queste differenze, immutate se non rafforzate dal controllo per bacino lavorativo degli insegnanti, sono notevolmente ridimensionate nel momento in cui si controlla per indirizzo di scuola superiore. La sovra-rappresentazione delle famiglie meno abbienti nelle scuole tecniche e professionali porta anche ad abbinamenti iniqui per quanto riguarda i predittori di qualità dell'insegnamento da noi considerati. Del resto, come già si è evidenziato nella tab. 3, gli insegnanti di italiano in possesso di quelli che abbiamo definito tratti non desiderabili sono una quota minima nei licei, per poi aumentare negli istituti tecnici e ancora di più nei professionali.

Tab. 3 – Associazione tra le caratteristiche degli insegnanti di italiano e il livello di istruzione familiare degli studenti (rif.: figli di laureati) per grado scolastico

Italiano	Modello 0		Modello 1		Modello 2	
	Diploma	Licenza media o meno	Diploma	Licenza media o meno	Diploma	Licenza media o meno
<i>Primaria</i>						
Temporaneo	-0,68	-0,98	-0,54	-0,21	–	–
Non radicato	-1,32	-1,40	-1,18	-0,91	–	–
Privo di laurea	3,38**	6,02**	2,20*	3,62**	–	–
<i>Secondaria di I grado</i>						
Temporaneo	0,49	-0,70	-0,63	-0,20	–	–
Non radicato	1,41	2,77	1,18	2,73	–	–
Out of field	0,72	1,29	0,85	-0,59	–	–
Voto di laurea	3,00*	7,71**	1,79	4,31*	–	–
<i>Secondaria di II grado</i>						
Temporaneo	0,96	1,85	1,10	3,22**	-0,04	0,36
Non radicato	6,07**	12,94**	6,00**	12,64**	2,18**	4,40**
Out of field	3,15**	7,04**	2,57**	5,64**	0,89	1,96*
Voto di laurea	6,60**	13,28**	4,43**	7,42**	1,01	2,50*

* p < 0,1; ** p < 0,05.

Passando agli insegnanti di matematica, il quadro è in parte differente. In primo luogo, si attenuano le differenze tra gradi scolastici: anche per la scuola secondaria di II grado l'abbinamento è contenuto. Inoltre, cambiano le variabili per cui si rileva un abbinamento iniquo insegnanti-studenti. La tab. 4 mostra anche in questo caso un'associazione più frequente tra figli di laureati e insegnanti laureati alle elementari, così come esiste un abbinamento tra origini sociali e radicamento (anziché voto di laurea) nella scuola secondaria di I grado.

Tab. 4 – Associazione tra le caratteristiche degli insegnanti di matematica e il livello di istruzione familiare degli studenti (rif.: terzile alto) per grado scolastico

Matematica	Modello 0		Modello 1		Modello 2	
	Diploma	Licenza media o meno	Diploma	Licenza media o meno	Diploma	Licenza media o meno
<i>Primaria</i>						
Temporaneo	-0,58	-2,10*	-0,26	-0,82	–	–
Non radicato	-1,68	-2,26	-0,82	-0,69	–	–
Privo di laurea	3,68**	6,98**	2,33**	3,70**	–	–
<i>Secondaria di I grado</i>						
Temporaneo	0,97	1,06	0,66	2,42*	–	–
Non radicato	3,46**	6,86**	3,46**	7,34**	–	–
Out of field	0,81	1,49	0,59	1,18	–	–
Voto di laurea	1,00	-0,25	-0,45	-2,82	–	–
<i>Secondaria di II grado</i>						
Temporaneo	0,26	1,69	0,30	2,13*	0,07	0,96
Non radicato	1,14	2,29	1,59	3,27*	1,57	2,39*
Out of field	2,98**	6,68**	2,60**	5,81**	0,47	1,39
Voto di laurea	1,27	3,24	0,24	1,86	-0,13	-0,33

* p < 0,1; ** p < 0,05.

Nelle scuole secondarie di II grado, si osservano differenze nell'esposizione a insegnanti temporanei, non radicati e *out of field* a scapito dei figli dei non laureati, ma si tratta di differenze più contenute rispetto a quelle identificate per gli insegnanti di italiano. Anche in questo caso le differenze tra studenti di diversa estrazione si riducono una volta che il tipo di scuola superiore è preso in considerazione.

5. Osservazioni conclusive

Ci siamo soffermati su alcune caratteristiche degli insegnanti che, secondo la letteratura, potrebbero essere non desiderabili in termini di apprendimento degli studenti e abbiamo guardato all'associazione tra queste grandezze e il background familiare degli studenti. Ci interessava capire se vi sia uniformità o meno nelle caratteristiche del personale docente che incontrano studenti di diversa estrazione sociale.

In primo luogo, abbiamo osservato che le caratteristiche potenzialmente non desiderabili degli insegnanti sono presenti su quote non trascurabili del campione e sono tendenzialmente più intense per matematica che per italiano. Inoltre, tali caratteristiche tendono a crescere con il grado scolastico e sono più concentrate negli indirizzi professionali e tecnici che nei licei.

In secondo luogo, abbiamo rilevato che gli insegnanti, in Italia, sono effettivamente distribuiti in modo diseguale tra studenti di diverso background familiare. A grandi linee, gli insegnanti dei figli di laureati sono più radicati nel proprio contesto scolastico, hanno avuto voti migliori all'università, si sono laureati nella medesima materia che insegnano e, nella primaria (ossia l'unico livello in cui questo indicatore conta), hanno più spesso un titolo di studio terziario.

L'entità di questi fenomeni è comunque relativamente contenuta nella scuola primaria e secondaria di I grado, fino al momento in cui i destini degli alunni si diversificano in base all'indirizzo secondario superiore prescelto. A quel punto i divari tra gli studenti di diversa estrazione nell'esposizione a insegnanti con profili diversi assumono proporzioni notevoli, soprattutto nel caso degli insegnanti di italiano. Il fatto che queste differenze siano spiegate largamente dall'indirizzo scolastico frequentato conferma le considerazioni sul *tracking* tra indirizzi come meccanismo istituzionale di rafforzamento delle disuguaglianze di origine sociale. Si osserva infatti che ai ben noti fenomeni di segregazione degli studenti dovuti alla ramificazione degli istituti secondari di II grado si accompagnano processi di segregazione simili tra insegnanti, dovuti al loro sistema di reclutamento e allocazione. Sulla base dei dati a nostra disposizione, gli insegnanti con caratteristiche che in letteratura sono associate a una maggiore efficacia si concentrano nei licei, a discapito degli istituti tecnici e professionali. Non sembra esagerato parlare quindi di "disuguaglianze sistemiche", nonostante la pretesa del sistema scolastico centralizzato di fornire un'istruzione uniforme a tutti gli studenti.

Riferimenti bibliografici

- Allensworth E., Ponisciak S., Mazzeo C. (2009), *The Schools Teachers Leave: Teacher Mobility in Chicago Public Schools*, Chicago, Consortium on Chicago School Research, testo disponibile al sito: http://ccsr.uchicago.edu/publications/CCSR_Teacher_Mobility.pdf, data di consultazione: 12 gennaio 2017.
- Argentin G., Barone C. (2016), "La riforma della buona scuola: innovazione e retorica del cambiamento", in M. Carbone, S. Piattoni (a cura di), *Politica in Italia*, il Mulino, Bologna.
- Argentin G., Triventi M. (2015), "The North-South divide in school grading standards: new evidence from national assessments of the Italian student population", *Italian Journal of Sociology of Education*, 7, 2: 157-185.
- Barbieri G., Cipollone P., Sestito P. (2007), "Labour market for teachers: demographic characteristics and allocative mechanisms", *Giornale degli Economisti e Annali di Economia*, 66, 1: 335-373.
- Barbieri G., Rossetti C., Sestito P. (2013), "Teacher mobility and student learning", *INVALSI Working Papers*, 18, testo disponibile al sito: http://www.invalsi.it/invalsi/ri/improving_education/Papers/rossetti/46.pdf, data di consultazione: 10 gennaio 2017.
- Barbieri G., Rossetti C., Sestito P. (2011), "The determinants of teacher mobility: Evidence using Italian teachers' transfer applications", *Economics of Education Review*, 30, 6: 1430-1444.
- Barone C., Ruggera L. (2015), "Le disuguaglianze sociali nell'istruzione in una prospettiva comparativa. Il rompicapo del caso italiano", *Scuola democratica*, 2: 321-342.

-
- Boyd D., Lankford H., Loeb S., Wyckoff J. (2005), "The draw of home: how teachers' preferences for proximity disadvantage urban schools", *Journal of Policy Analysis and Management*, 24, 1: 113-132.
- Bukodi E., Goldthorpe J.H. (2013), "Decomposing 'social origins': the effects of parents' class, status, and education on the educational attainment of their children", *European Sociological Review*, 29, 5: 1024-1039.
- Checchi D., Flabbi L. (2013), "Intergenerational mobility and schooling decisions in Germany and Italy: the impact of secondary school tracks", *Rivista di politica economica*, 3: 7-57.
- Clotfelter C., Ladd H., Vigdor J. (2002), "Who teaches whom? Race and the distribution of novice teachers", *Economics of Education Review*, 24, 4: 377-392.
- Colombo M., Argentin G. e Pandolfini V. (2016), "La sociologia dell'educazione in Italia: prospettive di ricerca di fronte alla "riformabilità" dei sistemi educativi", in F. Corbisiero, E. Ruspini (a cura di), *Sociologia del futuro. Studiare la società del ventunesimo secolo*, Cedam, Padova.
- Creemers B., Kyriakides L. (2012), *Improving Quality in Education: Dynamic Approaches to School Improvement*, Routledge, London.
- DeAngelis K.J., Presley J.B., White B.R. (2005), "The distribution of teacher quality in Illinois", *Policy Research Report: IERC 2005-1*, Illinois Education Research Council, Edwardsville (IL).
- Ferrer-Esteban G. (2011), "Beyond the traditional territorial divide in the Italian education system. Effects of system management factors on performance in school", *FGA Working Papers*, 43.
- Gasperoni G. (1997), *Il rendimento scolastico in Italia*, il Mulino, Bologna.
- Glazerman S., Max J. (2011). "Do low-income students have equal access to the highest-performing teachers?", *NCEE Evaluation Brief, NCEE 2011-4016*, testo disponibile al sito: <http://eric.ed.gov/?id=ED517966>, data di consultazione: 15 gennaio 2017.
- Goldhaber D., DeArmond M., DeBurgomaster S. (2011), "Teacher attitudes about compensation reform: implications for reform implementation", *Industrial & Labor Relations Review*, 64, 3: 441-463.
- Goldhaber D., Lavery L., Theobald, R. (2015), "Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students", *Educational Researcher*, 44, 5: 293-307.
- Goldhaber D., Liddle S., Theobald R., Walch, J. (2010), "Teacher effectiveness and the achievement of Washington's students in mathematics", *Center for Education Data and Research Working Papers*, 6, testo disponibile al sito: [https://www.cedr.us/papers/working/CEDR%20WP%202010-6_Teacher%20Effectiveness%20in%20WA%20\(12-7-10\).pdf](https://www.cedr.us/papers/working/CEDR%20WP%202010-6_Teacher%20Effectiveness%20in%20WA%20(12-7-10).pdf), data di consultazione: 3 gennaio 2017.
- Hanushek E. (1992), "The trade-off between child quantity and quality", *Journal of Political Economy*, 100, 1: 84-117.
- Hanushek E.A., Kain J.F., O'Brien D.M., Rivkin S.G. (2005), "The market for teacher quality", *NBER Working Paper Series*, 11154, testo disponibile al sito: <http://www.nber.org/papers/w11154.pdf>, data di consultazione: 3 gennaio 2017.
- Hanushek E.A., Luque J. (2000), "Smaller classes, lower salaries? The effects of class size on teacher labor markets", in S. Laine, J. Ward (eds.), *Using What We Know: A Review of The Research on Implementing Class-Size Reduction Initiatives for State and Local Policymakers*, North Central Regional Educational Laboratory, Oak Brook (IL).
- Hornig E.L. (2009), "Teacher tradeoffs: Disentangling teachers' preferences for working conditions and student demographics", *American Educational Research Journal*, 46, 3: 690-717.
- INVALSI (2010), *Servizio nazionale di valutazione a.s. 2009/2010. Rilevazione degli apprendimenti – SNV. Prime analisi. Parte I-II*, testo disponibile al sito: http://www.invalsi.it/download/rapporti/snv2010/Rapporto_SNV_2009_2010_-ParteI_II.pdf, data di consultazione: 3 gennaio 2017.
- INVALSI (2016), *Rilevazioni nazionali degli apprendimenti 2015-16. Rapporto risultati*, testo disponibile al sito: https://invalsi-areaprove.cineca.it/docs/file/07_Rappo-rto_Prove_INVALSI_2016.pdf, data di consultazione: 3 gennaio 2017.
- Isenberg E., Max J., Gleason P., Potamites L., Santillano R., Hock H., Hansen M. (2013), *Access to Effective Teaching for Disadvantaged Students*, National Center for Education Evaluation and Regional Assistance, Washington (DC).
- Jackson C.K. (2009), "Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation", *Journal of Labor Economics*, 27, 2: 213-256.
- Kalogrides D., Loeb S. (2013), "Different teachers, different peers: The magnitude and effects of student sorting within schools", *Educational Researcher*, 42, 6: 304-316.
- Kalogrides D., Loeb S., Beteille T. (2013), "Systematic sorting: Teacher characteristics and class assignments", *Sociology of Education*, 86, 2: 103-123.
- Pavolini E., Argentin G., Barbieri G., Falzetti P., Ricci R. (2015), "L'influenza delle scuole e del contesto locale sui divari territoriali delle competenze degli studenti", in F. Asso, E. Pavolini (a cura di), *L'istruzione difficile. I divari nelle competenze tra Nord e Sud*, Donzelli, Roma.
- Romito M. (2014), "Social origin and school guidance: Guidance practices in student's high school choices", *Etnografia e ricerca qualitativa*, 7, 3: 481-504.

-
- Sanders W., Rivers J. (1996), *Cumulative and Residual Effects of Teachers on Future Academic Achievement*, University of Tennessee Value-Added Research and Assessment Center, Knoxville.
- Sass T.R., Hannaway J., Xu Z., Figlio D.N., Feng L. (2010), "Value added of teachers in high-poverty schools and lower poverty schools", *Journal of Urban Economics*, 72, 2: 104-122.
- Scafidi B., Sjoquist D., Stinebricker T. (2007), "Race, poverty, and teacher mobility", *Economics of Education Review*, 26, 2: 145-159.
- Wayne A., Youngs P. (2003), "Teacher characteristics and student achievement gains: a review", *Review of Educational Research*, 73, 1: 89-122.

18. Differenze di genere e di status socio-economico nel rendimento scolastico: evidenze empiriche nella scuola primaria

Gender and socio-economic gaps in educational achievement: empirical evidence from primary school

di Clelia Cascella, Elisa Cavicchiolo

Sebbene sia il genere sia lo status socio-economico (SES) siano stati spesso utilizzati per spiegare alcune differenze nel rendimento scolastico, lo studio dell'effetto che l'interazione tra questi due fattori può avere sulle performance scolastiche è ancora poco esplorato, nella letteratura sia nazionale sia internazionale. Studi recenti suggeriscono infatti che esiste un effetto di interazione tra SES e genere: contesti di apprendimento caratterizzati da SES basso possono favorire lo sviluppo di atteggiamenti e comportamenti anti-scolastici, con impatti più negativi sulle performance dei maschi che delle femmine. Obiettivo di questo lavoro è quello di controllare empiricamente se e come la relazione tra SES e genere osservata a livello internazionale sia presente anche nella scuola primaria italiana (per la quale assai scarna è la letteratura sul tema). Inoltre, poiché studi precedenti hanno già dimostrato che l'effetto dei fattori socio-demografici può cambiare in funzione dell'età degli studenti, abbiamo analizzato e confrontato le risposte date dai bambini della classe seconda e quinta della scuola primaria ai test INVALSI di italiano e matematica. I punteggi degli studenti in matematica e in italiano calcolati con il modello di Rasch, per entrambi i livelli scolastici, sono stati utilizzati come variabile dipendente in un'Analisi della varianza (ANOVA) fattoriale. Coerentemente con quanto già emerso da studi simili su studenti italiani, dall'analisi dei dati relativi a entrambi i gradi scolastici, emerge un effetto statisticamente significativo tra genere e SES sulle performance in italiano e matematica.

Although both gender and socio-economic status (SES) have frequently been proposed to explain differences in academic achievement, there has been relatively little national or international research to evaluate the interactions between these two variables on students' performances. Recent studies suggest that the interaction between SES and gender may have an effect on academic achievement and that a learning environment with a low SES could be linked to the development of anti-academic attitudes and behavior patterns, with a more negative impact on the performance of males than that of females. This paper aims to assess the existence of this effect in Italian primary school. In addition, given that the influence of social factors tends to increase gradually as students get older, we analyzed and compared answers given by pupils attending the 2nd and the 5th grade level of primary school to INVALSI achievement tests in Italian and Maths. Students' scores calculated by using the Rasch model were used as dependent variable in a two-way factorial ANOVA. Consistently with similar studies carried out on Italian students, the data analyses have revealed a statistically significant effect of gender and SES on the results achieved in Italian and Mathematics for both grades.

1. Introduzione

La relazione tra status socio-economico (SES) e performance scolastiche è ben nota e studiata in letteratura (Coleman, 1988; Brooks-Gunn e Duncan, 1997; McLoyd, 1998; Coley e Morris, 2002; Ensminger *et al.*, 2003; Sirin, 2005; Aikens e Barbarin, 2008; Ream e Palardy, 2008; Morgan *et al.*, 2009; Dahl e Lochner, 2012; Igbo, Onu e Obiyo, 2015). Benché alcuni studi non rilevino una relazione significativa tra risultati scolastici e status socio-economico (Seyfried, 1998; Ripple e Luthar, 2000), la maggior parte delle ricerche empiriche degli ultimi anni ha messo in luce un effetto da

medio a forte della relazione tra SES e performance scolastiche (Sirin, 2005; White, 1982). In particolare, gli studenti con bassi livelli di status socio-economico sembrerebbero sviluppare competenze e/o abilità scolastiche – in una o in diverse discipline – più lentamente rispetto a quelli che provengono da famiglie con un più alto status socio-economico (Morgan *et al.*, 2009). Inoltre, il livello scolastico risulta essere anche un moderatore significativo delle correlazioni tra SES e performance (Sammons, 1995; Sirin, 2005; Lindberg *et al.*, 2010) e tra genere e performance (Mullis *et al.*, 2000). Studi recenti (per es. Legewie e Di Prete, 2012; Stoet e Geary, 2013) ribadiscono infatti l'importanza di comprendere le relazioni tra SES e performance scolastiche e approfondiscono i meccanismi che presiedono le differenze di genere in ambito educativo: contesti di apprendimento caratterizzati da SES basso possono favorire lo sviluppo di atteggiamenti e comportamenti anti-scolastici con impatti diversi sulle performance di maschi e femmine.

Per quanto riguarda, invece, il genere, alcuni studi, effettuati partendo dai dati relativi a diversi gradi scolastici, hanno messo in evidenza che, in generale, le ragazze conseguono spesso performance scolastiche migliori rispetto ai ragazzi (Kenney-Benson *et al.*, 2006; Lindsay e Muijs, 2006; Voyer e Voyer, 2014) mentre permangono differenze, a volte in favore dei maschi, altre volte in favore delle femmine, in relazione all'acquisizione di competenze specifiche (Buckingham, 1999; Else-Quest, Hyde e Linn, 2010; Lindberg *et al.*, 2010; Voyer e Voyer, 2014). Le differenze di genere si presentano a favore dei maschi rispetto alle competenze matematiche e scientifiche e a favore delle femmine in particolare per le competenze di italiano, anche se il gap sembrerebbe essere in riduzione. Le differenze di genere sono ancora più evidenti non tanto in relazione ai risultati di apprendimento, quanto rispetto all'atteggiamento e all'interesse verso la matematica o all'italiano. Per esempio, i ragazzi si dimostrano più sicuri di sé e meno ansiosi rispetto alle loro abilità matematiche e mostrano una motivazione maggiore, sia intrinseca sia estrinseca rispetto alle ragazze (Else-Quest, Hyde e Linn, 2010). D'altro canto le studentesse mostrano un atteggiamento più favorevole verso la lettura e comprensione del testo sia nella scuola primaria sia nell'istruzione secondaria di II grado (Logan e Johnston, 2009; Lynn e Mikk, 2009).

In questo quadro, la relazione tra SES e genere è certamente molto interessante: non solo il genere e il SES mostrano, presi singolarmente, effetti significativi sulle performance scolastiche, ma l'interazione tra le due variabili mette in luce *patterns* caratteristici. Le studentesse con un alto livello di SES ottengono risultati migliori rispetto ai maschi (Ma, 2000; Cook, 2006) mentre le performance dei maschi sembrano peggiorare più rapidamente delle femmine mano a mano che il livello di SES si abbassa (Teese *et al.*, 1995; Considine e Zappala, 2002; Legewie e Di Prete, 2012).

Le ricerche comparative internazionali sugli apprendimenti, quali IEA-TIMSS (*Trends in International Mathematics and Science Study*) e PIRLS (*Progress in International Reading Literacy Study*) hanno sottolineato l'importanza sia dei fattori di genere sia di quelli socio-economici nei risultati scolastici rispettivamente in matematica e italiano. L'indagine PIRLS, nella sua ultima edizione del 2011, ha confermato a livello internazionale le migliori performance delle femmine rispetto ai maschi, per la classe quarta della scuola primaria, con una piccola riduzione del *gender gap* (Mullis *et al.*, 2012), ma non ha messo in luce alcuna significativa differenza per il contesto italiano. L'indagine TIMSS ha mostrato come per la classe quarta della primaria sia presente un'equità di genere nei risultati in matematica in molti Paesi, mentre il *pattern* per la classe terza della scuola secondaria di I grado si è caratterizzato per un'eterogeneità molto più evidente tra i diversi Paesi (Mullis *et al.*, 2012).

Anche l'indagine internazionale PISA (*Programme for International Student Assessment*), benché condotta su un campione di studenti di età sensibilmente diversa da quella presa in considerazione da questo studio (i quindicenni per PISA mentre, in questo lavoro, gli studenti della classe seconda e quinta della scuola primaria), ha studiato e confermato, nelle sue varie edizioni, l'importanza della relazione tra background socio-economico e performance degli studenti (OECD, 2010). I risultati a livello internazionale di PISA 2012 hanno messo in luce come maschi e femmine differiscano nei livelli di performance in matematica, lettura e comprensione del testo e scienze, ma con *patterns* specifici in relazione alle differenze all'interno del gruppo più che tra i due gruppi, quello dei maschi e quello delle femmine (OECD, 2013). In particolare, le differenze si fanno più evidenti per la prova di italiano, nella quale le studentesse ottengono punteggi superiori in quasi tutti i Paesi partecipanti. Per l'Italia, i dati confermano le evidenze del campione internazionale, con una performance significativamente migliore dei maschi in matematica (con un divario maggiore rispetto alla media OCSE) e delle femmine in italiano (INVALSI, 2015).

I dati raccolti dall'Istituto nazionale per la valutazione del sistema educativo di istruzione e formazione (INVALSI) per il campione italiano (INVALSI, 2015), mettono in luce differenze tra gli alunni rispetto al loro status socio-economico per la classe quinta della scuola primaria (livello 5) e la classe seconda della scuola secondaria di II grado (livello

10). I risultati mostrano un aumento dei punteggi, sia in italiano sia in matematica, all'aumentare del livello socio-economico di partenza e differenze significative nelle performance tra maschi e femmine: a parità di altre condizioni (status socio-economico, background migratorio e regolarità negli studi), per la classe quinta della scuola primaria il punteggio di italiano cresce di 1,6 punti percentuali per le femmine (e di 3,2 punti percentuali per il livello 10), mentre quello in matematica decresce di 4 punti percentuali per le femmine che frequentano l'ultimo anno della scuola primaria (e di 8,3 punti percentuali per il livello 10).

2. Ipotesi della ricerca

Dalle analisi riportate nei rapporti di ricerca che l'INVALSI pubblica al termine di ogni anno scolastico¹, emerge chiaramente che, in linea con la letteratura di settore, sia lo status socio-economico sia il genere hanno un effetto statisticamente significativo sul punteggio conseguito alle prove INVALSI, già a partire dalla scuola primaria. Sebbene lo studio di queste due variabili e degli effetti che essi hanno sulla performance scolastica sia stato oggetto di frequenti approfondimenti, a oggi relativamente meno esplorato è invece lo studio dell'interazione tra SES e genere (Strand, 2014), in particolare con riferimento a studenti della scuola primaria. L'obiettivo di questo lavoro è, quindi, quello di controllare empiricamente se esiste ed è statisticamente significativa l'interazione tra SES e genere.

Le nostre ipotesi di ricerca sono:

- *HP1*. Esiste un effetto d'interazione tra SES e genere in grado di "spiegare" la differenza nei punteggi ottenuti dagli studenti della scuola primaria alle prove INVALSI di italiano e matematica;
- *HP2*. L'azione congiunta di SES e genere ha effetti diversi sul punteggio ottenuto dagli studenti al test in gradi scolastici differenti.

3. Metodologia

3.1. I dati: un approccio pseudo-longitudinale

L'INVALSI somministra agli studenti di classe seconda (livello 2) e quinta (livello 5) della scuola primaria, terza secondaria di I grado (livello 8) e seconda secondaria di II grado (livello 10), ogni anno, e censuariamente, due test psicometrici tesi alla valutazione dell'abilità in matematica e in italiano.

Sebbene i dati raccolti da INVALSI non siano longitudinali, in questo studio, essi sono stati comunque analizzati in prospettiva diacronica.

Ogni anno, infatti, parallelamente alla rilevazione censuaria, l'INVALSI somministra, per ciascun livello scolastico coinvolto nella rilevazione, i test di italiano e di matematica a un campione di studenti statisticamente rappresentativo dell'intera popolazione nazionale. Sebbene quindi non sia possibile seguire, nel tempo, l'evoluzione della performance scolastica del singolo studente, è però possibile confrontare i risultati ottenuti dal campione nazionale di un certo anno con quello di un altro anno, sapendo che i due campioni sono statisticamente rappresentativi della stessa popolazione studentesca.

Sebbene quest'approccio certamente non consenta un'analisi longitudinale, esso rende comunque possibile il confronto delle performance ottenute da gruppi di studenti statisticamente equivalenti e rappresentativi della medesima popolazione studentesca.

A questo scopo, abbiamo quindi utilizzato le risposte date sia alla prova di italiano sia a quella di matematica dagli studenti che, nel 2012, componevano le classi campione della seconda primaria, e, nel 2015, quelle della quinta primaria (tab. 1).

¹ I rapporti INVALSI sono scaricabili dall'area prove del sito istituzionale, al link <http://www.invalsi.it/areaprove/index.php>.

Tab. 1 – I dati INVALSI in una prospettiva pseudo-longitudinale

Livello scolastico	2012		2013		2014		2015	
	Popolazione (N)	Campione (%)	Popolazione (N)	Campione (%)	Popolazione (N)	Campione (%)	Popolazione (N)	Campione (%)
2	504.342	6,3	497.813	5,0	497.300	5,3	418.456	5,0
5	489.580	6,3	483.921	5,1	477.944	5,3	412.743	5,1
8	519.010	100,0	520.918	100,0	520.917	100,0	520.920	100,0
10	413.847	10,1	418.243	9,1	410.609	9,0	288.348	9,5

3.2. Il modello di Rasch e l'analisi della varianza

Le risposte date dagli studenti agli item contenuti in ciascuna delle prove sono state analizzate con il modello di Rasch per stimare sia l'abilità in matematica sia in lettura e comprensione del testo.

Il modello di Rasch è particolarmente adeguato per le finalità di questo studio per almeno due principali ragioni: l'invarianza della misurazione e la qualità dei dati.

L'invarianza della misurazione è una caratteristica esclusiva del modello di Rasch, la quale postula l'esistenza di una misura *non-case-sensitive*. Essa cioè assicura che la valutazione dell'abilità dei soggetti sia effettuata indipendentemente dalla difficoltà degli item (*test free*) e, viceversa, che la valutazione della difficoltà della prova sia effettuata indipendentemente dalla composizione del campione (*sample free*). In altri termini, uno studente che possieda un certo livello di abilità in una disciplina conseguirà un certo punteggio a qualsiasi possibile batteria di item indipendentemente dalla difficoltà della domanda e, simmetricamente, ciascun item avrà un certo livello di difficoltà indipendentemente dall'abilità dello studente a cui è somministrato.

Una volta accertata la coerenza tra dati e modello², la rilevazione (dell'abilità) perde la casualità dovuta alle possibili variazioni nell'abilità degli studenti o nella difficoltà della prova, e cioè è invariante, legittimando quindi il confronto tra le performance ottenute a un certo test da gruppi indipendenti di studenti quale che sia la variabile di stratificazione.

Un'altra caratteristica del modello di Rasch, assai utile ai fini della nostra analisi, è che l'abilità stimata è espressa su una scala a intervalli equivalenti, *quasi-metrica*. Tale caratteristica consente di utilizzare le stime dei parametri effettuate con il modello di Rasch come dati di input in altre tecniche di analisi quantitativa, come per esempio l'analisi della varianza.

Dopo aver verificato la significatività statistica delle differenze medie tra i punteggi su scala di Rasch tramite un t-test per campioni indipendenti, è stata utilizzata un'ANOVA fattoriale (a due vie) per controllare la significatività statistica del possibile effetto d'interazione tra il genere e lo status socio-economico sull'abilità stimata con il modello di Rasch.

3.3. Un indice alternativo di status socio-culturale

Similmente a quanto fatto nelle indagini OCSE-PISA, INVALSI calcola l'indice ESCS, un acronimo che sta per *Economic and Socio-Cultural Status*. L'ESCS fa infatti riferimento a tre dimensioni:

- 1) il livello d'istruzione dei genitori espresso in anni d'istruzione formale calcolata secondo standard internazionali;
- 2) lo status occupazionale del padre e della madre;
- 3) il possesso di alcuni beni materiali intesi come variabili di prossimità di un contesto economico-culturale favorevole all'apprendimento, rilevato attraverso la somministrazione del Questionario studente.

² Il controllo empirico del *fit* è una fase particolarmente importante perché il modello di Rasch gode di alcune proprietà se e soltanto se la congruenza tra i dati e il modello è adeguata. In questo studio, come sempre accade con campioni di grandi dimensioni, il controllo del *fit* ha seguito procedure in parte diverse rispetto a quelle tradizionali. Com'è noto, infatti, per dati di grande dimensione nessun set di dati è in grado di *fit* perfettamente gli assunti teorici di qualsivoglia modello e, quindi, quasi automaticamente, ma non per questo a ragione, ci si troverebbe a rifiutare praticamente qualsiasi modello, rendendo di fatto impraticabile qualsiasi tipo di analisi (Gustafson, 1980). Per questa ragione, in questi casi, la verifica del *fit* non può passare, esclusivamente, attraverso l'uso tradizionale degli indici di *infit* e *oufit*, ma deve piuttosto ricorrere a intervalli di tolleranza entro cui verificare che tali indici ricadano.

Il Questionario studente però è somministrato solo agli studenti di quinta primaria e di seconda secondaria di II grado. Informazioni sul benessere economico della famiglia di origine sono quindi disponibili solo per questi due livelli. Cionondimeno, per tutti i gradi scolastici sono disponibili informazioni relative al livello di istruzione e allo status professionale, sia del padre sia della madre. Queste variabili sono state utilizzate per costruire un indice di status socio-culturale chiamato SCINDEX, un indice tipologico che combina le informazioni relative all'istruzione dei genitori con il loro status professionale. Similmente a quanto fatto per la costruzione dell'ESCS, anche per lo SCINDEX si combina il più alto livello di istruzione tra quello del padre e quello della madre (definito in base alla classificazione internazionale ISCED) con il più alto status professionale tra quello del padre e quello della madre (definito in base alla classificazione adottata da INVALSI – Campodifiori *et al.*, 2010). Dalla combinazione di queste due informazioni derivano tre categorie in cui lo SCINDEX è articolato (tab. 2).

Tab. 2 – SCINDEX

Livello di istruzione	Status professionale				
	Disoccupato	Casalinga	Operaio	Impiegato	Imprenditore/ lavoratore autonomo
Basso	Basso	Basso	Basso	Medio	Medio
Medio	Basso	Basso	Basso	Medio	Alto
Alto	Medio	Medio	Medio	Alto	Alto

4. Risultati

Sulla base delle risposte date dagli studenti della classe seconda e quinta primaria, rispettivamente nel 2012 e nel 2015, agli item contenuti nelle prove, è stata stimata l'abilità sia in matematica sia in lettura e comprensione del testo e, conseguentemente, creata una scala con media 200 e deviazione standard 40, similmente a quanto fatto dall'INVALSI nei rapporti di ricerca pubblicati da INVALSI al termine di ciascun anno scolastico. I punteggi conseguiti dagli studenti su scala di Rasch sono stati utilizzati come variabile dipendente nell'analisi della varianza fattoriale.

Prima di procedere con quest'ultima, per ciascuna delle quattro prove analizzate, abbiamo controllato che non ci fossero *outliers* e/o punti estremi, che le risposte degli studenti si distribuissero in modo almeno approssimativamente normale, e infine abbiamo verificato l'ipotesi di omoschedasticità. L'ispezione grafica dei box-plot ha rilevato solo alcuni punti estremi, ma in numero talmente tanto esiguo (meno di dieci casi, complessivamente, per ciascuna prova) da ritenere opportuno procedere semplicemente eliminando tali studenti. Lo studio della distribuzione delle risposte ha inoltre confermato, per ciascun dataset, un andamento approssimativamente normale. Infine, abbiamo controllato, per ciascuno dei quattro dataset, che le varianze fossero omoschedastiche (Faraway, 2014).

Dalle analisi effettuate è emerso che la performance dei maschi è inferiore rispetto a quella delle femmine, in lettura e comprensione del testo, in ciascuna delle tre classi di status socio-culturale, sia in seconda sia in quinta primaria (tab. 3). Le differenze osservate, anche se modeste, sono tutte statisticamente significative, come confermato dal t-test per campioni indipendenti.

Differenze simili ma di segno opposto, che cioè indicano un vantaggio dei maschi rispetto alle femmine, si osservano in relazione ai punteggi medi stimati dal modello di Rasch per la prova di matematica, sia nel 2012 sia nel 2015 (tab. 3). Anche in questo caso, le differenze sono modeste ma statisticamente significative.

Questo risultato, pienamente coerente con quanto riportato nei rapporti INVALSI (2012 e 2015), è stato approfondito attraverso lo studio dell'interazione tra genere e status socio-culturale.

Tab. 3 – Abilità media intra-gruppo

Genere	SCINDEX	2012 italiano			2015 italiano			2012 matematica			2015 matematica		
		Media	Dev. std.	N	Media	Dev. std.	N	Media	Dev. std.	N	Media	Dev. std.	N
Maschio	Basso	202,06	37,40	3.381	188,18	39,73	94.067	185,42	40,75	21.197	196,17	40,89	3.826
	Medio	201,02	39,03	10.426	199,94	38,97	89.489	189,33	37,41	54.182	205,00	39,93	3.527
	Alto	205,07	39,17	3.141	212,20	39,42	64.480	204,24	37,57	175.253	217,06	40,02	2.542
	Totale	201,72	38,54	17.639	198,16	40,53	285.951	199,43	38,53	250.632	204,68	41,15	9.895
Femmina	Basso	208,85	34,26	2.858	192,34	39,32	92.831	184,13	37,54	21.127	187,76	37,43	3.670
	Medio	205,09	39,41	10.453	202,47	37,21	84.514	187,09	35,06	52.691	197,83	36,70	3.425
	Alto	204,42	41,78	2.510	215,14	37,80	61.091	200,24	35,25	169.428	210,18	38,04	2.396
	Totale	205,55	38,61	16.352	201,61	39,24	275.065	195,99	36,00	243.246	197,05	38,34	9.491
Totale	Basso	205,17	36,15	6.239	190,25	39,58	186.898	184,78	39,18	42.324	192,05	39,46	7.496
	Medio	203,01	39,29	20.902	201,17	38,15	174.003	188,22	36,29	106.873	201,47	38,54	6.952
	Alto	204,78	40,35	5.651	213,63	38,66	125.571	202,27	36,50	344.681	213,72	39,22	4.938
	Totale	203,53	38,63	34.014	199,85	39,95	561.080	197,73	37,35	493.878	200,95	39,98	19.386

Nota: Differenze statisticamente significative allo 0,01.

Sia per l'italiano (tab. 4) sia per la matematica (tab. 5), l'interazione tra genere e status socio-culturale risulta statisticamente significativa.

Tab. 4 – Test di effetti tra soggetti (prova di italiano, anni 2012, 2015)

Origine	2012 italiano				2015 italiano			
	gl	F	Sign.	η^2 parziale	gl	F	Sign.	η^2 parziale
Modello corretto	8	24,3	0,00	0,006	8	3.607,9	0,00	0,05
Intercetta	1	11.811,3	0,00	0,258	1	3.2861,6	0,00	0,06
Genere	2	37,5	0,00	0,002	2	626,7	0,00	0,00
SCINDEX	3	14,1	0,00	0,001	3	9.182,2	0,00	0,05
Genere * SCINDEX	3	10,0	0,00	0,001	3	34,9	0,00	0,00
Errore	34.005				561.071			
Totale	34.014				561.080			
Totale corretto	34.013				561.079			

Tab. 5 – Test di effetti tra soggetti (prova di matematica, anni 2012, 2015)

Origine	2012 italiano				2015 italiano			
	gl	F	Sign.	η^2 parziale	gl	F	Sign.	η^2 parziale
Modello corretto	5	3.783,4	0,00	0,037	5	222,4	0,0	0,054
Intercetta	1	6.863.418,6	0,00	0,933	1	507.593,0	0,0	0,963
Genere	1	294,3	0,00	0,001	1	173,9	0,0	0,009
SCINDEX	2	8.859,3	0,00	0,035	2	462,3	0,0	0,046
Genere * SCINDEX	2	42,7	0,00	0,000	2	0,7	0,5	0,000
Errore	493.872				19.380			
Totale	493.878				19.386			
Totale corretto	493.877				19.385			

L'ultima colonna (η^2 parziale) riporta l'effect size, cioè la quota parte di varianza osservata nella variabile dipendente che può essere attribuita alla variabile indipendente.

Secondo la classificazione di Cohen (1988), l'*effect size* è piccolo quando prossimo allo zero, medio quando vicino a 0,059 e ampio quando invece prossimo a 0,138. In ciascuna delle analisi effettuate, il valore dell' η^2 parziale è sempre molto basso e indica quindi che l'interazione tra il genere e lo status socio-culturale è in grado di "spiegare" poco della varianza nel punteggio conseguito dagli studenti ai test, sia per l'italiano sia per la matematica. Gli *effects size* sono bassi anche relativamente ai singoli fattori, sia per l'italiano sia per la matematica. Valori leggermente più alti si osservano solo in relazione allo status socio-culturale il quale dà conto della varianza nel punteggio conseguito dagli studenti nella misura del 3,5% per gli studenti di seconda primaria ($\eta^2 = 0,035$) e del 4,6% per gli studenti di quinta ($\eta^2 = 0,046$).

Questo risultato è certamente coerente con quanto precedentemente osservato in tab. 3 e con quanto riportato anche nei rapporti INVALSI (2012 e 2015): la varianza nei punteggi conseguiti da maschi e femmine, anche se piccola, è statisticamente significativa così come statisticamente significativo è l'effetto di interazione tra status socio-culturale e genere.

I grafici riportati di seguito mostrano la differenza nei punteggi ottenuti sia nelle prove di italiano (figg. 1 e 2) sia in quelle di matematica (figg. 3 e 4), approfondendo i risultati riportati nelle tabelle precedenti e consentendoci di comprendere meglio le differenze nei punteggi conseguiti ai test in funzione del genere e dello status socio-culturale. Ciascuna tavola è divisa in due colonne: nella prima, possiamo osservare la differenza nei punteggi di maschi e femmine in funzione del livello di status socio-culturale; nella seconda, invece, possiamo confrontare i punteggi ottenuti da studenti con status socio-culturale basso, medio e alto in funzione del genere.

Per le prove di italiano, le differenze nel punteggio su scala di Rasch (con media 200 e deviazione standard 40) mostrano, per la prova di seconda primaria, differenze molto contenute tra maschi e femmine con basso livello di status socio-culturale. Tali differenze tendono poi a ridursi fino ad annullarsi per livelli di status socio-culturale via via più alti. È inoltre assai interessante osservare che per il livello socio-culturale più alto, la performance dei maschi risulta leggermente superiore a quella delle femmine, in netta contro-tendenza rispetto alla letteratura di settore. In questo caso, si osserva inoltre un chiaro effetto di interazione, di tipo *disordinal*, mentre le interazioni osservate nelle altre tavole (figg. 2-4) sono invece di tipo ordinale. Cionondimeno, la differenza nei punteggi osservati è estremamente contenuta e occorrerebbe formulare ipotesi precise per l'interpretazione di un risultato che, in termini strettamente numerici, sembrerebbe invece di rilevanza trascurabile.

Nel passaggio dalla classe seconda alla classe quinta primaria, le differenze tra maschi e femmine si amplificano, segnando inoltre, in linea con la letteratura di settore, un netto, seppure sempre contenuto, vantaggio delle femmine rispetto ai maschi.

Per quanto concerne invece la prova di matematica (figg. 3 e 4), l'Analisi della varianza produce risultati molto più netti. Oltre a evidenziare una differenza nei punteggi a vantaggio, stavolta, dei maschi, anche lo status socio-culturale ha un effetto molto più netto. Esso è infatti in grado di spiegare, in particolare in seconda primaria, una differenza nei punteggi chiaramente più ampia rispetto a quella osservata per l'italiano, tra gli studenti con status socio-culturale medio e basso rispetto a quelli con SCINDEX alto. Tale differenza resta confermata anche in quinta primaria, ma in modo certamente meno marcato. La differenza nei punteggi osservata per maschi e femmine resta, però, comunque sostanzialmente costante, indipendentemente dallo status socio-culturale, sia in seconda sia in quinta primaria.

Fig. 1 – Effetti di interazione tra genere e status socio-culturale osservato per la prova di italiano del 2012

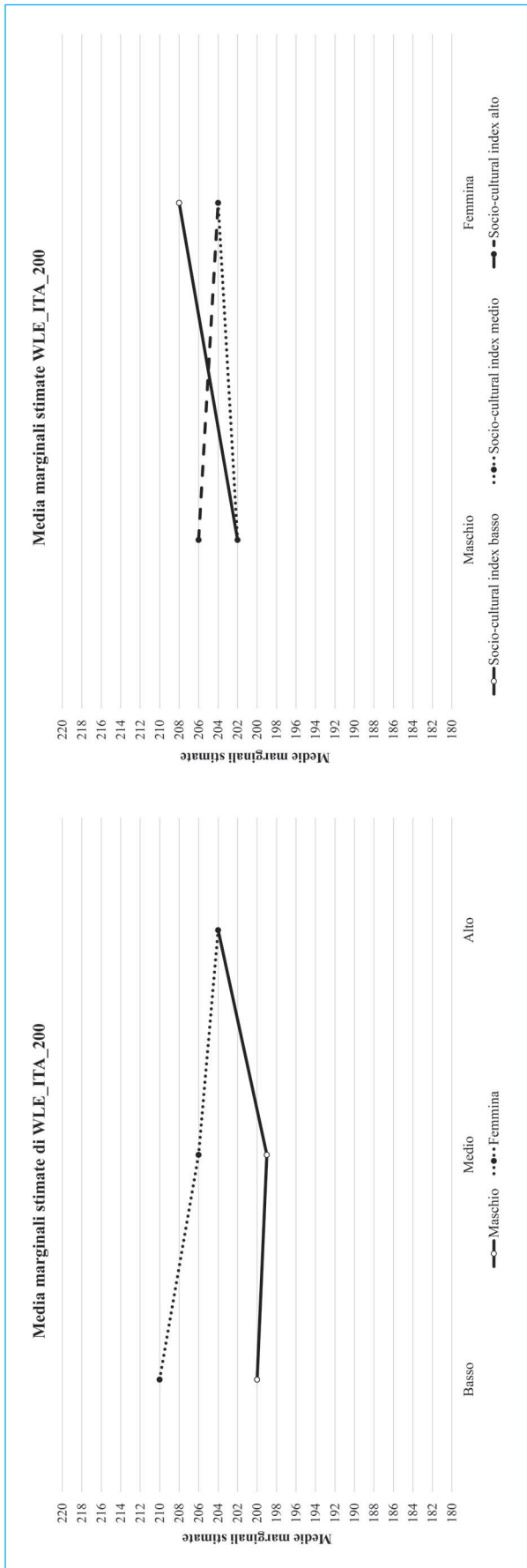


Fig. 2 – Effetti di interazione tra genere e status socio-culturale osservato per la prova di italiano del 2015

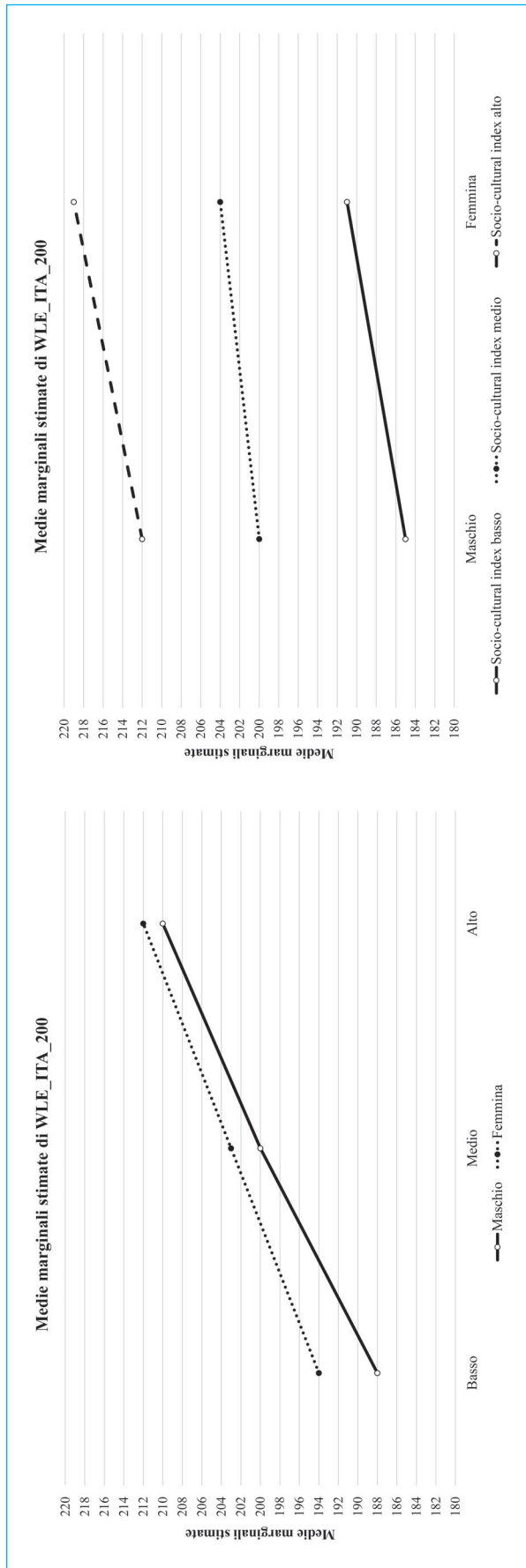


Fig. 3 – Effetti di interazione tra genere e status socio-culturale osservato per la prova di matematica del 2012

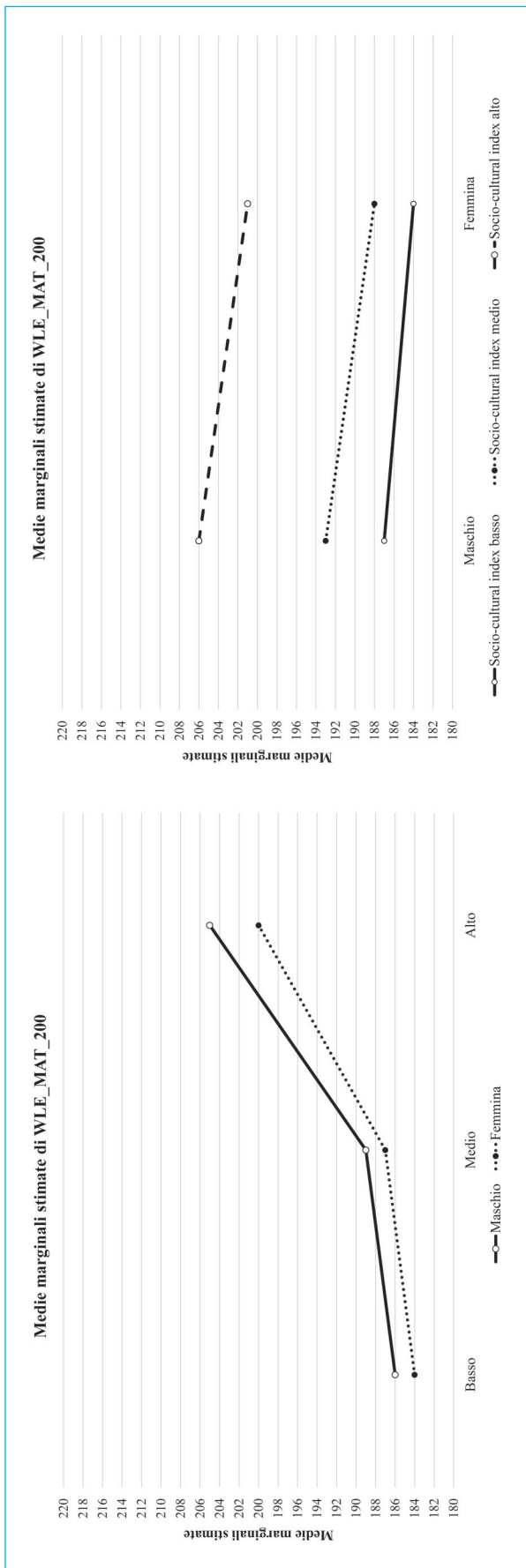
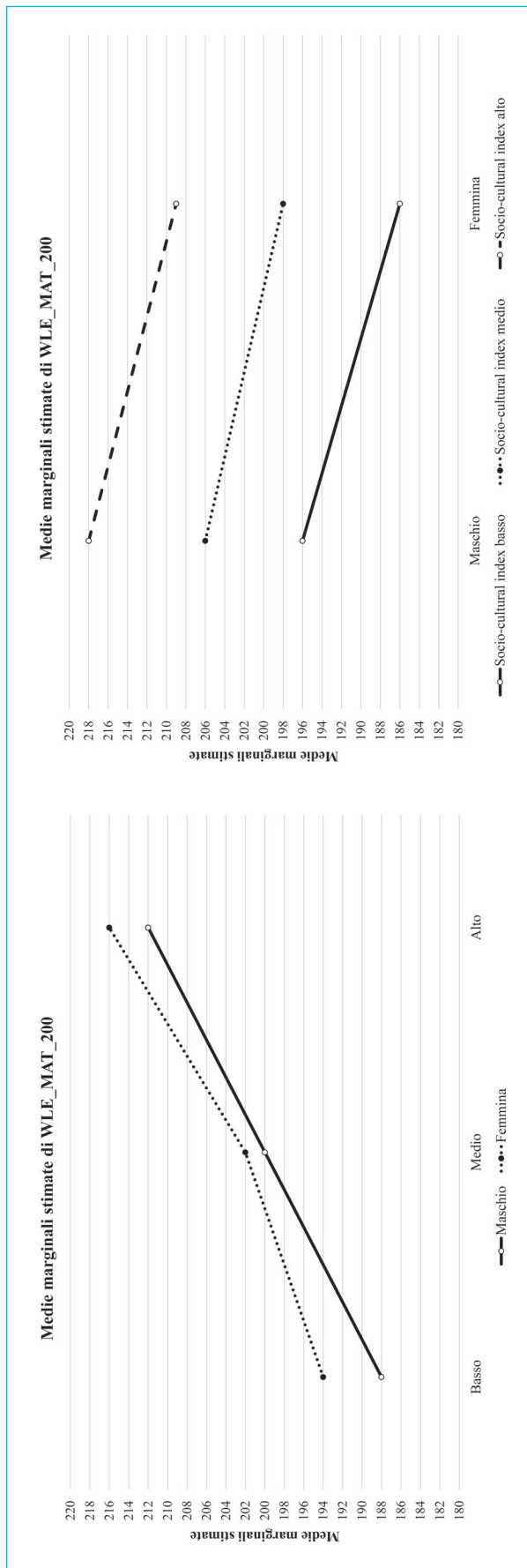


Fig. 4 – Effetti di interazione tra genere e status socio-culturale osservato per la prova di matematica del 2015



5. Conclusioni

Il nostro studio ha inteso verificare su dati italiani l'esistenza di un effetto d'interazione tra il background familiare e il genere utile a spiegare la differenza nei punteggi ottenuti dagli studenti della scuola primaria alle prove INVALSI di italiano e matematica; se e come l'interazione tra background familiare e genere dello studente abbia effetti diversi sul punteggio ottenuto al test in gradi scolastici differenti.

L'analisi dei dati sulla scuola primaria ha confermato, in linea con la più recente letteratura di settore, la significatività statistica di un effetto di interazione tra lo status socio-culturale della famiglia di origine dello studente e il genere.

Dal confronto dei risultati ottenuti sui dati del 2012 (seconda primaria) con quelli del 2015 (quinta primaria) emerge molto chiaramente, e in linea con la letteratura di settore (Sammons, 1995; Eurydice, 2010), che sia le differenze di genere sia di status socio-culturale aumentano nel corso del tempo.

Per la matematica, per esempio, in seconda primaria, le differenze maggiori si osservano soprattutto in funzione dello status socio-culturale della famiglia di origine dello studente, con effetti più marcati e positivi per i maschi che per le femmine. Le differenze di genere sono invece più contenute per il livello medio e basso di status socio-culturale, per i quali, oltre a osservarsi un drastico calo di performance medie, le differenze tra maschi e femmine si attestano, mediamente, intorno ai dieci punti sulla scala di Rasch.

Per lo stesso grado scolastico (livello 2), l'Analisi della varianza fattoriale ha evidenziato, per l'italiano, un chiaro effetto di interazione non ordinale, con differenze tra i punteggi di maschi e femmine statisticamente significativi ma di entità abbastanza contenuta (meno di 10 punti su scala di Rasch). Cionondimeno, occorrerebbe indagare più a fondo il risultato ottenuto, come spesso si fa in letteratura per effetti di interazione statisticamente significativi e non ordinali, per riuscire a interpretare correttamente differenze che, in termini strettamente numerici, sembrerebbero invece di rilevanza trascurabile.

Ciò che fin dalla nostra analisi emerge con chiarezza è il rapporto in cui il genere e lo status socio-culturale si pongono nello spiegare le differenze nell'abilità in lettura e comprensione del testo. Infatti, sebbene, in coerenza con la letteratura di settore, le differenze osservate tra maschi e femmine in funzione delle singole variabili tendano ad ampliarsi nel corso del tempo, non si può dire lo stesso anche in merito agli effetti di interazione. Prendiamo per esempio il caso della prova di italiano, somministrata nel 2012 agli studenti di seconda primaria. Dall'analisi emerge infatti che il leggero vantaggio delle bambine tende a ridursi progressivamente, fino ad annullarsi completamente, all'aumentare del livello socio-culturale della famiglia di origine. Questo risultato non viene però confermato in quinta primaria, in cui pur mostrando ancora uno svantaggio rispetto alle femmine, i maschi conseguono livelli di performance appena inferiori, in ciascuno dei tre livelli di status socio-culturale.

Queste tendenze emergono ancora più chiaramente in relazione alle prove di matematica somministrate sia nel secondo sia nel quinto grado della scuola primaria: per entrambi i gradi scolastici, diversamente da quanto osservato in seconda primaria per la prova di italiano, i punteggi conseguiti dai maschi e dalle femmine sono sostanzialmente uguali, a meno di trascurabili differenze.

Lo studio presentato in questo lavoro ha avuto come principale obiettivo quello di "importare" schemi di analisi e domande della ricerca che hanno interessato, soprattutto negli ultimi anni, il dibattito scientifico, ma pare aver avuto anche il pregio di mettere in evidenza alcuni aspetti della realtà scolastica italiana che, seppure in parte, si discostano dalle evidenze empiriche internazionali. Inoltre, sebbene di modesta entità, le interazioni tra status socio-culturale e genere risultano tutte statisticamente significative e sembrano suggerire alcune prime riflessioni, come per esempio la capacità della scuola primaria italiana di poter ridurre, nel corso del tempo, le differenze di genere in lettura e comprensione del testo per studenti con un background socio-culturale svantaggiato. Queste prime evidenze indirizzano quindi verso ulteriori approfondimenti che possano realmente chiarire sia il rapporto tra background socio-culturale e performance sia il modo in cui la scuola opera su tali differenze per costruire reali condizioni di equità tra gli studenti.

Riferimenti bibliografici

- Aikens N.L., Barbarin O. (2008), "Socioeconomic differences in reading trajectories: the contribution of family, neighborhood, and school contexts", *Journal of Educational Psychology*, 100, 2: 235-251.
- Bornstein M.H., Bradley R.H. (eds.) (2003), *Socioeconomic Status, Parenting, and Child Development*, Erlbaum, Mahwah (NJ).
- Brooks-Gunn J., Duncan G.J. (1997), "The effects of poverty on children", *The Future of Children*, 7, 2: 55-71.

-
- Buckingham J. (1999), "The puzzle of boys educational decline: a review of the evidence", *Issue Analysis*, 9, Centre for Independent Studies, Sydney.
- Campodifiori E., Figura E., Papini M., Ricci R. (2010), "Un indicatore di status socio-economico-culturale degli allievi della quinta primaria in Italia", *INVALSI Working Papers*, 2, testo disponibile al sito: http://www.invalsi.it/download/wp/wp02_Ricci.pdf, data di consultazione: 3 aprile 2017.
- Cascella C. (2016), "Differenze di genere nell'apprendimento matematico e variabili di contesto: l'effetto dello status socio-economico sui punteggi di maschi e femmine ai test INVALSI", in P. Paoloni (a cura di), *I mondi delle donne. Percorsi interdisciplinari*, Edicusano, Roma: 53-75.
- Cohen J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Erlbaum, Hillsdale (NJ), 2nd ed.
- Coleman J.S. (1988), "Social capital and schools", *Education Digest*, 53, 8: S95-S120.
- Coley R.L., Morris J.E. (2002), "Comparing father and mother reports of father involvement among low-income minority families", *Journal of Marriage and Family*, 64, 4: 982-997.
- Considine G., Zappala G. (2002), "Factors influencing the educational performance of students from disadvantaged background", in T. Eardley, B. Bradbury (eds.), *Competing Visions: Refereed Proceedings of the National Social Policy Conference 2001*, SPRC Report 1/02, Social Policy Research Centre, University of New South Wales, Sydney.
- Cook G. (2006), "Boys at risk: the gender achievement gap", *American School Board Journal*, 193: 4-6.
- Dahl G.B., Lochner L. (2012), "The impact of family income on child achievement: evidence from the earned income tax credit", *American Economic Review*, 102, 5: 1927-1956.
- Else-Quest N.M., Hyde J.S., Linn M.C. (2010), "Cross-national patterns of gender differences in mathematics: a meta-analysis", *Psychological Bulletin*, 136: 103-127.
- Ensminger M.E., Fothergill K.E. (2003), "A decade of measuring SES: what it tells us and where to go from here", in M.H. Bornstein, R.H. Bradley (eds.), *Socioeconomic Status, Parenting, and Child Development*, Erlbaum, Mahwah (NJ): 13-27.
- Eurydice E.C. (2010), *Gender Differences in Educational Outcomes: Study of the Measures Taken and the Current Situation in Europe*, Education, Audio-visual and Culture Executive Agency, Brussels.
- Faraway J.J. (2014), *Linear Models with R*, CRC Press, Taylor & Francis, Broken Sound Parkway (NW).
- Gustafson J.E. (1980), "Testing and obtaining fit of data to the Rasch model", *British Journal of Mathematical and Statistical Psychology*, 33: 205-233.
- Igbo J.N., Onu V.C., Obiyo N.O. (2015), "Impact of gender stereotype on secondary school students' self-concept and academic Achievement", *Sage Open*, 5, 1, testo disponibile al sito: <http://journals.sagepub.com/doi/full/10.1177/2158244015573934>, data di consultazione: 3 aprile 2017.
- INVALSI (2012), *Rilevazioni nazionali sugli apprendimenti 2011-2012. Rapporto tecnico*, INVALSI, Frascati, testo disponibile al sito: http://www.invalsi.it/snv2012/documenti/Rapporti/Rapporto_rilevazione_apprendimenti_2012.pdf, data di consultazione: 3 aprile 2017.
- INVALSI (2015), *Rilevazioni nazionali degli apprendimenti 2014-2015*, INVALSI, Frascati, testo disponibile al sito: http://www.invalsi.it/invalsi/doc_evidenza/2015/034_Rapporto_Prove_INVALSI_2015.pdf, data di consultazione: 3 aprile 2017.
- Kenney-Benson G.A., Pomerantz E.M., Ryan A.M., Patrick H. (2006), "Sex differences in math performance: the role of children's approach to schoolwork", *Developmental Psychology*, 42, 1: 11-26.
- Legewie J., Di Prete T.A. (2012), "School context and gender gap in educational achievement", *American Sociological Review*, 77, 3: 463-485.
- Lindberg S.M., Hyde J.S., Petersen J.L., Linn M.C. (2010), "New trends in gender and mathematics performance: a meta-analysis", *Psychological Bulletin*, 136: 1123-1135.
- Lindsay G., Muijs D. (2006), "Challenging underachievement in boys", *Educational Research*, 48, 3: 313-332.
- Logan S., Jonhston R. (2009), "Gender differences in reading ability and attitudes: examining where these differences lie", *Journal of Research in Reading*, 32, 2: 199-214.
- Lynn R., Mikk J. (2009), "Sex differences in reading achievement", *Trames*, 13: 3-13.
- Ma X. (2000), "Socioeconomic gaps in academic achievement within schools: are they consistent across subject areas?", *Educational Research and Evaluation*, 6, 4: 337-355.
- McLoyd V.C. (1998), "Socioeconomic disadvantage and child development", *The American Psychologist*, 53, 2: 185-204.
- Morgan P.L., Farkas G., Hillemeier M.M., Maczuga, S. (2009), "Risk factors for learning-related behavior problems at 24 months of age: Population-based estimates", *Journal of Abnormal Child Psychology*, 37, 38: 401-413.
- Mullis I.V.S., Martin M.O., Fierros E.G, Goldberg A.L. Stemler S.E. (2000), *Gender Differences in Achievement. IEA's Third International Mathematics and Science Study (TIMSS)*, TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill (MA).
- Mullis I.V.S., Martin M.O., Foy P., Arora A. (2012), *TIMSS 2011 International Results in Mathematics*, TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill (MA).
-

-
- Mullis I.V.S., Martin M.O., Foy P., Drucker K.T. (2012), *PIRLS 2011 International Results in Reading*, TIMSS & PIRLS International Study Center, Boston College, Chestnut Hill (MA).
- OECD (2010), *PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes*, vol. II, OECD Publishing, Paris.
- OECD (2013), *PISA 2012 Results: What Students Know and can do: Student Performance in Mathematics, Reading and Science*, vol. I, OECD Publishing, Paris.
- Ream R.K., Palardy G.J. (2008), “Reexamining social class differences in the availability and the educational utility of parental social capital”, *American Educational Research*, 45, 2: 238-273.
- Rippla C.H., Luthar S.S. (2000), “Academic risk among inner-city adolescents: The role of personal attributes”, *Journal of School Psychology*, 38 (3): 277-298.
- Sammons P. (1995), “Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years”, *British Educational Research Journal*, 21, 4: 465-485.
- Seyfried S.F. (1998), “Academic achievement of African American preadolescents: The influence of teacher perceptions”, *American Journal of Community Psychology*, 26, 3: 381-402.
- Sirin S.R. (2005), “Socioeconomic status and academic achievement: a meta-analytic review of research”, *Review of Educational Research*, 75, 3: 417-453.
- Stoet G., Geary D.C. (2013), “Sex differences in mathematics and reading achievement are inversely related: within-and across-nation assessment of 10 years of PISA data”, *PloS One*, 8, 3: e57988.
- Strand S. (2014), “School effects and ethnic, gender and socio-economic gaps in educational achievement at age 11”, *Oxford Review of Education*, 40, 2: 223-245.
- Teese R., Davies M., Charlton M., Polesel J. (1995), *Who Wins at School? Boys and Girls in Australian Secondary Education*, Macmillan, Sydney.
- Voyer D., Voyer S. (2014), “Gender differences in scholastic achievement: a meta-analysis”, *American Psychological Association*, 140, 4: 1174-1204.
- White K. (1982), “The relation between socioeconomic status and academic achievement”, *Psychological Bulletin*, 91: 461-481.

Gli autori

Giovanni Abbiati è attualmente Research Fellow presso l'istituto di ricerca valutativa FBK-IRVAPP. I suoi principali interessi di ricerca sono costituiti dalla valutazione delle politiche pubbliche in campo educativo e dallo studio dei processi di stratificazione sociale.

Gianluca Argentin è ricercatore presso il Dipartimento di Sociologia dell'Università Cattolica del Sacro Cuore di Milano. Sociologo dell'educazione si occupa di valutare l'impatto delle politiche educative, soprattutto mediante sperimentazioni controllate, e di analizzare le condizioni di lavoro degli insegnanti italiani.

Giorgio Bolondi, PhD in Matematica, è professore di Matematica all'Università di Bologna, dove insegna Didattica della matematica e Storia della matematica. Svolge attività di ricerca e divulgazione in Didattica della matematica e ha pubblicato oltre 100 articoli e diversi libri. È stato presidente della Commissione italiana per l'insegnamento dell'UMI.

Rebecca Boninsegna, dopo aver conseguito la maturità scientifica si è laureata in Matematica presso l'Università di Bologna con una tesi riguardante la validazione di uno strumento statistico al fine di misurare l'impatto delle variabili redazionali di un item matematico sulla performance degli studenti.

Giulia Bovini è dottoranda in Economia presso la London School of Economics and Political Science. Ha conseguito la laurea specialistica e triennale presso l'Università di Bologna. Ha trascorso un periodo di ricerca presso la Banca d'Italia e ha lavorato, per un breve periodo, presso la Banca europea per la ricostruzione e lo sviluppo. I principali interessi di ricerca sono legati all'economia del lavoro, all'economia dell'istruzione e all'economia pubblica. Attualmente sta trascorrendo un periodo di ricerca presso l'INPS.

Nicola Luigi Bragazzi, LM in Medicina e Chirurgia, PhD in Nanobioteconologie, specializzando in Igiene e medicina preventiva, Università degli studi di Genova. Svolge ricerche in differenti aree, in particolare in ambito psico-sociale.

Laura Branchetti è laureata in Matematica e PhD in Didattica della matematica. È docente nella scuola secondaria di II grado e tiene laboratori e corsi di Didattica della matematica per studenti universitari e per insegnanti. È membro dei gruppi di ricerca in Didattica della matematica e della fisica dell'Università di Bologna.

Maria Brutto, nata a Catanzaro, il 18 dicembre 1969, docente di materie letterarie, latino e greco presso il liceo classico "P. Galluppi" di Catanzaro. Progettista fondi europei. Valutatrice e formatrice regionale SNV e dei piani nazionali Poseidon, Didatec, PNS. Appassionata di scrittura creativa e arti orientali.

Roberto Capone è laureato in Ingegneria chimica e in Matematica, insegna Matematica e Fisica nei licei ed è professore a contratto di Didattica della matematica all'Università del Molise. Collabora con il nucleo di Ricerca di didattica della matematica dell'Università di Salerno dove svolge anche attività didattica presso i Dipartimenti di Ingegneria e di Matematica. Collabora con l'USR Campania tenendo seminari e corsi di formazione rivolti ai docenti sulla didattica per competenze.

Clelia Cascella, PhD in Metodologia della ricerca (Università di Firenze) e in Economia (Università Federico II), è ricercatrice psicometrista presso l'INVALSI. Interessi di ricerca: modelli per l'analisi longitudinale (anche nell'ambito della network analysis) e di *Item Response Theory Models*.

Mario Castoldi, docente presso l'Università degli studi di Torino, svolge attività di formazione e consulenza su problematiche didattiche e valutative in collaborazione con scuole e reti di scuole.

Elisa Cavicchiolo, PhD in Sociologia e Scienze sociali applicate. Interessi: immigrazione, metodi misti. Pubblicazione: con F. Alivernini, S. Manganelli (2016), "Immigrants are like... The representation of immigrants in Italy: the metaphors used by students and their family backgrounds", *ECPS Journal*, 13, pp. 161-188.

Cristina Coppola è ricercatrice di Mat04 presso l'Università degli studi di Salerno e si occupa da molti anni di formazione docenti. La sua attività di ricerca è rivolta principalmente allo studio della rilevanza degli aspetti linguistici nell'apprendimento della matematica e al ruolo dei fattori non cognitivi nella formazione professionale matematica dei docenti di scuola dell'infanzia e primaria. È autore di vari lavori scientifici su rivista, su libro e in atti di convegni e ha tenuto varie comunicazioni scientifiche in convegni internazionali e nazionali.

Antonella Costanzo è dottore di ricerca in Economia d'impresa e analisi quantitative presso l'Università degli studi di Cassino. Dal 2014 è collaboratore tecnico di ricerca presso INVALSI. I suoi interessi di ricerca riguardano la modellistica statistica, con applicazioni in campo economico e sociale.

Marta De Philippis è economista presso il Dipartimento di Economia e statistica della Banca d'Italia, dove si occupa di mercato del lavoro e dei sistemi di istruzione. In passato ha lavorato per periodi brevi presso la London School of Economics e il Centre of Economic Performance, presso la fondazione Rodolfo De Benedetti e presso la Commissione europea. Ha conseguito il dottorato in economia politica alla London School of Economics nel marzo del 2016 e la laurea specialistica e triennale presso l'Università Bocconi. I suoi interessi di ricerca sono legati all'economia del lavoro, in particolare all'economia dell'istruzione e dell'immigrazione.

Simone Del Sarto ha conseguito il dottorato di ricerca in Statistica presso l'Università degli studi di Perugia. I suoi interessi di ricerca riguardano lo studio dei modelli IRT multidimensionali e la loro applicazione in campo educativo e i modelli bayesiani spazio-temporali.

Marta Desimoni è dottore di ricerca in Psicologia dinamica, clinica e dello sviluppo. Ricercatore presso INVALSI, la sua attività si colloca nel settore della psicometria e include la costruzione e validazione di test e l'applicazione di modelli statistici alla ricerca in ambito psicologico e educativo.

Giselda Di Cesare, docente di scuola primaria presso l'IC di Loreto Aprutino. Si dedica con passione e creatività all'insegnamento e partecipa in modo attivo alla vita scolastica nel proprio Paese. Collabora con l'Università di Chieti per implementare il progetto RTI promuovendo percorsi di buone prassi.

Sergio Di Sano, dal 2004 ricercatore per il settore M-PSI/04 "Psicologia dello sviluppo e psicologia dell'educazione", e docente di "Psicologia dello sviluppo cognitivo", presso l'Università di Chieti-Pescara G. d'Annunzio, svolge attività di ricerca nell'ambito dei disturbi di apprendimento.

Federica Ferretti, PhD in Matematica, è assegnista di ricerca presso il Dipartimento di Matematica e il Dipartimento di Scienze della formazione di UNIBO. Svolge da diversi anni ricerca in Didattica della matematica ed è professore a contratto presso l'Università di Bologna e l'Università di Bressanone in corsi di Didattica della matematica.

Michela Freddano, dal 2013 ricercatore all'INVALSI, nel 2012 consegue il titolo di dottore di ricerca in Valutazione dei processi e dei sistemi educativi presso l'Università degli studi di Genova, con la quale collabora da diverso tempo. Cultore della materia di Sociologia dei processi culturali all'Università LUMSA, i suoi interessi scientifici si rivolgono principalmente all'analisi sociologica dei sistemi e delle istituzioni scolastiche e alla valutazione delle politiche educative e scolastiche.

Alessandro Gambini, laurea in Matematica, svolge da anni attività di formazione e divulgazione nel campo della Matematica. È docente di Matematica generale del corso di laurea magistrale a ciclo unico Chimica e tecnologia farmaceutiche ed è project manager del progetto d'Ateneo "Alma Mathematica" di Unibo.

Tiziano Gerosa è assegnista di ricerca presso il Dipartimento di Sociologia dell'Università di Milano-Bicocca. Si occupa di processi educativi e di civiness, con un particolare interesse per le relative questioni misuratorie.

Chiara Giberti, laureata in Matematica e dottoranda in Didattica della matematica presso l'Università degli studi di Trento. Insegnante nella scuola secondaria di I grado, abilitata classe A059. Collabora con INVALSI e si interessa all'uso e all'interpretazione dei risultati delle prove standardizzate nell'insegnamento della matematica.

Michela Gnaldi, ricercatrice al Dipartimento di Scienze politiche, Università di Perugia, abilitata alla II fascia in attesa di chiamata. È stata responsabile scientifico di due convenzioni con INVALSI. È co-autrice del libro *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*.

Alice Lemmo, laurea in Matematica, svolge ricerca in Didattica della matematica, membro del Nucleo di ricerca in Didattica della Matematica di Bologna. Frequenta il dottorato in Didattica della matematica presso l'Università di Palermo. Ha partecipato a progetti di ricerca INVALSI e collabora all'analisi e alla codifica delle domande dell'SNV. Abilitata classe A059.

Andrea Maffia, laureato in Matematica, è dottorando in Didattica della matematica presso il Dipartimento di Educazione e scienze umane dell'Università di Modena e Reggio Emilia. Svolge ricerca sull'apprendimento dell'aritmetica nel primo ciclo scolastico con una prospettiva semiotica e collabora come autore per libri scolastici.

Graziella Marrone, docente di scuola primaria presso l'IC di Loreto Aprutino, laureata in Psicologia. Cura percorsi e progetti di inclusione, disagio scolastico, difficoltà e disturbi di apprendimento. Referente di rete tra le scuole del territorio, l'USR Abruzzo e L'Università di Chieti per il progetto RTI.

Angela Martini si è laureata in Filosofia e successivamente in Psicologia sperimentale presso l'Università di Padova. Ha maturato un'esperienza pluriennale come docente e dirigente nella scuola e si occupa da vari anni di ricerca nell'ambito della valutazione oggettiva degli apprendimenti, della valutazione delle scuole, della comparazione e valutazione dei sistemi scolastici e dell'analisi delle politiche dell'istruzione, temi su cui ha pubblicato numerosi saggi e articoli. Collabora sistematicamente con l'INVALSI e altri enti di ricerca e formazione.

Antonella Mastrogiovanni è ricercatrice INVALSI in ambito psicometrico. È responsabile delle prove di valutazione di italiano, e si occupa dello sviluppo e della validazione dei test. Inoltre coordina i gruppi che si occupano della costruzione degli item per le indagini nazionali.

Alessia Mattei è laureata in Scienze dell'educazione indirizzo Esperto nei processi formativi. Responsabile delle prove di italiano presso INVALSI, la sua attività principale consiste nella costruzione di prove standardizzate e nel coordinare l'intero processo dalla fase di pre-test a quella di somministrazione censuaria.

Luca Oneto, LM in Ingegneria elettronica, PhD in Scienze e tecnologie per l'informazione e la conoscenza. Ricercatore a tempo determinato presso il DIBRIS, Università degli studi di Genova. Ha sviluppato particolare interesse verso la teoria dell'apprendimento statistico.

Serafina Pastore, Fulbright Research Scholar, è ricercatore in Didattica generale presso l'Università di Bari. Si occupa di valutazione dei processi formativi, formative assessment, valutazione della qualità didattica in università e assessment literacy degli insegnanti.

Marianna Rasetta, docente di scuola primaria presso l'IC di Loreto Aprutino; laureata in Scienze della formazione primaria. Promuove e organizza progetti su tematiche di sport e salute. Collabora con l'Università degli studi di Chieti per implementare il progetto RTI promuovendo percorsi di buone prassi.

Paolo Sestito dirige il Servizio struttura economica della Banca d'Italia. In passato commissario e presidente dell'INVALSI (2012-13) e chair del Board del *Programme for International Assessment of Adult Competencies* (2008-13) dell'OCSE. Autore di numerosi articoli e volumi su questioni economiche e sociali, il suo libro più recente tratta di scuola (*La scuola imperfetta*, il Mulino, Bologna, 2014).

Anna Siri, LM in Economia, PhD in Valutazione dei processi e dei sistemi educativi. Componente del gruppo di ricerca della cattedra UNESCO in Antropologia della salute. Biosfera e sistemi di cura, Università degli studi di Genova. Svolge da anni ricerche sul problema della dispersione scolastica e universitaria.

VAI SU: www.francoangeli.it

**PER SCARICARE (GRATUITAMENTE)
I CATALOGHI DELLE NOSTRE PUBBLICAZIONI
DIVISI PER ARGOMENTI E CENTINAIA DI VOCI:
PER FACILITARE LE TUE RICERCHE.**

Management & Marketing
Psicologia e psicoterapia
Didattica, scienze della formazione
Architettura, design, territorio
Economia
Filosofia, letteratura, linguistica, storia
Sociologia
Comunicazione e media
Politica, diritto
Antropologia
Politiche e servizi sociali
Medicina
Psicologia, benessere, auto aiuto
Efficacia personale, nuovi lavori

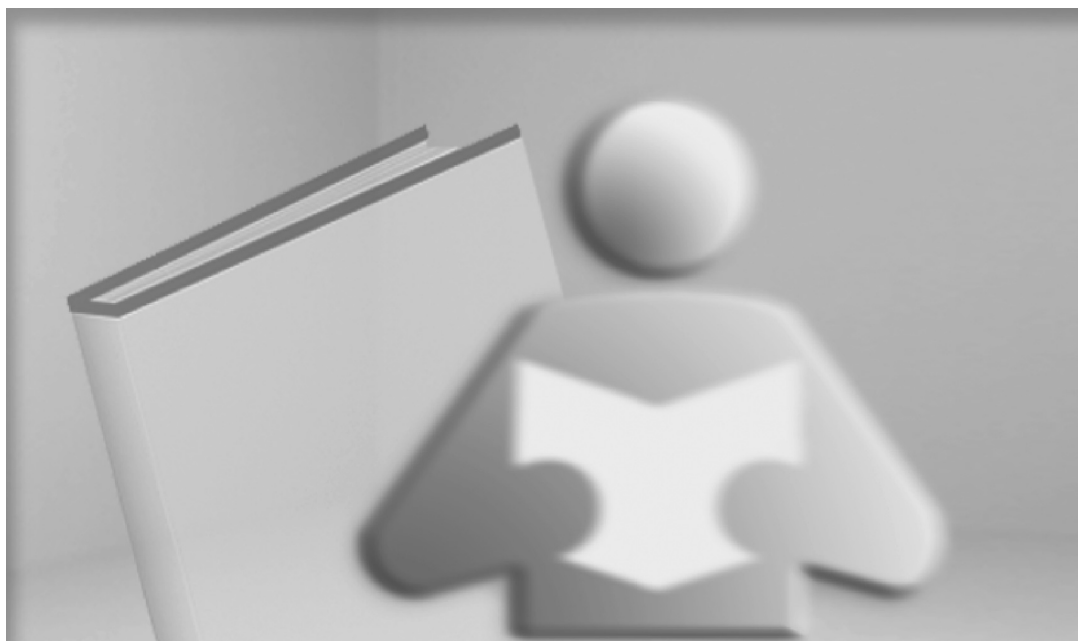


FrancoAngeli

QUESTO LIBRO TI È PIACIUTO?



Comunicaci il tuo giudizio su:
www.francoangeli.it/latuaopinione.asp



**VUOI RICEVERE GLI AGGIORNAMENTI
SULLE NOSTRE NOVITÀ
NELLE AREE CHE TI INTERESSANO?**



Seguici in rete



Sottoscrivi
i nostri feed RSS



Iscriviti
alle nostre newsletter

FrancoAngeli

Il volume ospita i contributi presentati al Seminario “I dati INVALSI: uno strumento per la ricerca”, svoltosi a Roma il 20 settembre 2016. In questa occasione, studiosi appartenenti a differenti ambiti disciplinari e provenienti da diversi contesti professionali (scuola, università, enti di ricerca) hanno presentato lavori scientifici condotti grazie ai dati INVALSI e selezionati tramite una *call for paper*.

Il volume è organizzato in due parti: la prima dedicata all'utilizzo dei dati INVALSI nella ricerca didattica e la seconda focalizzata sui dati INVALSI come strumento per il sostegno e l'orientamento delle scelte politiche sui temi dell'istruzione. I diciotto capitoli presentano approfondimenti che, muovendo da differenti angolazioni e utilizzando molteplici metodologie, offrono al lettore un panorama ricco e articolato dei possibili utilizzi dei dati raccolti annualmente dall'Istituto.

Patrizia Falzetti è Responsabile del Servizio Statistico INVALSI che gestisce l'acquisizione, l'analisi e la restituzione dei dati, riguardanti le rilevazioni nazionali e internazionali sugli apprendimenti, alle singole istituzioni scolastiche e al MIUR. Tali dati, inoltre, sono resi disponibili dal Servizio per quanti abbiano interesse al loro utilizzo ai fini di ricerca scientifica e divulgazione.

L'INVALSI è un ente pubblico di ricerca soggetto alla vigilanza del MIUR, che ne individua le priorità strategiche. L'Istituto ha il compito di promuovere, attraverso le attività di valutazione nazionali e internazionali, il miglioramento dei livelli di istruzione e della qualità del capitale umano, contribuendo allo sviluppo e alla crescita del sistema di istruzione e dell'economia italiana, nel quadro degli obiettivi fissati in sede europea e internazionale.