IMPLEMENTAZIONE E MIGLIORAMENTO DEL DATO

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

a cura di
Patrizia Falzetti

FrancoAngeli
OPEN ACCESS
ISBN 9788835101802

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

ISBN 9788835101802

5. Formulazione della domanda e funzionalità psicometrica: evidenze empiriche su un campione di studenti della terza secondaria di primo grado

di Giorgio Bolondi, Clelia Cascella, Chiara Giberti

In questo capitolo presentiamo uno studio sull'impatto che una variazione linguistica nella formulazione di una domanda di Matematica ha sulla sua funzionalità psicometrica. A questo scopo, partendo da un test già somministrato a livello censuario da INVALSI nell'anno scolastico 2010/11, sono state costruite tre ulteriori prove che testano la funzionalità di item variati secondo il quadro teorico di riferimento, scegliendo le formulazioni in maniera da includere, ove possibile, quelle più utilizzate nelle prove INVALSI, senza modificarne il *question intent*.

Le nostre tre prove sono state somministrate insieme alla versione originale del 2011, con un meccanismo di somministrazione a spirale, a un campione di 2040 studenti della terza classe della scuola secondaria di I grado, estratto con metodo probabilistico e stratificato in funzione della Regione e dello status socio-culturale.

Le analisi sinora effettuate confermano alcune rilevanti ipotesi circa l'associazione tra formulazione della domanda e funzionalità psicometrica suggerendo anche la possibilità di pervenire a un elenco di tali associazioni che possono diventare un utile strumento di supporto alla stesura di nuovi item.

1. Introduzione

1.1. 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

ISBN 9788835101802

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 problem* (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, Wolska, Meurers e Nuerk (2015).

Risulta complesso indicare quali siano le caratteristiche nella formulazione di un quesito che influenzano maggiormente la risposta degli studenti; nell'intento di classificare le tipologie di variazioni, Nesher, nel 1982, ha individuato tre principali componenti che possono variare in un *word problem*:

– componente logica (operazioni, la mancanza o sovrabbondanza di dati ecc.);
– componente sintattica (posizione della domanda nel testo del problema, numero di parole ecc.);
– componente semantica (relazioni contestuali, indicazioni implicite ecc.).

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 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). Questa problematica insorge particolarmente in ricerche in cui una o più versioni dello stesso quesito vengono proposte allo stesso gruppo di studenti (tra gli altri Lepik, 1990; Cummins, Kintsch, Reusser e Weimer, 1988; De Corte, Verschaffel e De Win, 1985; Thevenot, Devidal, Barrouillet

ISBN 9788835101802

e Fayol, 2007) e, in alcuni casi, viene parzialmente risolta cambiando l'ordine in cui i quesiti vengono sottoposti agli studenti (tra gli altri 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 (tra gli altri Spranos *et al.*, 1988).

1.2. Gli obiettivi della ricerca

Variazioni 2 è un programma di ricerca che si propone di raccogliere ed elaborare dati per perseguire una pluralità di obiettivi:

- raccogliere dati sulla funzionalità psicometrica di un item di Matematica in funzione di varianti di formulazione linguistica, testuale, grafica, e di contenuto;
- raccogliere dati che consentano una riflessione: 1) sul formato di risposta più appropriato in ragione dello scopo e della natura della domanda; 2) sulla scelta degli item da riferire a uno stesso stimolo; e, 3) sull'equilibrio degli stimoli presentati all'interno di una stessa prova;
- spiegare la relazione che esiste tra la formulazione dei quesiti e funzionalità psicometrica degli item anche in ragione del (possibile) nesso di causa ed effetto tra formulazione della domanda e l'attivazione dei processi cognitivi utili per fornire una risposta al quesito. In particolare, si è scelto di utilizzare quesiti che mettessero in luce fenomeni didattici già studiati in letteratura e operare variazioni mirate all'analisi del fenomeno stesso;
- approfondire tali relazioni in una prospettiva comparata, su sottoinsiemi specifici della popolazione studiata. Per esempio, partendo da quesiti che hanno mostrato un forte gap di genere nelle Rilevazioni nazionali, sono state proposte variazioni atte a studiare le possibili cause di questo divario nella funzionalità psicometrica dell'item;
- infine, esplorare la relazione che intercorre tra variazioni nella formulazione di una domanda e self-efficacy e/o Math anxiety, anche in una prospettiva di genere.

In questo capitolo presentiamo la prima parte del progetto discutendo alcune prime evidenze empiriche emerse dall'analisi dei dati in relazione ai primi due obiettivi, iniziando a condividere anche alcune prime ipotesi in

ISBN 9788835101802

relazione all'individuazione dei processi cognitivi che solo alcune formulazioni attivano.

2. L'impianto metodologico

Per perseguire gli obiettivi del presente lavoro, a partire da un test matematico che INVALSI ha somministrato nel 2011 agli studenti della III secondaria di I grado (III media), sono stati sviluppati tre ulteriori test per la valutazione della competenza matematica (*alternative forms*), ciascuno dei quali propone variazioni nella formulazione degli item, nel tentativo di non modificarne il *question intent*.

I fascicoli sono stati costruiti in modo che tutti avessero una parte consistente di item in comune, cioè invariati sia nella forma che nel contenuto in ciascuna delle forme, e rappresentativa in termini sia di contenuto sia di funzionalità psicometrica dell'intera prova. Abbiamo definito *Core Test* (CT) questo insieme di quesiti, composto da due sub-set di item rispettivamente utilizzati come àncora interna e àncora esterna. Per ciascuna prova sono state individuate domande, in parte provenienti da prove precedenti e in parte costruite ex novo, afferenti a diversi ambiti della Matematica, con diversi livelli di difficoltà (presunte)[1] e diverse dimensioni di riferimento. Ogni domanda è stata variata secondo il quadro teorico di riferimento, scegliendo le formulazioni in maniera da includere ove possibile quelle più utilizzate nelle prove INVALSI, e proponendone di nuove quando opportuno.

Le tre forme realizzate sono state somministrate insieme alla versione originale del 2011, con un meccanismo a spirale, a un campione probabilistico di 2040 studenti della III classe della secondaria di I grado. In questo modo, gli studenti che hanno risposto alle diverse versioni di un item non sono gli stessi, ma le loro risposte alle diverse versioni possono essere confrontate grazie al comportamento di risposta degli studenti osservato in relazione alla parte in comune del test.

[1] Per avere una misura (seppur presunta) della difficoltà degli item, sono stati seguiti due criteri. Per gli item che sono stati ripresi (e poi modificati) da precedenti test INVALSI, è stata analizzata la funzionalità di tali item all'interno delle prove dalle quali sono stati estratti. Si tratta comunque di una funzionalità presunta perché, come è noto, il comportamento di un item e, quindi, la sua difficoltà relativa, sono fortemente influenzati dal contesto entro il quale essi vengono proposti allo studente, e cioè dagli item che lo precedono. Per quanto riguarda invece i quesiti di nuova costruzione, non presenti in precedenti prove INVALSI, si è fatto riferimento alla letteratura di settore. Per gli stessi motivi appena esposti, anche in questo caso è da considerarsi presunta la misura di difficoltà a essi attribuita.

ISBN 9788835101802

In questo capitolo presentiamo un'analisi condotta su 800[2] studenti. La numerosità dei casi, consente di esplorare e confrontare la funzionalità degli item utilizzando il modello di Rasch, il quale ipotizza che la probabilità che uno studente fornisca una risposta corretta a un item sia governata dalla sua abilità relativa, cioè dall'abilità intrinseca dello studente confrontata con la difficoltà dell'item cui risponde. L'analisi di Rasch ha consentito il confronto della funzionalità non solo dei singoli item ma anche della funzionalità delle prove nel loro complesso.

Per l'analisi degli item, oltre a presentare alcune misure di sintesi indicative della loro funzionalità (e ottenute 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) (Cascella, 2016; Bolondi e Cascella, 2017).

2.1. La struttura delle forme e criteri di costruzione

Nella tabella seguente è riepilogata la struttura dei quattro fascicoli somministrati (tab. 1).

Gli item identificati con la lettera A costituiscono l'àncora esterna mentre gli item *Anch* costituiscono invece l'àncora interna, la prima posta all'inizio del test per evitare che effetti legati alla stanchezza potessero negativamente incidere sulla probabilità di una risposta corretta, la seconda composta da item collocati in punti diversi del test, entrambe inserite a garanzia della robustezza del *Core Test* (Kolen e Brennan, 2004).

Abbiamo evidenziato in grigio gli altri item per indicare che sono stati variati e che saranno quindi oggetto di studio. In particolare, sono stati evidenziati in grigio scuro le versioni originali degli item tratti da altre prove INVALSI e in grigio più chiaro le diverse versioni di quello stesso item (a gradazioni di grigio diverse corrispondono versioni diverse dell'item). Il nome dell'item riportato nella tabella fornisce il riferimento all'item origi-

[2] Il gruppo di studenti su cui abbiamo lavorato ai fini della presente indagine è una parte del campione totale (composto da 2.040 studenti) a cui sono state somministrate le prove sviluppate per il progetto Variazioni_2. Le somministrazioni relative alla ricerca hanno coperto un lungo arco temporale; per questo motivo i dati presentati al convegno e in questo capitolo, corrispondono ai risultati dei primi 800 fascicoli disponibili prima del convegno.

ISBN 9788835101802

nale; nel caso in cui il quesito non provenga da una passata prova INVALSI, nell'etichetta sarà riportato un nome contrassegnato da *NEW*.

Tab. 1 – Struttura dei fascicoli somministrati

Item	Fascicolo 1	Fascicolo 2	Fascicolo 3	Fascicolo 4
A1a	D1a_PN2013	D1a_PN2013	D1a_PN2013	D1a_PN2013
A1b	D1b_PN2013	D1b_PN2013	D1b_PN2013	D1b_PN2013
A2	D18_PN2014	D18_PN2014	D18_PN2014	D18_PN2014
A3	D22_PN2013	D22_PN2013	D22_PN2013	D22_PN2013
A4	D10a_PN2012	D10a_PN2012	D10a_PN2012	D10a_PN2012
A5	D20_PN2010	D20_PN2010	D20_PN2010	D20_PN2010
A6	E18_PN2012	E18_PN2012	E18_PN2012	E18_PN2012
Anch_1	D7_PN2011	D7_PN2011	D7_PN2011	D7_PN2011
D1	D13_PN_2011_v4	D13_PN_2011_originale	D13_PN_2011_v2	D13_PN_2011_v3
D2	D19_PN2011_originale	D7_PN2011_v4	D7_PN2011_v3	D7_PN2011_v2
D3	E15_PN2012_originale	E15_PN2012_v4	E15_PN2012_v3	E15_PN2012_v2
Anch_3	D18_PN2011	D18_PN2011	D18_PN2011	D18_PN2011
D5	D12_PN2011_originale	D12_PN2011_v2	D12_PN2011_v3	D12_PN2011_v4
D6	D4_L052010_originale	D4_L052010_v2	D4_L052010_v4	D4_L052010_v3
D7	D6_PN2011_originale	D6_PN2011_v4	D6_PN2011_v3	D6_PN2011_v2
D8	D7b_L062013_v1	D7b_L062013_originale	D7b_L062013_originale	D7b_L062013_v1
Anch_7	D27_PN2013	D27_PN2013	D27_PN2013	D27_PN2013
D9	1CG_NEW_v1	1CG_NEW_v1	1CG_NEW_v2	1CG_NEW_v2
Anch_4	D17_PN2011	D17_PN2011	D17_PN2011	D17_PN2011
D10	1LG_NEW_v1	1LG_NEW_v2	1LG_NEW_v3	1LG_NEW_v4
Anch_8	D26_PN2015	D26_PN2015	D26_PN2015	D26_PN2015
Anch_5	D25_PN2011	D25_PN2011	D25_PN2011	D25_PN2011
D11	E6_PN2012_v3	E6_PN2012_v1	E6_PN2012_v2	E6_PN2012_v4
Anch_2	D9b_PN2011	D9b_PN2011	D9b_PN2011	D9b_PN2011
D12	D5_PN2011_originale	D5_PN2011_v2	D5_PN2011_v4	D5_PN2011_v3
D13	E7_PN2012_v1	E7_PN2012_originale	E7_PN2012_v4	E7_PN2012_v3
D14	D3_L062012_v2	D3_L062012_v3	D3_L062012_originale	D3_L062012_v1
Anch_6	D22_PN2011	D22_PN2011	D22_PN2011	D22_PN2011
D15	3CG_NEW_v1	3CG_NEW_v1	3CG_NEW_v2	3CG_NEW_v2
D16	E16a_PN2012_originale	E16a_PN2012_v2	E16a_PN2012_v1	E16a_PN2012_v3
D17	D8ab_PN2011_originale	D8ab_PN2011_originale	D8ab_PN2011_v3	D8ab_PN2011_v3

Viste le molteplici finalità del progetto, che vede l'intreccio di interessi legati alla didattica e altri legati all'analisi dei quesiti relativamente al loro funzionamento psicometrico, sono stati diversi anche i criteri con cui sono state individuate le domande da variare e il modo in cui sono state effettuate le variazioni.

Alcune delle domande sono state selezionate tra quelle di prove passate che mostravano comportamenti devianti rispetto alle attese del modello. Le variazioni sono quindi state costruite per capire le ragioni di tali deviazioni che potrebbero essere, per esempio:

ISBN 9788835101802

– la formulazione della domanda potrebbe essere poco chiara e gli studenti potrebbero sbagliare non tanto perché non hanno raggiunto il *question intent* ma perché fraintendono la richiesta;
– il contenuto matematico della domanda e i processi cognitivi richiesti potrebbero essere distanti da quelli indagati con le altre domande del test e per questo non ci sarebbe una coerenza con il tratto latente misurato;
– potrebbero intervenire particolari fenomeni didattici che agiscono in modo trasversale rispetto all'abilità matematica e che potrebbero portare a sbagliare anche studenti molto bravi.

Si è scelto inoltre di intrecciare questo progetto relativo alle variazioni nella formulazione con altre ricerche svolte, sempre a partire dalle prove INVALSI, dagli stessi autori. In particolare, recenti studi (Bolondi, Cascella e Giberti, 2017; Giberti, Bolondi e Zivelonghi, 2016) hanno mostrato che le differenze di genere in Matematica a favore dei maschi risultano particolarmente marcate su alcuni quesiti e, per questo motivo, hanno indagato le caratteristiche che dei quesiti che possono creare un maggiore differenza nel rendimento di maschi e femmine. A partire da queste domande, abbiamo costruito variazioni specifiche con lo scopo di neutralizzare i fattori ritenuti responsabili del gender gap, senza però modificare il *question intent* della domanda.

Infine, abbiamo aggiunto alle prove alcuni item costruiti ex novo per testare ipotesi diverse: i quesiti D10 e D11, per esempio, hanno l'obiettivo di verificare la validità di un nuovo formato di risposta, mentre i quesiti D9 e D15 sono tratti da un articolo di didattica della Matematica (Sbaragli, 2012) al fine di indagare una particolare misconcezione.

2.2. Analisi pretest

Per la messa a punto dei test, abbiamo somministrato i quattro fascicoli di prova (F1, F2, F3, e F4), a 96 studenti della classe terza della scuola secondaria di primo grado, escludendo successivamente dalle analisi gli studenti con bisogni educativi speciali (tab. 2).

Tab. 2 – Numerosità del campione

	F1	*F2*	*F3*	*F4*	*Totale*
Numero di studenti	2	22	19	21	82
Numero di studenti con bisogni educativi speciali	2	2	7	3	14
Totale	22	24	26	24	96

ISBN 9788835101802

La scarsa numerosità campionaria non ha consentito di effettuare analisi con modelli IRT e, in particolare con il modello di Rasch, solitamente impiegato in tutte le rilevazioni INVALSI. Sono state quindi calcolate, per ciascun fascicolo, misure di statistica descrittive e poi misure afferenti alla Teoria classica dei test la quale ipotizza una relazione lineare e additiva tra il punteggio osservato X (il numero di risposte corrette fornite dallo studente agli item che compongono la prova), il punteggio vero V (il valore di abilità/competenza reale dello studente) e la componente erratica E (l'errore non sistematico che cambia da una prova all'altra essendo esso non imputabile a caratteristiche intrinseche dello strumento quanto piuttosto a naturali fluttuazioni campionarie) (X=V+E). L'analisi è stata condotta con una pluralità di obiettivi: avere una prima panoramica di insieme sulla funzionalità misuratoria degli item, in una prospettiva comparativa tra i quattro fascicoli; selezionare gli item in modo che il test fornisca una stima attendibile dell'abilità dei soggetti (cioè sia in grado di rilevare effettivamente ciò per la cui misurazione sono stati concepiti, minimizzando quindi la quantità di errore di rilevazione); verificare che i fascicoli e gli item siano di difficoltà adeguata (al livello scolare target); e, infine, che fascicoli e item abbiano un buon potere discriminante (cioè siano in grado di differenziare i soggetti in funzione della quantità di proprietà – abilità/competenza – posseduta). L'analisi in pretest con gli strumenti TCT viene solitamente articolata in quattro fasi, rispettivamente tese a esplorare, di ciascun item, la difficoltà, la discriminatività e il contributo alla coerenza interna del test, e la dimensionalità della prova.

Le statistiche classiche consentono un primo confronto tra i comportamenti di risposta degli studenti agli item inclusi nei quattro fascicoli. Dal confronto, osserviamo innanzitutto che la media delle risposte corrette è sostanzialmente invariante nei quattro fascicoli. Secondo la Teoria dell'Errore, infatti, la difficoltà di una prova è data, per item dicotomici, dalla proporzione di risposte corrette sul totale di risposte date, che può essere ponderata per un fattore di correzione per tener conto della probabilità che ciascuno studente ha di dare per caso una risposta corretta quando gli item sono a risposta multipla (tab. 3). L'indice ha campo di variazione [0; +1], quindi più l'indice si approssima a zero, maggiore è il suo livello di difficoltà e viceversa. Un coefficiente pari o prossimo a .50 indica invece un item di media difficoltà.

È stata inoltre calcolata la discriminatività, ossia la capacità del singolo item o dell'intera prova di differenziare i soggetti in funzione del loro livello di abilità/competenza, solitamente supposto uguale, nell'ambito della TCT, al punteggio conseguito all'intero test. Per calcolare la discriminatività, si ricorre quindi a misure di associazione tra il punteggio osservato in relazione al singolo item e il punteggio totale del test.

ISBN 9788835101802

Tab. 3 – Misure descrittive della funzionalità degli item e delle prove

	Fascicolo 1					*Fascicolo 2*					*Fascicolo 3*					*Fascicolo 4*				
Item	*D*	*Miss*	*M*	*V*	*C*	*D*	*Miss*	*M*	*V*	*C*	*D*	*Miss*	*M*	*V*	*C*	*D*	*Miss*	*M*	*V*	*C*
ANCH_1	0,840	0,000	28,250	310,408	+0,095	0,808	0,000	24,640	200,433	-0,288	0,810	0,000	22,370	163,468	-0,022	0,875	0,000	21,810	168,562	-0,146
D1	0,680	0,000	28,400	312,779	-0,062	0,808	0,000	24,730	198,208	-0,058	0,857	0,000	22,370	164,801	-0,161	0,708	0,000	21,950	164,348	+0,234
D2	0,440	0,160	26,850	287,818	+0,098	0,846	0,000	24,640	196,433	+0,115	0,714	0,000	22,530	160,374	+0,231	0,458	0,042	21,760	175,390	-0,231
D3	0,320	0,040	28,350	282,029	+0,385	0,500	0,000	24,950	198,141	-0,049	0,286	0,000	22,950	159,830	+0,294	0,417	0,042	21,810	144,062	+0,422
ANCH_3	0,600	0,000	28,350	317,608	-0,351	0,615	0,000	24,770	199,708	-0,173	0,619	0,000	22,530	157,374	+0,484	0,417	0,000	22,290	165,714	+0,107
D4	0,240	0,160	27,000	220,842	+0,733	0,192	0,000	25,320	195,180	+0,213	0,286	0,190	21,050	134,386	+0,185	0,292	0,042	21,900	145,690	+0,380
D5	0,680	0,000	28,350	311,924	-0,012	0,577	0,000	24,860	196,695	+0,055	0,381	0,000	22,840	165,585	-0,191	0,917	0,000	21,760	166,190	+0,136
D6	0,600	0,000	28,550	318,787	-0,388	0,808	0,000	24,730	198,398	-0,074	0,619	0,048	22,160	164,363	-0,096	0,375	0,000	22,330	165,633	+0,118
D7	0,200	0,240	26,600	234,779	+0,518	0,115	0,115	24,550	179,974	+0,153	0,190	0,190	21,160	97,140	+0,720	0,333	0,208	20,190	139,062	+0,159
D8	0,400	0,000	28,650	312,871	-0,066	0,577	0,000	24,910	201,229	-0,266	0,619	0,000	22,630	166,135	-0,229	0,208	0,000	22,480	169,262	-0,200
ANCH_7a	0,240	0,040	28,150	287,187	+0,312	0,423	0,000	25,450	198,450	-0,134	0,429	0,000	23,050	163,497	-0,025	0,500	0,083	21,710	134,314	+0,413
ANCH_7b	0,080	0,240	26,750	235,461	+0,497	0,192	0,231	23,320	169,370	+0,140	0,143	0,143	21,630	102,023	+0,750	0,250	0,167	21,140	119,929	+0,542
D9	0,840	0,000	28,200	314,063	-0,173	0,654	0,000	24,860	192,409	+0,369	0,857	0,000	22,370	163,357	-0,010	0,792	0,000	21,860	166,129	+0,101
ANCH_4	0,480	0,000	28,500	312,684	-0,055	0,308	0,038	25,140	201,171	-0,267	0,429	0,000	22,790	164,731	-0,122	0,542	0,000	22,140	166,129	+0,071
D10	0,360	0,000	28,600	315,200	-0,194	0,423	0,000	25,050	190,712	+0,477	0,381	0,000	22,840	162,251	+0,071	0,458	0,000	22,290	168,414	-0,103
D11	0,000	0,280	26,350	251,503	+0,317	0,000	0,308	22,640	129,957	+0,504	0,048	0,333	19,840	136,474	+0,071	0,042	0,125	21,330	135,133	+0,293
ANCH_5	0,480	0,040	28,200	283,537	+0,367	0,385	0,038	24,640	165,766	+0,585	0,524	0,000	22,740	165,316	-0,165	0,583	0,000	22,100	170,390	-0,250
D12	0,280	0,000	28,750	311,882	-0,010	0,269	0,000	25,180	198,156	-0,052	0,381	0,000	22,840	164,585	-0,113	0,375	0,000	22,290	167,514	-0,033
ANCH_2	0,560	0,000	28,450	306,997	+0,267	0,500	0,000	24,950	195,188	+0,157	0,429	0,000	22,790	160,731	+0,187	0,667	0,000	22,000	166,800	+0,024
D13	0,480	0,000	28,600	308,147	+0,197	0,692	0,038	24,410	169,968	+0,513	0,667	0,000	22,530	164,152	-0,080	0,875	0,000	21,810	169,762	-0,274
D14	0,160	0,000	28,900	313,463	-0,127	0,038	0,000	25,450	198,165	-0,087	0,143	0,000	23,110	164,099	-0,099	0,083	0,000	22,570	167,557	-0,040
D15a	0,920	0,000	28,150	314,871	-0,277	0,808	0,077	23,230	197,232	-0,021	0,571	0,000	22,680	158,784	+0,337	0,542	0,000	22,140	163,029	+0,309
D15b	0,440	0,000	28,700	309,379	+0,135	0,423	0,077	23,860	144,600	+0,819			*					*		
ANCH_6	0,400	0,000	28,650	311,503	+0,011	0,346	0,000	24,680	140,418	+0,801	0,429	0,048	22,370	134,690	+0,520	0,333	0,000	22,330	171,433	-0,343
D16a	0,960	0,000	28,100	309,674	+0,282	0,962	0,000	24,500	197,690	+0,000	0,952	0,000	22,210	163,398	+0,000	0,875	0,042	21,330	174,033	-0,211
D16b	0,640	0,000	28,450	308,261	+0,194	0,615	0,000	24,820	198,251	-0,059	0,571	0,000	22,630	158,135	+0,392	0,417	0,083	21,330	146,033	+0,232
D16c	0,360	0,040	28,250	291,355	+0,245	0,231	0,038	24,860	184,600	+0,181	0,286	0,000	22,950	165,497	-0,198	0,333	0,042	21,860	145,029	+0,397
D17	0,440	0,280	25,250	231,882	+0,397	0,423	0,154	23,360	182,052	+0,050	0,714	0,048	22,050	138,275	+0,465	0,417	0,167	21,330	128,833	+0,539

* Item non presente nel fascicolo; D = indice di difficoltà; Miss = percentuale di risposte non date; M = media scala se l'item viene eliminato; V = varianza di scala se l'item viene eliminato; C = correlazione elemento-totale corretta; α = Alpha di Cronbach se viene eliminato l'elemento.

ISBN 9788835101802

Tra i diversi indici disponibili, abbiamo riportato nella tab. 3 il coefficiente di correlazione item-totale corretto, che si calcola escludendo l'item oggetto di valutazione dal punteggio totale in modo da: 1) evitare che il valore del coefficiente sia artificialmente gonfiato dalla correlazione di un item con se stesso e 2) avere informazioni circa il contributo di ciascun item al potere discriminante dell'intera prova. Solitamente, consideriamo soddisfacente il contributo di quell'item il cui coefficiente di correlazione item-totale corretto sia almeno pari a 0.25 (Barbaranelli e Natali, 2005, p. 80). Il coefficiente di correlazione item-totale corretto è utilizzato anche come misura dell'attendibilità di un singolo item (cioè come la misura in cui la sua somministrazione consente una buona misurazione del tratto da rilevare). Quindi, maggiore è l'omogeneità degli item in termini di contenuto, più alto è il valore di questo coefficiente.

Il potere discriminante di un item è fortemente influenzato dalla sua difficoltà. Tutti i profili di risposta estremi, che identificano item troppo difficili (quelli a quali cioè tutti i soggetti inclusi nel campione hanno fornito una risposta errata) o quelli troppo semplici (tutti hanno fornito una risposta corretta), annullano la varianza associata a tali item e non danno un reale contributo alla misurazione delle abilità/competenze degli studenti (cioè al loro posizionamento lungo il tratto latente) perché essi sono rispettivamente più difficili e più facili di tutti gli altri ma non sappiamo in che misura e, quindi, non sono in grado di discriminare tra i soggetti. Il potere discriminante è invece massimo quando gli item hanno un livello di difficoltà pari a .5 perché in questo caso «la varianza dell'item è massima, si può concludere che gli item risultano più informativi e più utili quando il livello di difficoltà è intermedio» (Barbaranelli e Natali, 2005, p. 81). In generale, a mano a mano che la correlazione aumenta, aumenta anche la variabilità possibile in termini di difficoltà degli item e quindi la capacità della prova di scalare soggetti e item lungo il tratto latente con uno strumento sempre più preciso.

Infine, per confrontare la funzionalità degli item nelle quattro prove, è stata costruita una matrice delle covarianze, particolarmente utile per confrontare la funzionalità delle domande del *Core Test* (non variate) e delle domande variate nelle quattro prove somministrate. Le varianze in essa contenuta lungo la diagonale principale consentono di: 1) valutare l'adeguatezza degli item del *Core Test* perché più simile è la variazione di ogni item in ciascuna delle quattro forme, più stabile risulta la loro funzionalità tra le forme, quindi più stabile è la funzionalità degli item àncora e meglio questi si prestano a svolgere il compito per il quale sono stati inseriti nei fascicoli (tabb. 4 e 5); 2) valutare l'effetto delle variazioni sulla funzionalità dell'item perché maggiore è la varianza, maggiore è l'effetto che la variazione ha apportato al comportamento di risposta degli studenti a quell'item (tab. 6).

ISBN 9788835101802

Tab. 4 – Varianza delle risposte corrette osservate per ciascun item dell'àncora esterna nei quattro fascicoli di prova

	A1	*A2*	*A3*	*A4*	*A5*	*A6*
F1	3,11	0,26	0,26	0,19	0,25	12,25
F2	0,26	0,26	0,26	0,25	0,25	0,25
F3	0,26	0,26	0,26	0,25	0,21	10,06
F4	0,26	3,25	0,25	0,23	0,25	6,11

Tab. 5 – Varianza delle risposte corrette osservate per ciascun item dell'àncora interna (Core Test) nei quattro fascicoli di prova

	ANC1	*ANC2*	*ANC3*	*ANC4*	*ANC5*	*ANC7a*	*ANC7b*
F1	0,14	0,26	0,25	0,26	3,14	3,25	15,11
F2	0,14	0,26	0,24	3,10	3,07	0,26	14,29
F3	0,16	0,26	0,25	0,26	0,26	0,26	10,16
F4	0,11	0,23	0,25	0,26	0,25	5,94	11,15

Abbiamo infine esplorato possibile interconnessioni tra gli item inclusi nelle prove calcolando la covarianza di ciascun item con ciascuno degli altri quesiti contenuti nello stesso fascicolo. Questa informazione è stata utilizzata per la composizione dei fascicoli finali perché esprime il grado di "interdipendenza" tra gli item e, quindi, dà, seppure in primissima approssimazione, qualche indicazione circa possibili violazioni dell'indipendenza locale, uno degli assunti teorici del modello di Rasch. Sulla base di tali risultati, abbiamo pertanto escluso alcuni item e modificato degli altri. In ogni caso, la valutazione delle modifiche da apportare, anche in considerazione della bassa numerosità campionaria, si è informata a criteri di opportunità che hanno mediato tra il dato quantitativo e le indicazioni teoriche provenienti dalla letteratura di settore. I restanti item sono stati inseriti nei quattro fascicoli di prova (F1, F2, F3 e F4) senza ulteriori modifiche.

ISBN 9788835101802

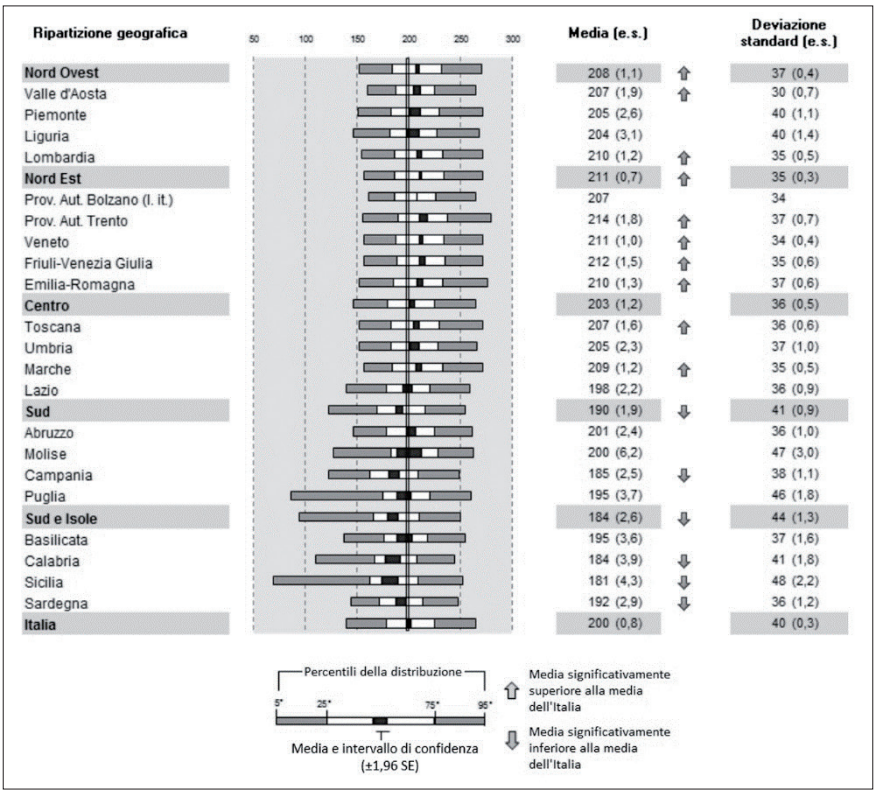
Tab. 6 – Varianza delle risposte corrette osservate per ciascun item nei quattro fascicoli di prova

	D1	*D2*	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*	*D9*	*D10*	*D11*	*D12*	*D13*	*D14*	*D15a*	*D15b*	*ANC6*	*D16a*	*D16b*	*D16c*	*D17*
F1	0,2	10,3	3,2	10,8	0,2	0,3	14,7	0,3	0,1	0,2	17,0	0,2	0,3	0,1	0,1	0,3	0,3	0,0	0,2	3,2	15,0
F2	0,2	0,1	0,3	0,2	0,2	0,2	8,5	0,2	0,2	0,3	17,9	0,2	2,8	0,0	4,9	5,6	0,2		0,2	3,1	
F3	0,1	0,2	0,2	12,3	0,2	3,5	12,6	0,2	0,1	0,2	18,6	0,2	0,2	0,1	0,3		3,7	0,0	0,3	0,2	3,4
F4	0,2	3,3	3,3	3,4	0,1	0,2	12,9	0,2	0,2	0,3	9,2	0,2	0,1	0,1	0,3		0,2	2,8	6,1	3,3	10,7

ISBN 9788835101802

2.3. Il campionamento

Il disegno di campionamento adottato in questo studio ricalca quello normalmente condotto da INVALSI nella selezione delle classi campione a cui somministrare la prova in *main study* (Falorsi, 2007), e ha preso in considerazione alcune tra le Regioni considerate maggiormente rappresentative al livello nazionale: Lazio, Campania, Lombardia ed Emilia Romagna, sia in termini di medie dei risultati osservate nelle precedenti Rilevazioni nazionali condotte da INVALSI, sia in termini di eterogeneità dal punto di vista socio-economico.



Fonte: Rapporto risultati INVALSI (2017).

Fig. 1 – Distribuzione dei punteggi di Matematica – Classe III secondaria di primo grado

Per ciascuna di queste Regioni, è stato costruito un campionamento stratificato a due livelli. All'interno del secondo strato sono stati selezionati i grappoli

ISBN 9788835101802

(cioè le classi). Per ragioni di convenienza, similmente a quanto fatto per la costruzione del campione nazionale INVALSI (Falorsi, 2007), abbiamo inoltre imposto una regola per la quale sono estraibili solo i grappoli di numerosità maggiore o uguale a 16 studenti. La stratificazione ha consentito di garantire dimensioni del campione adeguate per i sottogruppi desiderati e di aumentare la precisione delle stime complessive. Essa inoltre ha tenuto conto del background sociale, economico e culturale delle famiglie di provenienza degli studenti, da anni oramai considerato variabile imprescindibile negli studi sulle performance scolastiche (vedi anche INVALSI, 2017). Inoltre, poiché le variazioni includono anche aspetti linguistici, questa variabile è stata inserita con l'obiettivo di apprezzare l'elasticità del comportamento psicometrico degli item in gruppi socio-culturali diversi oltre a garantire una maggiore omogeneità all'interno di ciascuno strato e, quindi, migliorare la qualità delle stime prodotte.

2.3.1. Una misura di status socio-culturale

L'INVALSI, in coerenza con quanto fatto dall'OCSE nelle indagini internazionali, propone una misura di status socio-economico-culturale fondato su tre dimensioni: l'istruzione dei genitori, la loro professione e alcune misure del benessere economico della famiglia di origine indirettamente rilevate mediante *proxies* (come per esempio la disponibilità di uno spazio per studiare esclusivamente dedicato allo studente, il numero di libri presente in casa, l'accesso alla rete internet ecc.). Queste informazioni vengono rilevate da INVALSI mediante la somministrazione del questionario studente nelle classi quinte della scuola primaria e seconde della scuola secondaria di secondo grado. Per tutti gli altri livelli attualmente inclusi nelle rilevazioni INVALSI (e cioè la classe seconda della scuola primaria e la classe terza della scuola secondaria di primo grado), sono disponibili solo informazioni relative al grado di istruzione e alla professione dei genitori. Per questi ultimi due livelli scolastici, l'INVALSI non fornisce un indice di status socio-economico-culturale. In assenza di tali informazioni ma nella consapevolezza della rilevanza che il background familiare ha sulle performance degli studenti, è stata sviluppata una misura alternativa di status socio-culturale basata sull'istruzione dei genitori e il loro status professionale, che d'altra parte include anche qualche indicazione circa il benessere economico della famiglia dello studente.

Nelle rilevazioni INVALSI, la prima variabile è articolata in nove livelli (tab. 7) e raggruppata in sei classi in ragione del prestigio sociale e del livello di remunerazione di ciascuna professione (tab. 8).

Il grado di istruzione è invece classificato in sei livelli ISCED (tab. 9).

ISBN 9788835101802

Tab. 7 – Elenco delle professioni e loro descrizione

Professioni		*Descrizione*
1	Disoccupato	Molto basso (nessun prestigio sociale; nessun reddito prodotto)
2	Casalingo/a	Molto basso (nessun prestigio sociale; nessun reddito prodotto)
3	Dirigente, docente universitario, funzionario o ufficiale militare	Molto alto (prestigio sociale molto alto, reddito molto alto)
4	Imprenditore/proprietario agricolo	Alto (prestigio sociale alto, reddito alto)
5	Professionista dipendente, sottufficiale militare o libero professionista (medico, avvocato, psicologo, ricercatore ecc.)	Medio alto (status sociale alto; livello culturale alto; reddito medio-alto, e sicuro perché di natura dipendente)
6	Lavoratore in proprio (commerciante, coltivatore diretto, artigiano, meccanico ecc.)	Medio basso (status sociale medio-basso; reddito medio)
7	Insegnante, impiegato, militare graduato	Medio basso (status sociale medio; reddito sicuro di media entità)
8	Operaio, addetto ai servizi/socio di cooperativa	Basso (basso prestigio sociale; reddito prodotto basso)
9	Pensionato/a	Basso (nessun prestigio sociale; reddito basso/molto basso)

Fonte: ns. adattamento da Campodifiori *et al.* (2010).

Tab. 8 – Raggruppamento delle professioni in classi omogenee in termini di reddito e prestigio professionale

Gruppo 1	*Gruppo 2*	*Gruppo 3*	*Gruppo 4*	*Gruppo 5*	*Gruppo 6*
Disoccupato/a [1]	Operaio [8]	Lavoratore in proprio [6]	Professionista dipendente [5]	Imprenditore/ proprietario agricolo [4]	Dirigente/ docente universitario ecc. [3]
Casalingo/a [2]	Pensionato/a [9]	Insegnante, impiegato [7]			

Fonte: ns. adattamento da Campodifiori *et al.* (2010).

Tab. 9 – Classificazione dei livelli d'istruzione

Etichette	*Livelli di istruzione*	*Classificazione ISCED*	*Classificazione*
1	Licenza elementare	ISCED_1	Basso
2	Licenza media	ISCED_2	Medio basso
3	Qualifica professionale triennale	ISCED_3	Medio basso
4	Diploma di maturità	ISCED_4	Medio
5	Altro titolo di studio superiore al diploma	ISCED_5	Medio alto
6	Laurea o titolo superiore (dottorato/master...)	ISCED_6_7_8	Alto/Molto alto

ISBN 9788835101802

Per le finalità di questo studio, similmente a quanto fatto per l'ESCS da INVALSI, l'indice di status socio-culturale (SC-index, Cascella e Cavicchiolo, 2017; Stringher e Cascella, in preparazione) è stato costruito combinando il più alto livello di istruzione tra quello del padre e quello della madre e il più alto status professionale tra quello del padre e quello della madre (tab. 10).

Tab. 10 – Articolazione in classi dello SC-index

		Status professionale				
		Disoccupato	*Casalinga*	*Operaio*	*Impiegato*	*Imprenditore/ lav. autonomo*
Livello di istruzione	Basso	Basso	Basso	Basso	Medio	Medio
	Medio	Basso	Basso	Basso	Medio	Alto
	Alto	Medio	Medio	Medio	Alto	Alto

3. Risultati

3.1. Esempi di Variazioni

Il quesito riportato di seguito (D14) è tratto dalla prova INVALSI di livello 06 (prima classe della scuola secondaria di I grado) del 2012 ed è stato inserito nella versione originale nel fascicolo 3 (F3) e in tre versioni diverse negli altri fascicoli (F1, F2, F4).

Il quesito risulta particolarmente interessante sia da un punto di vista didattico perché chiama in causa una misconcezione molto studiata nella ricerca in didattica, sia da un punto di vista misuratorio in termini di andamento della risposta corretta e dei distrattori.

Nel quesito sono rappresentati su un foglio quadrettato diversi rettangoli e si chiede di confrontarne le aree e i perimetri. Il quesito non dovrebbe dare troppi problemi a studenti della scuola secondaria in quanto è possibile misurare, servendosi della quadrettatura, l'area e il perimetro dei diversi rettangoli. Osservando i risultati del main study, però, la percentuale di studenti che risponde correttamente è meno del 37%. Questa difficoltà risiede probabilmente nel fatto che il confronto tra le aree dei rettangoli risulta abbastanza immediato anche evitando il conteggio dei quadretti (dalla prima all'ultima figura infatti l'area dei rettangoli aumenta) e questo porta buona parte degli studenti a non adoperare una strategia di conteggio anche nel confronto dei perimetri. Questo possibile approccio degli studenti viene anche supportato da una misconcezione ampiamente studiata in Didattica della Matematica:

ISBN 9788835101802

molti studenti sono convinti che nel caso di figure piane, sussistano relazioni tra area e perimetro delle figure secondo cui se la figura A ha area maggiore della figura B, allora la figura A deve avere anche un perimetro maggiore della figura B e viceversa (D'Amore e Fañdino Pinilla, 2005). I dati relativi al main study confermano questa ipotesi: il 35% degli studenti sceglie il distrattore D e risponde, probabilmente in modo intuitivo e senza operare verifiche, che anche i perimetri dei rettangoli aumentano, incorrendo nella misconcezione descritta.

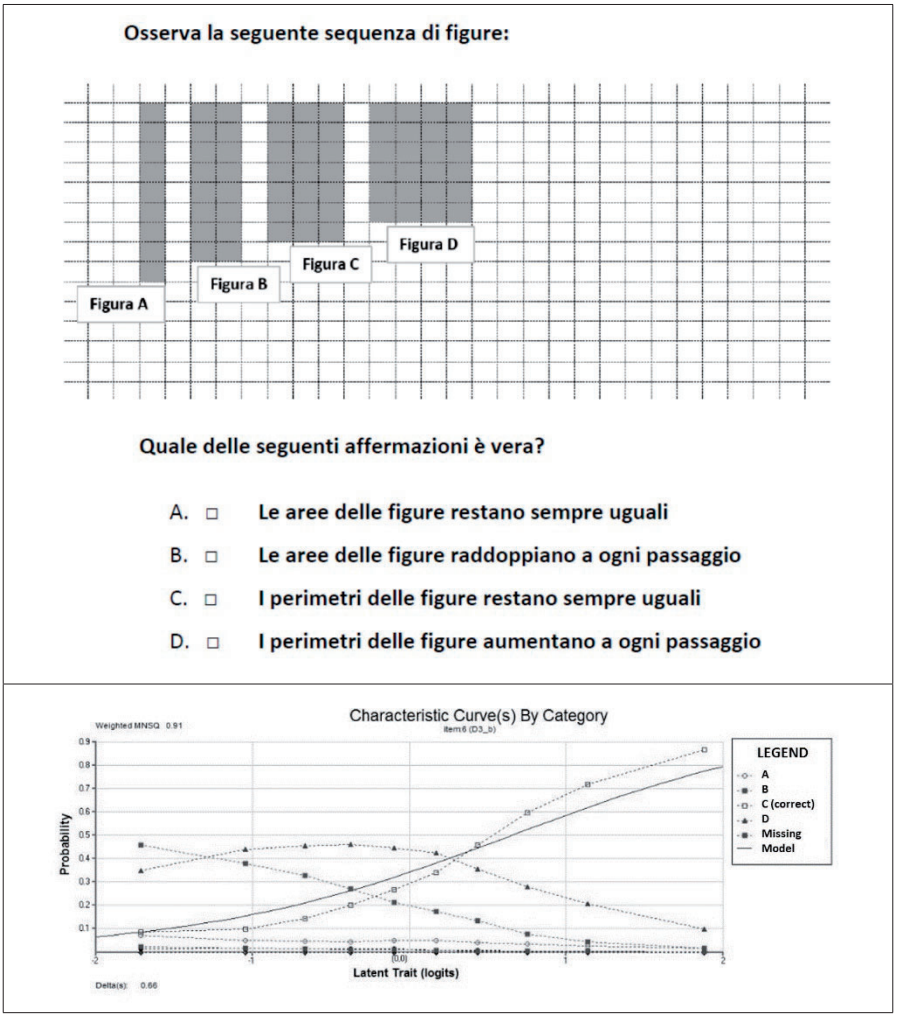


Fig. 2 – Analisi dell'item D14 versione originale: D3b prova INVALSI livello 6 del 2012

ISBN 9788835101802

Risulta interessante anche osservare l'andamento dei distrattori riportati nel grafico, gli studenti che scelgono la risposta D sono infatti principalmente studenti con abilità medie e medio basse, mentre per i livelli più bassi risulta molto attrattivo anche il distrattore B.

Per quanto riguarda l'andamento della risposta corretta invece si osserva generalmente un buon *fit* con il modello ma una tendenza a sovrastimare gli studenti con bassi livelli di abilità e sottostimare quelli con livelli alti di abilità.

Nell'articolo di D'Amore e Fañdino Pinilla (2005), incentrato sull'analisi delle misconcezioni sulle relazioni tra area e perimetro, gli autori sottolineano anche che, nel caso di figure isoperimetriche ma con diverse aree, è importante anche tenere in considerazione su cosa gli studenti focalizzano prima l'attenzione. Se infatti si chiede prima di confrontare l'area e poi il perimetro, la quasi totalità degli intervistati risponde in modo intuitivo che anche il perimetro cambia, incorrendo quindi nella misconcezione; se invece si inverte l'ordine delle domande, e si chiede prima di confrontare i perimetri e poi le aree, allora la misconcezione risulta meno forte, risponde correttamente un numero maggiore di studenti e molti di questi rispondono operando un conteggio.

Per questo motivo abbiamo scelto di riproporre nella sperimentazione del progetto Variazioni 2, questa domanda eseguendo una variazione nella tipologia e trasformandola in due risposte aperte univoche, una sulle aree dei rettangoli e una sui perimetri, quindi le altre due versioni sono nate invertendo l'ordine dei distrattori nella risposta originale e delle domande nella seconda versione.

Di seguito sono riportate le diverse versioni del quesito confrontate tra loro e i risultati ottenuti per ogni versione.

I risultati del fascicolo 3, in cui è inserita la domanda nella forma originale, confermano quanto osservato nel main study in termini di andamento della risposta corretta rispetto al modello e andamento dei distrattori. Risulta quindi interessante confrontare la versione originale con la versione inserita nel fascicolo 4 in cui i distrattori sono stati invertiti e gli studenti leggono prima le affermazioni che riguardano i perimetri e poi quelle relative all'area. In questo caso il quesito presenta caratteristiche simili al primo in termini di curva caratteristica e anche per quanto riguarda l'adattamento dei dati al modello (sovrastima dei livelli bassi e sottostima degli alti) ma la domanda nel suo complesso risulta più semplice (F4: locn = 0.363; F3: locn = 0.596). Seguendo l'esempio di D'Amore e Fandiño (2005) è possibile interpretare questa differente difficoltà con il fatto che gli studenti, incontrando prima le informazioni relative al perimetro che non permettono una risposta immediata, sono portati maggiormente a verificare le affermazioni lavorando sulla figura. Nel caso in cui invece l'attenzione è posta prima sulla relazione tra

ISBN 9788835101802

le aree, essendo evidente l'aumento delle aree, gli studenti sono portati a rispondere in modo immediato anche nel caso dei perimetri e l'influenza della misconcezione li porta maggiormente a sbagliare.

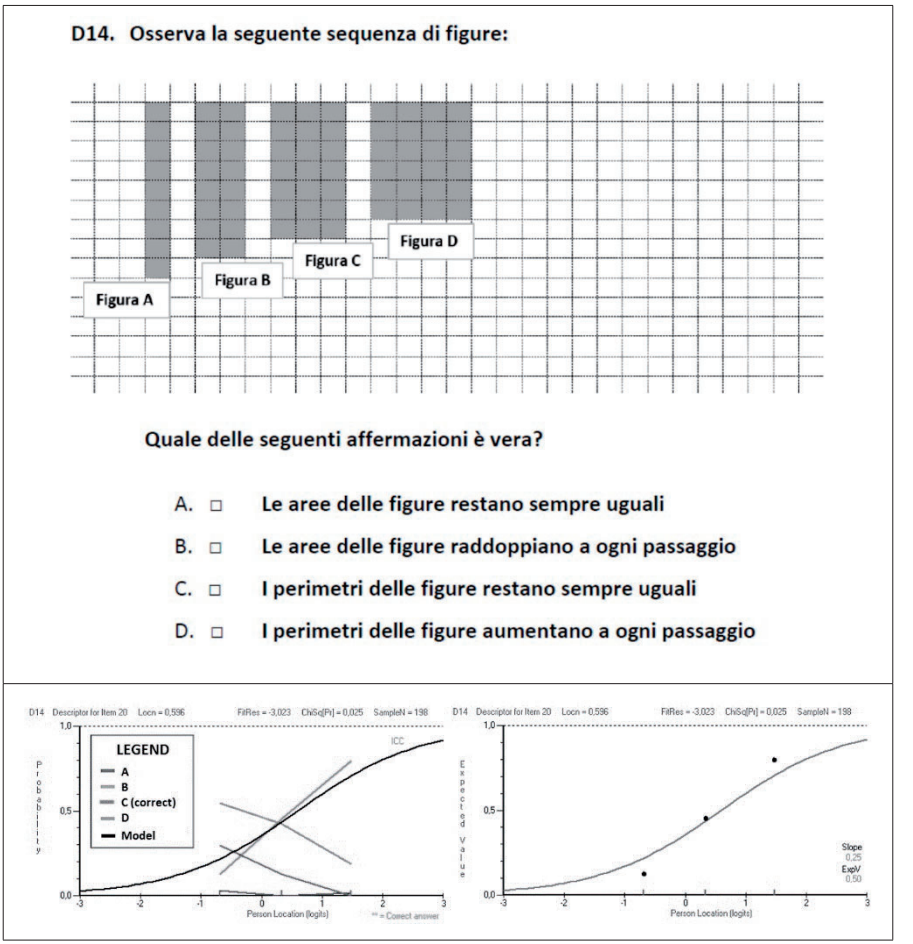


Fig. 3 – Analisi dell'item D14 versione originale inserita nel fascicolo F3

La riprova dell'immediatezza nel rispondere alle aree viene confermata dal primo item della versione del quesito riportata nel fascicolo F1 in cui le domande su perimetri e aree sono separate e viene proposta *in primis* quella sulle aree. Il primo item in questo caso risulta infatti molto semplice (F1 Aree: locn = -2.208) e poco discriminativo: anche rispondenti con livelli di abilità bassi e medi hanno una alta probabilità di rispondere correttamente. Il secondo item (F1 Perimetri: locn = 0.063), relativo ai perimetri, anche in

ISBN 9788835101802

questo caso presenta caratteristiche simili a quelle evidenziate dagli item a risposta multipla (under-discrimination del modello rispetto ai dati empirici), probabilmente causate proprio dalla incidenza della misconcezione. La stessa domanda però presentata per prima, mostra un buon funzionamento e anche un miglioramento in termini di adattamento al modello. Infatti nel fascicolo F2 la domanda relativa ai perimetri è presentata per prima e viene in questo modo ridotta l'influenza della misconcezione legata alla relazione tra area e perimetro (F2 Perimetri: locn = 0.199; F2 Aree: locn = - 0.992).

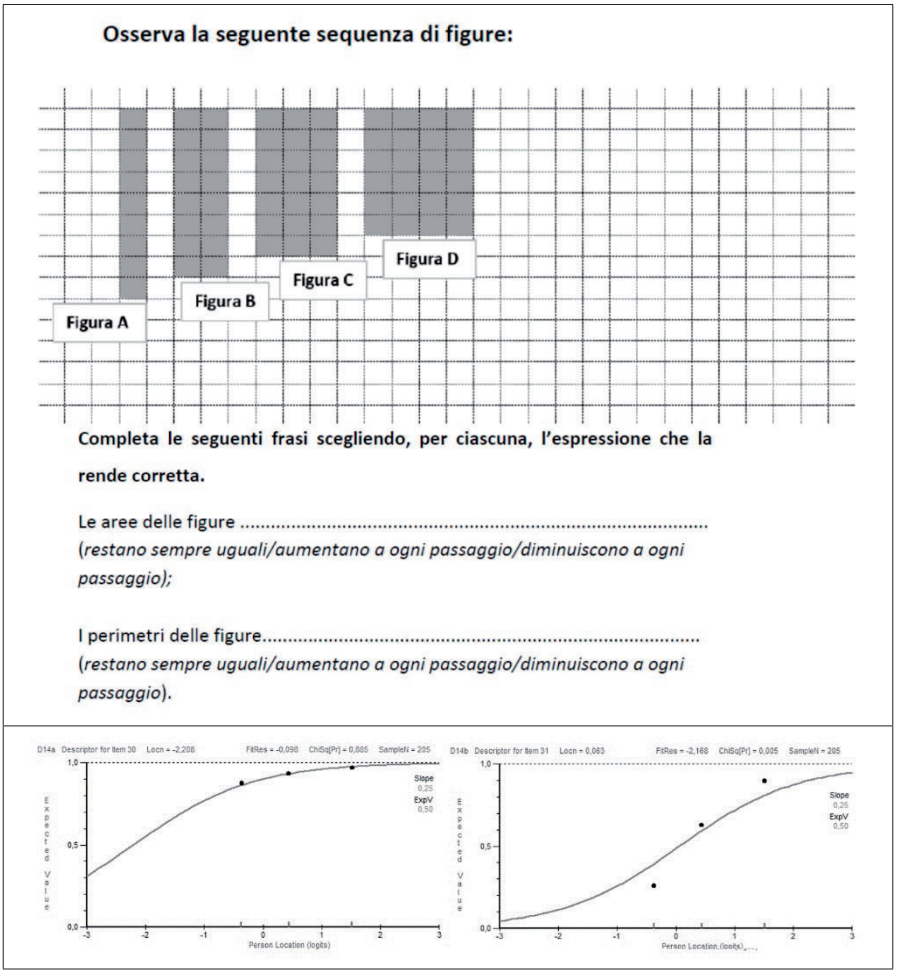
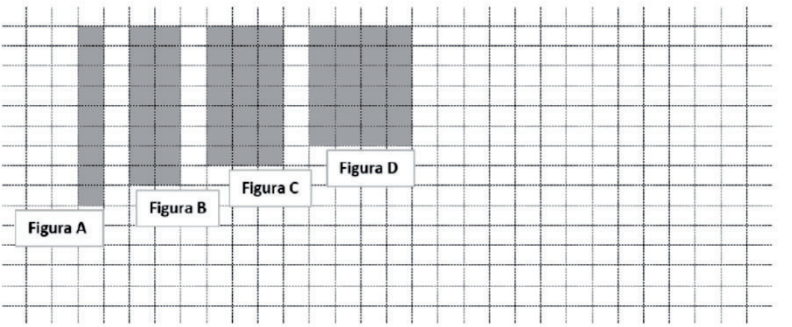


Fig. 4 – Analisi dell'item D14 versione variata inserita nel fascicolo F1 (variazione tipologia)

ISBN 9788835101802

D14. Osserva la seguente sequenza di figure:



Quale delle seguenti affermazioni è vera?

A. □ **I perimetri delle figure restano sempre uguali**

B. □ **I perimetri delle figure aumentano a ogni passaggio**

C. □ **Le aree delle figure restano sempre uguali**

D. □ **Le aree delle figure raddoppiano a ogni passaggio**

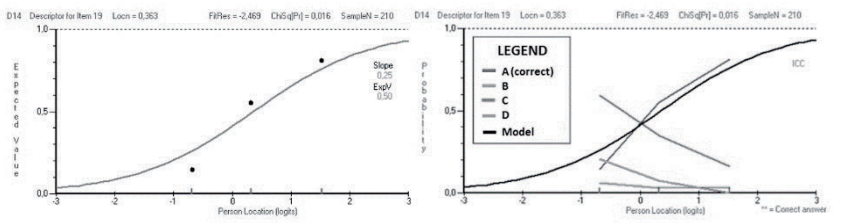
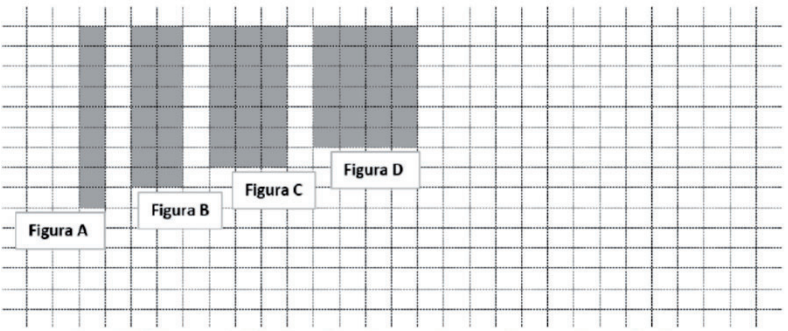


Fig. 5 – Analisi dell'item D14 versione variata inserita nel fascicolo F4 (variazione ordine)

ISBN 9788835101802

D14. Osserva la seguente sequenza di figure:



Completa le seguenti frasi scegliendo, per ciascuna, l'espressione che la rende corretta.

I perimetri delle figure.....
(*restano sempre uguali/aumentano a ogni passaggio/diminuiscono a ogni passaggio*).

Le aree delle figure .....
(*restano sempre uguali/aumentano a ogni passaggio/diminuiscono a ogni passaggio);*

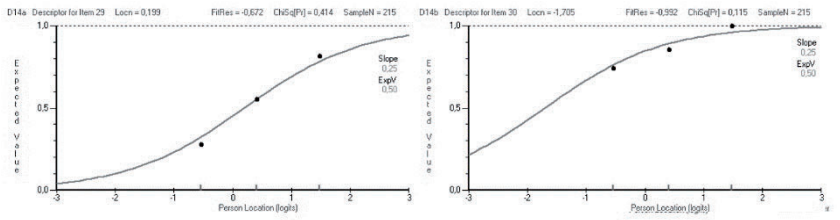


Fig. 6 – Analisi dell'item D14 versione variata inserita nel fascicolo F2 (variazione tipologia e ordine)

4. Conclusioni

Variazioni 2 è un programma di ricerca che si pone una pluralità di obiettivi ciascuno dei quali condivide però una finalità comune: studiare l'impatto che variazioni nella formulazione linguistica apportate a uno stesso item possano avere sulla sua funzionalità psicometrica, senza variare il suo *question intent*, cioè senza modificare l'obiettivo per il quale l'item è stato costruito e inserito in una prova.

ISBN 9788835101802

Per la valutazione della funzionalità psicometrica degli item, abbiamo confrontato non solo (e non tanto) il parametro di difficoltà stimato dal modello per ciascuno degli item variati, ma abbiamo lavorato, oltre che con gli indici di *infit* stimati per ciascun item, anche sull'esplorazione e sul confronto delle curve caratteristiche degli item costruite nel framework della Rasch analysis (Bolondi e Cascella, 2017). La curva caratteristica degli item (ICC) esprime infatti la probabilità di dare una risposta corretta a un certo item in funzione del livello di abilità posseduto dallo studente. Le analisi riportate in questo paper, oltre a tener conto delle variazioni in termini di difficoltà percepita dallo studente e del migliore o peggiore adattamento dei dati al modello globalmente valutato attraverso l'indice di *infit*, si sono basate principalmente sul confronto tra la curva teorica stimata dal modello (sulla quale giacciono le probabilità di dare una risposta corretta a un certo item in funzione del livello di abilità stimato dal modello per ciascuno studente) con la spezzata empirica (data dall'insieme di tutte le risposte effettivamente date da tutti gli studenti a ciascun item) (Cascella, 2016). Per capire quindi l'effetto delle variazioni in termini di funzionalità degli item abbiamo confrontato la curva teorica con quella empirica di tutti gli item contenuti nei quattro fascicoli, appositamente costruiti e somministrati per le finalità di Variazioni 2.

I risultati riportati in questo lavoro, e ottenuti su una parte del campione che abbiamo estratto con metodo probabilistico e raggiunto nei mesi di marzo e aprile 2017, risultano in linea con il quadro teorico definito da Daroczy in relazione all'impatto delle variabili di formulazione su un *word problem* e con lo schema Duval-Laborde per categorizzare le varianti utilizzate, ma suggeriscono anche nuove linee interpretative che possono essere utilizzate per comprendere fenomeni molto rilevanti per la costruzione dei test in campo educativo, come per esempio fenomeni legati al funzionamento differenziale di un item in ragione di variabili che il modello di Rasch considera spurie rispetto alla stima delle abilità (per esempio le variabili personali dello studente, come il genere).

Le possibilità di impiego dei risultati del presente studio sembrano essere molteplici. Questa ricerca, infatti, consente non solo di mettere a fuoco le ragioni per le quali un item – il cui *question intent* è invariante tra le formulazioni e in linea con il quadro di riferimento – possa mostrare una funzionalità coerente con il modello di Rasch in alcuni casi e non in altri, ma suggerisce anche ulteriori piste di ricerca, attraverso l'implementazione del nostro impianto metodologico, per l'esplorazione dei processi cognitivi attivati dallo studente e per formulare alcune indicazioni sul come costruire versioni alternative di uno stesso item senza tradirne il *question intent*.

ISBN 9788835101802

Riferimenti bibliografici

Alagumalai S., Curtis D. (2005), "Classical Test Theory", in S. Alagumalai, D. Curtis, N. Hungi, *Applied Rasch Measurement: A book of exemplars*, Springer, Dordrecht (The Netherlands), pp. 1-14.

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 *I dati INVALSI: uno strumento per la ricerca*, FrancoAngeli, Milano.

Bolondi G., Cascella C., Giberti C. (2017), "Highlights on gender gap from Italian standardized assessment in mathematics", in J. Novotnà, H. Moravà (*eds.*), *SEMT 17 proceedings – International Symposium Elementary Maths Teaching*, Universita Karlova Press, Prague, pp. 81-90.

Branchetti L., Viale M. (2015), "Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico", in M. Ostinelli (2015), La didattica dell'italiano. Problemi e prospettive, *Proceedings della conferenza Quale didattica dell'italiano? Problemi e prospettive, Locarno, ottobre 2014*, pp. 139-148.

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., Fandiño Pinilla M.I. (2005), "Area e perimetro Relazioni tra area e perimetro: convinzioni di insegnanti e studenti", *La matematica e la sua didattica*, 2, pp. 165-190.

D'Amore B. (2014), *Il problema di matematica nella pratica didattica*, Digital Index, Modena.

D'Amore B., Fandiño Pinilla, M. I. (2005), "Relazioni tra area e perimetro: convinzioni di insegnanti e studenti", *La matematica e la sua didattica*, 2, pp. 165-190.

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.

Falorsi D. (2007), "Nota metodologica sulla strategia di campionamento del sistema nazionale di valutazione delle competenze", *Working paper INVALSI*, testo disponibile al sito: http://www.INVALSI.it/download/INVALSI_indagine_SNV_strategia.pdf, data di consultazione 27/1/2020.

ISBN 9788835101802

Giampaglia G. (1990), *Lo scaling unidimensionale nella ricerca sociale*, Liguori, Napoli.
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. Szitanyi (*eds.*), *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education*, 2, IGPME, Szeged, pp. 275-282.
INVALSI (2017), *Rilevazioni nazionali degli apprendimenti 2016-2017. Rapporto risultati*, Roma.
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.
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.
Nesher P. (1982), "Levels of description in the analysis of addition and subtraction word problems", *Addition and subtraction: A cognitive perspective*, pp. 25-38.
Sbaragli S. (2012), "Il ruolo delle misconcezioni nella didattica della matematica", in B. Bolondi, M.I. Fandiño Pinilla (2012), *I quaderni della didattica. Metodi e strumenti per l'insegnamento e l'apprendimento della matematica*, pp. 121-139.
Sirin S.R. (2005), "Socioeconomic Status and Academic Achievement: A Meta-Analytic Reviwe of Research", *Review of Educational Research*, 75 (3), pp. 417-453.
Spranos G., Rhodes N.C., Dale T.C., J. Crandall (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*, Lawrence Erlbaum Associates, Hillsdale (NJ), pp. 221-240.
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.
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.

ISBN 9788835101802