

I DATI INVALSI PER INDAGARE E MIGLIORARE L'INSEGNAMENTO DELLA MATEMATICA

III Seminario "I dati INVALSI:
uno strumento per la ricerca"

a cura di
Patrizia Falzetti

FrancoAngeli
OPEN  ACCESS


2014-2020


INVALSI

INVALSI PER LA RICERCA
STUDI E RICERCHE



INVALSI PER LA RICERCA

La collana Open Access INVALSI PER LA RICERCA si pone come obiettivo la diffusione degli esiti delle attività di ricerca promosse dall'Istituto, favorendo lo scambio di esperienze e conoscenze con il mondo accademico e scolastico.

La collana è articolata in tre sezioni: "Studi e ricerche", i cui contributi sono sottoposti a revisione in doppio cieco, "Percorsi e strumenti", di taglio più divulgativo o di approfondimento, sottoposta a singolo referaggio, e "Rapporti di ricerca e sperimentazioni", le cui pubblicazioni riguardano le attività di ricerca e sperimentazione dell'Istituto e non sono sottoposte a revisione.

Direzione: Roberto Ricci

Comitato scientifico:

- Tommaso Agasisti (Politecnico di Milano);
- Cinzia Angelini (Università Roma Tre);
- Giorgio Asquini (Sapienza Università di Roma);
- Carlo Barone (Istituto di Studi politici di Parigi);
- Maria Giuseppina Bartolini (Università di Modena e Reggio Emilia);
- Giorgio Bolondi (Libera Università di Bolzano);
- Francesca Borgonovi (OCSE•PISA, Parigi);
- Roberta Cardarelli (Università di Modena e Reggio Emilia);
- Lerida Cisotto (Università di Padova);
- Patrizia Falzetti (INVALSI);
- Michela Freddano (INVALSI);
- Martina Irsara (Libera Università di Bolzano);
- Paolo Landri (CNR);
- Bruno Losito (Università Roma Tre);
- Annamaria Lusardi (George Washington University School of Business, USA);
- Stefania Mignani (Università di Bologna);
- Marcella Milana (Università di Verona);
- Paola Monari (Università di Bologna);
- Maria Gabriella Ottaviani (Sapienza Università di Roma);
- Laura Palmerio (INVALSI);
- Mauro Palumbo (Università di Genova);
- Emmanuele Pavolini (Università di Macerata);
- Donatella Poliandri (INVALSI);
- Roberto Ricci (INVALSI);
- Arduino Salatin (Istituto Universitario Salesiano di Venezia);
- Jaap Scheerens (Università di Twente, Paesi Bassi);
- Paolo Sestito (Banca d'Italia);
- Nicoletta Stame (Sapienza Università di Roma);
- Roberto Trincherò (Università di Torino);
- Matteo Viale (Università di Bologna);
- Assunta Viteritti (Sapienza Università di Roma);
- Alberto Zuliani (Sapienza Università di Roma).

Comitato editoriale:

Andrea Biggera; Ughetta Favazzi; Simona Incerto; Francesca Leggi; Rita Marzoli (coordinatrice); Enrico Nerli Ballati; Veronica Riccardi.



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 massimizza la visibilità, favorisce facilità di ricerca per l'utente e possibilità di impatto per l'autore.

Per saperne di più:

http://www.francoangeli.it/come_pubblicare/pubblicare_19.asp

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 PER INDAGARE E MIGLIORARE L'INSEGNAMENTO DELLA MATEMATICA

III Seminario "I dati INVALSI:
uno strumento per la ricerca"

a cura di
Patrizia Falzetti



FrancoAngeli
OPEN  ACCESS

Le opinioni espresse nei lavori sono riconducibili esclusivamente agli autori e non impegnano in alcun modo l'Istituto. Nel citare i contributi contenuti nel volume non è, pertanto, corretto attribuirne le argomentazioni all'INVALSI o ai suoi vertici.

Grafica di copertina: Alessandro Petrini

Copyright © 2021 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 4.0 Internazionale (CC-BY-NC-ND 4.0)

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
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.it>

ISBN 9788835131045

10. Formulazione degli item matematici e strategie di soluzione: alcuni esempi da uno studio empirico condotto al grado 8

di Giorgio Bolondi, Clelia Cascella, Chiara Giberti

Questo lavoro è parte del progetto “Variazioni 2”, che ha permesso di studiare l’impatto di una variazione nella formulazione di un quesito di Matematica, sulle risposte degli studenti. Le analisi sono basate sulla somministrazione a un campione di 1.648 studenti, di quattro varianti di una stessa prova INVALSI. Le diverse forme del test contengono gli stessi quesiti: alcuni di questi sono stati variati nella formulazione mentre una parte consistente di quesiti è inserita invariata in tutti i fascicoli in modo da permettere una procedura di ancoraggio dei risultati. In questo lavoro, presentiamo l’analisi di due quesiti, entrambi proposti a seguito di variazioni del contesto. Le analisi permettono di studiare l’effetto che può avere il contesto in cui viene presentato un quesito matematico non solo sulla difficoltà percepita dagli studenti ma anche sulle strategie di soluzione che sembrano cambiare in funzione della presenza/assenza di un contesto reale.

This paper reports on some preliminary results from the research project “Variazioni 2” aimed at understanding if and how variations of items phrasing affect students’ answers.

For the purposes of this paper, four anchored tests consisting of the same stem-items (i.e. items with the same question intent and the same math content but different phrasing) were administered to a sample of 1648 students attending grade 8. In this paper, we have analysed two items with different context.

Our analysis allows studying the effect of variations relative to the context on item difficulty, and students’ solving strategy. Results confirmed that variations in items context strongly affect both students’ solving strategies and item difficulty.

1. Introduzione

Quali e quanti fattori influiscono sulla probabilità che uno studente dia una risposta corretta a un item e che ruolo essi hanno nell'attivazione di una specifica strategia di soluzione anziché di un'altra?

Per rispondere a questa domanda nel progetto denominato “Variazioni 2” (Bolondi, Cascella e Giberti, 2020), è stato condotto uno studio sperimentale su un campione probabilistico di studenti di grado 8, stratificato sia per regione sia per background socio-culturale della famiglia dello studente. A partire da una prova INVALSI somministrata a livello censuario nello stesso grado scolastico (che ha coinvolto circa 500.000 studenti), sono state costruite 3 ulteriori prove, distribuite agli studenti, insieme a quella originale INVALSI, secondo un meccanismo di somministrazione a spirale che consente di somministrare le quattro prove agli studenti in modo casuale, sia all'interno della stessa classe, sia tra classi diverse. Questo meccanismo di somministrazione rende probabilisticamente equivalenti – cioè direttamente comparabili – i gruppi di soggetti che rispondono a test diversi: le differenze osservate in relazione a due o più prove possono quindi essere considerate un'indicazione diretta della differenza in termini di difficoltà tra di esse. Inoltre, al fine di garantire la comparabilità diretta tra le risposte date dagli studenti alle prove, ciascuna di esse contiene un nucleo di item invarianti e comuni a tutte, utilizzati per ancorare ed equalizzare i test su una sola metrica, uguale per tutte le prove, attraverso la quale esprimere lungo lo stesso tratto latente le difficoltà di tutti gli item e le abilità di tutti gli studenti coinvolti nella rilevazione. Infine, per capire quali elementi della formulazione attivino strategie di soluzione differenti, ogni item (a meno di quelli utilizzati per ancoraggio) è stato proposto in ciascuno dei quattro fascicoli variando un solo elemento della formulazione da un fascicolo all'altro.

Ogni variazione è stata concepita per investigare l'effetto sulla performance: 1) di specifiche misconcezioni; 2) del contesto (reale o matematico); 3) dei dati numerici (per es. diversi registri semiotici o diversi valori numerici); 4) del formato di risposta; 5) delle rappresentazioni grafiche.

In questo articolo, la metodologia e i risultati del progetto “Variazioni 2” verranno utilizzati per studiare l'effetto che può avere il contesto in cui viene presentato un quesito matematico non solo sulla difficoltà percepita dagli studenti ma anche sulle strategie di soluzione messe in atto in funzione della presenza/assenza di un contesto reale. Questo studio, sebbene certamente necessiti di un approfondimento qualitativo (per es., interviste dirette a studenti e insegnanti), consente già di osservare l'effetto diretto di ciascuna variazione in termini di performance e fornisce un ricco set di informazioni

attraverso cui verificare le ipotesi fatte in sede di costruzione delle domande circa gli errori e le strategie risolutive più frequenti.

2. Quadro teorico

2.1. *Variazioni nella formulazione di un quesito matematico*

Ogni volta che gli studenti affrontano una domanda in Matematica, molteplici fattori influiscono sulle loro risposte perciò, soprattutto quando una domanda fa parte di una prova standardizzata, il *question intent* deve essere ben definito e preciso; solo così si può affermare che lo studente che risponde correttamente ha raggiunto la conoscenza/competenza per la cui valutazione l'item è stato costruito. Restano però diverse componenti che condizionano la risoluzione di un quesito e che esulano dal *question intent*: prima fra tutti la formulazione. Se un quesito risulta complesso come formulazione, lo studente potrebbe avere difficoltà a comprenderlo, rispondendo in maniera errata per questo motivo.

Le ricerche sul tema della formulazione di un quesito in Matematica e quelle specifiche sui cosiddetti *word problems* (problemi verbali) sono numerose nel campo della didattica: per esempio, una recente review della letteratura di queste ricerche nel campo dell'aritmetica è stata proposta da Daroczy *et al.* (2015).

La letteratura del settore non si è occupata solo di variazioni nella formulazione nel caso di *word problems* e in contesto aritmetico. L'influenza della comprensione del testo e della maggiore o minore reperibilità delle informazioni è fondamentale nella risoluzione di un qualsiasi problema. Si può pensare che anche piccole variazioni del modo in cui un problema viene posto, possano quindi modificare sensibilmente le risposte degli studenti, andando a incidere anche sulle strategie risolutive adottate (D'Amore, 2014). A tal proposito, Duval nel 1991 definisce queste modifiche nella formulazione usando il termine variabili redazionali, termine ripreso poi da Colette Laborde (1995) con l'intento di includere in questa categoria di variazioni anche quelle di tipo non verbale, come per esempio l'introduzione/modifica di immagini.

Il problema maggiore che queste ricerche si trovano ad affrontare è come confrontare due diverse formulazioni di uno stesso quesito: non è possibile, infatti, chiedere a uno studente di rispondere a due versioni di una stessa domanda senza che la risposta alla seconda versione somministrata sia influenzata dall'aver già risposto alla prima (Branchetti e Viale, 2015; Bolondi,

Branchetti e Giberti, 2018; Bolondi, Cascella e Giberti, 2020). Questa problematica insorge particolarmente in ricerche in cui una o più versioni dello stesso quesito vengono proposte allo stesso gruppo di studenti (e.g. Lepik, 1990; Cummins *et al.*, 1988; De Corte, Verschaffel e De Win, 1985; Thevenot *et al.*, 2007) e, in alcuni casi, viene parzialmente risolta cambiando l'ordine in cui i quesiti vengono sottoposti agli studenti (e.g. Vicente, Orrantia e Verschaffel, 2007) oppure lasciando trascorrere del tempo tra il momento in cui gli studenti affrontano la prima versione e il momento in cui affrontano la seconda (De Corte *et al.*, 1985). Un altro approccio a questa problematica, riscontrato in diverse ricerche, consiste nel somministrare le diverse versioni a diversi gruppi di studenti (Nesher, 1976) perdendo però così in termini di comparabilità dei risultati, oppure svolgendo ricerche qualitative basate su interviste e analisi di protocolli (e.g. Spanos *et al.*, 1988).

2.2. L'influenza del contesto nella risoluzione di un quesito matematico

La formulazione di un quesito viene sicuramente influenzata dalla presenza/assenza di un contesto reale in cui viene posto il problema matematico. L'influenza della presenza di un contesto reale e della tipologia del contesto stesso sulle risposte degli studenti è già stata ampiamente studiata nella ricerca in didattica della Matematica.

Nel 1992 Stern e Lehrndorfer hanno mostrato che un problema posto in un contesto familiare per gli studenti può essere di grande aiuto per la comprensione e la risoluzione; tematica trattata anche approfonditamente da Roth nel 1996 proprio focalizzandosi il suo studio su studenti frequentanti il grado 8. Diversi studi, più recenti, hanno analizzato anche quanto l'introduzione di un contesto potesse cambiare la difficoltà di un problema (e.g. Léger *et al.*, 2002). Van Den Heuvel-Panhuizen nel 2005 analizzò tre principali funzioni che la presenza di un contesto potrebbe avere sul processo risolutivo degli studenti. In particolare, il contesto potrebbe: migliorare l'accessibilità ai problemi, contribuire alla trasparenza e all'elasticità dei problemi, e suggerire strategie risolutive agli studenti.

Anche nelle valutazioni standardizzate il contesto in cui viene proposto un quesito risulta molto importante e la vicinanza/lontananza del contesto dalle esperienze reali dello studente devono essere tenute fortemente in considerazione, visto che la stessa prova viene svolta da studenti con background anche molto diversi. Nelle indagini OECD-PISA, per esempio, il ruolo del contesto nei quesiti matematici è fondamentale, e viene esplicitato anche nella prima frase della definizione di *mathematical literacy*: «Mathematical

literacy is an individual's capacity to formulate, employ and interpret mathematics in a variety of contexts» (OECD, 2016).

Questo articolo propone l'analisi di due quesiti tratti dalle prove INVALSI di Matematica. Ogni quesito è stato proposto agli studenti in varie forme, a seguito di variazioni sul contesto del problema, ma senza modificarne il contenuto matematico e il *question intent*. Attraverso la metodologia utilizzata (Biondi, Cascella e Giberti, 2020) è possibile studiare l'influenza della presenza/assenza del contesto in un quesito e misurarne gli effetti sulle performance degli studenti. Inoltre, l'analisi quantitativa dei dati e lo studio dei *distractor plot* delle domande nelle diverse forme, permettono di fare delle ipotesi anche sul diverso approccio degli studenti nell'approccio a un problema in contesto o senza contesto e capire se questo porta a diverse strategie risolutive.

3. Metodo: il progetto “Variazioni 2”

Il progetto “Variazioni 2” si propone di studiare variazioni nella formulazione di alcuni quesiti superando le limitazioni descritte precedentemente attraverso una metodologia basata sull'*Item Response Theory* e l'ancoraggio di diverse prove.

Per studiare l'impatto di determinate variazioni nella formulazione di un item, nel progetto “Variazioni 2”, si è scelto di partire da un test di Matematica che INVALSI ha somministrato nel 2011 agli studenti della 3^a secondaria di I grado. A partire da questa prova, sono stati sviluppati tre ulteriori test, contenenti una parte di domande invariate rispetto al test originale, che costituiscono il *Core Test* (CT), e altri quesiti inseriti nei diversi fascicoli in forme diverse, avendo effettuato variazioni di una specifica caratteristica nella formulazione che non portasse però a modificarne il *question intent*.

Le tre forme realizzate sono state somministrate insieme alla versione originale del 2011, con un meccanismo a spirale, a un campione probabilistico di 2000 studenti della 3^a classe della secondaria di I grado. La somministrazione a spirale è stata concepita per fare in modo che tutti gli studenti affrontassero la prova nella stessa situazione, condizione questa necessaria «per giustificare l'affermazione secondo cui le calibrizioni degli item sono confrontabili da un gruppo di esaminandi a un altro e che gli esaminandi siano confrontabili anche quando rispondono a set diversi di item» (Wolfe, 2004).

Inoltre, sebbene gli studenti che hanno risposto alle diverse versioni di un item non siano gli stessi, le loro risposte alle diverse versioni possono essere comunque considerate confrontabili grazie alla presenza dei quesiti che formano il *Core Test*. Il *Core Test* è costituito da un corposo set di item

inseriti in ciascuno dei quattro fascicoli, invariati nella forma, nel contenuto e nella posizione all'interno di ciascuna delle prove, rappresentativo sia in termini di contenuto che di funzionalità psicometrica dell'intera prova (Kolen e Brennan, 2004). Abbiamo definito *Core Test* (CT) questo insieme di quesiti, composto da due sub-set di item, di cui uno somministrato all'inizio della prova e l'altro composto da item dislocati all'interno del fascicolo, nella stessa posizione in ciascuna delle quattro forme.

La necessità di ancorare le prove deriva dalla consapevolezza che anche piccole variazioni nella formulazione di un item cambiano l'item stesso e quindi possono alterarne la funzionalità (per es., Boninsegna *et al.*, 2017; Bolondi, Branchetti e Giberti, 2018). Inoltre, è sempre impossibile essere certi, *a priori*, dell'equivalenza in termini di abilità dei sotto-gruppi di studenti a cui vengono somministrate prove diverse. In questi casi, le differenze tra le medie (o tra qualsiasi altro parametro di interesse teso a investigare differenze nelle performance) potrebbero essere frutto di una combinazione di differenze fra le forme e/o tra il livello di abilità degli studenti piuttosto che un reale effetto della variazione apportata all'item sulla sua funzionalità. Abbiamo quindi inserito un nucleo di item invarianti e abbiamo equalizzato le prove nel framework del modello di Rasch. Infatti, l'equalizzazione (*equating*) è «una procedura statistica usata per aggiustare i punteggi ottenuti su prove diverse in modo che le forme possano essere utilizzate intercambiabilmente» (Kolen e Brennan, 2004, p. 2).

Per l'analisi degli item, oltre a presentare le misure di *infit* calcolate con ConQuest 4.0, abbiamo confrontato la curva caratteristica (plottata da RUMM2030) di ogni item in ciascuna delle quattro versioni e interpretato gli scostamenti tra la spezzata empirica (data dall'insieme degli *observed scores*, cioè delle risposte date dagli studenti al test) con la curva teorica calcolata dal modello (che, per ciascun item, stima la probabilità di una risposta corretta in funzione del livello di abilità degli studenti) (Casella, 2016; Bolondi e Casella, 2017).

In particolare, abbiamo confrontato, le quattro versioni di ciascun item, attraverso alcuni indici calcolati da ConQuest.

Ci riferiamo in particolare a:

- 1) l'indice di discriminatività, attraverso cui viene valutata la capacità dei singoli item di differenziare i soggetti lungo il tratto latente in ragione della quantità di proprietà da questi posseduta, cioè in funzione della loro abilità. Per la valutazione dell'indice di discriminazione (I_d), si considerano i seguenti intervalli (Barbaranelli e Natali, 2005):
 - a) $I_d > 0,40$ – ottimo (item da non revisionare);
 - b) $0,30 \leq I_d < 0,40$ – buono (revisioni minime);

- c) $0,20 \leq Id < 0,30$ – sufficiente (revisioni parziali);
 - d) $0,20 < Id$ Insufficiente (item da riformulare o da rimuovere);
- 2) l'indice di correlazione punto-biseriale, definito come la correlazione tra i punteggi ottenuti dai soggetti a un item e il punteggio totale dei rispondenti su tutti gli item;
 - 3) la percentuale di risposte corrette registrate per ciascuna opzione di risposta;
 - 4) l'indice di *infit*, ossia il weighted MNSQ, che restituisce una misura della bontà di adattamento di ciascun item al modello. Esso mostra quindi la quantità di casualità della misura o, detto in altri termini, la quantità di distorsione del sistema di misurazione. Il suo valore atteso è unitario ma la grandezza e la direzione di ogni spostamento fornisce importanti indicazioni circa la funzionalità di ciascun item (Bolondi e Cascella, 2017): valori inferiori a 1 indicano infatti osservazioni troppo predicibili e quindi ridondanza nel modello; mentre, valori superiori a 1 indicano invece scarsa predicibilità, dovuta spesso alla presenza di un rumore non modellato che compromette l'adattamento dei dati al modello.

Il contributo di RUMM2030 è, invece, prevalentemente legato agli output grafici forniti dal pacchetto. Ci riferiamo in particolare a:

- 1) la curva caratteristica (ICC) di ciascun item che fornisce una stima della probabilità di dare una risposta corretta a ciascun item in funzione dell'abilità relativa dello studente (ossia della sua abilità calcolata sull'intero test in relazione alla difficoltà di ciascun item). RUMM2030 ha il pregio di suddividere il campione totale di studenti in gruppi di abilità simile e di indicarli intorno alla curva caratteristica dell'item (sulla quale giacciono le stime della probabilità di una risposta corretta per ciascun livello di abilità) e alla spezzata empirica (sulla quale giacciono invece i punteggi osservati). L'insieme di tutte le ICC forma la TCC, ossia la funzione informativa dell'intero test;
- 2) i distractor plot che mostrano, per gli item a risposta multipla, l'andamento di ciascun distrattore in funzione del livello di abilità dei rispondenti.

4. Quesiti analizzati

Domanda D6

Il primo quesito analizzato in questo lavoro è tratto dalla prova INVALSI di Matematica somministrata in 5^a primaria nel 2010; considerando che la sperimentazione del progetto “Variazioni 2” è stata svolta al grado 8 (3^a secon-

daria di primo grado), ci si aspetta che risulti più semplice rispetto ai risultati del *main study* INVALSI.

Fascicolo F1 (versione originale)
D6. Per incorniciare una fotografia rettangolare è stato utilizzato 1 metro di cornice. Un lato della fotografia misura 20 cm. Quanto misura l'altro lato?
A. <input type="checkbox"/> 30 cm
B. <input type="checkbox"/> 50 cm
C. <input type="checkbox"/> 60 cm
D. <input type="checkbox"/> 80 cm

Fig. 1 – Quesito D6 nella versione originale tratta dalla prova INVALSI di Matematica somministrata in 5^a primaria nel 2010 (grado 5)

Il quesito appartiene all'ambito Spazio e figure e il *question intent* esplicitato nelle guide alla lettura è «Ricavare un lato del rettangolo dopo averne individuato il perimetro»; si tratta però di un problema in contesto reale. Per rispondere correttamente lo studente deve sia considerare le differenti unità di misura tra il testo e le possibili risposte (anche se la equivalenza di per sé è molto semplice $1\text{ m} = 100\text{ cm}$) e quindi tener conto del fatto che un rettangolo ha lati opposti uguali.

La difficoltà principale consiste nel comprendere il contesto e capire quindi che i lati opposti del rettangolo sono uguali e che ciascuno viene sommato due volte per ottenere il perimetro.

Nel *main study* la domanda risulta di difficoltà media e risponde correttamente il 47% degli studenti (fig. 2). Il 31% degli studenti sceglie la risposta D e sbaglia quindi eseguendo direttamente una sottrazione tra i due dati presenti nel problema. Questa risposta può essere data da studenti che non si soffermano abbastanza sulla comprensione del testo del problema stesso. Anche alcune pratiche didattiche, piuttosto diffuse, che portano gli studenti a concentrarsi principalmente sui dati numerici, sulla domanda e sulle cosiddette *parole chiave* del testo al fine di identificare l'operazione che permette di risolvere il problema, piuttosto che sulla comprensione vera del problema stesso nel suo complesso, possono portare a errori di questo tipo (Zan, 2012).

Gli altri due distrattori risultano meno attrattivi e vengono scelti entrambi da circa un 10% degli studenti. Il distrattore B corrisponde al semiperimetro del rettangolo mentre la risposta C è probabilmente scelta dagli studenti che sottraggono correttamente due volte il lato conosciuto dal perimetro, ma poi non dividono per due il risultato.

Tab. 1 – Risultati del main study relativi alla domanda D6 tratta dalla prova INVALSI di Matematica somministrata in 5^a primaria nel 2010 (grado 5)

Cases for this item	35.538	Discrimination	0,24
Delta	0,12	Weighted MNSQ	1,16
	Count	% of Tot	
A (correct)	16.795	47%	
B	3.388	10%	
C	3.587	10%	
D	10.871	31%	
Missing	808	2%	

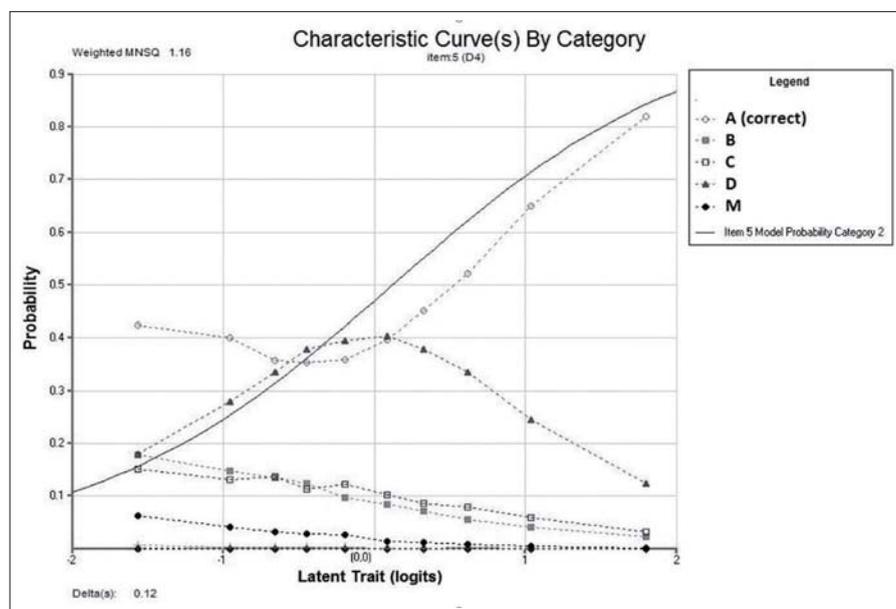


Fig. 2 – Distractor plot del main study relativo alla domanda D6 tratta dalla prova INVALSI di Matematica somministrata in 5^a primaria nel 2010 (grado 5)

Osservando i risultati del *main study* emerge che il fit dell'item inoltre è molto particolare: si ha una netta sottostima degli studenti più deboli e una sovrastima degli studenti con livelli di abilità medi e alti; inoltre distrattore D presenta un particolare andamento prima crescente e poi decrescente, risultando particolarmente attrattivo per gli studenti con livelli di abilità medi. In precedenti studi si è mostrato che andamenti così particolari relativi a uno specifico distrattore possono essere legati a costrutti ben noti in didattica della Matematica come effetti del contratto didattico e misconcezioni (Ferretti,

Giberti e Lemmo, 2018). In questo caso gli studenti con livelli di abilità medi potrebbero scegliere il distrattore D (sottrazione tra i due dati del testo) proprio perché eccessivamente legati a pratiche didattiche che spingono a analizzare il testo dei problemi in base a elementi chiave e a focalizzarsi sull'operazione risolutiva, piuttosto che a comprendere realmente la situazione problematica e le relazioni tra le informazioni (non solo numeriche) fornite nel testo.

L'obiettivo principale è quello di capire se e in che modo il contesto influisca sul processo di risoluzione e sulla tendenza degli studenti a rispondere facendo la differenza tra i dati. Inoltre, si è indagato se una diversa tipologia di risposta (risposta aperta) possa migliorare il funzionamento psicometrico dell'item. La domanda è stata quindi cambiata attraverso una variazione incrociata di contesto e tipologia di risposta (Risposta Multipla/Risposta Univoca). Il contesto è stato variato eliminando il riferimento al reale e trasformando il quesito in un problema puramente geometrico, mantenendo il *question intent* e la struttura sintattica del testo. Infine, nell'ultima versione dell'item si è agito sul testo stesso, mantenendo come contesto quello della fotografia ma eliminando la cornice.

Le forme in cui si presenta la domanda sono quindi 4:

- F1: versione originale;
- F2: variazione contesto (contesto geometrico);
- F3: variazione contesto (contesto geometrico) e tipologia risposta (Risposta univoca);
- F4: variazione testo (non si parla più di cornice ma solo di bordo ma si parla sempre di una fotografia) e tipologia risposta (RU).

Fascicolo F1 (versione originale)	Fascicolo F2
<p>D6. Per incorniciare una fotografia rettangolare è stato utilizzato 1 metro di cornice. Un lato della fotografia misura 20 cm. Quanto misura l'altro lato?</p> <p>A. <input type="checkbox"/> 30 cm B. <input type="checkbox"/> 50 cm C. <input type="checkbox"/> 60 cm D. <input type="checkbox"/> 80 cm</p>	<p>D6. Un rettangolo ha il perimetro di 1 metro. Un lato del rettangolo misura 20 cm. Quanto misura l'altro lato?</p> <p>A. <input type="checkbox"/> 30 cm B. <input type="checkbox"/> 50 cm C. <input type="checkbox"/> 60 cm D. <input type="checkbox"/> 80 cm</p>
Fascicolo F3	Fascicolo F4
<p>D6. Un rettangolo ha il perimetro di 1 metro. Il lato corto del rettangolo misura 20 cm. Quanto misura il lato lungo?</p> <p>Risposta: cm</p>	<p>D6. Il bordo di una fotografia rettangolare misura in tutto 1 metro. Il lato corto della fotografia misura 20 cm. Quanto misura il lato lungo?</p> <p>Risposta: cm</p>

Fig. 3 – Quesito D6 nella versione originale (fascicolo F1) e nelle versioni variate

In questo capitolo si analizzerà in particolare l'effetto della variazione legata al cambiamento di contesto. Nel paragrafo risultati saranno quindi principalmente confrontate la versione originale della domanda (fascicolo F1) con quella fornita in contesto puramente geometrico (fascicolo F2).

Domanda D16

Fascicolo F1 (originale)

D16. Il cavo (AB) di un ripetitore per telefonia cellulare è stato fissato a un palo a una distanza dal suolo di 9 m.
Una lampada di segnalazione (C) viene agganciata al cavo a 3 m di altezza e a 5 m dal punto di ancoraggio a terra (A).

Qual è la lunghezza del cavo AB?

Risposta: m

Fig. 4 – Quesito D16 nella versione originale tratta dalla prova INVALSI di Matematica somministrata in 3^a secondaria di I grado primaria nel 2012 (grado 8)

La domanda D16 è tratta dalla prova INVALSI di Matematica somministrata al grado 8 nel 2012; a differenza della domanda precedente, il grado della prova originale è il medesimo della sperimentazione “Variazioni 2” e ci si aspetta quindi che i risultati siano maggiormente confrontabili.

Questo quesito è stato inserito nel progetto proprio per studiare l'influenza del contesto sulla risoluzione del problema matematico: il contesto, che risulta

abbastanza complesso e non abituale per gli studenti, potrebbe avere influito negativamente sull'approccio degli studenti al quesito. Inoltre, anche in termini di quantità di testo il quesito risulta abbastanza pesante da affrontare per i ragazzi e tutto ciò potrebbe anche aver scoraggiato gli studenti più in difficoltà.

Questa ipotesi è supportata soprattutto dal fatto che nei risultati del *main study* (tab. 2) quasi il 30% degli studenti non ha proprio risposto al quesito, percentuale maggiore anche degli studenti che hanno risposto in modo sbagliato (22%). Nonostante questo, la percentuale di risposte corrette è quasi del 50% e la domanda ha quindi una difficoltà media all'interno della prova.

Tab. 2 – Risultati del main study relativi alla domanda D16 tratta dalla prova INVALSI di Matematica somministrata in 3^a secondaria di I grado nel 2012 (grado 8)

<i>Cases for this item</i>	25.556	<i>Discrimination</i>	0,46
<i>Delta</i>	0,23	<i>Weighted MNSQ</i>	0,95
	<i>Count</i>	<i>% of Tot</i>	
Correct	13.129	51	
Wrong	5.204	20	
Missing	7.223	28	

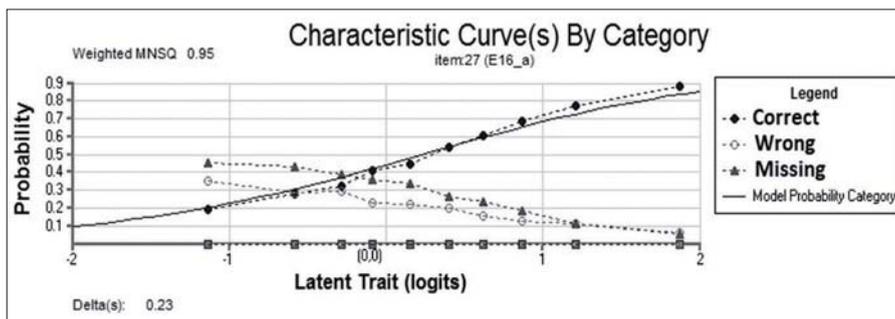


Fig. 5 – Distractor plot del main study relativo alla domanda D16 nella versione originale tratta dalla prova INVALSI di Matematica somministrata in 3^a secondaria di I grado primaria nel 2012

Risulta interessante notare che, in realtà, il contesto non è necessario nella risoluzione del quesito in quanto i dati sono riportati direttamente in figura. Se l'alta percentuale di mancate risposte dovesse essere effettivamente dovuta a una difficoltà legata alla comprensione del contesto reale e al testo del quesito, queste potrebbero essere superate dagli studenti concentrandosi esclusivamente sulla figura e sui dati inseriti.

Dall'analisi dei risultati del *main study*, inoltre, si osserva che il fit risulta generalmente accettabile, anche se il modello tende a sovrastimare gli studenti con livelli di abilità bassi e sottostimare gli studenti con livelli di abilità alti. L'item discrimina molto bene e le risposte mancanti superano le risposte sbagliate principalmente per i livelli di abilità bassi e medi.

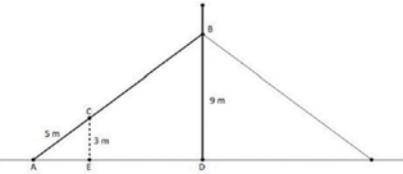
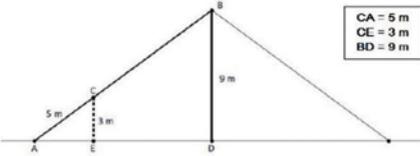
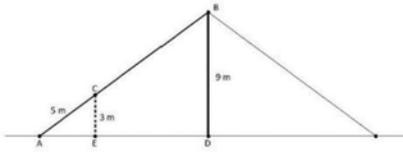
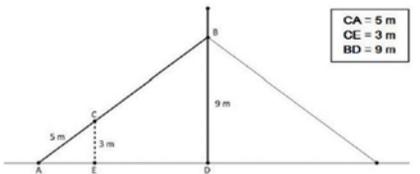
Fascicolo F1 (originale)	Fascicolo F2
<p>D16. Il cavo (AB) di un ripetitore per telefonia cellulare è stato fissato a un palo a una distanza dal suolo di 9 m. Una lampada di segnalazione (C) viene agganciata al cavo a 3 m di altezza e a 5 m dal punto di ancoraggio a terra (A).</p>  <p>Qual è la lunghezza del cavo AB?</p> <p>Risposta: m</p>	<p>D16. Osserva il disegno.</p>  <p>Qual è la lunghezza del segmento AB?</p> <p>Risposta: m</p>
Fascicolo F3	Fascicolo F4
<p>D16. Il cavo (AB) di un ripetitore per telefonia cellulare è stato fissato a un palo a una distanza dal suolo di 9 m. Una lampada di segnalazione (C) viene agganciata al cavo a 3 m di altezza e a 5 m dal punto di ancoraggio a terra (A).</p>  <p>Qual è la lunghezza del cavo AB?</p> <p>Risposta: m</p>	<p>D16. Osserva il disegno.</p>  <p>Qual è la lunghezza del segmento AB?</p> <p>Risposta: m</p>

Fig. 6 – Quesito D16 nella versione originale (fascicolo F1) e nelle versioni variate

Ai fini della ricerca, si è scelto di inserire questo quesito eliminando completamente il contesto reale e mantenendo esclusivamente le informazioni fornite già in figura. Eliminando il contesto è stato necessario inserire una legenda che specificasse a cosa si riferiscono le misure inserite in figura. Inoltre, per evitare possibili ambiguità dovute al fatto che, escludendo il contesto, non aveva più senso considerare il segmento centrale in alto, in due versioni questo segmento è stato eliminato.

Il quesito è quindi stato inserito nei fascicoli in 4 diverse forme:

F1: versione originale;

F2: versione con eliminazione del contesto (e per evitare ambiguità eliminato anche l'ultimo segmento centrale che non c'entra con la risoluzione);

F3: mantenuto il contesto (e per evitare ambiguità eliminato l'ultimo segmento centrale che non c'entra con la risoluzione);

F4: eliminato il contesto (ma mantenuto il segmento centrale).

5. Risultati

Domanda D6

L'item D6 presenta caratteristiche diverse nelle quattro formulazioni, sia in termini di difficoltà, sia in termini di funzionalità, che risulta molta buona nella formulazione proposta in F1 e in F2 e, invece, leggermente peggiore in F3 e F4, fascicoli nei quali l'item, a risposta aperta univoca, mostra un comportamento più marcatamente over-discriminante (tab. 3).

Per questo item è stata provata una doppia combinazione: due item a risposta multipla, di cui uno in contesto (fascicolo F1) e uno non in contesto (F2), e due item a risposta aperta univoca, di cui uno in contesto (F4) e uno non in contesto (F3).

Dal confronto, emerge chiaramente che le formulazioni con quattro opzioni di risposta sono più semplici rispetto a quelle aperte univoche per le quali si registra, invece, un più alto tasso di risposte mancanti. Inoltre, in linea con le rilevazioni OCSE-PISA, la contestualizzazione della domanda aumenta (sia per le domande aperte che per quelle chiuse) il livello di difficoltà della domanda perché il processo di matematizzazione che porta alla soluzione del problema si articola in due fasi: *formulating* ed *employing* (OECD, 2016).

In questa ottica, si osserva in particolare che tra le due versioni a risposta multipla risulta più semplice quella presente in F2 probabilmente proprio grazie all'assenza di un contesto concreto. L'errore più diffuso (distrattore D) in entrambe le versioni, consiste nel non considerare i lati come doppi e fare solo la sottrazione tra i dati presenti nel testo; si potrebbe pensare che il contesto aiuti a considerare la situazione nel suo complesso e a compiere meno questo tipo di errore, mentre accade esattamente il contrario: il secondo quesito risulta più semplice in quanto il distrattore D è meno attrattivo, in particolare per i rispondenti con livelli di abilità minori (figg. 7 e 8). Inoltre, tra le domande di tipo *multiple choice*, quelle contestualizzate, risultano più difficili ma anche più funzionali, cioè con un migliore fit tra dati e modello.

Tra le domande aperte (F3 e F4), non emerge lo stesso miglioramento ma questo potrebbe essere dovuto al formato della domanda cui si associa un elevato numero di risposte mancanti.

Tab. 3 – Output ConQuest 4.0 per l'item D6 nelle 4 forme inserite nel progetto “Variazioni 2”

<i>Fascicolo F1</i>						
<i>item: 14(D6)</i>						
<i>Discrimination</i>	0,43					
<i>Weighted_MNSQ</i>	0,99					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVI Avg: l</i>	<i>PVI SD: l</i>
1	1,00	61,24	0,43	9,71(,000)	0,34	0,87
2	0,00	7,18	-0,14	-2,79(,005)	-0,33	0,73
3	0,00	7,18	-0,11	-2,29(,023)	-0,27	0,62
4	0,00	23,68	-0,32	-6,84(,000)	-0,43	0,79
9	0,00	0,72	-0,13	-2,58(,010)	-1,06	0,17
<i>Fascicolo F2</i>						
<i>item: 14(D6)</i>						
<i>Discrimination</i>	0,42					
<i>Weighted_MNSQ</i>	0,95					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVI Avg: l</i>	<i>PVI SD: l</i>
1	1,00	64,71	0,42	9,26(,000)	0,17	0,82
2	0,00	9,31	-0,15	-3,07(,002)	-0,45	0,72
3	0,00	8,58	-0,18	-3,66(,000)	-0,57	0,70
4	0,00	14,71	-0,20	-4,21(,000)	-0,48	0,71
<i>Fascicolo F3</i>						
<i>item: 14(D6)</i>						
<i>Discrimination</i>	0,56					
<i>Weighted_MNSQ</i>	0,84					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVI Avg: l</i>	<i>PVI SD: l</i>
0	0,00	33,58	-0,38	-8,24(,000)	-0,50	0,86
1	1,00	56,69	0,56	13,77(,000)	0,43	0,82
9	0,00	9,73	-0,34	-7,30(,000)	-1,03	0,63
<i>Fascicolo F4</i>						
<i>item: 14(D6)</i>						
<i>Discrimination</i>	0,57					
<i>Weighted_MNSQ</i>	0,88					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVI Avg: l</i>	<i>PVI SD: l</i>
0	0,00	36,91	-0,44	-10,38(,000)	-0,54	0,81
1	1,00	55,48	0,57	14,77(,000)	0,49	0,88
9	0,00	7,61	-0,27	-5,96(,000)	-0,95	1,03

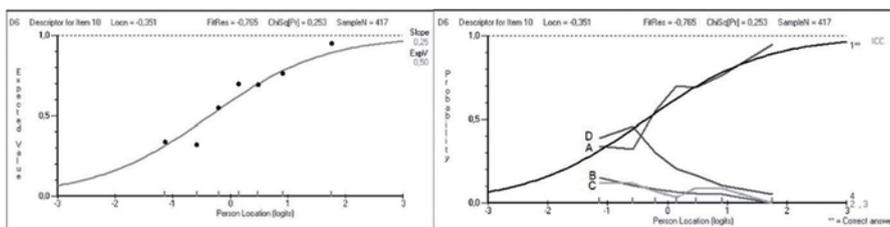


Fig. 7 – Curva caratteristica e distractor plot dell'item D6, nel fascicolo F1

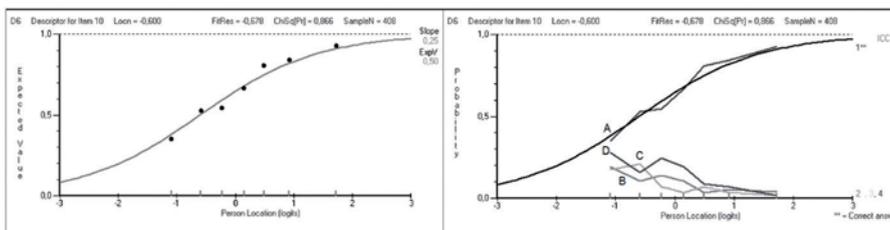


Fig. 8 – Curva caratteristica e distractor plot dell'item D6, nel fascicolo F2

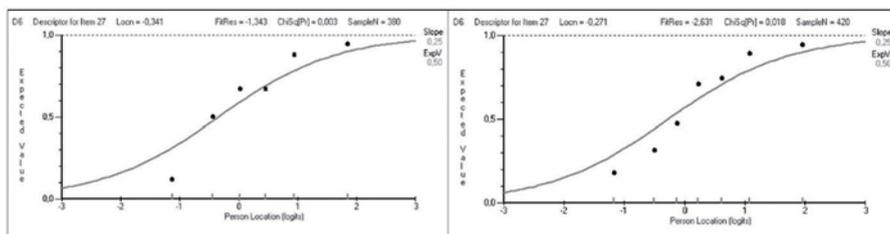


Fig. 9 – Curva caratteristica e distractor plot dell'item D6, nei fascicoli F3 e F4

Domanda D16

L'item D16 è stato formulato in tutti i fascicoli come un item a risposta aperta. Non si notano differenze in termini di difficoltà tra le versioni in F1 e F2, mentre risulta più semplice il quesito in F3 e più difficile in F4.

Anche questo item offre spunti di riflessione in relazione all'effetto che il contesto e le variazioni della figura hanno sulla difficoltà e sulla funzionalità degli item. Se si confrontano le domande con la medesima immagine (F1 con F4 e F2 con F3) si osserva che nelle due varianti senza contesto, l'item risulta significativamente più facile (come nella D6).

Tab. 4 – Output ConQuest 4.0 per l'item D16 nelle 4 forme inserite nel progetto “Variazioni 2”

<i>Fascicolo F1</i>						
<i>item:35(D16)</i>						
<i>Discrimination</i>	0,52					
<i>Weighted_MNSQ</i>	0,93					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVIAvg:1</i>	<i>PV1 SD:1</i>
0	0,00	30,86	-0,25	-5,27(.000)	-0,28	0,83
1	1,00	46,41	0,52	12,36(.000)	0,51	0,79
9	0,00	22,73	-0,34	-7,40(.000)	-0,44	0,74
<i>Fascicolo F2</i>						
<i>item:34(D16)</i>						
<i>Discrimination</i>	0,40					
<i>Weighted_MNSQ</i>	1,00					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVIAvg:1</i>	<i>PV1 SD:1</i>
0	0,00	36,52	-0,23	-4,81(.000)	-0,30	0,78
1	1,00	42,40	0,40	8,72(.000)	0,28	0,85
9	0,00	21,08	-0,21	-4,26(.000)	-0,43	0,75
<i>Fascicolo F3</i>						
<i>item:34(D16)</i>						
<i>Discrimination</i>	0,47					
<i>Weighted_MNSQ</i>	0,99					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVIAvg:1</i>	<i>PV1 SD:1</i>
0	0,00	39,66	-0,25	-5,19(.000)	-0,34	0,79
1	1,00	36,98	0,47	10,64(.000)	0,59	0,94
9	0,00	23,36	-0,24	-5,08(.000)	-0,46	0,85
<i>Fascicolo F4</i>						
<i>item:34(D16)</i>						
<i>Discrimination</i>	0,40					
<i>Weighted_MNSQ</i>	1,02					
<i>Label</i>	<i>Score</i>	<i>% of tot</i>	<i>Pt Bis</i>	<i>t (p)</i>	<i>PVIAvg:1</i>	<i>PV1 SD:1</i>
0	0,00	30,65	-0,20	-4,29(.000)	-0,25	0,92
1	1,00	48,77	0,40	9,16(.000)	0,36	0,97
9	0,00	20,58	-0,26	-5,80(.000)	-0,48	1,00

Le versioni non contestualizzate (F2 e F4), come già nelle altre domande, risultano più facili, rispetto alle omologhe in contesto reale (confronto F1 con F4 e F3 con F2). Tra le contestualizzate, risulta molto più facile la versione *con il palo eccedente* (F1) rispetto all'altra (F3); e così tra le non contestualizzate (F4 rispetto a F2).

Anche la funzionalità (buona in tutti i casi) risulta leggermente migliore nel non contestualizzato; inoltre la versione col palo eccedente è più facile, ma ha un fitting peggiore in entrambi i casi (contestualizzato e no). Si osserva infatti che nelle versioni contestualizzate, il fit è minore di 1, mentre in quelle non contestualizzate il fit è maggiore di 1, in entrambi i casi.

Nell'originale INVALSI (corrispondente a F1) la domanda aveva avuto una percentuale di risposte confrontabile, ed era risultata con un fit minore di 1. Inoltre, pare opportuno osservare che, ancora una volta, il comportamento osservato nel *main study* coincide perfettamente con quello osservato in “Variazioni 2”, a conferma della buona rappresentatività del campione e dell'inferibilità dei risultati alla popolazione.

Nello studio di questo item, sembra emergere che le domande non contestualizzate fittino meglio con il modello. Questo risultato può essere spiegato dal fatto che il processo del *formulating* è distinto dall'*employing*: una domanda in cui si toglie la componente del *formulating* risulta più pulita nei suoi risultati.

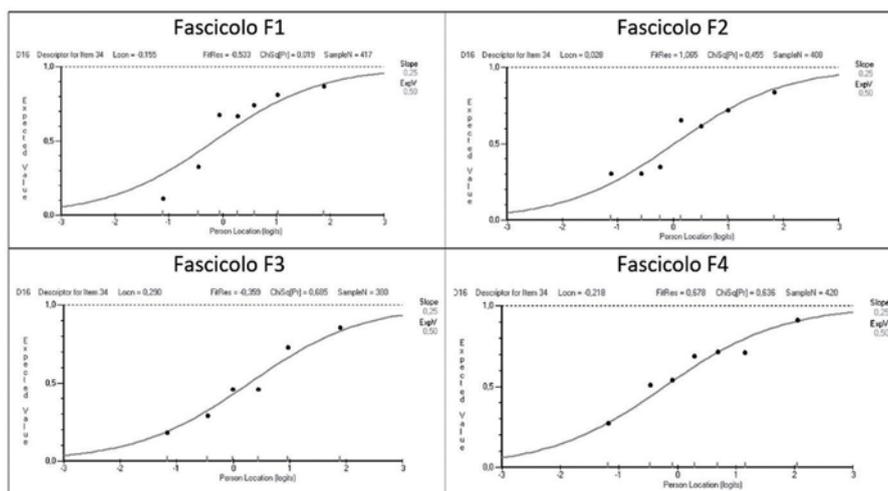


Fig. 10 – Curva caratteristica e distractor plot dell’item D16, nei fascicoli F1, F2, F3 e F4

6. Conclusioni

In questo articolo sono state presentate le analisi di due quesiti inseriti nel progetto “Variazioni 2” con lo scopo di comprendere l’impatto di una variazione del contesto sulle risposte degli studenti. Questo tema è già stato

affrontato da numerose ricerche in didattica della Matematica che hanno mostrato come, un contesto familiare per gli studenti possa aiutare, ma in generale la presenza di un contesto possa anche rendere il quesito più complesso.

I risultati delle analisi presentate in questo lavoro mostrano come il contesto, specialmente se complesso e lontano dall'esperienza degli studenti come nel caso del quesito D16, possa mettere maggiormente in difficoltà gli studenti e soprattutto portare a una percentuale maggiore di mancate risposte. Chiaramente questi risultati possono variare in quanto strettamente legati anche alla pratica didattica e all'abitudine degli studenti di affrontare problemi matematici all'interno di un contesto concreto. Si è osservato inoltre che per entrambi i quesiti le variazioni di contesto hanno portato anche a una variazione della funzionalità dell'item in termini di discriminazione e fit.

La metodologia utilizzata infatti permette non solo di capire se il contesto influenzi le risposte degli studenti ma anche di misurare l'effetto di una variazione di contesto e, per esempio nel caso dell'item D6, di analizzare in che modo cambino le risposte degli studenti e per quali livelli di abilità. In questo caso, infatti, l'uso dei *distractor plot* ha permesso di evidenziare che, avendo eliminato il contesto, un distrattore legato a uno specifico errore degli studenti diventa notevolmente meno attrattivo per gli studenti con livelli di abilità bassi.

Riferimenti bibliografici

- Barbaranelli C., Natali E. (2005), *I test psicologici: teorie e modelli psicometrici*, Carrocci, Roma.
- Bolondi G., Branchetti L., Giberti C. (2018), "A tool for analyzing the impact of the formulation on the performance of students answering a mathematical item", *Studies in Educational Evaluation*, 58, pp. 37-50.
- Bolondi G., Cascella C. (2017), "Somministrazione delle prove INVALSI dal 2009 al 2015: un patrimonio di informazioni tra evidenze psicometriche e didattiche", in P. Falzetti (a cura di), *I dati INVALSI: uno strumento per la ricerca*, FrancoAngeli, Milano, pp. 68-92.
- Bolondi G., Cascella C., Giberti C. (2020), "Formulazione della domanda e funzionalità psicometrica: evidenze empiriche su un campione di studenti della terza secondaria di primo grado", in P. Falzetti (a cura di), *Implementazione e miglioramento del dato. II Seminario "I dati INVALSI: uno strumento per la ricerca"*, FrancoAngeli, Milano, pp. 68-92.
- Boninsegna R., Bolondi G., Branchetti L., Giberti C., Lemmo A. (2017), "Uno strumento per analizzare l'impatto di una variazione nella formulazione di un quesito INVALSI di matematica", in P. Falzetti (a cura di), *I dati INVALSI: uno strumento per la ricerca*, FrancoAngeli, Milano, pp. 101-110.

- Cascella C. (2016), “Examining the relationship between items position and their functionality”, *Working paper INVALSI*, 29, pp. 1-19, testo disponibile al sito: http://www.invalsi.it/download2/wp/wp29_Cascella.pdf, data di consultazione 26/1/2021.
- 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), *La didattica dell'italiano. Problemi e prospettive*, Proceedings della conferenza “Quale didattica dell'italiano? Problemi e prospettive”, Locarno, ottobre 2014.
- Cummins D.D., Kintsch W., Reusser K., Weimer R. (1988), “The role of understanding in solving word problems”, *Cognitive Psychology*, 20, 4, pp. 405-438.
- D'Amore B. (2014), *Il problema di matematica nella pratica didattica*, Digital Index, Modena.
- Daroczy G., Wolska M., Meurers W.D., Nuerk H.C. (2015), “Word problems: A review of linguistic and numerical factors contributing to their difficulty”, *Frontiers in Psychology*, 6, pp. 1-13.
- De Corte E., Verschaffel L., De Win L. (1985), “Influence of rewording verbal problems on children's problem representations and solutions”, *Journal of Educational Psychology*, 77, 4, p. 460.
- Duval R. (1991), “Interaction des différents niveaux de représentation dans la compréhension de textes”, *Annales de Didactique et de sciences cognitives*, 4, pp. 136-193.
- Ferretti F., Giberti C., Lemmo A. (2018), “The Didactic Contract to Interpret Some Statistical Evidence in Mathematics Standardized Assessment Tests”, *EURASIA Journal of Mathematics, Science and Technology Education*, 14, 7, pp. 2895-2906.
- Kolen M., Brennan R. (2004), *Test Equating, Scaling, and Linking. Methods and Practices*, Springer, New York, 2nd ed.
- Laborde C. (1995), “Occorre apprendere a leggere e scrivere in matematica”, *La matematica e la sua didattica*, 9, 2, pp. 121-135.
- Léger L., Sander E., Richard J.F., Brissiaud R., Legros D., Tijus C. (2002), “Propriétés des objets et résolution de problèmes mathématiques”, *Revue française de pédagogie*, 139, pp. 97-105.
- Lepik M. (1990), “Algebraic word problems: Role of linguistic and structural variables”, *Educational Studies in Mathematics*, 21, 1, pp. 83-90.
- Nesher, P. (1976), “Three determinants of difficulty in verbal arithmetic problems”, *Educational Studies in Mathematics*, 7, 4, pp. 369-388.
- OECD (2016), *PISA 2015 assessment and analytical framework: Science, reading, mathematic and financial literacy*, OECD Publishing, Paris.
- Roth W.M. (1996), “Where is the Context in Contextual Word Problem? Mathematical Practices and Products in Grade 8 Students' Answers to Story Problems”, *Cognition and Instruction*, 14, 4, pp. 487-527.
- Spanos G., Rhodes N.C., Dale T.C., Crandall J. (1988), “Linguistic features of mathematical problem solving: Insights and applications”, in R.R. Cocking, J.P. Mestre (eds.), *Linguistic and cultural influences on learning mathematics*, Erlbaum, Hillsdale (NJ), pp. 221-240.

- Stern E., Lehrndorfer A. (1992), "The role of situational context in solving word problems", *Cognitive Development*, 7, 2, pp. 259-268.
- Thevenot C., Devidal M., Barrouillet P., Fayol M. (2007), "Why does placing the question before an arithmetic word problem improve performance? A situation model account", *The Quarterly Journal of Experimental Psychology*, 60, 1, pp. 43-56.
- Van Den Heuvel-Panhuizen M. (2005), "The role of contexts in assessment problems in mathematics", *For the learning of mathematics*, 25, 2, pp. 2-23.
- Vicente S., Orrantia J., Verschaffel L. (2007), "Influence of situational and conceptual rewording on word problem solving", *British Journal of Educational Psychology*, 77, 4, pp. 829-848.
- Wolfe E. (2004), "Equating and Item Banking with the Rasch model", in E. Smith, R. Smith (eds.), *Introduction to Rasch Measurement. Theory, Models and Applications*, JAM Press, Maple Grove, pp. 366-390.
- Zan R. (2012), "La dimensione narrativa di un problema: il modello CandD per l'analisi e la (ri) formulazione del testo", *L'insegnamento della matematica e delle scienze integrate*, 3, pp. 107-126.