PhD in Mathematics

Department of Mathematics

Funzioni e potenzialità dell'analisi statistica di test su larga scala in didattica della matematica

Chiara Giberti

UNIVERSITÀ DEGLI STUDI DI TRENTO

2017

Supervisors:

Giorgio Bolondi, University of Bolzano

Claudio Fontanari, University of Trento

Doctoral thesis in **Mathematics, XXIX cycle**

Department of Mathematics, **University of Trento**

Academic year **2016/2017**

Supervisor/s: **Giorgio Bolondi, University of Bolzano**

Claudio Fontanari, University of Trento

University of Trento

Trento, Italy

A Picci,
solida spalla per le mie insicurezze,
ottimo complice nei miei mille progetti,
sperando di condividere con te tante altre soddisfazioni
e momenti di "entusiasmo gratuito"!

Indice

Introduzione

Questa tesi si propone di mettere in luce come l'analisi statistica di prove standardizzate di matematica possa avere importanti ricadute per lo studio di fenomeni didattici.
In particolare, le ricerche presentate si riferiranno alle prove INVALSI studiate attraverso il modello di Rasch. Si mostrerà come questo approccio possa far emergere macro-fenomeni già osservati in didattica della matematica, studiandoli anche da un punto di vista quantitativo. In particolare verranno utilizzati per questi casi dei grafici, detti *distractor plot*, che mostrano l'andamento della risposta corretta e delle altre opzioni di risposta in funzione dell'abilità degli studenti. Questi grafici, applicati all'intera popolazione o a sottoinsiemi della stessa, permetteranno di evidenziare se una determinata risposta degli studenti, legata a un costrutto didattico, abbia una maggiore influenza su particolari livelli di abilità.

Negli ultimi due decenni sono stati introdotti test standardizzati a livello internazionale come le prove OECD-PISA e IEA-TIMSS che hanno portato a importanti riflessioni nel campo delle politiche educative, suscitando anche molto interesse nell'opinione pubblica.
A livello nazionale sono sempre più numerosi i test standardizzati proposti per monitorare gli apprendimenti degli studenti e in Italia, dal 2008, sono state introdotte le prove INVALSI che vengono proposte a tutti gli alunni in diversi momenti del loro percorso scolastico.
Nel primo capitolo della tesi saranno introdotte le principali rilevazioni nazionali e internazionali relative all'apprendimento della matematica, verranno quindi confrontati brevemente i Quadri di Riferimento (QdR) delle diverse indagini e le finalità che si pongono.

Dalla costruzione delle prove, seguendo i principi esplicitati nel QdR, si potrà approfondire il modo in cui i risultati di tali prove vengono analizzati da un punto di vista statistico.
Nel secondo capitolo saranno introdotti i principali metodi di analisi dei risultati dei test INVALSI attraverso gli strumenti forniti dalla Classical Test Theory e dall'Item Response Theory. Un approfondimento particolare sarà dedicato all'introduzione del modello di Rasch, principalmente utilizzato per l'analisi dei quesiti del *pretest* e del *main study* e alla base delle ricerche presentate in questa tesi.

Nonostante l'interesse verso queste indagini su larga scala sia crescente, risulta però ancora limitato l'uso delle prove standardizzate per la ricerca in didattica della matematica: in particolare in Italia, le prove INVALSI vengono utilizzate solo raramente, come evidenziato nell'articolo riportato nel terzo capitolo. Con l'obiettivo di colmare questa lacuna, si è deciso di partire osservando il modo in cui le prove internazionali PISA vengono usate per la ricerca in didattica. Dall'analisi emergeranno potenzialità ancora non sfruttate di queste prove e possibili piste di ricerca da seguire sfruttando i dati INVALSI.

Una delle tematiche più affrontate attraverso l'analisi delle prove PISA è l'*equità nell'educazione matematica,* ovvero le modalità attraverso le quali garantire a ogni studente di trarre il maggior profitto dall'educazione matematica.

Nel quarto capitolo verranno affrontati il tema dell'*equità nell'educazione matematica* e, in particolare, saranno approfondite le differenze che emergono nell'apprendimento della matematica considerando gli studenti suddivisi in base alla cittadinanza e al genere. Studi di questo tipo sono spesso condotti attraverso l'uso di prove standardizzate che garantiscono il confronto e la generalizzazione dei risultati; spesso però le ricerche si limitano a una analisi del gap complessivo sull'intera prova.

Una delle novità di questa tesi consisterà nel portare questo tipo di analisi a livello dei singoli quesiti, individuando le domande che mettono maggiormente in evidenza queste differenze, esaminandole attraverso l'uso di particolari grafici che mostrano l'evoluzione del gap in funzione del livello di abilità sull'intero test (*distractor plot)*.

L'analisi dei test standardizzati in didattica della matematica si limita spesso all'analisi complessiva dei dati, in questa tesi si utilizzeranno i dati complessivi della prova (a cui sarà applicato il modello di Rasch) per poi focalizzare le analisi sui singoli quesiti perché, come specificato anche da Leder e Lubienski (2015):

> *Item-level analyses can pinpoint the mathematics that students do and do not know, including which problems most students can and cannot solve, and which problems have the largest disparities between groups. This information can inform both textbook writers and teachers, as they strive to address curricular areas in need of additional attention. Hence, it is important for item-level analyses to be systematically conducted and reported.*

Il tema delle differenze di genere in matematica risulta di particolare interesse, sia perché le rilevazioni internazionali mostrano una situazione particolarmente critica in Italia, sia perché evidenzia caratteristiche molto particolari a confronto con il gap di cittadinanza. Nel quarto capitolo verranno approfonditi questi aspetti anche attraverso una revisione della letteratura su questo tema e saranno presentate le analisi effettuate sulle prove INVALSI con esempi di quesiti particolarmente significativi.

Nell'ultimo capitolo saranno descritte due ricerche che mirano ad analizzare l'impatto di una variazione nella formulazione di un quesito di matematica sulle risposte degli studenti. Nonostante siano numerose le ricerche in questa direzione, la nuova metodologia presentata in questa tesi, basata sul *modello di Rasch* e sulle *tecniche di ancoraggio* (paragrafo 2.6), fornirà la possibilità di superare i principali ostacoli nello studio di questo tema.
La metodologia presentata fornisce anche importanti indicazioni relative all'impatto di una variazione su determinati sottogruppi della popolazione. Nella prima ricerca presentata in questo capitolo si è evidenziato un diverso effetto di una variazione sulle performance di maschi e femmine che è stato esaminato anche in funzione del livello di abilità degli studenti. A partire da questo risultato, nella predisposizione dei fascicoli per la seconda ricerca, sono stati inseriti anche quesiti che avevano mostrato un gender gap particolarmente marcato alla luce delle analisi oggetto del quarto capitolo. Queste domande sono state variate cercando di eliminare ciò che, nell'analisi a priori, poteva essere considerata la causa del divario e confrontando il gap tra la domanda originale e quella variata è stato possibile convalidare o rigettare le ipotesi fatte.

Capitolo 1

Valutazioni su larga scala in didattica della matematica

L'utilizzo di valutazioni standardizzate, a livello nazionale e internazionale, potrebbe sembrare un fenomeno recente soprattutto perché in Italia, questo tema è diventato di rilevanza pubblica nell'ultimo decennio, dall'introduzione delle prove INVALSI. In realtà, le maggiori indagini a livello internazionale, che sono le prove OCSE PISA e IEA-TIMSS, sono nate circa vent'anni fa e hanno le radici in valutazioni standardizzate ben più antiche. In Cina, già 3000 anni fa vennero utilizzati i primissimi test per selezionare il personale per la burocrazia statale attraverso i cosiddetti *imperial examination* (Kenney & Schloemer, 2001) e, da allora, in Cina i test standardizzati hanno assunto un ruolo fondamentale anche in campo educativo.

In generale, i test standardizzati in ambito educativo si sono sviluppati principalmente nella seconda metà del secolo scorso e con caratteristiche diverse a seconda delle nazioni. L'Italia ha partecipato alle prove TIMSS e PISA fin dal loro inizio (rispettivamente nel 1995 e nel 2000) e dal 2008 ha sviluppato un sistema di prove standardizzate a livello nazionale a cura dell'Istituto INVALSI.

Con il termine "standardizzato" si vuole sottolineare che un test viene effettuato e valutato nello stesso modo per tutti gli studenti, la definizione di "standardized test" tratta dal "The Glossary of Education Reform" è la seguente:

> *A standardized test is any form of test that (1) requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same way, and that (2) is scored in a "standard" or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. While different types of tests and assessments may be "standardized" in this way, the term is primarily associated with large-scale tests administered to large*

> *populations of students, such as a multiple-choice test given to all the eighth-grade public-school students in a particular state, for example.* (Abbott et al., 2014)

Nel campo della ricerca in didattica della matematica negli ultimi anni si è messo in discussione il termine "large scale" e ci si è interrogati su quali ricerche possano rientrare nella categoria delle ricerche su larga scala. Middleton, Cai e Hwang (2015) riprendono gli studi di Anderson e Postlethwaite (2007) su questo problema e definiscono le ricerche su larga scala non solo in termini di dimensione del campione analizzato ma anche in base allo scopo della ricerca, alle modalità e alla complessità relative all'analisi dei dati, alla possibilità di generalizzare i risultati.

In questa tesi tutte le analisi saranno basate su test standardizzati che rientrano a pieno titolo in questa tipologia di ricerche e si tratterà quindi di ricerche standardizzate su larga scala.

1.1 Valutazioni standardizzate internazionali

Le principali indagini internazionali volte a misurare gli apprendimenti della matematica e di altre discipline sono le prove PISA promosse dall'OCSE e quelle TIMSS promosse dall'*International Association for Evaluation of Educational Achievement* (IEA).

Le due indagini, che considerano anche paesi diversi e che si riferiscono a diversi livelli scolastici, forniscono una panoramica molto ampia e approfondita relativamente all'apprendimento della matematica.

1.1.1 Indagine OECD-PISA

Le prove PISA (*Programme for International Student Assessment*) sono state somministrate per la prima volta nel 2000 e hanno cadenza triennale. Gli ambiti investigati in ogni test PISA sono Lettura, Matematica, Scienze e Financial Literacy ma ogni rilevazione è focalizzata principalmente su uno di questi (ad esempio nel 2015 l'approfondimento era relativo alla Matematica).

L'indagine è rivolta a studenti quindicenni scolarizzati (indipendentemente dalla classe frequentata e dal tipo di scuola) e il test viene somministrato a un campione di studenti per ogni paese partecipante con la finalità di valutare l'acquisizione delle competenze chiave per la piena partecipazione alla società moderna (OECD, 2016c). I paesi partecipanti nel 2015 sono stati 72 per un totale di circa 540 000 studenti.

Le finalità del PISA sono orientate alla valutazione delle competenze chiave per l'età adulta e il test risulta per questo motivo più slegato dai curricula scolastici rispetto al TIMSS.

I quesiti proposti dal PISA sono quindi orientati ad attività di problem solving in contesto reale e la definizione di mathematical licteracy fornita nel framework rispecchia questa concezione:

> ***Mathematical literacy*** *is an individual's capacity to formulate, employ, and interpret mathematics in a variety of contexts. It includes reasoning mathematically and using mathematical concepts, procedures, facts and tools to describe, explain and predict phenomena. It assists individuals to recognise the role that mathematics plays in the world and to make the well-founded judgments and decisions needed by constructive, engaged and reflective citizens.* (OECD, 2016c)

Nelle prove PISA i quesiti vengono classificati in base al *contesto* (personale, occupazionale, scientifico o pubblico), al *contenuto* matematico (quantità, spazio e forme, cambiamento e relazioni, incertezza e dati) e ai *processi* relativi alla risoluzione di un problema matematico (formulare, utilizzare e interpretare).

1.1.2 Indagine IEA-TIMSS

Le rilevazioni TIMSS (*Trends in International Mathematics and Science Study*) iniziano pochi anni prima (nel 1995) rispetto alle indagini PISA e si svolgono ogni quattro anni. I test sono rivolti studenti che frequentano la quarta primaria (livello 4) e la terza secondaria di primo grado (livello 8) e anche in questo caso il test non è censuario ma in ogni paese viene selezionato un campione di studenti per entrambi i livelli.

I paesi partecipanti all’ultima rilevazione, avvenuta nel 2015, sono stati 57 e sono, in parte, nazioni diverse rispetto a quelle che partecipano al PISA, motivo per cui a volte i risultati delle due indagini possono risultare contrastanti.

Gli ambiti presi in esame dai test TIMSS sono scienze e matematica e, in generale, le prove sono focalizzate sull’analisi dei curricula di queste due discipline. Le indagini TIMSS hanno infatti tra le finalità quella analizzare e confrontare i curricula ministeriali (*intended curriculum*), con i curricula concretamente svolti nelle scuole (*implemented curriculum*) e con ciò che gli studenti apprendono (*atteined curriculum*). Per fare ciò i test somministrati agli studenti sono relativi alle parti comuni dei diversi curricula e vengono svolti altri studi con lo scopo di indagare le prime due accezioni di curriculum descritte.

Nell’ambito della matematica i quesiti proposti nei test TIMSS vengono classificati in base a due caratteristiche principali: i domini di contenuto e i domini cognitivi. In base al contenuto matematico gli item vengono suddivisi in 3 categorie al livello 4: *Numeri*, *Geometria* e *Dati e probabilità*. A questi si aggiunge, al livello 8, anche il dominio *Algebra*.

Per quanto riguarda i domini cognitivi i quesiti vengono invece suddivisi in: *Knowing* (conoscere concetti e fatti), *Applying* (applicare procedure e conoscenze già acquisite) e *Reasoning* (risolvere problemi non routinari e complessi).

1.2 Valutazioni standardizzate nazionali: le prove INVALSI

In molti paesi, a fianco delle indagini internazionali descritte, sono state sviluppate anche indagini di carattere nazionale che permettono di approfondire le evidenze fornite da PISA e TIMSS, tenendo anche conto delle specificità del paese considerato. In Italia dall’anno scolastico 2007/2008 sono state somministrate in diversi livelli scolastici delle prove standardizzate di matematica e italiano (grammatica e comprensione del testo) a cura del Servizio Nazionale di Valutazione (SNV) che fa capo all’Istituto INVALSI.

Come evidenziato anche nel Rapporto Nazionale del 2017 (INVALSI, 2017), l’INVALSI ha l’incarico di *“attuare verifiche periodiche e sistematiche sulle conoscenze ed abilità degli studenti”* (in base al d. lgs. n. 286/2004).

L’obiettivo principale delle prove INVALSI è quello di fornire informazioni riguardanti gli apprendimenti degli studenti in matematica e italiano nella scuola italiana. Le prove devono essere considerate come una possibilità di avere una visione di sistema relativa agli apprendimenti, individuare particolari necessità e quindi poter intervenire efficacemente. Queste informazioni possono essere utili a diversi livelli a partire dalle decisioni politiche riguardanti il mondo della scuola, passando per le amministrazioni, le dirigenze e le singole scuole e arrivando fino ai singoli docenti, agli studenti e ai genitori. Inoltre, come già sottolineato, negli ultimi anni le prove INVALSI stanno mostrando le proprie potenzialità anche nel campo della ricerca in didattica.

Per questo motivo, risulta fondamentale la restituzione dei dati di tali prove in modo che i risultati siano approfonditi ma, al contempo, chiari e intellegibili per diverse tipologie di fruitori. In questa prospettiva è fondamentale il legame, esplicitato nel Quadro di Riferimento INVALSI, tra le prove e le Indicazioni Nazionali per il Curricolo che delineano le conoscenze e le competenze fondamentali per ogni grado di istruzione in Italia.

Le prove INVALSI nei diversi anni hanno interessato diversi livelli scolastici, come si può notare nella tabella sottostante e, negli ultimi anni, le classi II e V primaria, III secondaria di primo grado (Prova Nazionale facente parte dell'esame di Stato) e la II secondaria di secondo grado.

	LIVELLO 2 (II primaria)	LIVELLO 5 (V primaria)	LIVELLO 6 (I sec. 1°grado)	LIVELLO 8 (Prova Nazionale III sec. 1°grado)	LIVELLO 10 (II sec. 2°grado)
A.S. 2007/08					
A.S. 2008/09					
A.S. 2009/10					
A.S. 2010/11					
A.S. 2011/12					
A.S. 2012/13					
A.S. 2013/14					
A.S. 2014/15					
A.S. 2015/16					
A.S. 2016/17					

Tabella 1.1: Test INVALSI somministrati dal 2007/08 ad oggi.
Le caselle in grigio indicano una prova INVALSI in quell'anno e in quel livello.

Le rilevazioni hanno carattere censuario: le prove vengono somministrate a tutti gli studenti frequentanti la classe interessata sia nelle scuole pubbliche, sia in quelle paritarie. Ogni anno gli studenti che affrontano le prove INVALSI sono circa 500-600 mila per ogni livello scolastico interessato e i dati vengono restituiti alle singole scuole e alle istituzioni.

I dati che vengono però analizzati al fine della rilevazione degli apprendimenti a livello nazionale sono relativi a un campione di studenti rappresentativo dell'intera popolazione composto da circa 30-40 mila alunni per ogni livello scolastico. Nelle classi facenti parte del campione la somministrazione dei fascicoli e l'inserimento dei dati è effettuato direttamente da un osservatore esterno inviato dall'INVALSI.

La costruzione di ogni prova INVALSI è frutto di un anno e mezzo di lavoro dovuto alla necessità di pretestare le domande allo stesso livello scolastico e nello stesso momento dell'anno ma un anno prima rispetto al main study. Alla creazione delle domande partecipano insegnanti (circa 100 da tutta Italia), ricercatori e esperti di didattica e il principale riferimento per la stesura dei quesiti è il Quadro di Riferimento INVALSI (QdR). All'interno del Quadro di Riferimento sono esplicitate le linee guida da seguire nella stesura dei quesiti e nella creazione delle nuove prove. Il Quadro di Riferimento è strettamente connesso con le Indicazioni Nazionali per il Curricolo (MIUR, 2012): ogni quesito deve essere etichettato in base allo scopo che si pone, detto *question intent* (accezione simile a quella del *question intent* delle prove OECD-PISA) e in base agli obiettivi e ai traguardi presenti nelle Indicazioni Nazionali. Le domande possono essere di diverse tipologie: risposta aperta, risposta aperta univoca, risposta multipla, risposta chiusa e *cloze*[1]. Il QdR, inoltre, stabilisce gli ambiti di contenuto nei quali le domande devono essere suddivise e i processi cognitivi che l'allievo deve attivare per rispondere alla domanda.

A partire dai quesiti creati seguendo le indicazioni del QdR, vengono creati dei fascicoli che vengono sottoposti a una fase di pretest. L'analisi dei dati del pretest, effettuata con le metodologie descritte nel capitolo 3, permette di avere le prime indicazioni sul funzionamento dei singoli quesiti e della prova nel suo complesso. In seguito alla selezione, revisione e modifica dei quesiti andati in pretest vengono quindi composti i fascicoli definitivi che saranno utilizzati nel *main study*.

[1] Per approfondimenti riguardo alle diverse tipologie di quesiti si rimanda al Quadro di Riferimento INVALSI 2017 (http://www.invalsi.it/invalsi/doc_evidenza/2017/QdR2017_190417.pdf)

Come già specificato, la prova di Italiano è composta da due test distinti che intendono valutare la comprensione nella lettura di un testo e le conoscenze di grammatica. La prova di matematica, invece, è unica e nel paragrafo seguente ne viene approfondita la struttura.

1.2.1 Struttura delle prove INVALSI di Matematica

Le prove di Matematica INVALSI sono create seguendo le indicazioni appena descritte e specificate nel QdR. In particolare per ogni domanda deve essere definito un *question intent* ben determinato e deve essere individuato l'ambito di contenuto prevalente a cui la domanda appartiene. Gli ambiti di contenuto sono definiti sulla base delle Indicazioni Nazionali e in linea con i *frameworks* delle indagini internazionali IEA-TIMSS e OECD-PISA (OECD, 2016c; Mullis & Martin, 2013).

Gli ambiti di contenuto definiti nel Quadro di Riferimento INVALSI per la prova di Matematica sono *Numeri, Spazio e figure, Relazioni e funzioni, Dati e previsioni*; questo in tutti i livelli scolastici ad eccezione della seconda primaria in cui non viene considerato l'ambito *Relazioni e funzioni*. Al fine di costruire prove il più possibile confrontabili tra loro, negli anni la proporzione di item per ogni ambito è simile se si confrontano prove relative allo stesso livello scolastico.

Risulta più complessa la definizione delle categorie riguardanti i processi messi in atto dagli allievi nella risoluzione di un quesito di matematica; nei diversi anni sono stati analizzati i processi utilizzando categorie diverse che, nelle ultime prove, si è deciso di raggruppare in tre macro-processi: *Argomentare, Risolvere Problemi* e *Conoscere.*

Il numero di quesiti che costituiscono una prova varia a seconda del livello e a seconda del numero di item che costituiscono i singoli quesiti. In generale si cerca di non variare eccessivamente il numero di item di una prova di un determinato livello di anno in anno per non modificarne la lunghezza. Le prove della primaria, ovviamente, sono composte da un numero inferiore di item rispetto alle prove della secondaria e come tipologia di quesiti si predilige la risposta multipla. A titolo esemplificativo viene riportata di seguito la suddivisione di ogni prova del 2017 in base alle caratteristiche dei quesiti.

Classe	Ambiti di contenuto	N. quesiti per ambito	N. item per ambito	N. item per formato
II Primaria	- Numeri - Spazio e Figure - Dati e Previsioni	15 8 5	15 8 5	11 a scelta multipla semplice 0 a scelta multipla complessa 17 a risposta aperta univoca
	Totale	28	28	
V Primaria	- Numeri - Spazio e figure - Dati e Previsioni - Relazioni e funzioni	10 9 10 10	13 9 14 10	13 a scelta multipla semplice 9 a scelta multipla complessa 24 a risposta aperta univoca
	Totale	39	46	
III Sec. 1° grado	- Numeri - Spazio e figure - Dati e Previsioni - Relazioni e funzioni	10 10 9 11	10 13 12 15	14 a scelta multipla 14 a scelta multipla complessa 22 a risposta aperta univoca
	Totale	40	50	
II Sec. 2° grado	- Numeri - Spazio e figure - Dati e Previsioni - Relazioni e funzioni	12 9 9 10	18 9 11 15	14 a scelta multipla 19 a scelta multipla complessa 20 a risposta aperta univoca
	Totale	40	53	

Tabella 1.2: Composizione prove INVALSI 2017 in base al numero e alle caratteristiche degli item. Fonte: Rapporto Nazionale Prove INVALSI 2017.

Risulta infine interessante confrontare i diversi modi di classificare i quesiti delle prove INVALSI, PISA e TIMSS che come si può notare dalle tabelle riportate di seguito risultano in buona parte coerenti tra loro.

ARTICOLAZIONE DEI CONTENUTI			
SNV - INVALSI	PISA 2015	TIMSS 2015 IV Livello	TIMSS 2015 VIII Livello
Numeri	Quantità	Numero	Numero
Spazio e figure	Spazio e forme	Figure geometriche e misure	Geometria
Dati e previsioni	Incertezza e dati	Visualizzazione dati	Dati e probabilità
Relazioni e funzioni	Cambiamento e relazioni		Algebra

Tabella 1.3: Confronto tra la classificazione dei quesiti nei tre framework in base al contenuto matematico. Fonte: Rapporto Nazionale Prove INVALSI 2017.

ASPETTI TRASVERSALI		
PROVE INVALSI	**PISA 2015**	**TIMSS 2015**
DIMENSIONI	PROCESSI	DOMINI COGNITIVI
Conoscere Risolvere problemi Argomentare	Formulare Utilizzare Interpretare	Conoscenza Applicazione Ragionamento

Tabella 1.4: Confronto tra la classificazione dei quesiti nei tre framework in base agli aspetti trasversali. Fonte: Rapporto Nazionale Prove INVALSI 2017.

Capitolo 2

Strumenti statistici per l'analisi dei dati

Le analisi statistiche sviluppate in questa tesi per lo studio dei dati INVALSI hanno come punto di partenza quelle prodotte dal team dell'Istituto in fase di costruzione, validazione dei test e di studio dei risultati dei main study. Il campione di riferimento è lo stesso utilizzato per le rilevazioni nazionali, quindi composto approssimativamente da 30.000-40.000 studenti, appartenenti a classi dove i test si sono svolti con la supervisione di un osservatore esterno. Le caratteristiche misuratorie, dei test e dei singoli item all'interno di ogni test, sono già state ampiamente studiate e validate dall'INVALSI e tutte le analisi che seguiranno nella tesi partiranno dal presupposto di avere a disposizione una batteria di test "robusti" da un punto di vista psicometrico in quanto già analizzati, in fase di pretest e main study, su ampi campioni a livello nazionale.

Esistono diverse teorie e modelli che possono essere utilizzati per studiare un test psicometrico. La scelta degli indici e dei modelli statistici utilizzati nelle ricerche qui presentate risulta coerente con le analisi effettuate dall'INVALSI ed esplicitate all'interno del Rapporto Tecnico annuale. In particolare, per studiare la coerenza complessiva del test e alcune caratteristiche fondamentali degli item, verranno utilizzati alcuni strumenti afferenti alla Teoria Classica dei Test (CTT). Mentre il modello di Rasch che appartiene alla Teoria della Risposta all'Item (IRT) verrà impiegato per analizzare più approfonditamente le peculiarità degli item e applicare a test diversi procedure di ancoraggio.

2.1 Analisi dei test INVALSI

L'analisi delle caratteristiche psicometriche degli item che costituiscono una prova e del test nel suo complesso vengono svolte non solo al termine della rilevazione nazionale ma anche durante la creazione della prova stessa. La costruzione di un test INVALSI, infatti, prevede una prima fase di selezione dei quesiti e revisione qualitativa degli stessi in modo da formare uno o più fascicoli che saranno sottoposti a una fase di pretest. L'analisi del pretest ha lo scopo di identificare i quesiti che hanno un buon funzionamento da un punto di vista psicometrico, modificare i quesiti in cui si riscontrano lievi anomalie che possono essere risolte con una revisione della domanda ed eliminare le domande che presentano gravi anomalie da un punto di vista misuratorio.
Dalle domande così selezionate e analizzate, viene poi formato un fascicolo definitivo che sarà utilizzato per il main study. La fase di pretest permette così di costruire un fascicolo che deve risultare completo dal punto di vista dei contenuti disciplinari e dei traguardi che si vogliono andare a misurare e che, al contempo, mostri buone proprietà misuratorie.

In particolare nella costruzione di un test è necessario tenere conto di diversi aspetti relativi alla prova considerata nel suo complesso, tra cui:

1) La *unidimensionalità* del test: tutti gli item devono concorrere alla misurazione di un unico fattore o tratto latente (nelle prove INVALSI di matematica il tratto latente dovrebbe corrispondere alla "abilità matematica"). E' necessario quindi verificare che il costrutto latente che influenza le risposte ai quesiti (variabili osservate) sia sempre il medesimo all'interno del test e per farlo si può procedere attraverso una analisi fattoriale esplorativa[2] che consente di determinare il numero minimo di dimensioni latenti necessarie a descrivere la dipendenza statistica dei dati (Barbaranelli & Natali, 2005). Nell'ambito delle ricerche svolte in questa tesi non è stata necessaria una analisi di questo tipo in quanto sono stati utilizzati test INVALSI degli anni scolastici passati, la

[2] L'unidimensionalità del test può essere verificata in molteplici modi, in nel caso delle prove INVALSI si fa riferimento all'analisi fattoriale come uno dei metodi maggiormente utilizzati per indagare il numero minimo di dimensioni necessarie per descrivere il costrutto latente come specificato nel Rapporto Tecnico INVALSI (INVALSI, 2016b).

cui validità e robustezza era già stata ampiamente confermata dalle analisi statistiche compiute direttamente dall'Istituto nelle fasi di pretest e main study (INVALSI, 2016b).

2) La *validità* e *l'attendibilità* del test: sono due concetti diversi ma hanno come obiettivo comune quello di classificare le performances dei soggetti lungo una sola dimensione latente unidimensionale[3] (INVALSI, 2016b).
    - La *validità* del test è il grado in cui lo strumento misura effettivamente ciò che dovrebbe misurare. In particolare si parla di:
        - Validità di contenuto: gli item di un test devono coprire il più possibile l'argomento che il test vuole misurare. Ad esempio, un test di matematica non deve contenere solamente domande di geometria ma riguardare anche gli altri ambiti di contenuto.
        - Validità interna: tutti gli item che formano un test devono misurare lo stesso costrutto e quindi devono essere tra loro correlati.

    - L'*attendibilità* del test, ovvero la precisione con cui il test riesce a misurare il tratto latente.

Indicazioni importanti in questo senso si possono avere attraverso l'uso di molteplici e sofisticati strumenti statistici afferenti sia alla CTT sia alla IRT.
In questa tesi, potendo sempre contare sull'uso di test già validati anche da un punto di vista della validità e dell'attendibilità, si farà uso di un indice particolarmente importante per avere un feedback su questi fattori anche nel caso in cui, per le sperimentazioni di ricerca, i test fossero stati leggermente variati rispetto alla versione originale validata dall'INVALSI. L'indice in questione si chiama Alpha di Cronbach, è un metodo empirico di stima dell'attendibilità e si inquadra all'interno della CTT.

[3] A partire dall'ipotesi di unidimensionalità del costrutto latente operata dall'INVALSI in fase di analisi dei test, discendono le modalità di verifica empirica della validità degli indicatori e dell'attendibilità del test. Se fosse stato invece ipotizzato un costrutto multidimensionale, la procedura di verifica empirica della validità e dell'attendibilità sarebbe stata svolta diversamente.

In particolare l'α di Cronbach misura la coerenza interna degli item che compongono un test confrontando la somma delle varianze dei singoli item con la varianza del punteggio totale nel test.

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{j=1}^{n} \sigma_j^2}{\sigma_X^2}\right] \qquad (2.1)$$

Dove:

- n è il numero di item del test;
- σ_j^2 è la varianza dell'item j;
- σ_X^2 è la varianza totale del test.

Per interpretare il coefficiente di attendibilità si utilizzano i valori di riferimento riportati nella tabella sotto.

Valore α-Cronbach	**Interpretazione**
> 0,90	Ottimo
0,80 ≤ α < 0,90	Buono
0,70 ≤ α < 0,80	Discreto
0,60 ≤ α < 0,70	Sufficiente
α < 0,60	Inadeguato

Tabella 2.1: Valori di riferimento α-Cronbach.

Fonte: Rapporto Tecnico INVALSI 2016, adattamento da Barbaranelli e Natali (2005, p. 55).

L'Alpha di Cronbach non garantisce l'unidimensionalità di un test (che, come già specificato, deve essere indagata attraverso una analisi fattoriale[4]) ma fornisce a tal

[4] L'alpha di Cronbach è inadatta a rilevare la dimensionalità della prova perché è un indice che presuppone che le correlazioni tra gli indicatori siano riconducibili ad un unico tratto latente investigato. L'analisi fattoriale è uno dei possibili approcci che consentono di valutare la dimensionalità dei dati, ma non è l'unica strada che può essere perseguita. La scelta dell'uso di questo metodo per l'analisi dei dati INVALSI è ampiamente spiegata all'interno del Rapporto Tecnico (INVALSI, 2016), in particolare l'analisi dei risultati a

riguardo importanti indicazioni: nel caso in cui esistano item incoerenti con il resto della prova, si può infatti supporre che essi possano appartenere ad un'altra *dimensione* e che quindi l'unidimensionalità non sia verificata (INVALSI, 2016b). Ai fini di questa tesi, risulta anche necessario sottolineare che l'Alpha di Cronbach è un indice che risente di diversi fattori come ad esempio la numerosità e la rappresentatività del campione (Traub, 1994; Barbaranelli, 2005), la lunghezza del test e il limite di tempo per rispondere al test. Nel caso di ricerche in cui vengono analizzati i dati delle prove INVALSI somministrate a livello nazionale dall'Istituto questi vincoli sono già stati presi in considerazione e studiati nel processo di costruzione delle prove stesse, nelle ricerche che hanno previsto una nuova somministrazione delle prove, in parte anche modificate si è dovuto tenere conto di questi fattori e di come possano avere influito sui risultati.

Altri indici, sempre afferenti ai principali modelli psicometrici, portano invece a importanti informazioni sui singoli item e permettono in particolare di operarne la selezione e la modifica nella fase di analisi del pretest e costruzione del fascicolo per il main study.

In particolare la Teoria Classica dei Test fornisce i seguenti indici che vengono utilizzati per le analisi dei singoli item:

1) Indice di **difficoltà dell'item**: percentuale di risposte corrette rispetto al totale di rispondenti. In questo modo si ha una prima stima della difficoltà di ogni singolo item, che risulta però strettamente legata all'abilità dei rispondenti a cui il test è stato sottoposto. In generale, sono accettabili domande con percentuali di risposta corretta che vanno dal 10% al 90%, sono quindi da scartare domande eccessivamente facili ed eccessivamente difficili in quanto non sufficientemente informative.

2) Indice di **discriminatività dell'item**: è basato sulla relazione tra un singolo item e il punteggio totale del test in cui l'item è inserito e fornisce una informazione riguardo a quanto l'item stesso è in grado di distinguere tra rispondenti con alti livelli di abilità e

livello nazionale si basa sull'approccio UVA (Underlying Variable Approach, UVA; Moustaki, 2000) che consente di superare le principali limitazioni legate all'analisi fattoriale secondo il modello lineare (INVALSI, 2016).

rispondenti con bassi livelli di abilità. L'indice di discriminatività utilizzato nei test INVALSI viene anche detto indice di correlazione *punto-biseriale* e consiste nella correlazione tra il punteggio ottenuto da ogni studente a quell'item e il punteggio ottenuto sull'intera prova. In particolare:

$$r_{pbis} = \frac{M_j - M_T}{s_T}\sqrt{\frac{p_j}{1-p_j}} \qquad (2.2)$$

Dove:

- M_j indica la media dei punteggi ottenuti nell'intera prova dai soggetti che hanno risposto correttamente all'item j;
- M_T indica la media dei punteggi di tutti i rispondenti;
- S_T è la deviazione standard dal punteggio totale;
- p_j è la proporzione di risposte corrette date all'item j.

Un item con buone proprietà misuratorie deve essere sufficientemente discriminativo e presentare una discriminatività ≥ 0.25 (Barbaranelli & Natali, 2005).

In generale l'indice di correlazione punto-biseriale può essere utilizzato per studiare la correlazione tra ogni possibile opzione di risposta e il risultato complessivo ottenuto dagli studenti nel test. In questo caso l'indice dovrà assumere valori negativi per le opzioni di risposta sbagliate e un valore positivo sufficientemente alto per la risposta corretta.

Infine, sempre per quanto riguarda la Teoria Classica dei Test, si può considerare l'**abilità di uno studente** come la percentuale di risposte corrette fornite nel test (punteggio grezzo : numero domande totali). Sicuramente una indicazione di questo tipo può essere importante per operare confronti tra gli studenti che hanno risposto al test ma non bisogna trascurare un grande limite della CTT: l'abilità dello studente dipende sempre dal test a cui ha risposto e somministrando allo stesso studente un test differente che misuri lo stesso tratto latente si otterranno due abilità diverse. Questo limite viene superato nei modelli appartenenti alla IRT in cui il livello di abilità del rispondente risulta essere indipendente dal test che viene somministrato. Infatti grazie alle assunzioni teoriche dei modelli della IRT e attraverso procedure specifiche di ancoraggio, è possibile ordinare sullo stesso continuum di abilità anche studenti che hanno risposto a test differenti e lo stesso sarà possibile per quanto concerne la difficoltà degli item di due prove distinte.

Il modello di Rasch è uno dei modelli della IRT e permette di superare le principali limitazioni imposte dalla CTT. Questo modello è quello principalmente utilizzato dall'INVALSI per l'analisi dei test e sarà anche alla base di molte delle ricerche presentate in questa tesi.
Per questo motivo il prossimo paragrafo è interamente dedicato alla spiegazione del modello di Rasch e delle sue proprietà.

2.2 Modello di Rasch e Item Response Theory

Nonostante la Teoria Classica dei Test fornisca importanti strumenti statistici per lo studio dei test, le analisi che vengono presentate in questa tesi e quelle dei test INVALSI fanno uso anche della più moderna Teoria di Risposta all'Item. Quest'ultima comprende diversi modelli matematici per la misurazione di variabili latenti e permette di superare le principali limitazioni della CTT tra cui ad esempio la dipendenza tra la stima dell'abilità dei soggetti e la difficoltà delle domande del test.

In questo contesto prenderemo in considerazione il più semplice tra i modelli della IRT che viene anche utilizzato nelle principali analisi dei test INVALSI: il modello di Rasch (Rasch, 1960).

Il modello di Rasch è un modello logistico a un parametro ed è quindi il più semplice tra i modelli della IRT. Questo modello consente di calcolare la probabilità di rispondere correttamente a un determinato item, in funzione della abilità dello studente e delle caratteristiche psicometriche dell'item stesso (in particolare la difficoltà dell'item). Consideriamo quindi i dati di una prova con k item a cui hanno risposto N studenti; possiamo scrivere i dati all'interno di una matrice come segue:

ITEM

STUDENTI	**1**	**2**	**…**	**i**	**…**	**k**	
1	X_{11}	X_{12}	…	X_{1i}	…	X_{1k}	$\mathbf{R_1}$
2	X_{21}	X_{22}	…	X_{2i}	…	X_{2k}	$\mathbf{R_2}$
…	…	…	…	…	…	…	**…**
j	X_{j1}	X_{j2}	…	X_{ji}	…	X_{jk}	$\mathbf{R_j}$
…	…	…	…	…	…	…	**…**
N	X_{N1}	X_{N2}	…	X_{Ni}	…	X_{Nk}	$\mathbf{R_N}$
	$\mathbf{S_1}$	$\mathbf{S_2}$	**…**	$\mathbf{S_i}$		$\mathbf{S_k}$	

Tabella 2.2: Matrice di risposta a un test.

Dove:

- X_{ji} è una variabile aleatoria data dalla risposta dello studente j all'item i e assume valore 1 se la risposta è corretta, 0 se la risposta è sbagliata;
- $R_j = \sum_{i=1}^{k} X_{ji}$ è il punteggio grezzo del j-esimo studente, che corrisponde al numero di item a cui ha risposto correttamente;
- $S_i = \sum_{j=1}^{N} X_{ji}$ è il punteggio grezzo dell'item i-esimo, che corrisponde al numero di studenti che hanno risposto correttamente all'item.

Il modello di Rasch a partire da questa matrice opera una stima congiunta di due tipologie di parametri:

- a ogni studente j viene associato un parametro reale θ_j che rappresenta l'abilità del soggetto rispetto al tratto latente indagato dal test
- a ogni item i viene associato un parametro reale β_i che rappresenta la difficoltà dell'item.

La probabilità che lo studente j risponda correttamente all'item i dipende quindi dall'abilità del soggetto θ_j e dalla difficoltà dell'item β_i, in particolare:

$$P(X_{ji} = 1) = \frac{e^{\theta_j - \beta_i}}{1+e^{\theta_j - \beta_i}} \qquad (2.3)$$

$$P(X_{ji} = 0) = 1 - P(X_{ji} = 1) = \frac{1}{1+e^{\theta_j - \beta_i}} \qquad (2.4)$$

Le due forme possono essere rappresentate in una forma sintetica che le comprende entrambe:

$$P(X_{ji}) = \frac{e^{X_{ji}\,(\theta_j - \beta_i)}}{1+e^{\theta_j - \beta_i}} \qquad (2.5)$$

Infatti se la risposta è sbagliata, $X_{ji} = 0$ e si ottiene la (2.4) mentre se la risposta è esatta $X_{ji} = 1$ e si ottiene la (2.3).

Per ogni item i di difficoltà β_i , la funzione:

$$\varphi(\theta_j - \beta_i) = \frac{e^{\theta_j - \beta_i}}{1+e^{\theta_j - \beta_i}} \qquad (2.6)$$

permette di esprimere la probabilità di rispondere correttamente in funzione del livello di abilità degli studenti θ_j. La relazione tra abilità del soggetto e probabilità di rispondere correttamente all'item può essere rappresentata graficamente attraverso la curva teorica ipotizzata dal modello, detta anche *curva caratteristica dell'item* (Item Characteristic Curve – ICC).

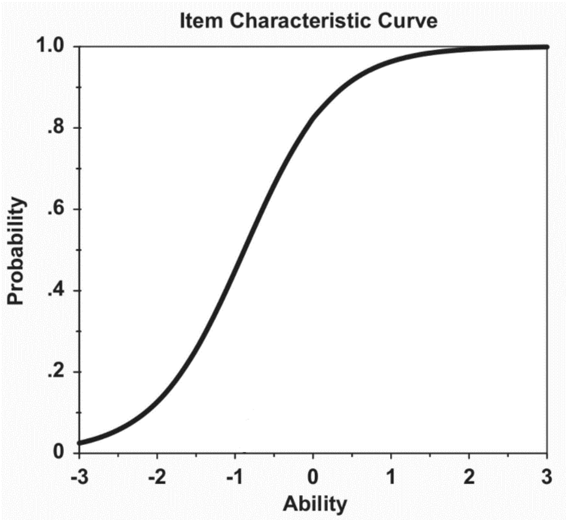


Figura 2.1: Curva Caratteristica di un item (Item Characteristic Curve -ICC)

Come si può vedere, si tratta di una funzione logistica monotona crescente.

Nel modello di Rasch la risposta di uno studente è governata dall'interazione di due componenti: l'abilità del rispondente e la difficoltà dell'item. Non si parlerà quindi più di uno studente "bravo" in assoluto o di un item "difficile" in assoluto.

Inoltre quando $\beta_i = \theta_j$ si ottiene:

$$P(X_{ji} = 1) = \frac{e^0}{1+e^0} = \frac{1}{2} = 0{,}5 \qquad (2.7)$$

Per questo motivo, un individuo con abilità θ_j ha una probabilità pari al 50% di rispondere correttamente a un item con livello di difficoltà $\beta_i = \theta_j$. In base a questa osservazione si definisce anche la difficoltà di un item nella IRT in termini probabilistici.

La **difficoltà di un item** nel modello di Rasch viene interpretata come quel punto nella scala di abilità in cui la probabilità di rispondere correttamente è pari al 50% (Barbaranelli & Natali, 2005). La difficoltà di un item viene anche definita *location* perché, da un punto di vista grafico, definisce lo spostamento della curva caratteristica sulla scala di abilità.

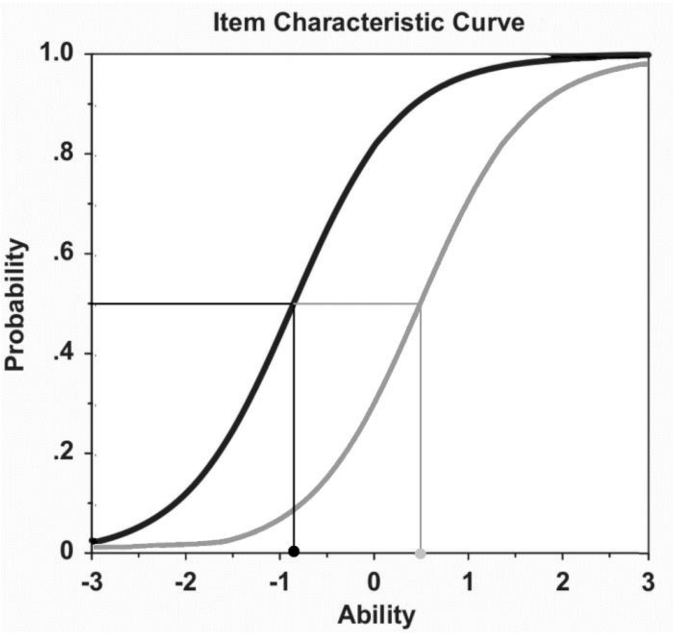


Figura 2.2: ICC di due item con diverso livello di difficoltà. I due contrassegni sull'asse delle ascisse indicano le difficoltà dei due item in quanto corrispondono al livello di abilità per il quale la probabilità di rispondere correttamente è pari al 50%.

La definizione della difficoltà di un item nella IRT è quindi molto diversa dall' interpretazione della difficoltà nella CTT (percentuale di soggetti nel campione che risponde correttamente) che risultava essere fortemente legata alle caratteristiche del campione a cui era stato somministrato il test.

2.3 Assunzioni del modello di Rasch

Il modello di Rasch può essere però applicato solo nel caso in cui siano verificate alcune assunzioni, che permettono l'applicazione del metodo e la stima dei parametri (Hambleton, Swaminathan & Rogers, 1991):

- *Unidimensionalità* del test: l'abilità misurata dai diversi item del test deve essere relativa ad un unico tratto latente. La dimensione dominante deve essere quindi unica e, come detto in precedenza, questa condizione può essere verificata attraverso una analisi fattoriale esplorativa (in accordo con quanto riportato nel Rapporto Tecnico INVALSI 2016).

- *Indipendenza Locale*: mantenendo costante il livello di abilità θ degli studenti, le risposte agli item sono eventi indipendenti.
  Quindi:
  $$P(X_1, X_2, \dots, X_k|\theta) = P(X_1|\theta)\ P(X_2|\theta) \dots P(X_k|\theta) \qquad (2.8)$$

  Indipendenza locale e unidimensionalità sono fattori strettamente legati e, avendo una conferma attraverso l'analisi dei dati empirici della unidimensionalità del test, allora anche l'ipotesi di indipendenza locale risulta essere verificata. Per avere un ulteriore riscontro però si possono svolgere analisi riguardanti le matrici di varianza-covarianza o la matrice di correlazione tra gli item.

- *Monotonicità*: per ogni item, la probabilità di rispondere correttamente deve crescere monotonicamente all'aumentare del livello di abilità dei rispondenti.
  La monotonicità può essere verificata attraverso la rappresentazione grafica dei dati empirici, in particolare si può rappresentare la percentuale di risposte corrette a ogni item in funzione del punteggio ottenuto sull'intera prova (punteggio di Rasch o punteggio grezzo).

Una volta verificate le assunzioni appena descritte il modello di Rasch permette la stima congiunta dei parametri di difficoltà β_i di ogni item e dei parametri di abilità θ_j di ogni studente. Per fare ciò si utilizza un processo iterativo basato sul metodo della *massima verosimiglianza* (Joint Maximum Likelihood – JML) il cui scopo è quello di determinare i valori dei parametri β_i e θ_j per i quali la matrice delle risposte empiriche ha la più alta probabilità di verificarsi.

2.4 Proprietà del modello di Rasch

In un test con k item, ogni studente può raggiungere un punteggio grezzo da 0 a k. A ciascun punteggio grezzo corrisponde un determinato valore di abilità dei soggetti θ (punteggio di Rasch degli studenti), questo accade perché il punteggio grezzo R_j è una statistica sufficiente per la stima dei parametri θ_j (proprietà di *sufficienza*). La trasformazione da punteggio grezzo a punteggio di Rasch è monotona ma non lineare, il legame tra le due misure è stretto ma i parametri di Rasch risultano essere delle misure migliori dell'abilità degli studenti. Infatti una delle proprietà fondamentali del modello è l'*invarianza:* le stime delle abilità dei soggetti risultano essere indipendenti, a meno di una costante, dalle caratteristiche degli item del test e, al contempo, le misure delle difficoltà degli item sono indipendenti dagli studenti che hanno risposto al test. Questa proprietà consente quindi di superare uno dei maggiori ostacoli della CTT in cui il livello di abilità stimato per ogni studente è legato al test che viene somministrato e viceversa.

Inoltre dalla proprietà di invarianza discende anche l'*oggettività specifica* (Rasch, 1960; Rasch, 1977) che rende possibile due tipologie di confronti che risultano essere invarianti:

- confronto tra soggetti, indipendentemente dagli item che formano il test;
- confronto tra item, indipendentemente dal campione al quale il test è stato somministrato.

L'invarianza della misurazione è verificata attraverso l'analisi del *fit* tra i dati empirici e il modello teorico. In particolare per questo tipo di analisi si può utilizzare un indice di fit detto *weighted MNSQ*[5] o indice di adattamento al modello. Il valore atteso per questo indice è 1 nel caso in cui la corrispondenza tra dati empirici e modello sia ottima ma sono accettabili valori compresi tra 0,80 e 1,20 (INVALSI, 2016b).

[5] L'indice *weighted MNSQ* è solo uno dei possibili indici di fit che può essere utilizzato tra quelli che afferiscono alla "classe" dei mean-square mean.

Infine il modello di Rasch crea una scala omogenea per item e soggetti (Barbaranelli & Natali, 2005) definendo i parametri θ_j e β_i sulla stessa scala (generalmente si considerano i valori tra -4 e +4). Questa proprietà consente di selezionare gli item in modo che all’interno di un test siano presenti item atti a misurare i diversi livelli di abilità dei soggetti.

Informazioni sul confronto tra abilità degli studenti e difficoltà degli item all’interno di un test possono essere ricavate direttamente anche da grafici come quello seguente, detti mappe di Wright.

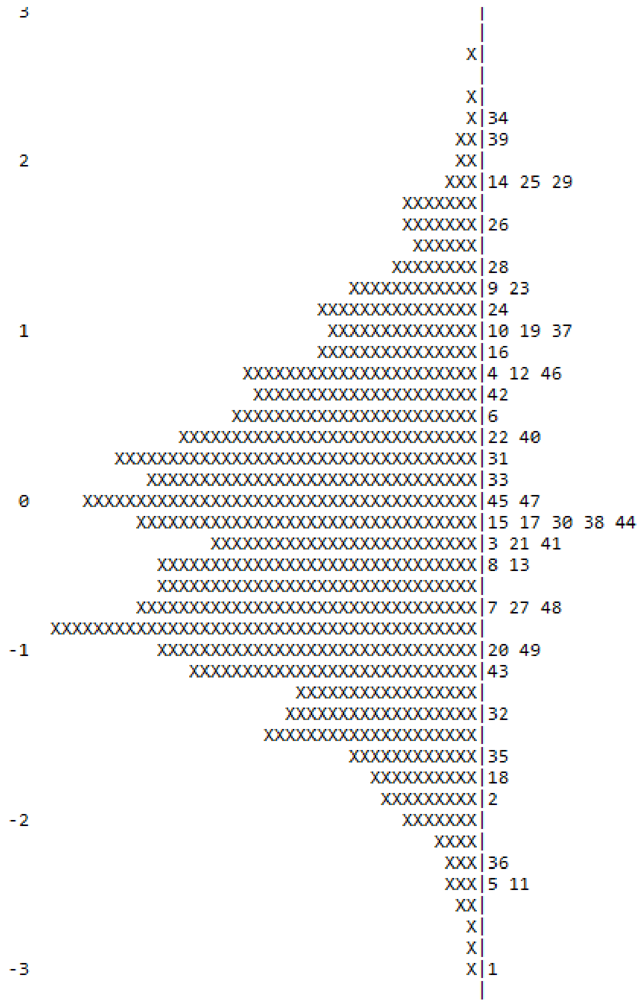


Figura 2.3: Mappa di Wright di una prova INVALSI di livello 6.

La mappa di Wright di una prova consente di osservare la distribuzione degli studenti (a sinistra della linea) in funzione del punteggio di abilità e la distribuzione degli item (sulla destra) in funzione del parametro di difficolta. Come già discusso precedentemente uno studente con un livello di abilità θ avrà una probabilità maggiore del 50% di rispondere correttamente a item con difficoltà $\beta<\theta$ e una probabilità minore del 50% di rispondere correttamente a item con difficoltà $\beta>\theta$.

Il modello di Rasch è il più semplice tra i modelli della IRT e la probabilità di rispondere correttamente dipende esclusivamente dall'abilità dei rispondenti e dalla difficoltà degli item, ovvero dai parametri θ_j e β_i. Modelli più complessi della IRT tengono in considerazione anche altre caratteristiche degli item come ad esempio la *discriminatività* (che nel modello di Rasch è considerata pari a 1 per tutti gli item) o il *guessing*. Pur analizzando il test attraverso il modello di Rasch, possiamo avere informazioni relativamente a queste caratteristiche utilizzando indici specifici che possono essere affiancati all'uso di questo modello, in particolare:

- Indice di discriminazione dell'item: tratto dalla CTT e già descritto nel paragrafo 2.1. Un item per essere sufficientemente discriminativo deve presentare un valore superiore a 0,25 e questo consente di affermare che l'item riesce a distinguere bene tra rispondenti con alti livelli di abilità e rispondenti con bassi livelli di abilità;

- Guessing: indice che misura l'incidenza del caso nella risposta all'item ovvero evidenzia il caso in cui studenti con livelli molto bassi hanno una probabilità maggiore di zero di rispondere correttamente fornendo risposte a caso (per approfondimenti: Lord & Novick, 2008; Linacre, 2004).

Un grafico esplicativo di tutte le caratteristiche degli item viste fin ora è il seguente.

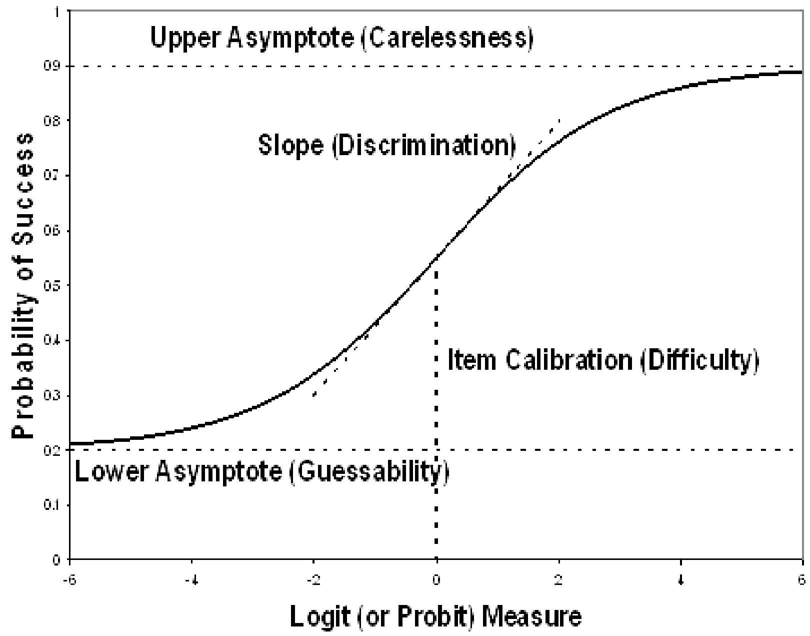


Figura 2.4: ICC di un item ottenuta con un modello IRT a 4 parametri (difficoltà, discriminazione, guessing e carelessness).

2.5 Distractor Plot

Una volta applicato il modello di Rasch, informazioni molto utili riguardanti i singoli item possono essere estrapolate osservando specifici grafici chiamati distractor plot. Nello stesso grafico in cui viene rappresentata la curva teorica ICC relativa alla risposta corretta (linea continua blu), vengono rappresentati anche i dati empirici relativi alla risposta corretta e alle altre opzioni di risposta. In questo modo è possibile osservare quanto la curva empirica della risposta corretta sia coerente con la curva teorica e, inoltre, si può analizzare l'andamento di ogni distrattore (inteso come risposta non corretta) in funzione del livello di abilità degli studenti.

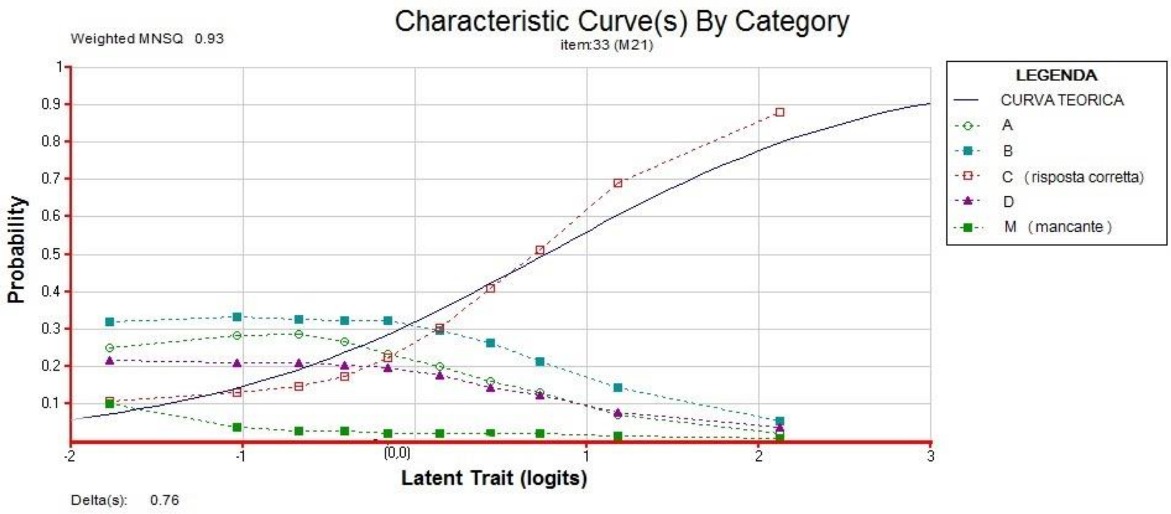


Figura 2.5: Esempio di *distractor plot* di un item.

Il grafico (fig.2.5), in ascissa, riporta il punteggio di Rasch in termini di abilità degli studenti θ sull'intera prova e la linea continua rappresenta la curva teorica del modello che esprime la probabilità di rispondere correttamente in funzione del livello di abilità. Le spezzate tratteggiate rappresentano i dati empirici ricavati dalla somministrazione dell'item: per tracciarle, gli studenti sono stati divisi in decili (corrispondenti ai simboli) in base al livello di abilità sull'intera prova e, per ogni decile, è stata riportata la percentuale di studenti che ha scelto ciascuna opzione di risposta.

In questo item, il confronto tra l'andamento della risposta corretta empirica e la curva teorica risulta essere accettabile (weighted = 0,93) anche se il modello tende a sovrastimare i livelli medio bassi e a sottostimare i livelli alti. Il distrattore più attrattivo risulta essere il B che viene scelto da una significativa percentuale di studenti anche nei livelli medi e medio-alti. Anche gli altri due distrattori comunque mostrano un buon funzionamento e vengono scelti da studenti con livelli di abilità bassi e medi. Infine si può notare che solo pochi studenti non rispondono all'item, quasi tutti appartenenti al decile più basso in termini di abilità.

2.6 Modello di Rasch e Test Equating

In ambito educativo i test standardizzati possono fornire anche importanti indicazioni riguardanti l'evoluzione delle abilità e delle competenze degli studenti nel corso degli anni e questo è solo uno dei possibili esempi di analisi che necessitano di procedure di *test equating*.

Volendo confrontare gli studenti di un certo grado scolare in diversi anni, non è possibile sottoporre agli studenti ogni anno il medesimo test. Bisogna quindi ricorrere a forme diverse di test che misurino lo stesso costrutto latente e che garantiscano inoltre la possibilità di confrontare i risultati di tutti gli studenti riportandoli sulla medesima scala di misura. L'applicazione del modello di Rasch, congiuntamente a sofisticate tecniche di test equating, consente di trovare una soluzione a questo problema e permette un confronto tra i punteggi ottenuti da differenti soggetti su test diversi che misurano lo stesso tratto latente.

Nonostante la proprietà dell'oggettività specifica del modello, grazie a cui è possibile confrontare l'abilità degli individui indipendentemente dagli item utilizzati per misurarla, nel caso di test differenti non è corretto confrontare direttamente i punteggi ottenuti ma è necessario applicare una procedura di equalizzazione dei punteggi (test equating). In particolare: "Se due test, X e Y, sono stati resi equivalenti e ad un soggetto viene somministrato il test X, è possibile determinare il punteggio equivalente sulla scala nella quale sono espressi i punteggi nel test Y" (Crocker & Algina, 1986).

Esistono diverse procedure di test equating, in questo contesto verrà illustrata una tecnica di ancoraggio di due o più prove che è stata utilizzata anche all'interno delle ricerche proposte in questa tesi. Seguendo questa procedura, ogni test viene somministrato a un gruppo diverso di studenti ma i due test devono essere costituiti in modo da avere una parte sufficientemente ampia di item in comune (test ancora o item di aggancio). Risulta quindi possibile utilizzare questa parte di test "comune" per riportare sulla medesima scala i punteggi di tutti gli studenti e i parametri di difficoltà di tutti gli item, potendo così operare un confronto diretto. Esistono due modi principali per stimare tali parametri, in questa tesi si adopererà la *concurrent calibration* che permette una stima congiunta di tutti i parametri ponendoli sulla medesima scala. Considerando i risultati dei due test contemporaneamente, questa procedura risulta essere più accurata rispetto a una calibrazione separata nel caso in cui il numero di item in comune ai test sia sufficientemente elevato e vi sia un buon adattamento dei dati al modello (Kolen & Brennan, 1995).

2.7 Differential Item Functioning

Un ulteriore strumento, sempre legato al modello di Rasch e utilizzato nelle ricerche esposte in questa tesi per analizzare le differenze nelle performance di sottogruppi di studenti, è il DIF (Differential Item Functioning).

Un item viene detto *unbiased* quando, a parità di livello di abilità, la probabilità di rispondere correttamente è la medesima per i soggetti appartenenti a diversi sottogruppi della popolazione (Osterlind, 1983). Un item invece viene detto *biased* nel caso in cui vi è una distorsione sistematica nella probabilità di rispondere correttamente, se il processo di misurazione dell'abilità avviene suddividendo la popolazione in sottogruppi in base a una variabile (Barbaranelli & Natali, 2005). Il concetto di item *biased* applicato nel contesto del modello di Rasch e della IRT rientra nel Funzionamento Differenziale dell'Item (Differential Item Functioning – DIF). In particolare un item ha un comportamento differenziale se, suddividendo la popolazione in funzione di una o più variabili (ad esempio il genere), cambia la probabilità di superare/non superare un certo item. Se un item presenta DIF significa quindi che la relazione con il tratto latente è differente all'interno dei due sottogruppi e quindi che la misurazione risulta essere "distorta" da altri fattori che incidono sulla unidimensionalità (Zumbo, 1999; Osterlind 1983; Cascella, 2014). In alcuni casi questo funzionamento differenziale non inficia gli assunti teorici e può essere trascurato, mentre in altri casi risulta più marcato ed è necessaria una analisi più approfondita dell'item. Il funzionamento differenziale può essere osservato tramite un test chi-quadrato per verificare l'uguaglianza o meno dei parametri relativi all'item all'interno dei due gruppi oppure attraverso l'analisi delle ICC prodotte separatamente per i due gruppi.

Nella prima immagine (fig. 2.6) ad esempio si può notare che l'item è unbiased e le due curve coincidono.

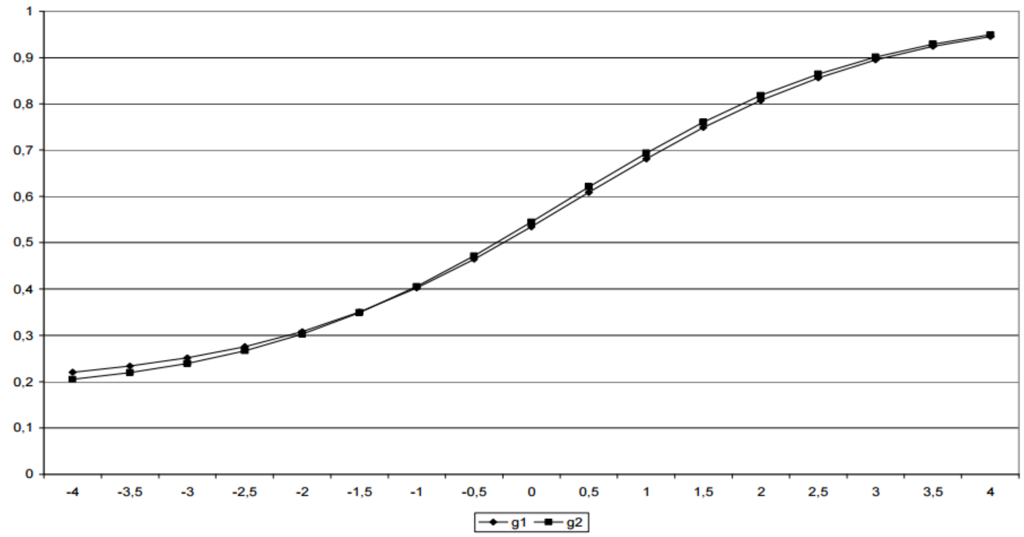


Figura 2.6: Esempio di item unbiased: le ICC dei due gruppi coincidono.
Adattamento da Barbaranelli & Natali (2005, p. 174).

Nel caso in cui uno dei due gruppi presenti una probabilità maggiore di rispondere correttamente rispetto all'altro gruppo per tutti i livelli di abilità, si parla di DIF uniforme e il grafico sotto ne è un esempio.

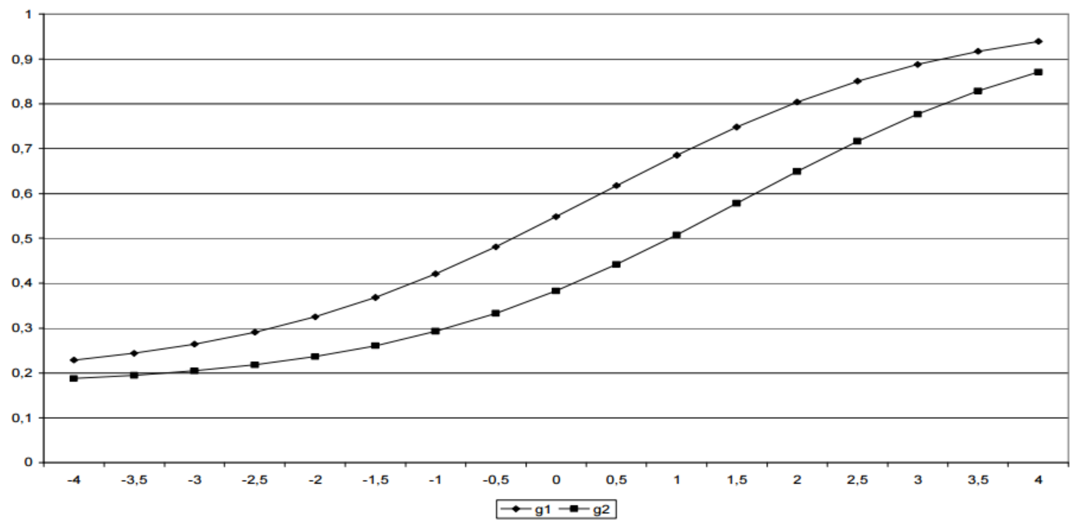


Figura 2.7: Esempio di item con DIF uniforme: le due ICC sono distinte e non si intersecano.
Adattamento da Barbaranelli & Natali (2005, p. 174).

Se invece l'influenza del DIF varia a seconda del livello di abilità e le due ICC si intersecano come nel grafico sotto, si parla di DIF non uniforme.

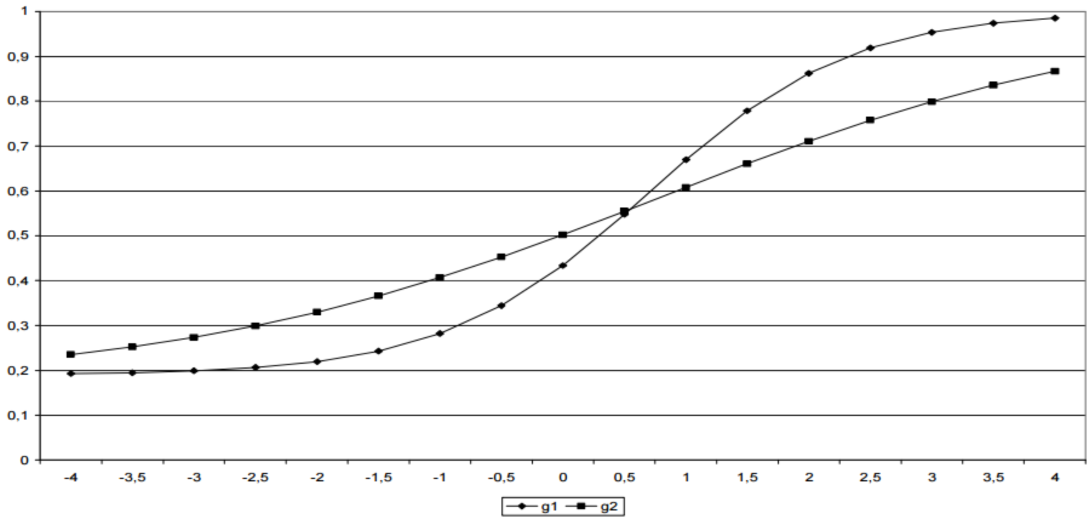


Figura 2.8: Esempio di item con DIF non uniforme: la differenza tra le due ICC varia a seconda del livello di abilità. Adattamento da Barbaranelli & Natali (2005, p. 174).

A prescindere dalle problematiche legate agli assunti teorici del modello, risulta molto interessante interpretare il funzionamento differenziale dell'item in particolari sottogruppi, ad esempio maschi e femmine. Nei principali test standardizzati nazionali (INVALSI) e internazionali (PISA, TIMSS) le differenze legate al genere nelle prove di lettura e in particolar modo nelle prove di matematica sono sempre presenti. In questa tesi si farà quindi uso del DIF per individuare gli item che mostrano un comportamento differenziale per maschi e femmine e capire quali possano essere i fattori che determinano queste differenze. Andando nello specifico delle ricerche presentate in questa tesi, il DIF può diventare quindi un importante strumento di analisi delle differenze per sottopopolazioni nell'apprendimento della matematica e questo risultato può essere raggiunto reinterpretando attraverso le lenti della didattica della matematica il funzionamento differenziale dell'item da un punto di vista psicometrico.

Capitolo 3

Uso delle prove standardizzate per la ricerca in didattica della matematica

L'introduzione, sia a livello internazionale sia a livello nazionale, di prove standardizzate come quelle PISA e INVALSI può fornire importanti informazioni per ricerca in didattica della matematica. Anche se l'obiettivo principale di queste indagini risulta essere la valutazione del sistema educativo, delle abilità e delle competenze raggiunte dagli studenti nei diversi livelli scolastici, non sono da sottovalutare le potenzialità che queste prove possono avere anche se utilizzate nel campo della ricerca. I risultati dei test possono, infatti, far emergere macro fenomeni molto interessanti, come ad esempio nuovi effetti di contratto didattico (Ferretti, 2015); questi potranno poi essere studiati approfonditamente attraverso un approccio *mixed-method*, passando quindi da una indagine quantitativa a una qualitativa (Johnson & Onwuegbuzie, 2004). Non sono solo i risultati di tali indagini però a destare interesse, oggetto di studio e di confronto sono anche i framework di riferimento delle diverse prove, si pensi ad esempio alla definizione di *Mathematical Literacy* definita nel framework PISA e alla concezione di modellizzazione matematica nota come *ciclo della matematizzazione* (OECD, 2015b). Infine anche gli stessi quesiti proposti all'interno di prove standardizzate possono costituire stimolo per una ricerca su un particolare argomento che interessa la didattica della matematica.

Nonostante le grandi potenzialità offerte da queste indagini su larga scala, l'uso che ne viene fatto all'interno della ricerca in didattica della matematica risulta ancora limitato, soprattutto per quel che riguarda le prove INVALSI a livello nazionale. L'obiettivo dell'articolo riportato di seguito (Maffia & Giberti, 2015) è di analizzare quale impatto abbia avuto l'introduzione di prove standardizzate sulle ricerche in questo campo e come esse siano state presentate nelle maggiori riviste di didattica della matematica.

Si è deciso di analizzare gli articoli selezionati dalle maggiori riviste internazionali in base al modo d'uso delle prove PISA (analisi del framework, uso dei dati, …) e in funzione dell'ambito in cui si inserisce la ricerca (differenze di genere, formazione insegnanti, confronto tra nazioni, …). In questo modo è stato possibile evidenziare come le prove PISA siano utilizzate a livello internazionale e quali settori specifici della disciplina ne facciano maggiormente uso. Dalla stessa indagine, compiuta anche su riviste nazionali, è emerso che l'uso delle prove INVALSI per la ricerca in didattica risulta ancora estremamente limitato (nonostante le potenzialità mostrate dalle ricerche internazionali sulle prove PISA).

Sarebbe quindi interessante anche mostrare come, in campo pedagogico e psicologico, le indagini internazionali hanno avuto una risonanza molto maggiore. Se si considera anche solo la rivista "Educational Studies" che si occupa di educazione in generale, attraverso una ricerca con le parole chiave "OECD" e "PISA" sono migliaia gli articoli individuati che fanno uso dell'indagine OECD PISA in campo educativo.

3.1 Ricerca in Didattica della matematica e PISA: percorsi battuti e nuove piste da esplorare[6]

Maffia Andrea, Università degli Studi di Modena e Reggio Emilia

Giberti Chiara, Università degli Studi di Trento

L'uso dell'indagine OCSE-PISA è estremamente limitato nella ricerca in Didattica della matematica, nonostante il forte interesse che i risultati di questi test hanno suscitato nell'opinione pubblica nei confronti dell'educazione matematica (Sfard, 2005).
Sono stati analizzati 107 articoli che fanno uso del PISA tratti dalle principali riviste internazionali di Didattica della matematica, selezionati all'interno di quelle riviste di classe A (secondo la classificazione ANVUR), il cui sito è dotato di uno strumento di ricerca per parole chiave. Gli articoli individuati sono stati classificati in base al tema di ricerca, determinando così otto principali categorie. Si osserva che la maggior parte degli articoli presenta ricerche relative alle specificità dell'insegnamento/apprendimento della matematica all'interno di uno o più Paesi (53% del totale). Fra questi si trova l'unico articolo italiano presente all'interno del campione analizzato (Boero & Dapueto, 2007).
Una seconda classificazione è stata realizzata a partire dal modo in cui le prove PISA sono utilizzate nelle diverse ricerche. In particolare sono state individuate quattro categorie: uso del quadro di rifermento, uso dei quesiti, uso dei risultati e analisi degli effetti delle prove.
Si è quindi osservato che nella maggior parte degli articoli viene fatto uso dei risultati (64%), tuttavia anche l'impiego del quadro di riferimento (26%) e dei quesiti (11%) è rilevante.
Incrociando le due classificazioni presentate, si è notato come alcune combinazioni siano piuttosto frequenti (per es. uso dei risultati per studiare le differenze di genere) mentre altre risultino assenti. Ciò nonostante alcune di queste possibilità rappresentano interessanti opportunità per future ricerche, si esplorano quindi eventuali nuovi impieghi degli ultimi risultati delle prove PISA e si ipotizzano alcuni usi analoghi, a livello italiano, per le prove INVALSI.

[6] Articolo presentato al convegno tenutosi nei giorni 26-27 Febbraio 2015 e organizzato dall'INVALSI "PISA 2012: contributi di approfondimento". L'articolo è stato successivamente pubblicato nel volume "OCSE Pisa 2012. Contributi di approfondimento" a cura di Laura Palmiero, Franco Angeli Edizioni (2016).

1 Introduzione

Sin dalla sua nascita, l'indagine OCSE-PISA è stata al centro dell'attenzione della comunità dei ricercatori in ambito didattico ed educativo. Si trovano riferimenti al test dell'Organizzazione per la Cooperazione e lo Sviluppo Economico in lavori dei primissimi anni del ventunesimo secolo (per es. Kramarski et al., 2002), compreso il documento di discussione del quattordicesimo studio della Commissione Internazionale sull'Istruzione Matematica (ICMI) in cui si rileva come

> a seguito della pubblicazione, nel 2001, dei risultati del primo ciclo PISA (del 2000), è iniziata un'intensa discussione, in diversi Paesi, sugli obiettivi e la progettazione dell'istruzione matematica nelle scuole, in particolare sul ruolo della modellizzazione matematica, le applicazioni e le relazioni col mondo reale. (Blum, 2002)

Tuttavia l'attenzione degli studiosi di Didattica della matematica si è concentrata su queste prove a partire dal 2003. In quell'anno il focus principale della prova era proprio la matematica e, inoltre, molti degli stati partecipanti furono coinvolti anche nelle prove organizzate dall'Associazione Internazionale per la Valutazione dei Risultati dell'Educazione (IEA) per la valutazione delle Tendenze Internazionali nello Studio della Matematica e delle Scienze (TIMSS). Negli anni successivi, i risultati delle prove PISA 2003 (realizzate al livello 10) insieme a quelli delle prove TIMSS 2003 (livello 4 e 8) permisero di avere una fotografia degli apprendimenti in matematica, nei diversi livelli scolastici, per quattordici stati distribuiti in tutti i continenti[7]. Un esempio di lavoro centrato proprio su questo tipo di analisi può essere quello di Ferrini-Mundy e Schmidt (2005).
L'insieme di queste diverse prove suscitò un crescente interesse da parte dell'opinione pubblica sui risultati degli apprendimenti in matematica, tuttavia l'impiego dei risultati delle rilevazioni nella ricerca in Didattica della matematica rimase scarso. Si trova testimonianza di questo fenomeno nel lavoro di Sfard (2005) che, intervistando 74 ricercatori provenienti da tutto il mondo, nota che

> Ed Silver, fino a poco tempo fa editore del *Journal for Research in Mathematics Education*, si meraviglia, in

[7] I Paesi che parteciparono sia al test PISA che a entrambe le prove TIMSS nel 2003 furono (in ordine alfabetico) Australia, Belgio (area fiamminga), Giappone, Gran Bretagna (Scozia), Hong Kong, Italia, Lettonia, Norvegia, Nuova Zelanda, Olanda, Russia, Stati Uniti, Tunisia, Ungheria.

uno dei suo editoriali, che "in questo periodo sembra che i didattici della matematica siano un gruppo di individui competenti nella ricerca quantitativa inclini a condurre studi qualitativi." Con l'aiuto di una metafora abilmente scelta implica che, per alcuni autori, "qualitativo" non significa nient'altro che "senza numeri".
Questa preferenza dei nostri intervistati verso il qualitativo è controbilanciata dall'aumento di popolari studi comparativi internazionali, come TIMSS e PISA, che si concentrano soprattutto sulle prestazioni misurabili degli studenti. Solo uno, fra i ricercatori nel nostro campione, sembra essere coinvolto in uno di questi progetti su larga scala.
In ogni caso, è abbastanza notevole il fatto che, nei nostri dati, non troviamo più di tre riferimenti a questi studi. I nostri intervistati non si aiutano con i risultati di TIMSS o PISA anche quando rispondono alle nostre domande sullo stato della didattica della matematica nei loro Paesi.

Nonostante ci siano state variazioni dal 2003 a oggi (si veda la prossima sezione), i dati OCSE-PISA sembrano avere ancora uno scarso impiego nella ricerca nazionale e internazionale in Didattica della matematica. Tuttavia i lavori pubblicati fino a oggi mostrano, attraverso i molteplici temi affrontati e le diverse modalità d'uso dei test, la notevole ricchezza che queste prove possono offrire. Tale ricchezza sembra essere stata recentemente ribadita dalla ICMI nel conferimento della Freudenthal Medal a Frederick Koon Shing Leung, dell'Università di Hong Kong, motivato anche facendo riferimento a come

il suo lavoro innovativo, per il quale è noto a livello internazionale, è l'uso della prospettiva della cultura confuciana per spiegare i risultati superiori in matematica degli studenti dell'est asiatico negli studi internazionali quali il [TIMSS] dello IEA e il [PISA] dell'OCSE.[8]

Scopo del presente lavoro è l'esplorazione delle varietà di impiego delle indagine OCSE-PISA nella ricerca in Didattica della matematica, anche al fine di evidenziare i modi in cui sono state maggiormente usate, e suggerire possibili nuovi impieghi delle prove stesse. Il modo in cui i test standardizzati internazionali hanno influenzato la ricerca sarà usato come spunto per indicare possibili lavori realizzabili anche con le prove nazionali (SNV-INVALSI) che si sono affermate negli ultimi anni.

[8] Il testo è tratto dal sito ufficiale di ICMI, http://www.mathunion.org/icmi/activities/awards/the-hans-freudenthal-medal-for-2013/

2 L'uso dell'indagine OCSE-PISA per la ricerca in Didattica della Matematica

Al fine di comprendere come viene utilizzata l'indagine OCSE PISA si è scelto di iniziare dall'analisi degli articoli di ricercatori in Didattica della matematica che basano i loro studi proprio su queste prove standardizzate.

Le riviste considerate nella selezione degli articoli sono state individuate tra quelle di classe A (secondo la classificazione ANVUR) che trattano di Didattica della matematica e il cui sito è dotato di uno strumento di ricerca per parole chiave.

Una prima ricerca è stata effettuata inserendo nel motore di ricerca di ogni rivista le parole chiave: "PISA" o "OECD". Successivamente, tra gli articoli così individuati, sono stati selezionati quelli che effettivamente facevano uso delle prove OCSE. L'indagine ha quindi portato a individuare 107 lavori, pubblicati entro il mese di febbraio 2014.

Come si osserva nell'areogramma sottostante (fig.1) la maggior parte degli articoli proviene dalla rivista tedesca *Zentralblatt für Didaktik der Mathematik* (ZDM), in Germania in effetti le prove standardizzate come TIMSS e PISA vengono studiate approfonditamente anche al fine di confrontarle con altre valutazioni presenti a livello nazionale (Lorenz, 2005).

Le altre riviste da cui sono stati tratti gli articoli sono *Educational Studies in Mathematics* (ESM) che copre il 15% dei lavori e, in parte minore, *The Journal of Mathematical Behaviour* (TJMB) e *Journal of Mathematics Teacher Education* (JMTE).

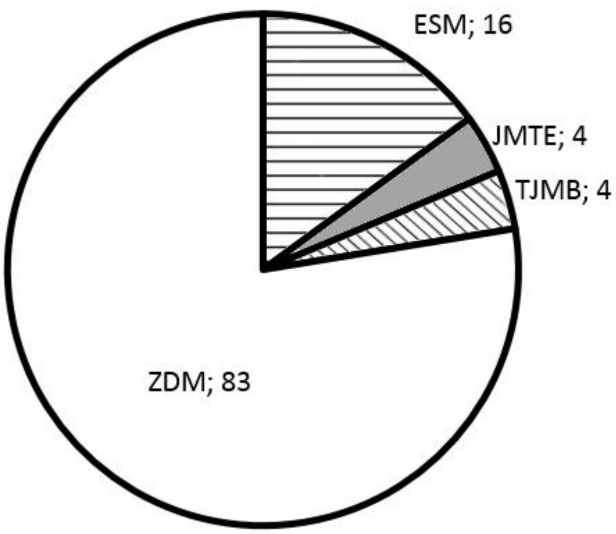


Fig.1: Distribuzione degli articoli selezionati nelle riviste.

La distribuzione degli articoli negli anni di pubblicazione, che si osserva nel grafico seguente (fig.2), mostra un crescente interesse nei confronti delle prove OCSE PISA per quel che riguarda la ricerca in Didattica della matematica.

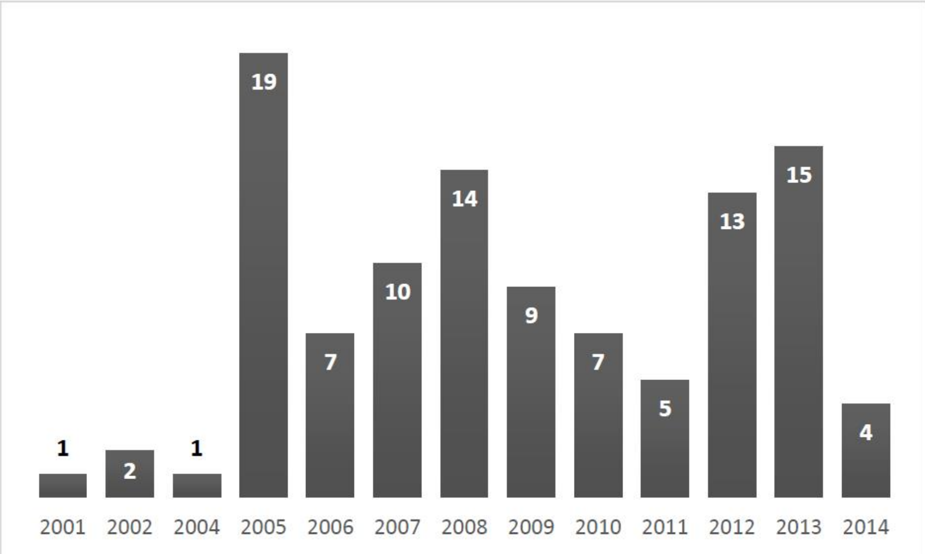


Fig.2: Distribuzione degli articoli selezionati negli anni.

Tenendo presente che gli anni delle rilevazioni sono stati 2000, 2003, 2006, 2009 e 2012, si nota che, due anni dopo ogni indagine OCSE, le pubblicazioni sono state più numerose. L'elevato numero di articoli pubblicati nel 2005 e nel 2008 tratta, rispettivamente, dei risultati delle rilevazioni del 2003 e del 2006; i due anni di differenza rappresentano il tempo necessario per la restituzione dei risultati e la pubblicazione dell'articolo.
Come si era osservato anche nel paragrafo precedente, alcuni articoli compaiono già dopo la prima rilevazione OCSE (2000) ma è da quella del 2003, anno che vede in concomitanza prove TIMSS e PISA, che l'attenzione dei ricercatori in Didattica della matematica risulta essere in netto aumento.
Negli ultimi anni il numero di articoli si è mantenuto a livelli piuttosto alti[9] infatti le pubblicazioni più recenti non si concentrano solamente sull'analisi dei dati delle prove precedenti, ma si occupano anche di studiare il quadro di riferimento PISA e utilizzano i quesiti delle prove stesse disponibili dopo la somministrazione.

[9] Si noti che i 4 articoli pubblicati nel 2014 non confutano questa osservazione in quanto di tale anno sono stati considerati solo i mesi di Gennaio e Febbraio, ci si aspetta quindi che entro la fine dell'anno il numero di articoli pubblicati risulti essere rilevante.

Gli articoli selezionati secondo i criteri illustrati in questo paragrafo vengono quindi analizzati, nelle sezioni seguenti, operando una classificazione in base all'argomento trattato e al modo di utilizzare le rilevazioni OCSE PISA.

2.1 Classificazione in base al tema

Una classificazione di articoli riguardanti l'indagine PISA è stata proposta da Owens (2013). Nel suo lavoro vengono analizzati 74 articoli in inglese tratti da riviste peer-reviewed, che si occupano di esaminare, attraverso i dati OCSE-PISA, le politiche educative dell'ultimo decennio (2000-2010).

Tema	Numero di articoli
Livello Studente	**50**
Genere	5
Immigrazione	8
Risultati d'apprendimento	6
Motivazione e Affettività	6
Panoramica nazionale	4
Background socio-economico	15
Tecnologia	6
Livello Scuola	**11**
Autonomia e amministrazione	5
Insegnante e Classe	6
Livello Sistema	**13**
Responsabilità	1
Efficienza	1
Ritenzione	2
Scelta scolastica	3
Età scolare	2
Tracciamento	4
Totale	**74**

Tab.1: Classificazione per temi proposta in Owens (2013).

Come si può vedere nella tabella 1, in questo caso, l'autrice ha individuato una categorizzazione in base agli argomenti trattati negli articoli, determinando 15 aree tematiche principali.

A partire dalla classificazione di Owens si è cercato di analizzare i 107 articoli che fanno uso delle prove PISA nelle riviste specializzate in Didattica della matematica, suddividendoli per argomento trattato.

Tra i temi individuati dalla Owens, diversi corrispondono a quelli trattati anche negli articoli qui analizzati, nonostante le pubblicazioni selezionate da questa autrice non trattino necessariamente di Didattica della matematica e le riviste da cui sono state prese non siano le stesse considerate in questo articolo.

Suddividendo i 107 articoli seguendo questa classificazione, si ha che buona parte rientra nella macro-categoria *livello studente*, che risulta essere la più consistente anche nel lavoro della Owens.

La categoria *panoramica nazionale* è stata suddivisa in due aree tematiche: *didattica della matematica in una nazione* e *confronto tra nazioni*. Relativamente alle categorie individuate a *livello di scuola* e a *livello di sistema* si nota che, nel nostro campione di articoli, sono presenti solo specifici temi (per es. si parla di insegnanti solo relativamente alla loro formazione), pertanto si è deciso di utilizzare queste particolari tematiche nella nostra classificazione. Sono state così determinate otto principali categorie (tab.2) entro le quali si riesce a classificare più dell'80% degli articoli individuati.

Si osserva che buona parte dei lavori (circa il 37%) si occupa della *didattica della matematica in una nazione*; leggendo questi articoli si nota come, molto spesso, i dati OCSE-PISA vengano utilizzati non solo per analizzare l'apprendimento della matematica nei suoi diversi aspetti, ma fungano soprattutto da stimolo per avviare nuovi progetti e ricerche. In Germania, per esempio, sembra che i risultati delle ultime prove standardizzate, integrate con altre prove presenti a livello nazionale, abbiano portato anche a notevoli cambiamenti nel curricolo di matematica (Lorenz, 2005).

L'uso dei test OCSE-PISA viene spesso legato alle capacità degli studenti di un determinato Paese nelle attività di problem-solving. Fra questi si trova anche l'unico articolo italiano presente all'interno del campione analizzato (Boero & Dapueto, 2007) nel quale, anche attraverso il confronto con altri Paesi, si evidenzia la difficoltà degli studenti italiani nella risoluzione dei problemi, si analizza il ruolo della ricerca in questo campo della didattica della matematica negli ultimi anni e come la ricerca ha influito sui programmi nazionali e sui curricoli.

Molti articoli appartenenti a questa prima categoria sono tratti dalla rivista ZDM e quindi si interessano di Didattica della matematica in Germania. L'uso delle prove in queste

pubblicazioni permette diversi approfondimenti tra i quali, per esempio, il confronto tra i risultati nelle diverse zone geografiche della Germania, cioè i sedici stati federali, oppure tra i risultati nei diversi indirizzi scolastici tedeschi.
Quest'ultimo confronto viene studiato da Wynands e Möller (2005) che si interessano alle differenze sui risultati in matematica di diversi gruppi di studenti provenienti dai differenti sistemi scolastici in Germania e per fare ciò fanno ampio uso dei risultati PISA:

> Diamo uno sguardo ai risultati in matematica degli studenti con alte prestazioni nella Hauptschule […]. Inoltre, confrontiamo questo gruppo con studenti provenienti da altri indirizzi in Germania (Gesamtschule, Realschule e Gymnasiums). Il nostro interesse è quello di trovare le differenze e le caratteristiche dei diversi gruppi. I risultati del test nazionale di PISA 2000 sono la base empirica della nostra analisi.

I dati PISA però non offrono solo la possibilità di confronti all'interno di una stessa nazione ma vengono spesso utilizzati anche per studiare le differenze nell'apprendimento della matematica fra diversi Paesi. Questa tipologia di articoli costituisce la seconda categoria: *confronto tra nazioni*.
Spesso gli articoli partono dalla constatazione di forti differenze nei risultati delle prove OCSE-PISA in due o più nazioni, per poi andarne a individuare le cause nelle caratteristiche dei sistemi educativi.
Numerosi studi comparano Paesi occidentali con Paesi orientali che, come è noto, ottengono risultati notevoli nelle prove standardizzate di matematica. Un esempio paradigmatico di questa tipologia di studi può essere individuato nell'articolo di Young e Leung (2011) nel quale viene mostrato come il contesto sociale e culturale influenzi notevolmente la didattica della matematica e, per questo motivo, l'approccio a questa disciplina risulta essere radicalmente differente se si confrontano Paesi occidentali e orientali.
Queste prime due categorie sono particolarmente studiate; si osserva, infatti, che gli articoli che presentano ricerche relative alle specificità dell'insegnamento/apprendimento della matematica all'interno di uno o più Paesi, risultano essere più del 50% del totale degli articoli.
La terza categoria riguarda la *formazione degli insegnanti* e comprende il 10% degli articoli.
Risulta, infatti, che le prove OCSE-PISA vengano utilizzate sia per valutare le difficoltà dei futuri insegnanti di matematica (si veda, nella sezione successiva, il riferimento a Sáenz, 2009) sia per sperimentare nuovi metodi di formazione che prevedono l'introduzione di nuove tecnologie (Goos & Geiger, 2012).

Inoltre, come sottolineano Büchter e Leuders (2005), l'uso di quesiti simili a quelli proposti dall'OCSE-PISA può essere uno stimolo per aiutare gli insegnanti a riflettere e lavorare sull'uso dei problemi in classe. Proporre agli studenti "problemi adeguati agli specifici processi di apprendimento" (*ibidem*) risulta essere una attività sempre più importante per gli insegnanti e la formazione, in questo senso, può essere d'aiuto proprio partendo da un lavoro di selezione e modifica dei quesiti OCSE-PISA.

In molti casi i test OCSE vengono affiancati ad altri test somministrati a livello locale: buona parte degli articoli (9%) si occupa proprio di confrontare le prove e i risultati dell'indagine OCSE-PISA con i test analoghi somministrati però a livello nazionale.

Particolare attenzione al *confronto con altre prove standardizzate* si può riscontrare in articoli riguardanti la Germania, dove è stato progettato un test comune a sette stati federali che valuti le competenze in matematica (Lorenz, 2005).

Sempre a livello nazionale i test OCSE-PISA e in particolare il quadro di riferimento, concorrono allo *sviluppo dei curricoli* di matematica e circa il 9% degli articoli analizzati trattano questo tema. In Svizzera, nell'ambito di un progetto denominato "Armonizzazione della scuola dell'obbligo", viene analizzato il concetto di matematica che è alla base dello studio PISA e più precisamente definito nel quadro di riferimento (Linneweber-Lammerskitten & Wälti, 2005).

La sesta categoria individuata comprende circa il 6% degli articoli e riguarda le ricerche sulle differenze di *genere*. L'interesse per questo argomento, che viene trattato sia a livello generale sia focalizzando l'attenzione su una singola nazione, risulta essere in aumento negli ultimi anni e questo anche a causa del ruolo crescente delle rilevazioni OCSE-PISA nelle ricerche in Didattica della matematica (Leder & Forgasz, 2008).

Un esempio significativo di uso delle prove per analizzare i differenti risultati in matematica di studentesse e studenti può essere riscontrato nel lavoro di Steinthorsdottir e Sriraman (2008); analizzando i risultati ottenuti in Islanda e confrontandoli con quelli di altri stati si osserva infatti che

> i risultati in matematica degli studenti in Islanda, come riportato in PISA 2003 hanno mostrato una significativa e (per confronto) insolita differenza in base al genere in matematica: l'Islanda è l'unico paese in cui il divario di genere in matematica favorisce le ragazze.

Le ultime due categorie, *modellizzazione* e *motivazione e affettività*, includono rispettivamente il 2% e il 3% degli articoli analizzati.
Alcune ricerche che studiano il ruolo dei fattori affettivi e metacognitivi nel processo di insegnamento/apprendimento della matematica, hanno fatto uso dei risultati delle prove OCSE-PISA proprio per evidenziare quanto le prestazioni degli studenti in matematica dipendano fortemente dal contesto (Roesken, Pepin & Toerner, 2011).

Tema	Numero di articoli
Didattica della matematica in una nazione	40
Confronto tra nazioni	17
Formazione degli insegnanti	11
Confronto con altre prove standardizzate	10
Sviluppo dei curricoli	10
Genere	7
Motivazione e affettività	3
Modellizzazione	2
Altro	18

Tab.2: Classificazione per temi (alcuni articoli compaiono due volte perché affrontano due temi).

Dalla lettura di questi articoli si è notato che l'importanza delle prove OCSE in Didattica della matematica non risiede solamente nell'analisi dei risultati: molti articoli infatti si concentrano sul quadro di riferimento PISA e altri studiano gli effetti delle prove sul contesto sociale, politico e culturale.
Il seguente paragrafo si concentra sulla classificazione dei medesimi articoli, non più in base al tema trattato ma in funzione del modo in cui vengono utilizzate le rilevazioni OCSE-PISA.

2.2 Classificazione in base all'uso

Nella sezione precedente si è visto come i dati che emergono dalle prove possano essere utilizzati per affrontare diversi temi di ricerca e, in Didattica della matematica, i più frequentemente scelti sono quelli centrati sullo studente (secondo la classificazione di Owens, 2013).
Una lettura più approfondita degli articoli selezionati ha permesso di mettere in evidenza un secondo fattore: molte delle pubblicazioni che citano le prove OCSE-PISA non usano in

nessun modo i risultati delle prove né i dati del questionario studente.
Si è quindi deciso di realizzare una seconda classificazione a partire dal modo in cui le prove sono utilizzate nelle diverse ricerche.
Non avendo riscontrato letteratura in merito a questa tipologia di classificazione, si è deciso di individuare delle categorie a partire direttamente dai dati a disposizione. In una prima analisi si è osservato che, oltre ai dati restituiti dopo la somministrazione delle prove, vengono usati tutti i documenti forniti dall'OCSE sulle prove PISA. Si nota quindi l'impiego sia dei documenti forniti prima della somministrazione, come il quadro di riferimento a partire dal quale le prove vengono realizzate, sia di quelli pubblicati successivamente, come i testi delle prove stesse. Si è deciso quindi di realizzare una catalogazione in base alla tipologia di documento utilizzato.
Seguendo l'ordine cronologico con cui i documenti sono presentati, il primo raggruppamento individuato è quello dei lavori che fanno *uso del quadro di riferimento*. In questi articoli l'oggetto di ricerca è proprio il quadro di riferimento delle prove, al fine di rivederne alcuni aspetti oppure di esplorarne le possibilità di utilizzo in altri contesti. Un esempio paradigmatico è dato dal lavoro di Linneweber-Lammerskitten e Wälti (2005) che analizzano il modo in cui la matematica è concepita nelle prove OCSE per studiare l'applicabilità di tale concezione in un progetto di valutazione a livello nazionale (come già evidenziato nella precedente sezione). Uno dei costrutti maggiormente citati e discussi è quello di *competenza matematica*[10] che viene spesso messo a confronto con definizioni analoghe in altri test standardizzati sia a livello nazionale che internazionale.

[10] Nell'articolo citato viene fatto uso della definizione data per le prove del 2003 in cui la competenza matematica è intesa come "la capacità di un individuo di identificare e comprendere il ruolo che la matematica gioca nel mondo reale, di operare valutazioni fondate e di utilizzare la matematica e confrontarsi con essa in modi che rispondono alle esigenze della vita di quell'individuo in quanto cittadino che esercita un ruolo costruttivo, impegnato e basato sulla riflessione." (OCSE, 2003). Nei lavori più recenti ci si riferisce invece alla definizione data per le prove del 2012 in cui la competenza matematica è definita come "la capacità di un individuo utilizzare e interpretare la matematica, di darne rappresentazione mediante formule, in una varietà di contesti. Tale competenza comprende la capacità di ragionare in modo matematico e di utilizzare concetti, procedure, dati e strumenti di carattere matematico per descrivere, spiegare e prevedere fenomeni. Aiuta gli individui a riconoscere il ruolo che la matematica gioca nel mondo, a operare valutazioni e a prendere decisioni fondate che consentano loro di essere cittadini impegnati, riflessivi e con un ruolo costruttivo." (OCSE, 2010).

Nel secondo gruppo di articoli viene fatto *uso dei quesiti* rilasciati. Questo avviene generalmente in due modi diversi: i testi dei problemi proposti vengono utilizzati (eventualmente con modifiche) per realizzare sequenze di insegnamento sul problem-solving rivolte agli studenti (per es. O'Shea & Leavy, 2013) oppure sono riuniti per comporre nuovi test rivolti a studenti di livelli diversi rispetto a quelli dell'indagine PISA o a insegnanti (per es. Sàenz, 2009). Nel primo caso viene solitamente fatto uso di un solo quesito, il secondo tipo di utilizzo può prevedere anche l'impiego di diverse domande.

La terza categoria di ricerche prevede l'*uso dei dati* restituiti dopo la somministrazione del test dell'OCSE. I dati utilizzati possono essere sia i risultati della prova di matematica, per verificare le prestazioni di un determinata categoria di studenti (per es. Hong & Choi, 2014), sia le risposte al questionario studente, eventualmente correlate col test di matematica (si vedano i dati citati in Heinze et al., 2005).

Sebbene queste categorie possano sembrare esaustive, ci si è resi conto che una piccola percentuale di articoli non rientrava in nessuna. Tali pubblicazioni, pur trattando delle prove realizzate dall'OCSE, non fanno nessun uso del materiale prodotto da PISA, si occupano invece di effettuare *un'analisi degli effetti* delle prove sul contesto scientifico, politico, economico o sociale. Un esempio di questa tipologia di articolo può essere identificato nel lavoro di Leder e Forgasz (2008) che, lavorando sulle differenze di genere, notano come

> i risultati delle rilevazioni internazionali su larga scala, incluso il programma per la valutazione internazionale degli studenti (PISA), hanno attratto diffusamente sia l'attenzione della comunità dei ricercatori in didattica della matematica generale sia di coloro che avevano un particolare interesse nelle differenze di genere nell'apprendimento della matematica.

Un altro lavoro in cui i test vengono citati con questo fine, in modo paradigmatico, è quello di Gellert et al. (2013) in cui è messo in evidenza il ruolo dei risultati delle prove sulle politiche cilene:

> Anche se il Cile è, dal punto di vista economico, tra i principali paesi dell'America Latina e ha mostrato una crescita economica significativa negli ultimi venti anni, i suoi dati sul successo accademico (PISA 2009, matematica: 421 punti; TIMSS 2011, nei livelli 4/8: 462/416 punti) rimangono una preoccupazione per il governo cileno e per la società. Sembra che il governo consideri il sistema di controllo della qualità dei risultati accademici come un importante mezzo per far progredire il successo scolastico.

Nel testo citato vengono riportati i punteggi degli studenti cileni nel test di matematica, ma il dato importante per le finalità della ricerca presentate nell'articolo è la preoccupazione che tali punteggi hanno sollevato nella società. Questo esempio mette in evidenza come, in alcuni articoli, si faccia uso delle prove in più di uno dei modi sopracitati; in questi casi l'articolo è stato catalogato inserendolo nella categoria relativa all'uso fatto per le finalità principali espresse nell'articolo.

Questa classificazione permette di osservare che nella maggior parte degli articoli viene fatto uso dei risultati (60%), tuttavia anche l'impiego del quadro di riferimento (24%) e dei quesiti (10%) è decisamente rilevante. Nella figura 3 sono riassunte le frequenze delle diverse categorie fra i 107 articoli selezionati.

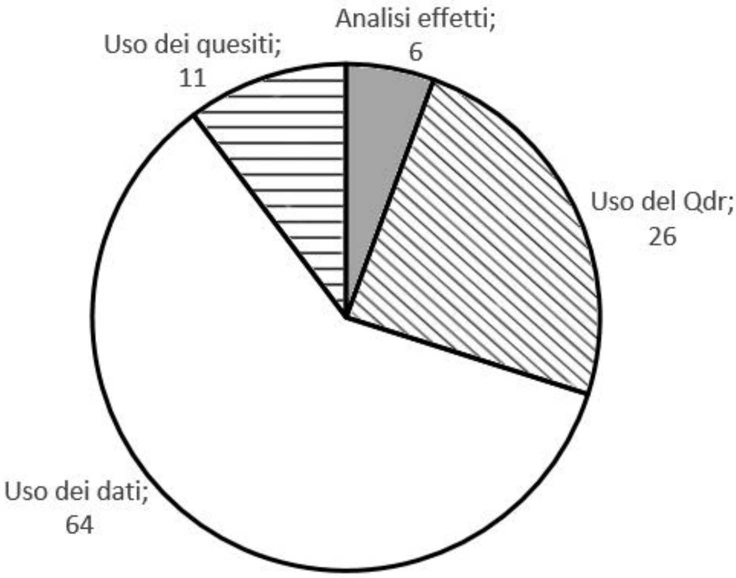


Fig.3: Distribuzione degli articoli selezionati nelle diverse categorie d'uso delle prove.

Nella prossima sezione, attraverso l'analisi di alcune riviste italiane, si esaminerà la frequenza dei diversi usi delle rilevazioni PISA e dei test standardizzati in generale all'interno della ricerca in Didattica della matematica a livello nazionale. Le due classificazioni proposte in questa sezione e in quella precedente saranno poi incrociate nell'ultima parte dell'articolo.

3 L'uso delle prove standardizzate nella ricerca italiana

Si è osservato che la maggior parte degli articoli analizzati riguarda ricerche relative alle specificità dell'insegnamento/apprendimento della matematica all'interno di uno o più Paesi e che l'unico articolo italiano del campione (Boero & Dapueto, 2007) cade all'interno di questa categoria. Ci si è quindi chiesti se la presenza di un solo articolo italiano indicasse uno scarso impiego dell'indagine OCSE-PISA nella ricerca in Didattica della matematica all'interno della nostra nazione. Al fine di dare una parziale risposta a questa domanda è stata eseguita una rassegna degli indici delle riviste italiane che si dedicano principalmente alla Didattica della matematica, scegliendo tra quelle che permettono di consultare gli indici online, ossia:

- Archimede
- Progetto Alice
- L'educazione matematica
- L'insegnamento della matematica e delle scienze integrate

In questo modo sono stati individuati nove articoli che di seguito classifichiamo secondo i criteri riportati nelle sezioni precedenti.
Nell'articolo di Ricci (2008) vengono commentati i risultati del PISA 2003 attraverso un'analisi multilivello dei dati della rilevazione fatta in Italia. Tale lavoro può quindi essere classificato nella categoria "Didattica della matematica in una nazione" ed è l'unico articolo tra quelli analizzati a fare esclusivamente uso dei dati.
A differenza di quanto riscontrato nelle riviste internazionali, le ricerche italiane fanno soprattutto uso dei quesiti e del quadro di riferimento. Per esempio, Pozio (Pozio, 2010; 2011; Pozio & Benvenuto, 2005) ripropone in classe alcuni quesiti per fare una analisi anche qualitativa degli errori e delle omissioni degli studenti italiani nella prova PISA 2003. Ancora più numerosi sono i lavori che fanno uso del quadro di riferimento per costruire attività didattiche e problemi (Fantini, 2009; Raimondi, 2009). Un esempio del tutto particolare è l'uso del quadro di riferimento per costruire quesiti "tipo PISA" per livelli scolastici diversi da quello a cui la rilevazione è destinata (Grasso & Mellone, 2009; Mellone, 2010).
Visto l'esiguo numero di articoli, l'uso delle prove PISA non sembra molto diffuso nella ricerca italiana in Didattica della matematica, tuttavia le prove standardizzate stanno acquisendo un ruolo crescente negli ultimi anni attraverso l'impiego dei test INVALSI. I

primi articoli riguardo alle prove nazionali che si possono trovare sulle riviste sopracitate si occupano di presentare il quadro di riferimento (Perelli D'Argenzio, 2006; Bolondi, 2010), senza però trattare nessuno dei temi rilevati nella sezione 2.1.

Mangini (2010) e Chimetto et al. (2010) riportano le riflessioni emerse dal seminario organizzato dal centro Morin in merito alle prove INVALSI. Lo stesso fa Zottarel (2010) riportando anche diversi esempi presi direttamente dalle prove con le relative risposte. In questi ultimi tre articoli si ipotizza di poter utilizzare le prove per migliorare la pratica didattica.

Così come si è visto, a livello internazionale, per le prove PISA, il rapporto fra test e sviluppo del curricolo è un tema piuttosto dibattuto. Lo stesso vale anche relativamente al Servizio Nazionale di Valutazione (SNV). A partire dal 2010 Brunelli (2010a-b; 2013) ha presentato le sue riflessioni in merito ai risultati delle proprie classi nella Prova Nazione fornendo spunti per un ripensamento della pratica didattica. Mayer (2011) entra nel merito del tema realizzando una analisi longitudinale degli obiettivi delle Indicazioni Nazionali attraverso i risultati dei test del SNV. Più recentemente De Virgilis e Pesci (2014) hanno fornito un esempio di ripensamento dell'azione didattica e dei contenuti da insegnare a partire sia dai risultati delle prove INVALSI sia dalla sperimentazione in classe.

Relativamente agli altri temi indicati nella sezione 2.1, il lavoro di Paola (2005; 2013) costituisce l'unico esempio di confronto fra prove standardizzate (in questo caso PISA-INVALSI) ed è anche l'unico a utilizzarle soprattutto per evidenziarne gli effetti sulla società.

Per quel che riguarda le prove INVALSI, si rileva quindi una totale mancanza di articoli che trattino dei risultati nell'intero Paese nonché del confronto con altri Paesi, delle differenze di genere, di motivazione o affettività e di modellizzazione. Si nota che il tema della formazione degli insegnanti viene giudicato importante in molti degli articoli analizzati, tuttavia nessuno di essi focalizza su questo aspetto.

4 Conclusioni

Le classificazioni proposte hanno permesso di mettere in evidenza sia i diversi temi affrontati dalla ricerca in Didattica della matematica facendo uso delle prove standardizzate, sia i materiali legati alle prove che vengono utilizzati. In ultima battuta appare interessante osservare come l'incrocio delle due classificazioni possa fornire ulteriori informazioni.

Nella tabella 3 è riportata la classificazione incrociata per i 107 articoli del campione internazionale.

	Uso dati	Uso quesiti	Uso QdR	Analisi effetti
Didattica della matematica in una nazione	2	3	6	2
Confronto tra nazioni	16		1	
Formazione degli insegnanti	5	4	2	
Confronto con altre prove standardizzate			9	
Sviluppo dei curricoli	3	2	5	
Genere	6			
Motivazione e affettività	3			
Modellizzazione			2	

Tab. 3: Classificazione incrociata del campione internazionale.

Incrociando le due classificazioni presentate, si nota come alcune combinazioni siano piuttosto frequenti mentre altre risultino assenti. Per esempio, le differenze di genere sono studiate esclusivamente attraverso l'uso dei dati, tuttavia potrebbe essere possibile utilizzare i quesiti per ulteriori analisi di tipo qualitativo. Viceversa, in nessuno degli articoli analizzati, vengono utilizzati i dati per confrontare diverse prove standardizzate o analizzare i processi di modellizzazione degli studenti.

In accordo con Owens (2013) si rileva, inoltre, un forte squilibrio sullo studio dei risultati dei diversi Paesi: le nazioni che vengono maggiormente esaminate negli articoli risultano essere appartenenti a Europa, Nord America, Oceania, parte dell'Asia e parte dell'America Latina con una totale assenza del continente africano (fig. 4) che effettivamente ha sempre contato il minor numero di Paesi partecipanti alle rilevazioni.

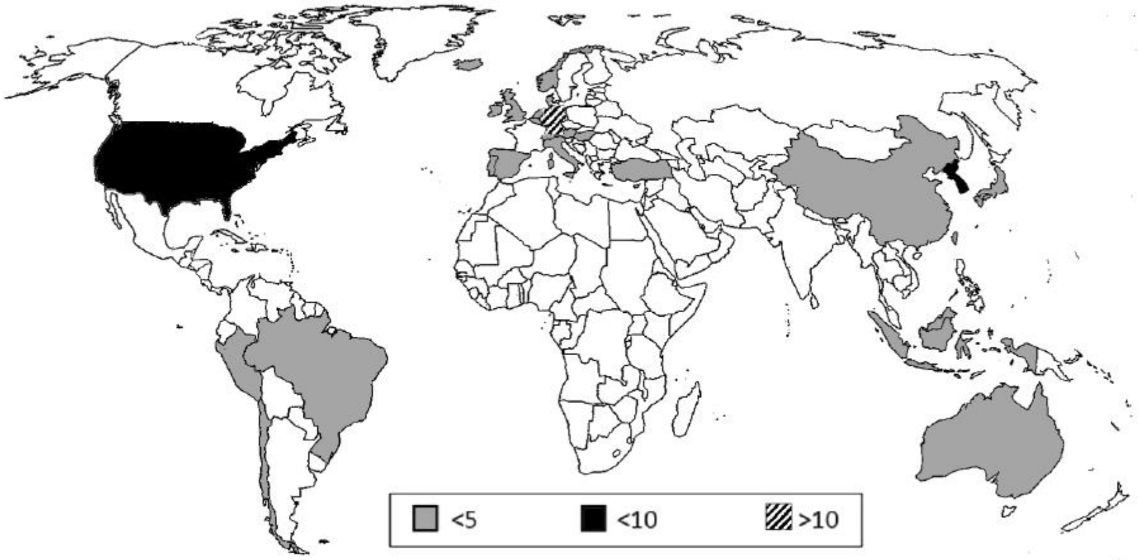


Fig.4: Distribuzione geografica degli articoli selezionati per i temi

"Didattica della matematica in una nazione" e "Confronto tra nazioni".

Per quanto riguarda le specificità del nostro Paese, gli articoli italiani analizzati che fanno uso del PISA, sebbene siano un numero ristretto, coprono le diverse possibilità di utilizzo dei materiali ma concentrandosi solo su alcuni dei temi individuati. Per esempio, nessuno degli articoli tratta il tema dell'affettività e motivazione, nonostante esista un cospicuo corpo di ricerca italiana in questo campo. Un altro tema che non trova riscontro negli articoli italiani, ma che potrebbe essere studiato a partire dai risultati delle prove PISA, risulta essere quello delle differenze di genere.

Conclusioni simili possono essere tratte dallo studio degli articoli che riguardano le prove nazionali INVALSI. Per quanto riguarda gli usi, buona parte degli articoli si riferiscono ai quesiti, per parlare di miglioramenti del curricolo; si rileva invece una quasi totale assenza di pubblicazioni relative alla formazione insegnanti, che invece conta diversi rappresentanti nella ricerca internazionale sul PISA. A livello nazionale lo studio degli effetti politici e sociali delle prove standardizzate (siano esse nazionali o internazionali) non compare in nessuno degli articoli di Didattica della matematica del campione analizzato.

Tali mancanze potrebbero costituire nuovi temi di ricerca e nuove modalità di utilizzo dei test non ancora esplorati ma non per questo infruttuosi, così come ci suggerisce la ricerca a livello internazionale.

Nella seconda sezione abbiamo notato come il 2005 sia stato un anno particolarmente produttivo per gli articoli relativi al PISA in quanto anno successivo alla restituzione dei risultati sia della prova dell'OCSE, sia del TIMSS. Nel 2015 tale coincidenza si ripeterà nuovamente (anche se il dominio principale del PISA saranno le scienze) e l'Italia sarà uno dei Paesi[11] che parteciperà a entrambe le rilevazioni. Sarà quindi la prima volta in cui sarà possibile confrontare i risultati delle prove INVALSI con ben due indagini internazionali, una buona occasione per riflettere sulle nostre pratiche, anche con una prospettiva internazionale.

Riferimenti

Blum, W. (2002). ICMI Study 14: Applications and modelling in mathematics education – Discussion document. *Educational Studies in Mathematics*, 51(1-2), 149-171.

Boero, P., & Dapueto, C. (2007). Problem Solving in Mathematics Education in Italy: Dreams and Reality. ZDM – *The International Journal on Mathematics Education*, 39(5-6), 383-393.

Bolondi, G. (2010). Come usare in classe le prove Invalsi. *L'insegnamento della matematica e delle scienze integrate*, 33(6A-B), 686-702.

Brunelli, F. (2010). Un quadrato metà di un altro: riflessioni sulla prova INVALSI di matematica per il primo anno della Scuola Secondaria di Primo Grado del 13 maggio 2010. *Progetto Alice*, XI(33), 407-426.

Brunelli, F. (2010). La prova nazionale di matematica per l'esame di stato della scuola secondaria di primo grado nell'anno scolastico 2009 2010. *Archimede*, 1-2010, 6-15.

Brunelli, F. (2013). Considerazioni in margine alle prove Invalsi di matematica per l'esame di licenza media del giugno 2012. *L'insegnamento della matematica e delle scienze integrate*, 36(2), 169-184.

Buchter, A., & Leuders, T. (2005). Quality development in mathematics education by

[11] Oltre all'Italia parteciperanno a entrambe le indagini Argentina, Australia, Austria, Belgio, Bulgaria, Canada, Cile, Cina (Hong Kong e Taipei), Corea del Sud, Croazia, Danimarca, Emirati Arabi, Finlandia, Francia, Georgia, Germania, Giappone, Giordania, Indonesia, Inghilterra, Irlanda, Israele, Kazakistan, Libano, Lituania, Malesia, Malta, Norvegia, Nuova Zelanda, Olanda, Polonia, Qatar, Repubblica Ceca, Repubblica Slovacca, Russia, Singapore, Slovenia, Spagna, Stati Uniti, Svezia, Tailandia, Turchia, Ungheria.

focussing on the outcome: new answers or new questions? *ZDM – The International Journal on Mathematics Education*, 37(4), 263-266.

Chimetto, M.A., Tomasi, L., & Zoccante, S. (2010). Le prove di valutazione di matematica INVALSI per la Scuola secondaria di II grado. *L'insegnamento della matematica e delle scienze integrate*, 33(6A-B), 739-749.

De Virgilis, R., & Pesci, A. (2014). I quesiti di matematica INVALSI 2013 sulle percentuali: dall'analisi degli errori al ripensamento dell'azione didattica. *L'insegnamento della matematica e delle scienze integrate*, 37(2B), 133-158.

Fantini, R. (2009), Metodologie e strategie didattiche per lo sviluppo di competenze matematiche in un'ottica OCSE-PISA. *L'educazione matematica*, 2009-3, 35-44.

Ferrini-Mundy, J., & Schmidt W.H. (2005). International Comparative Studies in Mathematics Education: Opportunities for Collaboration and Challenges for Researchers. *Journal for Research in Mathematics Education*, 36(3), 164-175.

Gellert U., Espinoza L., & Barbé J. (2013), Being a mathematics teacher in times of reform. *ZDM – The International Journal on Mathematics Education*, 45(4), 535-545.

Goos, M., & Geiger, V. (2012). Connecting social perspectives on mathematics teacher education in online environments. *ZDM – The International Journal on Mathematics Education,* 44(6), 705-715.

Grasso N., Mellone M (2009). Un problema OCSE-PISA per bambini di 8 anni: le scelte dell'insegnante. *L'insegnamento della matematica e delle scienze integrate,* 2-2009, 43-50.

Heinze, A., Reiss, K., & Franziska, R. (2005). Mathematics achievement and interest in mathematics from a differential perspective. *ZDM – The International Journal on Mathematics Education,* 37(3), 212-220.

Hong, D.S., & Mi Choi, K. (2014). A comparison of Korean and American secondary school textbooks: the case of quadratic equations. *Educational Studies in Mathematics*, 85(2), 241-263.

Kramarski, B., Mevarech, Z.R., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics*, 49(2), 225-250.

Leder, G., & Forgasz, H. (2008). Mathematics education: new perspectives on gender. *– The International Journal on Mathematics Education,* 40(4), 513-518.

Linneweber-Lammerskitten, H., & Wälti, B. (2005). Is the definition of mathematics as used in the PISA assessment Framework applicable to the HarmoS Project? *ZDM – The International Journal on Mathematics Education,* 37(5), 402-407.

Lorenz, J. H. (2005). Zentrale Lernstandsmessung in der Primarstufe—Vergleichsarbeiten Klasse 4 (VERA) in sieben Bundesländern. *ZDM – The International Journal on Mathematics Education,* 37(4), 317-323.

Mangini, P. (2010). Le prove dell'INVALSI per i vari livelli scolastici. *L'insegnamento della matematica e delle scienze integrate,* 33(6A-B), 735-738.

Mayer, G. (2011). L'influenza delle prove Invalsi nel curricolo di Matematica. *Progetto Alice*, 35(XII), 195-220.

Mellone, M. (2010). La gestione di un'attività del tipo pisa per bambini di 8 anni: il ruolo del mediatore semiotico. *L'insegnamento della matematica e delle scienze integrate*, 33(2), 159-180.

OCSE (2003). PISA 2003 *Assesment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*.

OCSE (2010). PISA 2012 *Mathematics Framework.*
http://www.oecd.org/pisa/pisaproducts/46961598.pdf

O'Shea, J., & Leavy, M.L. (2013). Teaching mathematical problem-solving from an emergent constructivist perspective: the experiences of Irish primary teachers. *Journal of Mathematics Teacher Education,* 16(4), 293-318.

Owens, T.L. (2013). Thinking beyond League Tables: a review of key PISA research questions. PISA, Power, and Policy: the emergence of global educational governance, 27-49.

Paola, D. (2005). Possibili conseguenze didattiche dell'uso dei test strutturati per la valutazione delle competenze matematiche: i casi delle prove PISA e INVALSI. *Progetto Alice*, VI(18), 493-518.

Paola, D. (2013). Test INVALSI e valutazione degli apprendimenti: otto anni dopo. *Progetto Alice*, II, 14, 41, 307-330

Perelli D'Argenzio, M. P. (2006). La valutazione esterna degli apprendimenti: le prove di valutazione INVALSI. Le prove di Matematica. *L'insegnamento della matematica e delle scienze integrate*, 29(1A), 31-46.

Pozio, S. (2010). Le difficoltà degli studenti italiani nelle prove di matematica del PISA 2003. *L'insegnamento della matematica e delle scienze integrate*, 33B(5), 511-532.

Pozio, S. (2011). La risoluzione di prove di competenza matematica. Analisi dei risultati italiani nell'indagine OCSE-PISA 2003. Nuova Cultura. Roma.

Pozio, S., & Benvenuto, G. (2005). La valutazione delle abilità matematiche e l'indagine OCSE-PISA (2003). *Progetto Alice*, VI (17), 357-382.

Raimondi, R. (2009), Un'esperienza innovativa in classe con l'utilizzo nella didattica di quesiti sul modello OCSE-PISA. *L'educazione matematica*, 1-2009, 59-64.

Ricci, R. (2008). La valutazione della competenza matematica sulla base dei dati OCSE-PISA. *L'insegnamento della matematica e delle scienze integrate*, 31(6A-B), 633-643.

Roesken, B., Pepin, B., & Toerner, G. (2011). Beliefs and beyond: affect and the teaching and learning of mathematics. *ZDM – The International Journal on Mathematics Education*, 43(4), 451-455.

Sàenz, C. (2009). The role of contextual, conceptual and procedural knowledge in activating mathematical competencies (PISA). *Educational Studies in Mathematics,* 71(2), 123-143.

Sfard, A. (2005). What Could be More Practical than Good Research?. *Educational Studies in Mathematics*, 58(3), 393-413.

Steinthorsdottir, O. B., & Sriraman, B. (2008). Exploring gender factors related to PISA 2003 results in Iceland: a youth interview study. *ZDM – The International Journal on Mathematics Education*, 40(4), 591-600.

Wynands, A., & Möller G. (2005). High-performing students in the 'Hauptschule'—A comparison of different groups of students in secondary education within Germany. *ZDM – The International Journal on Mathematics Education*, 37(5), 437-444.

Yang, X., & Leung, F. K. S. (2011). Mathematics teaching expertise development approaches and practices: similarities and differences between Western and Eastern countries. *ZDM – The International Journal on Mathematics Education*, 43(6-7), 1007-1015.

Zottarel, L. (2010). Le prove dell'INVALSI per i vari livelli scolastici. Istruzioni per l'uso. Scuola primaria. *L'insegnamento della matematica e delle scienze integrate*, 33(6A-B), 725-734.

Capitolo 4

Equity in math education:

Analisi dei dati INVALSI per sottogruppi di studenti

Un tema molto discusso negli ultimi decenni è l'equità nell'educazione, ovvero la possibilità per tutti gli studenti di poter raggiungere un alto livello di istruzione e mettere a frutto le proprie potenzialità. Le società odierne multiculturali, plurilingue e soggette a continui cambiamenti devono sempre di più tenere in considerazione l'importanza di una educazione che possa raggiungere tutti tenendo conto delle differenze e delle peculiarità tipiche di ogni gruppo di persone e di ogni individuo. Diversi studi, anche basati su prove standardizzate internazionali come PISA e TIMSS, hanno evidenziato come siano tutt'ora presenti disuguaglianze nell'istruzione degli individui non solo nei paesi in via di sviluppo ma anche nei paesi sviluppati, come l'Italia (OECD, 2013a).

Equity in education is more than an issue of fairness and distributive justice especially in the current period when many countries are trying to develop their human resources as one element in enhancing growth and international competitiveness in the job market. Unequal education implies that human potential is being wasted, and that some individuals do not have the competence to perform well in a modern society. (Wößmann & Schutz, 2006; Ongaki & Musa, 2014)

Il compito di una politica educativa che mira all'equità è quello di permettere a ogni cittadino di sviluppare a pieno le proprie abilità, partecipare attivamente allo sviluppo di una società sempre più interconnessa e globale e, infine, garantire condizioni lavorative che consentano un migliore livello di vita.

L'OECD, relativamente alle prove PISA del 2012 e del 2015, ha dedicato un intero report proprio al tema dell'equità:

PISA defines equity in education as providing all students, regardless of gender, family background or socio-economic status, with opportunities to benefit from education. Defined in this way, equity does not imply that everyone should have the same results. It does mean, however, that students' socio-economic status or the fact that they have an immigrant background has little or no impact on their performance, and that all students, regardless of their background, are offered access to quality educational resources and opportunities to learn. (OECD, 2013a)

Politiche educative che non riescono a garantire equità nel diritto d'accesso all'istruzione non gravano in modo casuale su alcuni studenti della popolazione ma rischiano di penalizzare particolarmente determinati gruppi che risultano essere già socialmente svantaggiati. Per questo motivo gli studi sull'equità sono spesso affrontati collegando le performance degli studenti con fattori come il genere, le condizioni socio economiche, la cittadinanza, la lingua madre, la regione geografica e il livello di istruzione dei genitori (Zhu, 2016). Questi fattori non incidono ciascuno singolarmente sulle disuguaglianze nell'educazione ma, spesso, concorrono nel determinare un aumento dello svantaggio di alcuni gruppi (Morley, Leach, & Lugg, 2009; UNESCO, 2008).

4.1 Equità nell'educazione matematica

Anche nell'ambito dell'educazione matematica si riscontrano differenze tra gruppi di persone legate a diversi fattori. Il primo obiettivo delle politiche educative deve essere quello di garantire equità nell'accesso all'educazione matematica perché come sottolineato da Bishop e Forgasz (2007) "*without access to mathematics education there can be no equity*". Bisogna quindi fare in modo che tutti gli studenti abbiano le stesse possibilità di imparare la matematica, indipendentemente da genere, status socio-economico, cittadinanza o altro.

Le diversità che si riscontrano, sempre più marcate nelle società moderne e multiculturali, devono essere considerate e analizzate per garantire l'equità nell'educazione. Quest'ultima deve essere intesa non come uguaglianza ma nel senso suggerito da Don Milani di *"Non fare parti uguali tra disuguali"*.

Negli ultimi decenni in molti paesi ci sono stati miglioramenti notevoli dal punto di vista dell’equità ma divari di diversa natura nell’apprendimento della matematica persistono anche nelle società attuali e risultano marcate anche in Italia (OECD, 2013a; OECD, 2016a). Le ricerche in questo campo hanno lo scopo di evidenziare le differenze, studiarle e interpretarle per ricavarne ulteriori indicazioni che possano servire per migliorare le pratiche e le politiche educative.

Prove standardizzate come PISA e INVALSI che vengono somministrate a popolazioni molto numerose e diversificate, permettono di studiare le performance degli studenti proprio tenendo conto delle tipiche classificazioni in base al genere, allo status socio economico o della cittadinanza.

Gli studi presentati nei prossimi paragrafi riguardano in particolare lo studio delle differenze legate alla cittadinanza e al genere degli studenti e sono svolti utilizzando i risultati delle prove INVALSI.

4.2 Analisi dei dati INVALSI per sottopopolazioni di studenti

Nel quadro italiano lo studio dei dati relativi alle prove INVALSI, risulta essere una ricca fonte di informazioni per studiare le performance degli studenti in italiano e in matematica e confrontare questi risultati con quelli delle indagini a livello internazionale. Le prove di matematica e italiano INVALSI vengono accompagnate da un approfondito "questionario studente" che ha lo scopo di raccogliere informazioni relative al contesto familiare e sociale in cui vive l'alunno, alle attività scolastiche ed extrascolastiche svolte, alle convinzioni e agli atteggiamenti relativi allo studio e alla scuola. L'analisi congiunta del "questionario studente" e dei risultati ottenuti nelle prove disciplinari può essere di grande aiuto nell'individuazione di correlazioni tra l'andamento scolastico, il background socio-economico-culturale e le convinzioni degli studenti.

Nei prossimi paragrafi saranno analizzate, attraverso lo studio dei risultati delle prove INVALSI, le diverse performance in matematica in base alle differenze di cittadinanza e di genere. In analogia con quanto si osserva nelle prove standardizzate internazionali come PISA e TIMSS (OECD, 2015a; OECD, 2016b; OECD, 2012a; Mullis et al., 2016), anche dai risultati INVALSI emerge negli anni un costante gap nelle performance in matematica tra maschi e femmine, a favore dei maschi, e tra studenti italiani e stranieri, a favore dei primi (INVALSI, 2015; INVALSI, 2016a). Per quanto riguarda le prove di italiano invece, le ragazze ottengono generalmente risultati migliori dei ragazzi e questo risulta essere sempre in accordo con i risultati delle indagini internazionali. Gli studenti stranieri sono, come presumibile, maggiormente in difficoltà rispetto ai coetanei italiani nella prova di italiano.

4.2.1 Strumenti di analisi

L'uso dei risultati delle prove standardizzate per lo studio di differenze di genere e di cittadinanza risulta essere diffuso, specialmente a livello internazionale (Maffia & Giberti, 2016). L'analisi quantitativa di grandi gruppi di studenti permette non solo di verificare la presenza di un gap, ma anche di studiare approfonditamente in che modo si formano tali differenze per ipotizzarne le cause. L'analisi di un gap di performance viene solitamente riportata, all'interno dei rapporti e dei principali studi che fanno uso delle prove standardizzate, in termini di punteggio medio ottenuto nel test (INVALSI, 2016a; OECD, 2016a). Negli studi presentati in questa tesi sono stati utilizzati i risultati delle prove

INVALSI per individuare differenze non solo sull'intero test, ma sui singoli item che lo compongono; il fine è quello di verificare se vi siano alcune caratteristiche degli item (tipologia, formulazione, contesto) o alcuni contenuti matematici che abbiano una influenza maggiore sul gap.

Gap Index

Gender gap Index

Il divario tra le performance tra maschi e femmine sui singoli item è stato inizialmente analizzato come differenza tra le percentuali di risposte corrette agli item per i due gruppi.

Sarebbe più corretto considerare la differenza tra le due percentuali anche in base alla difficoltà della domanda e quindi alla percentuale complessiva di risposte corrette. Un 10% di differenza su una domanda molto semplice che ottiene, ad esempio, una percentuale media di risposte pari all'80%, dovrebbe avere un peso minore rispetto alla stessa differenza su un quesito più complesso a cui solo il 20% degli studenti risponde correttamente.

A tal fine è stato creato un apposito indice che permette di analizzare il gender gap sui singoli item tenendo conto di questo fattore:

$$I_{GG_k} = \frac{M_k - F_k}{P_k} \qquad (4.1)$$

Dove:

- M_k è la percentuale di risposte corrette dei maschi all'item *k*
- F_k è la percentuale di risposte corrette delle femmine all'item *k*
- P_k è la percentuale di risposte corrette dell'intera popolazione all'item *k*.

Individuare, attraverso questo indice, le domande in cui il gender gap è più marcato permette di studiarne le caratteristiche comuni e quindi di comprendere maggiormente le specificità di questo fenomeno.

Indice gap di cittadinanza

Allo stesso modo chiaramente è possibile considerare il gap di cittadinanza tenendo conto della difficoltà dei quesiti. In questo caso quindi l'indice sarà applicato come segue per analizzare il gap di cittadinanza:

$$I_{CIT\,k} = \frac{I_k - S_k}{P_k} \qquad (4.2)$$

Dove:

- I_k è la percentuale di risposte corrette degli italiani all'item k
- S_k è la percentuale di risposte corrette degli stranieri all'item k
- P_k è la percentuale di risposte corrette dell'intera popolazione all'item k.

Analisi di coorti di studenti

Le prove INVALSI vengono somministrate in diversi livelli scolastici ormai da quasi 10 anni e questo rende possibile lo studio dei risultati in due prospettive particolarmente interessanti:

- Studio dei risultati degli studenti in uno specifico livello nel corso degli anni
- Analisi dei dati longitudinale, considerando quindi i risultati ottenuti da una coorte[12] di studenti che negli anni ha sostenuto più prove di diversi livelli.

In tutti gli studi che seguono, i dati considerati sono quelli del campione INVALSI, composto da circa 30000-40000 studenti, che risulta essere rappresentativo dell'intera popolazione di

[12] Con il termine *coorte* in statistica si indica "A cohort is a group of persons who experience a certain event in a specified period of time. For example, the birth cohort of 1985 would be the people born in that year." (fonte: OECD, Glossary of Statistical Terms; *Handbook of Vital Statistics Systems and Methods*, Volume 1: Legal, Organisational and Technical Aspects, United Nations Studies in Methods, Glossary, Series F, No. 35, United Nations, New York 1991). In questa tesi, come specificato anche nel testo, il termine coorte è utilizzato con una accezione leggermente più ampia in quanto, per ogni prova, non sono considerati i dati relativi all'intera popolazione ma solamente i dati relativi al campione di studenti rappresentativo di quella popolazione. Se consideriamo quindi una coorte di studenti che ha svolto più prove nel corso del proprio percorso scolastico, gli studenti facenti parte del campione nei diversi anni non saranno sempre gli stessi ma saranno rappresentativi della stessa popolazione.

studenti per quel determinato livello e anno scolastico. I risultati del campione risultano essere particolarmente controllati e affidabili in quanto la somministrazione e l'inserimento dei dati viene effettuato direttamente da esperti INVALSI.

Possiamo quindi considerare le prove somministrate dall'INVALSI dall'anno scolastico 2008/09 ad oggi:

		Livello scolastico				
		L_02	L_05	L_06	L_08	L_10
Anno scolastico	2008/09	A	B			
	2009/10			B		
	2010/11					
	2011/12		A		B	
	2012/13			A		
	2013/14					B
	2014/15				A	
	2015/16					
	2016/17					A

Tabella 4.1: Prove INVALSI suddivise per livelli scolastici dal 2008/09 ad oggi.
In grigio le caselle relative alle prove che sono state somministrate.

Come possiamo osservare nella tabella, il livello 6 e il livello 10 non erano presenti il primo anno e inoltre la prova di livello 6 non è più stata svolta dall'anno 2013/14 ad oggi.

Le prove INVALSI esistono quindi da quasi 10 anni e su diversi livelli scolastici, questo permette anche studi di carattere longitudinale che analizzino l'evoluzione di una coorte di studenti sottoposta a più prove nel corso degli anni. In particolare le coorti considerate per due studi presentati in questo capitolo sono le seguenti:

- Coorte A, comprende gli studenti che nel 2008/09 hanno svolto la prova di livello 2, nel 2011/12 la prova di livello 5, nel 2012/13 la prova di livello 6, nel 2014/15 la prova di livello 8 (Prova Nazionale) e che nel 2016/17 hanno svolto la prova del livello 10. Questa coorte è stata analizzata, ad esclusione del livello 10 i cui dati non sono ancora disponibili, per lo studio del gender gap e alcuni dei risultati ottenuti sono presentati negli articoli "Highlights on gender gap from Italian standardized assessment in Mathematics" (par. 4.4.5) e "Gender Gap in Mathematics and Misconceptions: a study based on large-scale results" (par. 4.4.6).
- Coorte B, comprende gli studenti che nel 2008/09 hanno svolto la prova di livello 5, nel

2009/10 la prova di livello 6, nel 2011/12 la prova di livello 8 (Prova Nazionale) e che nel 2013/14 hanno svolto la prova del livello 10. Questa coorte invece è stata oggetto di studio per le analisi svolte nell'articolo riguardante le differenze di cittadinanza "Lo studente straniero di fronte al testo delle prove INVALSI di italiano e matematica: dall'analisi dei dati agli spunti di intervento" (par. 4.3.1).

Per essere precisi, negli studi sopra citati, per ogni prova sono stati analizzati i risultati del campione INVALSI rappresentativo della popolazione degli studenti di quel livello scolastico. Il campione varia di anno in anno e quindi gli studenti che formano la coorte non sono esattamente gli stessi negli anni, gli studi però prendono in considerazione il campione di studenti che risulta, per ogni singola prova, rappresentativo della medesima popolazione di studenti che può essere quindi seguita durante la sua evoluzione negli anni.

Distribuzione degli studenti in funzione del punteggio di Rasch

Come già spiegato nel primo capitolo, il modello di Rasch applicato all'analisi di una prova standardizzata, opera una stima congiunta di due tipologie di parametri, posti entrambi su una scala da -4 a +4: un parametro di difficoltà per ogni item (delta) e un parametro di abilità per ogni studente, considerando il suo risultato sull'intera prova.

Una prima informazione riguardante lo studio delle differenze nei risultati tra due gruppi di studenti può essere ricavata dalla distribuzione della numerosità percentuale dei due gruppi in funzione del punteggio di Rasch. Ad esempio, nello studio del gender gap in matematica, se si applica il modello di Rasch sull'insieme di tutti i dati di una prova si potrà avere per ogni studente un parametro che indica la sua "abilità matematica", ponendo tutti gli studenti, indipendentemente dal sesso, sulla medesima scala di abilità. Così si possono suddividere gli studenti in base al genere e, visto che il numero di studenti maschi e studentesse femmine non è esattamente lo stesso all'interno del campione, riportare la numerosità percentuale dei due gruppi in funzione del punteggio di abilità ottenuto.

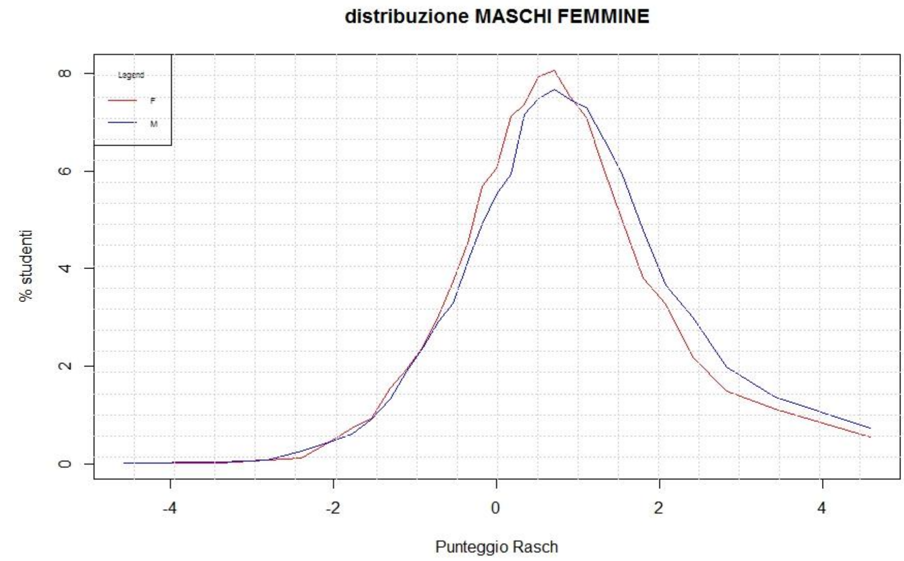
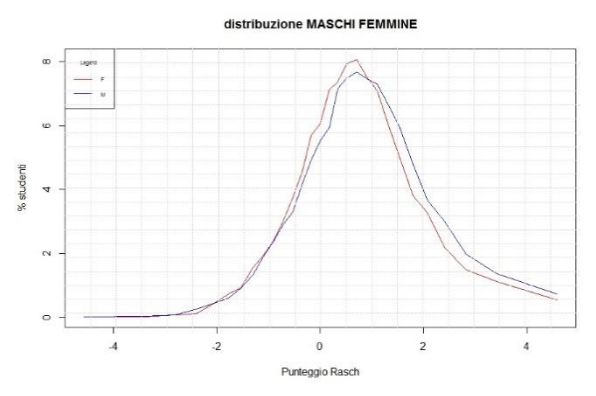
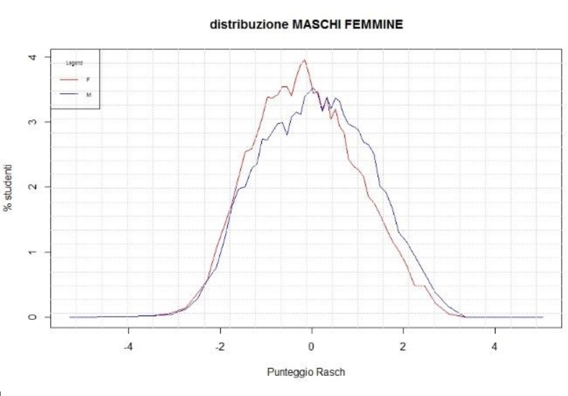
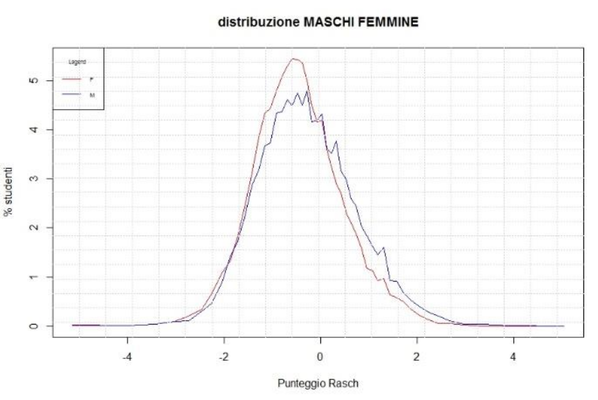


Figura 4.1: Distribuzione numerosità percentuale di maschi e femmine in funzione del punteggio di Rasch nella prova di livello 2 del 2008/09.

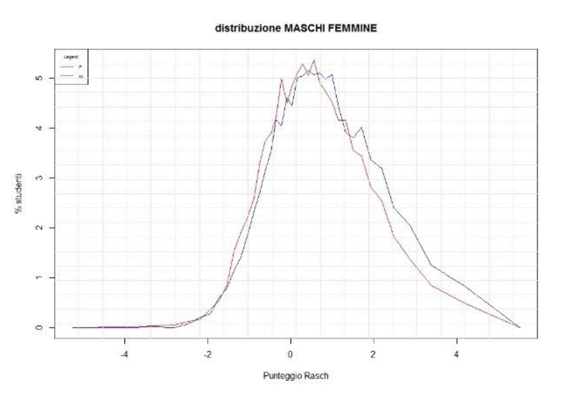
In questo modo si può osservare che nel caso del gap di genere in matematica, tendenzialmente sembra che non vi siano particolari differenze per i livelli bassi di abilità mentre, come confermato anche da molti studi basati non solo sui dati INVALSI ma anche sui dati PISA, un numero minore di alunne raggiunge risultati medio-alti e alti, rispetto ai coetanei maschi (Leder & Forgasz, 2008; González de San Román & De La Rica, 2012; Fryer & Levitt, 2010; Di Tommaso, Mendolia & Contini, 2016).

I grafici seguenti riportano la distribuzione percentuale di studenti e studentesse in funzione del punteggio di Rasch ottenuto nelle prove di matematica dei diversi livelli relative alla coorte A:

	
L02 del 2008/09	L05 del 2011/12



L06 del 2012/13



L08 del 2014/15

Figura 4.2: Distribuzione numerosità percentuale di maschi e femmine in funzione del punteggio di Rasch nelle prove di matematica relative all’analisi della coorte A.

Dall’osservazione dei grafici relativi alla coorte A, sembrerebbe confermato il fatto che a raggiungere i livelli alti nelle prove di matematica vi siano più maschi che femmine in tutti i livelli scolastici. La distribuzione grafica relativa alle femmine però non risulta simile a quella dei compagni maschi (semplicemente traslata verso livelli di abilità più bassi): per quanto si può notare da questi grafici, infatti, per i livelli bassi le due distribuzioni sono pressoché coincidenti, dopo di che la curva delle ragazze presenta un picco più alto per livelli medi di abilità. Questo proprio perché sono poche le ragazze che riescono a raggiungere livelli alti e medio alti, in cui invece è maggiore la numerosità dei ragazzi.

Queste prime osservazioni andrebbero però approfondite maggiormente attraverso studi più rigorosi da un punto di vista statistico e ampliando la ricerca anche includendo altre coorti di studenti e altre prove. Risulta però particolarmente interessante osservare le differenze tra i grafici delle distribuzioni di maschi e femmine e quelli ottenuti, nello stesso modo, confrontando gli studenti in base alla cittadinanza.

Nella pagina seguente sono riportati i grafici relativi alla distribuzione della numerosità percentuale degli studenti italiani e stranieri nelle stesse prove.

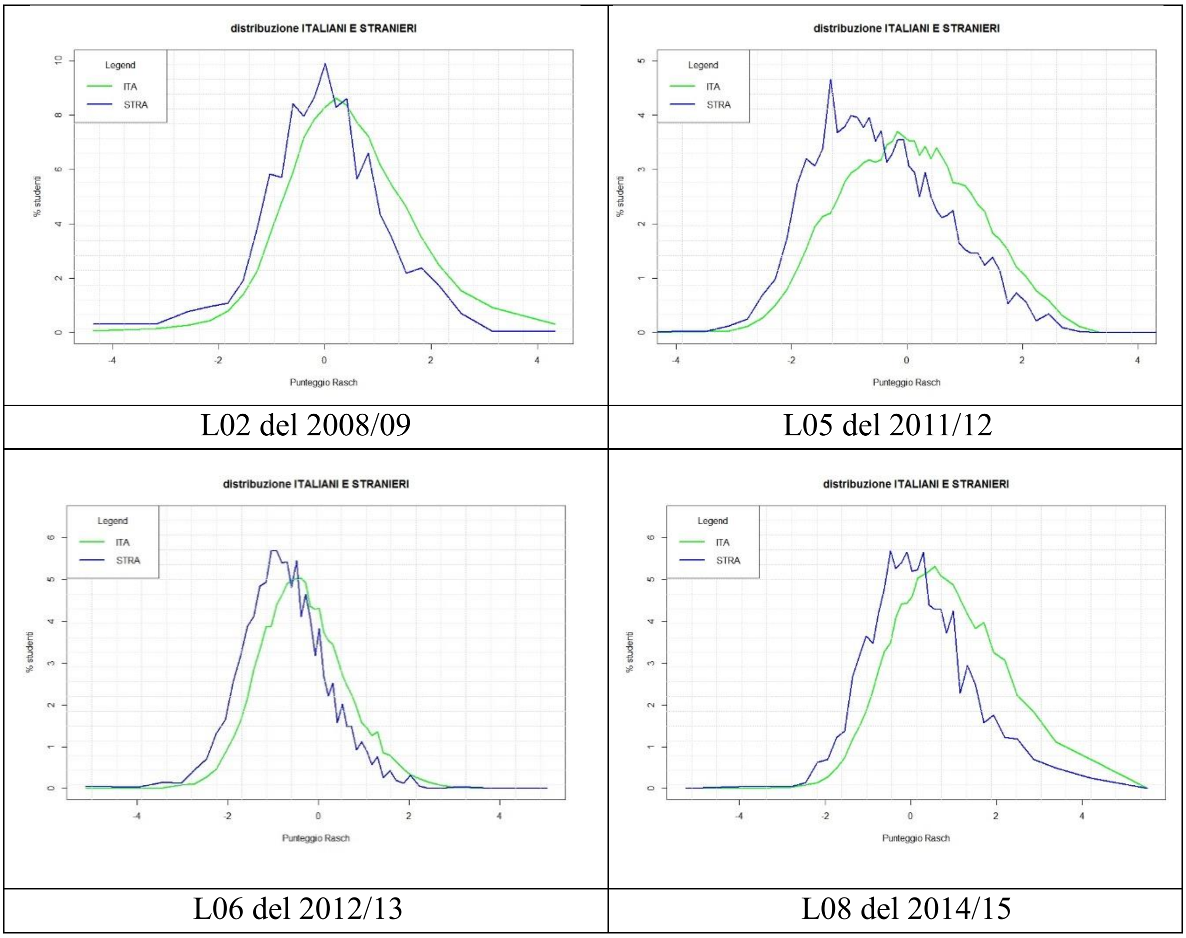


Figura 4.3: Distribuzione numerosità percentuale di maschi e femmine in funzione del punteggio di Rasch nelle prove di matematica relative all'analisi della coorte A.

In questo caso si può notare che le due curve si distanziano per tutti i livelli e quella relativa ai risultati degli studenti stranieri sembra somiglia a quella degli italiani ma traslata verso livelli di abilità più bassi. Al contrario di quanto avveniva suddividendo la popolazione in base al genere, in questo caso il gap tra italiani e stranieri risulta evidente anche per i livelli di abilità più bassi.

La diversa tipologia di gap (italiani e stranieri - maschi e femmine) andrebbe ulteriormente analizzata ma una ipotesi che si potrebbe avanzare relativamente alle differenze osservate è quella che la "natura" dei due gap sia diversa. Il gap tra italiani e stranieri potrebbe essere strettamente legato alle difficoltà nella lingua, che come vedremo nel prossimo paragrafo continua a influenzare anche studenti stranieri di seconda generazione. La diversa padronanza della lingua italiana potrebbe quindi costituire il maggiore ostacolo per gli studenti stranieri e lo studio della matematica in una lingua diversa dalla lingua madre potrebbe influenzare le performances degli studenti stranieri in tutti i livelli di abilità, creando quindi un gap "uniforme"[13].

Per quanto riguarda le differenze di genere la letteratura riporta una enormità di cause che possono essere alla base del gap nelle performance. La natura di questo gap risulta molto complessa e composta da fattori di natura biologica, sociale, culturale, psicologica che probabilmente hanno una diversa influenza in base all'abilità degli studenti e che fanno sì che il gap si evidenzi particolarmente per livelli medi e alti.

[13] Con "gap uniforme" in questo caso si considera un divario presente su tutti i livelli di abilità.

4.3 Differenze di cittadinanza

I flussi migratori che hanno interessato il nostro paese e buona parte dell'Europa negli ultimi decenni, hanno fatto sì che la presenza di alunni stranieri nella scuola italiana aumentasse notevolmente. Dalle ultime rilevazioni ISTAT (2015) e INVALSI (2016a) risulta infatti che se nel 2001/2002 solo il 2% degli alunni era di origine immigrata, negli ultimi anni questa percentuale è incrementata fino a raggiungere il 10-11% in tutti i livelli scolastici. Nella scuola primaria la maggioranza degli alunni stranieri risulta essere di seconda generazione[14], mentre nella scuola secondaria la percentuale di studenti stranieri di prima e seconda generazione è all'incirca la stessa (INVALSI, 2016a).

I risultati delle prove INVALSI di italiano e matematica confermano un significativo gap nelle performance tra gli studenti italiani e stranieri a sfavore di questi ultimi, in linea con quanto osservato anche negli anni passati. Inoltre si registra, sia in italiano sia in matematica, una maggiore difficoltà da parte degli studenti stranieri di prima generazione rispetto a quelli di seconda generazione; questa differenza risulta statisticamente significativa tranne per la prova di matematica della seconda primaria (INVALSI, 2016a). Anche i risultati delle indagini ISTAT relativi all'anno 2015 (ISTAT, 2015) confermano un sostanziale divario nei risultati di italiano e matematica tra gli studenti italiani e i compagni di origine immigrata: nella scuola secondaria gli studenti stranieri hanno voti mediamente inferiori di mezzo punto in entrambe le discipline. Caso a parte sono gli studenti di origine cinese, che in matematica ottengono risultati superiori agli studenti stranieri provenienti da altri paesi e che, nella scuola secondaria di secondo grado, ottengono risultati superiori anche a quelli degli alunni italiani (ISTAT, 2015).

Rivolgendo uno sguardo verso il panorama internazionale, i report forniti dall'OECD relativi alle prove PISA forniscono informazioni relative all'apprendimento in matematica e italiano degli studenti di origine immigrata. La media di alunni stranieri nei paesi OECD risulta essere vicina a quella italiana e nel 2015 l'11% dei quindicenni che hanno risposto ai test PISA era

---

[14] La suddivisione degli alunni in base alla cittadinanza segue i criteri riportati dall'INVALSI (2013) e conforme ai criteri di classificazione internazionali (OECD - Pisa Technical Report 2006, 2009). Sono ritenuti studenti italiani, gli alunni nati in Italia da genitori nati anch'essi in Italia. Sono studenti di origine immigrata di prima generazione, alunni nati all'estero da genitori nati all'estero. Si considerano studenti di origine immigrata di seconda generazione gli allievi nati in Italia da genitori nati all'estero.

straniero: il 6% di seconda generazione e il 5% di prima generazione (OECD, 2015c). Nella prova del 2015 si nota un ulteriore aumento della percentuale di alunni stranieri che sale al 12,5% (OECD, 2016a).

In generale nei paesi OECD le prove PISA di matematica dal 2002 al 2013 mostrano che i gap tra studenti autoctoni e stranieri stanno diminuendo grazie a un lieve miglioramento delle performance degli studenti stranieri, seguito anche da un lieve peggioramento delle performance degli studenti locali. Questo non vale però l'Italia che dal 2002 al 2013 ha incrementato tale divario: gli studenti italiani hanno migliorato i loro risultati in matematica mentre gli studenti stranieri rimangono fortemente svantaggiati (OECD, 2015c).

Performance differences between immigrant and non-immigrant students

Performance gap between immigrant and non-immigrant students in 2012 (Score-point difference)	Change between 2003 and 2012 in the performance gap between immigrant and non-immigrant students (2012-2003 score-point difference)	
-31	-35	Hungary
-26	-29	Australia
-16		Macao-China
	-70	Slovak Republic
		Turkey
	-14	New Zealand
		Ireland
		Canada
		Latvia
		Hong Kong-China
13		United States
22		Russian Federation
26		Czech Republic
40		Luxembourg
37	-10	**OECD average 2003**
44		Portugal
46		Norway
48	26	Italy
50		Liechtenstein
51		Greece
52		Iceland
52		Spain
54	-27	Germany
57		Netherlands
58		Sweden
59		Austria
63	-12	Switzerland
66		Denmark
67		France
73		Mexico
75	-25	Belgium
85		Finland

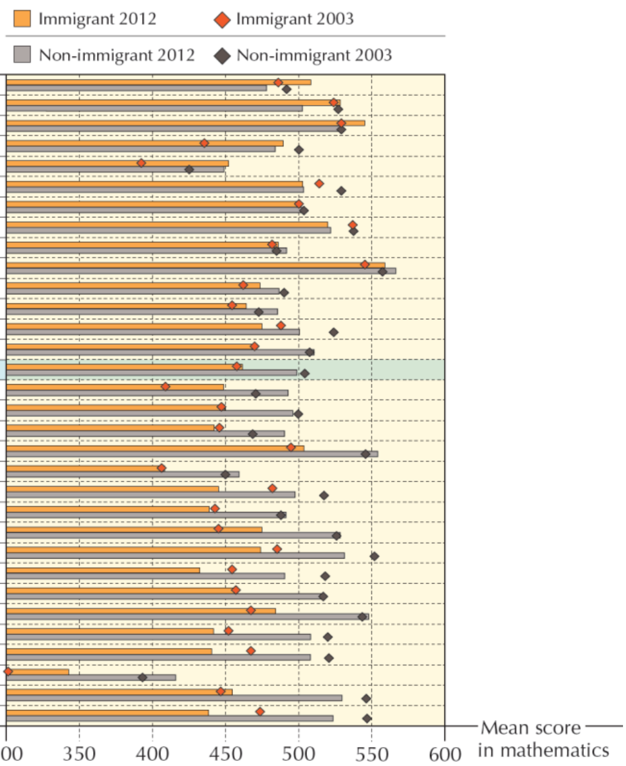


Figura 4.4: Differenze di performance in base alla cittadinanza nelle prove PISA dal 2003 al 2012 (fonte: OECD, 2015c).

Le difficoltà degli studenti stranieri legate a una minore conoscenza della lingua del test e quindi anche della lingua di apprendimento della matematica, sono sicuramente un fattore determinante non solo per gli esiti delle prove di comprensione del testo ma anche nelle prove di matematica. Dai rapporti relativi al test PISA 2015 emerge, infatti, che nei paesi OECD il

67% degli studenti stranieri di prima generazione non parla la lingua del test a casa e anche nelle seconde generazioni questa percentuale rimane molto elevata e pari al 45% degli studenti (OECD, 2016a).

Le cause di questo notevole divario nelle performance in "reading" e in matematica, sono chiaramente di natura linguistica ma anche dovute al differente background socio-economico-culturale. Dai risultati delle rilevazioni PISA (OECD, 2015c) emerge però che solo meno della metà del divario nelle performance in matematica può essere attribuito a fattori legati a un differente background sociale, economico e culturale; l'Italia risulta essere tra i paesi in cui il gap tra italiani e stranieri in matematica risulta essere statisticamente significativo anche al netto degli effetti dovuti al differente background.

The performance gap in mathematics related to immigrant background shrinks by less than half after accounting for differences in socio-economic status (from 37 to 23 score points across OECD countries with data for 2003 and 2012) and remains significant in most countries. This suggests that countries need to do more than fine-tune their immigrant selection mechanisms; they need to strengthen the capacity of their education systems to unleash the potential of all immigrant students. Subsidies for all-day schools or structured language instruction for immigrant students would help them and their families reap the full benefits of education and ensure that immigrant students can contribute to their host country's economic and social well-being. (OECD, 2015c)

In questo contesto, l'analisi delle prove INVALSI può fornire importanti indicazioni al fine di individuare le principali difficoltà riscontrate dagli stranieri e intervenire con una didattica mirata che permetta a tutti gli studenti, indipendentemente dalla propria provenienza, di esprimere al meglio il proprio potenziale. Lo studio dei risultati in relazione alle informazioni di contesto, permette di capire in che misura, nei diversi livelli scolastici e nelle diverse realtà, fattori di natura sociale, economica e culturale continuino a influenzare i risultati degli studenti stranieri.

Inoltre le prove INVALSI permettono di studiare i risultati dello stesso campione di studenti in tre prove distinte: prova di grammatica, comprensione del testo e prova di matematica. La comparazione tra i risultati nelle diverse prove e lo studio delle correlazioni tra le performance, potrebbe essere un grande aiuto anche per capire in che modo la comprensione della lingua italiana influenzi l'apprendimento in matematica e quanto le difficoltà legate alla lingua italiana per gli studenti immigrati possano essere causa dei minori risultati in matematica. Si potrebbe pensare infatti che l'apprendimento della matematica possa prescindere, più delle altre discipline, dalla lingua di insegnamento ma sono molteplici gli studi che dimostrano quanto una scarsa comprensione della lingua e le difficoltà legate al lessico e alla sintassi disciplinare possano influire notevolmente (Sfard, 2008; Zan, 2007; Branchetti & Viale, 2015; D'Amore, 2000; Bolondi & Viale, in stampa; Bolondi, Branchetti & Ferretti, in stampa).

L' articolo presentato nelle pagine seguenti è frutto della collaborazione con il professor Viale dell'Università di Bologna che si occupa di linguistica italiana ed educazione linguistica. Il lavoro è stato presentato a Siena nel marzo 2016 in occasione del XIX convegno GISCEL dal titolo "L'italiano dei nuovi italiani" ed è in fase di pubblicazione negli atti del convegno.

4.3.1 Lo studente straniero di fronte al testo delle prove INVALSI di italiano e matematica: dall'analisi dei dati agli spunti di intervento

Chiara Giberti, Università di Trento
Matteo Viale, Università di Bologna[*]

1. Numerosità e *performance* complessive degli studenti non madrelingua italiana nelle prove standardizzate

Da un decennio a questa parte, gli studenti delle scuole italiane sono sottoposti in diversi momenti del loro percorso a prove standardizzate del Servizio Nazionale di Valutazione tese a misurare i progressi degli apprendimenti in italiano e matematica[15]. L'ingente quantità di dati resi disponibili da queste prove – puntualmente presentati dall'Istituto nazionale per la valutazione del sistema educativo di istruzione e formazione (INVALSI) nei periodici rapporti tecnici e in documenti di approfondimento – fanno emergere le minori prestazioni degli studenti stranieri immigrati rispetto agli studenti italiani. Un dato generalmente interpretato come il riflesso di una minore competenza linguistica, ma per il quale sono possibili approfondimenti da vari punti di vista.

Questo contributo si propone di passare in rassegna i dati delle prove di italiano e di matematica di una specifica coorte di studenti assunta a esempio[16], per mettere in evidenza, grazie all'impiego sistematico di tecniche di analisi statistica, le diverse *performance* di

[*] Il contributo è frutto del lavoro comune dei due autori. Chiara Giberti ha curato la redazione dei paragrafi 2.1 e 2.3, Matteo Viale dei paragrafi 1, 2.2 e 3.

Gli autori ringraziano il Servizio statistico dell'Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione (INVALSI) per aver messo a disposizione i dati alla base delle elaborazioni presentate in questo contributo.

[15] Si rimanda al sito dell'INVALSI (www.invalsi.it) per il recupero dei testi completi delle prove citate, dei quadri di riferimento, dei rapporti tecnici e dei documenti di approfondimento. Per una presentazione delle prove INVALSI nel contesto scolastico si rimanda a Castoldi, 2014.

[16] Si tratta della coorte formata dagli studenti che hanno sostenuto la prova INVALSI in V primaria (L5) nell'anno scolastico 2008/2009, in I secondaria di primo grado (L6) nel 2009/2010 (dal 2013/2014 la prova di L6 non è più somministrata), in III secondaria di primo grado (L8) nel 2011/2012 e in II secondaria di secondo grado (L10) nel 2013/2014. Il lavoro sistematico su tutte le coorti, attualmente in corso, non sembra mostrare differenze di rilievo tra coorti diverse per quanto riguarda il comportamento degli studenti stranieri.

studenti italiani e stranieri e trarre alcune conclusioni.

Si deve innanzitutto osservare che gli studenti stranieri che hanno sostenuto le prove INVALSI a partire dal 2009/2010 sono circa il 10% degli studenti complessivi e rappresentano una percentuale sostanzialmente stabile nel tempo e, a seconda dei cicli scolastici, con variazioni minime[17]. Pur secondo tendenze non sempre facilmente identificabili, in II primaria (L2) gli stranieri di prima generazione rappresentano circa i due terzi; con l'avanzare dei cicli scolastici questa proporzione si inverte progressivamente, fino alla II secondaria di secondo grado (L10).

Se si guarda il risultato complessivo nelle prove di italiano e matematica degli studenti italiani e stranieri appartenenti alla coorte analizzata, è possibile mettere in evidenza l'entità del *gap* nei risultati degli studenti italiani e stranieri. L'analisi dei risultati delle prove sostenute, nei diversi anni, da campioni di studenti rappresentativi della medesima popolazione, ha quindi permesso di evidenziare possibili variazioni del *gap* nei diversi livelli scolastici. Nel caso della prova di italiano si è scelto inoltre di riportare sia i risultati complessivi della prova, sia i risultati distinti per comprensione del testo e riflessione sulla lingua, allo scopo di osservare se le difficoltà degli studenti stranieri fossero concentrate particolarmente su una delle due.

Limitandosi per ragioni di spazio all'esemplificazione dei dati delle prove di una coorte, nelle tabelle 1 (L6), 2 (L8) e 3 (L10) sono riportati i punteggi medi (in termini di percentuali di risposte corrette) di ogni prova e la differenza tra la percentuale di risposte corrette degli studenti italiani e stranieri nelle diverse prove.

		Totale studenti	Stud. italiani	Stud. stranieri	***Gap***	Str. I gen.	Str. II gen.
Italiano	Punteggio medio	61%	62%	52%	**10%**	49%	57%
	Punt. medio comprensione	61%	62%	53%	**9%**	50%	57%
	Punt. medio rifl. sulla lingua	62%	63%	52%	**11%**	48%	57%
Matem.	Punteggio medio	51%	52%	46%	**8%**	44%	48%

Tabella 1: punteggi medi delle *performance* di studenti italiani e stranieri nelle diverse

[17] Non si prende qui in considerazione la diversa percentuale di studenti stranieri nelle zone d'Italia; su questo aspetto si vedano i dati forniti da Lo Duca, 2014, 3-5.

prove INVALSI di livello 6 dell'anno scolastico 2009/2010.

		Totale studenti	Stud. italiani	Stud. stranieri	***gap***	Str. I gen.	Str. II gen.
Italiano	Punteggio medio	74%	75%	66%	**9%**	64%	70%
	Punt. medio comprensione	74%	75%	67%	**8%**	65%	71%
	Punt. medio rifl. sulla lingua	73%	74%	63%	**11%**	61%	68%
Matem.	Punteggio medio	55%	56%	47%	**9%**	46%	50%

Tabella 2: punteggi medi delle *performance* di studenti italiani e stranieri nelle diverse prove INVALSI di livello 8 dell'anno scolastico 2011/2012.

		Totale studenti	Stud. italiani	Stud. stranieri	***gap***	Str. I gen.	Str. II gen.
Italiano	Punteggio medio	65%	66%	57%	**9%**	55%	61%
	Punt. medio comprensione	67%	68%	59%	**9%**	57%	63%
	Punt. medio rifl. sulla lingua	60%	61%	50%	**11%**	47%	55%
Matem.	Punteggio medio	55%	56%	50%	**6%**	49%	52%

Tabella 3: punteggi medi delle *performance* di studenti italiani e stranieri nelle diverse prove INVALSI di livello 10 dell'anno scolastico 2013/2014.

Il primo dato che emerge con forza dall'osservazione delle tabelle è che gli studenti stranieri ottengono risultati inferiori rispetto ai colleghi italiani in tutte le prove e in tutti i livelli.

Per quanto riguarda l'italiano, il *gap* non sembra variare sensibilmente nel passaggio a un diverso ciclo scolastico, soprattutto per quanto riguarda la prova di italiano, come se la scuola faticasse a colmare – anche nel lungo periodo – le differenze di prestazione legate alla minore conoscenza della lingua. Inoltre, si può notare che il *gap* è più marcato per le domande che riguardano la riflessione sulla lingua, in cui accanto alla competenza linguistica trovano spesso posto le conoscenze di specifici contenuti grammaticali.

Anche nelle prove di matematica si può osservare un divario tra studenti italiani e stranieri ma, per quanto si può osservare dalle prove di questa coorte di studenti, sembra che il *gap* possa lievemente diminuire durante il percorso scolastico.

Si deve infine evidenziare, sia nelle prove di italiano sia in quelle di matematica, una discrepanza tra i risultati degli studenti stranieri di prima generazione e seconda generazione.

Gli studenti di origine immigrata di seconda generazione, nati quindi in Italia da genitori nati all'estero[18], ottengono risultati superiori agli studenti stranieri di prima generazione, che mostrano maggiori difficoltà con la lingua. Si può notare anche che gli studenti di seconda generazione, che sono nati e cresciuti in Italia e quindi dovrebbero padroneggiare la nostra lingua, continuano tuttavia a evidenziare un *gap* nelle prove sia di italiano sia di matematica. Questo *gap* potrebbe però essere dovuto infatti, oltre che alla padronanza linguistica, all'incidenza di fattori sociali, la cui decifrazione va oltre gli obiettivi di questo lavoro.

Il dato percentuale complessivo rischia però di appiattire la corretta rappresentazione della situazione, che per essere compresa nella sua complessità richiede un'analisi approfondita con adeguati strumenti statistici che tenga conto di come i risultati si differenziano a seconda dei tipi di domanda e dei livelli di competenza.

2. L'impatto di domande specifiche per livello di competenza

2.1. Uno sguardo d'insieme

Per un'analisi dettagliata delle *performance* nella prova degli studenti stranieri che vada oltre il semplice dato percentuale, risulta utile il cosiddetto "modello di Rasch", un modello statistico ampiamente utilizzato nell'analisi di test standardizzati, come i test PISA e INVALSI[19]. L'uso di questo strumento, congiuntamente ad altri strumenti statistici afferenti alla "Teoria classica dei test", permette di stimare per ogni studente un parametro di abilità ottenuto sull'intera prova. Allo stesso tempo, a ogni *item* del test viene associato un parametro di difficoltà (delta) in base a quanti studenti hanno risposto correttamente a quell'*item*. L'importanza dell'uso di questo modello sta nel fatto che la difficoltà degli *item* e l'abilità degli studenti sono definite sulla stessa scala e possono assumere valori da -4 a +4. Questo permette quindi di confrontare direttamente la probabilità di rispondere correttamente a un certo *item* per uno studente con un determinato livello di abilità: ad esempio, se consideriamo un *item* con indice di difficoltà pari a -1, quindi piuttosto facile, ci si attende che tutti gli

---

[18] I dati sulla base dei quali è operata la distinzione sono riferiti a quanto dichiarato dagli stessi studenti nel questionario associato al codice meccanografico.

[19] Cfr. Rasch, 1980 per una presentazione dettagliata del modello. Per un inquadramento generale delle metodologie di analisi si rimanda a Barbaranelli, Natali, 2005.

studenti con abilità pari a -1 avranno una probabilità del 50% di rispondere correttamente, gli studenti con livelli di abilità superiori a -1, avranno una probabilità di rispondere correttamente superiore al 50%, mentre gli studenti con abilità inferiore a -1 avranno una probabilità inferiore al 50%.

Applicando questo modello a ogni singola prova analizzata è stato quindi possibile stimare i parametri di abilità di ogni studente e rappresentare la distribuzione percentuale degli studenti italiani e stranieri rispetto al punteggio di Rasch.

A fronte della notevole quantità di dati in esame, si forniranno di seguito esempi riferiti alla coorte già presentata nel paragrafo precedente per introdurre i dati percentuali medi. Per quanto si è potuto finora verificare, le tendenze di seguito presentate sono esemplificative di quelle generali e si ripetono anche per altre coorti.

Nei grafici si può osservare la numerosità percentuale degli studenti per ogni livello di abilità in italiano (grafico 1) e matematica (grafico 2); si nota che la percentuale di studenti stranieri che riesce a raggiungere punteggi alti è decisamente inferiore rispetto agli studenti italiani[20]. Per tutte le prove analizzate, sia di italiano sia di matematica, la distribuzione degli studenti stranieri risulta simile a quella degli studenti italiani, ma con una curva di distribuzione traslata verso abilità più basse.

Il *gap* tra studenti italiani e stranieri è quindi presente in tutte le prove di italiano e matematica analizzate e riguarda tutti i livelli di abilità degli studenti. Dai grafici esemplificati si nota che la curva traslata verso valori mediamente più bassi si riproduce sia per le prove di italiano sia per quelle di matematica, nonostante la diversa forma dalle due curve, con punteggi più concentrati attorno a valori medio-alti per italiano e più distribuiti per matematica. Inoltre, è interessante rilevare che questo tipo di *gap*, con una traslazione della curva dei punteggi degli studenti stranieri verso valori più bassi rispetto agli studenti italiani, non si verifica nel confronto tra altri tipi di popolazioni, come il cosiddetto *gender gap*, cioè le diverse prestazioni nelle prove standardizzate tra maschi e femmine, in cui le differenze si concentrano solo attorno ad alcuni livelli di competenza (Giberti, Zivelonghi, Bolondi, 2016).

[20] Questi dati e tutti quelli di seguito presentati si riferiscono al campione di scuole scelte dall'INVALSI e rappresentativo della popolazione studentesca nei diversi contesti sociali e geografici, in cui la somministrazione della prova avviene in forma controllata per evitare distorsioni dei risultati.

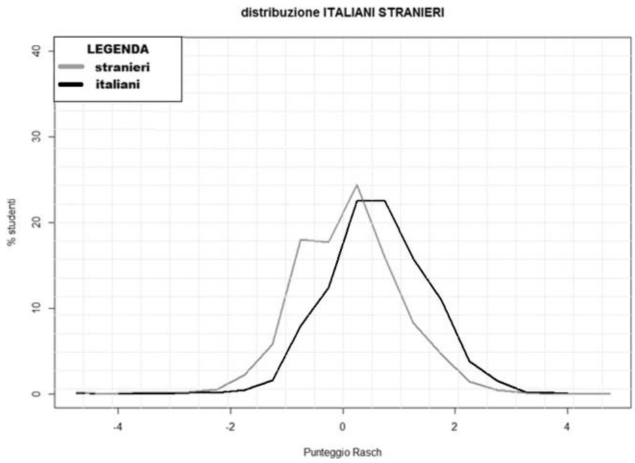


Grafico 1: distribuzione degli studenti italiani e stranieri rispetto al punteggio di Rasch nella prova di italiano di livello 6 del 2009/2010.

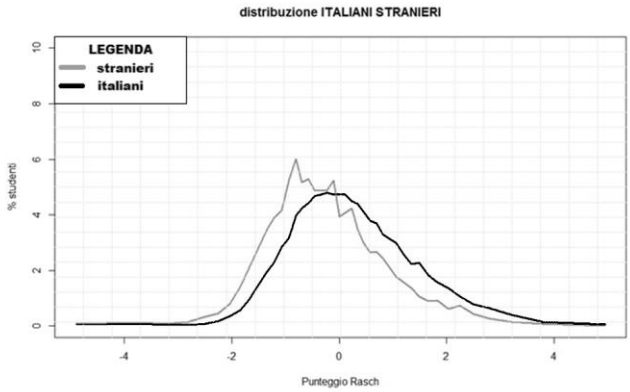
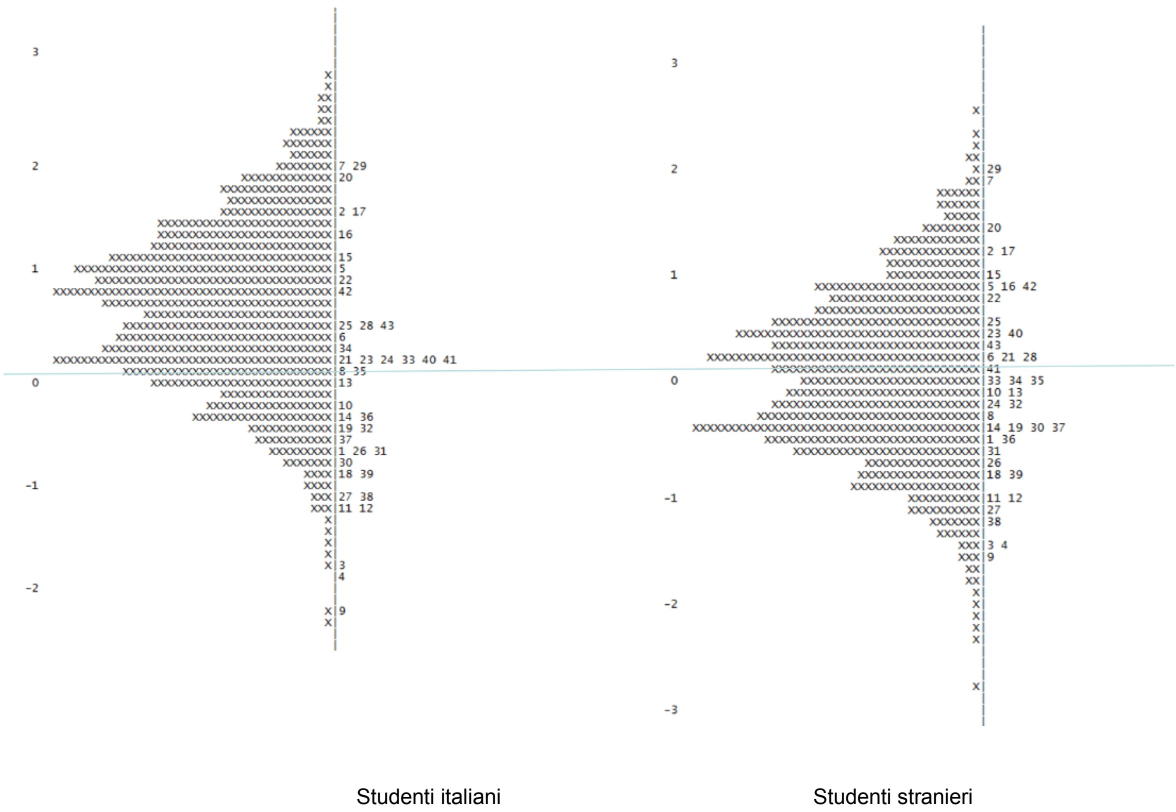


Grafico 2: distribuzione degli studenti italiani e stranieri rispetto al punteggio di Rasch nella prova di matematica di livello 6 del 2009/2010.

Un'altra possibilità data dall'uso del modello di Rasch per l'analisi dei dati consiste nel poter confrontare la distribuzione delle domande e degli studenti rispetto al punteggio ottenuto con le cosiddette "mappe di Wright". Ogni mappa di Wright riporta a sinistra dell'asse verticale la distribuzione degli studenti rispetto al punteggio di Rasch, mentre a destra i numeri indicano ciascuno un *item* della prova e si può quindi osservare la distribuzione degli *item* rispetto al punteggio. In questo modo è stato possibile costruire le mappe di Wright della medesima prova distinguendo tra studenti italiani e stranieri e confrontare le diverse distribuzioni delle domande per livello di difficoltà. In particolare, dal confronto è stato possibile evidenziare le differenze legate ai singoli *item*, riscontrare se il *gap* evidenziato sull'intera prova fosse legato in particolare alle risposte ad alcune domande e quindi verificare in quali casi le differenze rispetto a singoli *item* risultino più marcate.



Corrispondenza numeri mappa- *item*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A1	A2	A3	A4	A5	A6	A7	A8_a	A8_b	A8_c	A8_d	A8_e	A9	A10	A11
16	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
A12	A13	A14	A15	A16	A17	A18	A19	B1	B2	B3	B4	B5	B6	B7
31	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**		
B8	B9	B10	B11	B12	B13_a	B13_b	B13_c	B13_d	B14	B15	B16	B17		

Grafico 3: Confronto tra le mappe di Wright di studenti italiani e stranieri per la prova di italiano di livello 6 del 2009/2010.

L'osservazione delle mappe di Wright (Grafico 3) suggerisce che per comprendere la natura del *gap* in italiano e matematica è necessario uno studio più specifico sulle singole domande dei test allo scopo di verificare per quali *item* emergano maggiori difficoltà per gli studenti stranieri. A tal scopo sono state calcolate le percentuali di risposta corretta di ogni *item* delle prove analizzate, distinguendo gli studenti italiani e stranieri. Pur nell'impossibilità di dar conto dei dati riferiti a ogni domanda per ragioni di spazio, dal confronto delle percentuali è già possibile in prima istanza osservare se il *gap* è distribuito sulla totalità degli *item* oppure se si concentra solo su alcuni.

Per misurare con maggiore precisione il *gap* su ogni *item* si è scelto di utilizzare un indice così costruito:

$$I_{GAP} = \frac{\%risp\ corrette\ ita - \%risposte\ corrette\ stra}{\%media\ risposte\ corrette}$$

In questo modo l’analisi del *gap* risulta essere più accurata: l’indice così calcolato tiene conto del fatto che una differenza percentuale del 10% su una domanda molto facile che ottiene una media di risposte corrette del 70% dovrà essere considerata meno incisiva rispetto alla stessa differenza (10%) su una domanda molto più difficile che ottiene solamente un 30% di risposte corrette.

Utilizzando questo indice sulle prove di italiano e matematica della coorte scelta è stato possibile osservare che il *gap* dovuto alla diversa cittadinanza degli studenti è presente in quasi tutti gli *item* delle prove, ma risulta essere particolarmente marcato in alcune. Si è quindi scelto di selezionare e analizzare in modo approfondito per ogni prova analizzata le domande che creano un *gap* maggiore e quelle che evidenziano un *gap* minore e discuterne alcuni esempi per italiano (2.2.) e per la matematica (2.3.).

2.2. *Prova di lettura: alcuni esempi*

Se si esaminano i dati riferiti a ogni singola domanda della coorte esemplificata in questo lavoro, si nota innanzitutto che un forte discrimine è dato dal formato della domanda, indipendentemente dai contenuti e dalla difficoltà. Di fronte alle domande aperte, anche di facile risposta, lo studente straniero appare sempre in maggiore difficoltà.

L’applicazione del modello di Rasch ha anche permesso di analizzare ogni singola domanda con particolari grafici (*distractor plot*) che mostrano l’andamento della risposta corretta e dei distrattori in funzione del punteggio di Rasch ottenuto sull’intera prova. I *distractor plot*, costruiti suddividendo la popolazione in italiani e stranieri, permettono quindi un confronto specifico che mostra il modo in cui si è formato il *gap* in quella particolare domanda.

Se si osserva, a titolo di esempio, la prova di italiano di livello 6 del 2010, si nota che la domanda a scelta multipla con il maggior *gap* tra italiani e stranieri è la B6, con indice di *gap* di ben 37 punti:

B6. Per quale ragione esistono dei buoi muschiati in Norvegia?

☐ A. Hanno continuato a vivere qui fin da tempi antichissimi

☐ B. Anche se erano scomparsi, sono poi stati reintrodotti dall'uomo

☐ C. Alcuni esemplari di buoi muschiati sono arrivati qui dalla Siberia

☐ D. In questo paese i buoi muschiati possono vivere isolati oltre che in branco

Si tratta di una domanda di comprensione testuale che richiede di «individuare informazioni date esplicitamente nel testo» (Ambito 2 del *Quadro di riferimento della prova di italiano*) e che chiama in causa competenze linguistiche molto alte. Per rispondere correttamente con l'alternativa B, lo studente deve fare riferimento a un passo del testo che compare solo a riga 25, in cui si dice che in Norvegia gli esemplari di bue muschiato sono «frutto di ripopolamenti artificiali», espressione che per essere ricollegata alla risposta corretta richiede una parafrasi non immediata. D'altro canto, i distrattori presenti fanno riferimento a informazioni centrali e ben ribadite del testo (come il riferimento al branco del distrattore D), ma non pertinenti rispetto alla domanda.

Il confronto tra l'alternativa scelta da italiani e stranieri a seconda del livello di competenza complessiva testato sull'intera prova, attraverso i due *distractor plot* (Grafico 4), consente alcune osservazioni di interesse.

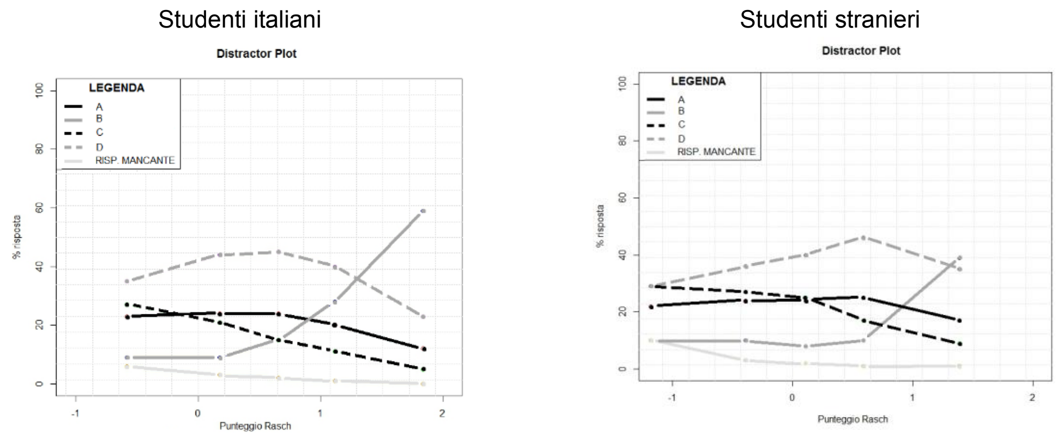


Grafico 4: Confronto tra *distractor plot* delle risposte di studenti italiani e stranieri alla domanda B6 della prova di italiano di livello 6 del 2009/2010.

La domanda risulta innanzitutto molto difficile anche per gli studenti italiani. In particolare, il distrattore D risulta molto attrattivo anche per studenti con livelli di competenza medio-alti e attrae soprattutto nei livelli di competenza medi.

Nonostante la forte attrattività del distrattore, per gli studenti italiani la risposta corretta è scelta comunque in modo proporzionale al livello di competenza complessivo. Ciò non si verifica per gli studenti stranieri, per i quali l'andamento della risposta corretta resta invariato fino a livelli molto alti, quando cessa l'influenza del distrattore e la scelta della risposta corretta inizia a crescere in base al livello di competenza. In un certo senso, è come se la domanda rappresentasse per lo studente straniero una soglia superabile solo per chi ha raggiunto un livello di competenza molto elevato.

È interessante notare che la domanda della stessa prova in cui il *gap* tra italiani e stranieri risulta minore è la B4, con indice di *gap* di 9 punti, che riguarda lo stesso testo e chiama in causa sempre l'individuazione di informazioni date esplicitamente nel testo (Aspetto 2). In questo caso, però, l'informazione da ritrovare è dove si trova l'autore e può essere ricollegata a un'informazione molto esplicita presente nelle primissime righe del testo («Mi trovo in Norvegia»).

In questo caso, sia per italiani sia per stranieri, la scelta dei distrattori decresce e la risposta corretta cresce con il progredire dei livelli di competenza complessivi. La differenza è data dal fatto che la linea della risposta corretta degli studenti stranieri risulta traslata verso i livelli più bassi, come già osservato per l'andamento complessivo.

Il fatto che le domande B6 e B4 facciano riferimento allo stesso aspetto di comprensione della lettura sembra suggerire che la difficoltà per lo studente straniero non è solo legata al processo cognitivo sotteso alla domanda, ma anche alla natura del materiale linguistico oggetto della prova.

A differenza di quanto avviene per la prova di matematica, in cui la formulazione linguistica dei quesiti può essere semplificata per concentrare l'attenzione sulle conoscenze e le competenze matematiche, in casi del genere la complessità dell'*input* va di pari passo con la competenza linguistica richiesta. Il dato, letto in rapporto al comportamento degli studenti italiani, diventa una spia del punto del percorso in cui lo studente si trova e rappresenta uno spunto molto importante per azioni didattiche mirate tese a colmare il *gap*.

Nelle domande di riflessione sulla lingua, invece, il *gap* sembra dipendere più dalle conoscenze grammaticali in gioco, specie nei casi in cui il rapporto tra competenza linguistica e metalinguistica appare meno stretto[21].

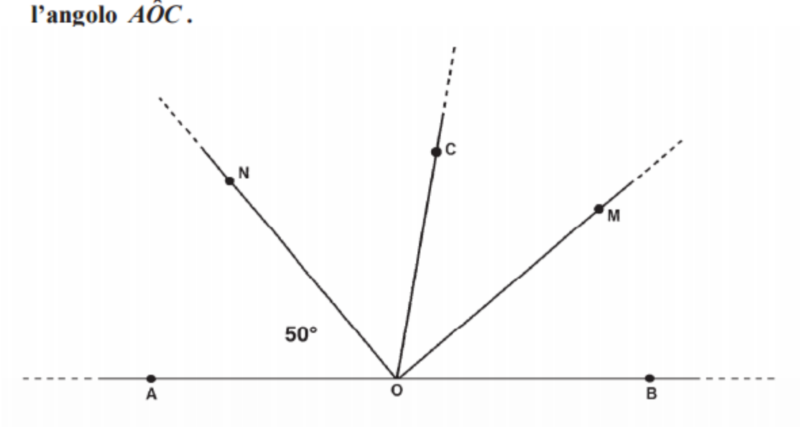
2.3. *La prova di matematica: alcuni esempi*

Se si osservano, come per la prova di italiano, gli indici del *gap* tra italiani e stranieri nelle diverse domande di matematica, si può notare che gli studenti stranieri mostrano maggiori difficoltà non solo in domande difficili, ma anche in domande di difficoltà facile e media. Le domande con *gap* maggiore non appartengono in particolare a un singolo ambito di contenuto ("Numeri", "Spazio e Figure", "Dati e Previsioni", "Relazioni e funzioni") e, pur essendo in prevalenza a risposta univoca, anche domande a risposta multipla hanno creato un *gap* notevole.

Ad esempio, nella prova di livello 6 del 2009/2010, tra le domande con maggiore *gap* tra italiani e stranieri vi è il quesito D12, che appartiene all'ambito "Spazio e Figure" e richiede di interpretare le informazioni contenute nel testo collegandole all'immagine sottostante per risolvere un problema legato agli angoli della figura.

[21] Per un'analisi del diverso comportamento di studenti italiani e stranieri di fronte alla prova INVALSI di riflessione sulla lingua si rimanda allo stimolante studio di Lo Duca, 2014.

D12. Nella seguente figura i punti A, O e B giacciono sulla stessa retta. OM divide in due parti uguali l'angolo $B\hat{O}C$ e ON divide in due parti uguali l'angolo $A\hat{O}C$.



a. Qual è la misura dell'angolo $M\hat{O}B$?

Risposta: …………………… gradi

b. Scrivi i calcoli che hai fatto per trovare la risposta.

………………………………………………………………………………

………………………………………………………………………………

Il quesito è suddiviso in due *item*: il primo a risposta univoca in cui si chiede di riportare il risultato, mentre nel secondo *item* si chiede di riportare i calcoli svolti per giungere al risultato. Entrambi gli *item* risultano molto difficili e la percentuale di risposte corrette è molto bassa per tutti gli studenti. Gli studenti stranieri, però, mostrano difficoltà ancora maggiori rispetto ai compagni italiani e risponde correttamente solo il 14% al primo *item* e il 22% al secondo. L'indice del *gap* tra italiani e stranieri è di 28 punti nel primo *item* (D12a) e di 34 nel secondo (D12b).

Al di là dei contenuti matematici in gioco, le maggiori difficoltà degli studenti stranieri sono probabilmente dovute soprattutto alla fase di comprensione e "decifrazione" del testo del problema, in cui compaiono termini specifici del linguaggio matematico non essenziali rispetto al quesito, come l'uso del verbo *giacere*; anche la struttura sintattica della frase risulta complessa, come tipico nella tradizione dei problemi in matematica, e impatta negativamente sulla prestazione degli studenti (Branchetti, Viale, 2015).

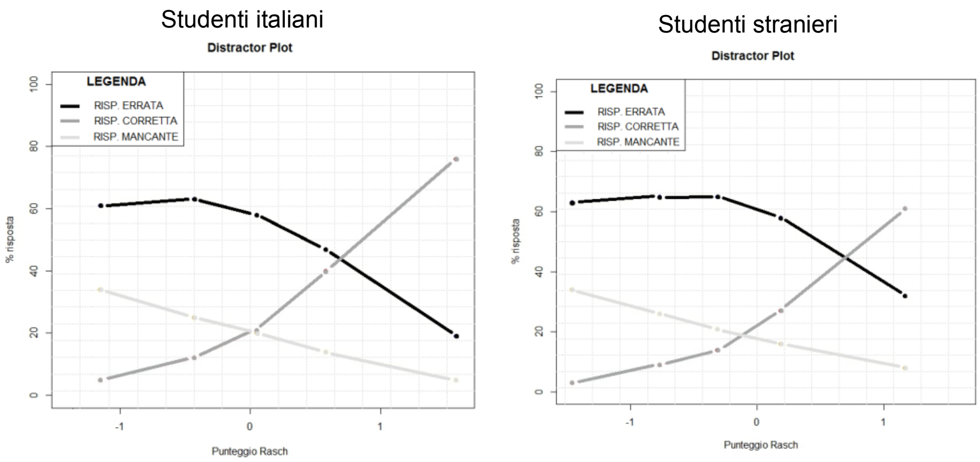


Grafico 5: Confronto tra *distractor plot* delle risposte di studenti italiani e stranieri alla domanda D12 della prova di matematica di livello 6 del 2009/2010.

Nel grafico 5, i *distractor plot* mostrano un andamento abbastanza simile per quanto riguarda la risposta corretta di italiani e stranieri, indice che probabilmente che le difficoltà incontrate dai due gruppi di studenti sono simili. Si può però notare che il forte *gap* si attesta per i livelli di abilità alti, ai quali gli studenti stranieri non riescono ad arrivare. Si può quindi affermare che, a parità di livello di abilità, uno studente italiano e uno straniero hanno circa la stessa probabilità di rispondere correttamente a questa domanda, ma sono pochi gli studenti stranieri che riescono a raggiungere livelli molto alti e per questo motivo si ha un forte *gap* in termini di percentuale di risposte corrette.

Il *gap* si ripresenta ampio anche in casi di domande tendenzialmente più semplici, come la D13, caratterizzata dall’uso di subordinate implicite al gerundio. Anche in questi casi sembra avere un ruolo determinante nella risoluzione del quesito la componente testuale, segno che anche in matematica il linguaggio è una fonte di maggiore difficoltà per gli studenti stranieri.

Non è un caso che le domande in cui il divario tra studenti italiani e stranieri risulta molto lieve siano quelle in cui la componente testuale esercita un ruolo secondario, come la domanda D1, in cui l’indice del *gap* è di soli 5 punti.

D1. Angela guarda lo scontrino del supermercato e si accorge che una macchia ha coperto il prezzo del detersivo.

	EURO
PASTA	2,50
DETERSIVO	
FRAGOLE	5,20
TOTALE EURO	**9,80**

Quanto è costato il detersivo?

☐ A. 1,10 euro
☐ B. 2,10 euro
☐ C. 2,70 euro
☐ D. 3,10 euro

La domanda è generalmente molto semplice e le percentuali di risposta corretta superiori all’80 % per entrambi i gruppi di studenti. Si può osservare che in questo caso la componente testuale, già ridotta, non è nemmeno strettamente necessaria per intuire quale sia la richiesta del quesito. Inoltre, il contesto reale a cui si riferisce la domanda ha probabilmente facilitato la comprensione.

3. Prospettive di ricerca

Il tipo di analisi proposta in questo contributo ha consentito di passare in rassegna tutte le domande delle prove di una coorte esemplificativa di studenti per individuare l’entità del *gap* e capire in quali livelli di abilità gli studenti stranieri hanno ottenuto risultati inferiori rispetto ai colleghi italiani. Ciò ha reso possibile evidenziare particolari livelli di abilità in cui il *gap* è massimo o minimo, mostrare se il *gap* è dovuto alla scelta di un particolare distrattore o se è dovuto a una percentuale maggiore di risposte mancanti e, infine, confrontare l’andamento delle risposte corrette e dei distrattori di italiani e stranieri in base al medesimo livello di abilità.

Nella prospettiva di ampliare e sistematizzare le osservazioni qui proposte, in successivi lavori sarà possibile procedere a un’analisi sistematica delle risposte ai quesiti delle prove di tutte le coorti disponibili, per trarre indicazioni ancor più generali circa la natura delle domande che creano *gap* tra italiani e stranieri e verificare in modo più approfondito se vi sono stati cambiamenti nel tempo.

Un'altra interessante prospettiva di ricerca riguarda la possibilità di studiare con strumenti statistici e cognitivi più raffinati il rapporto tra i risultati ottenuti dagli studenti stranieri e italiani in matematica e in italiano. Anche se i dati e gli esempi qui presentati hanno consentito di rilevare differenze di comportamento tra italiano e matematica, un'analisi incrociata dei risultati ottenuti nelle due prove consentirebbe di studiare la correlazione tra il punteggio in comprensione del testo degli studenti e le risposte ai singoli *item* di matematica. Ciò renderebbe possibile capire quali quesiti di matematica sono maggiormente influenzati dalle difficoltà nella comprensione del testo e di verificare se si tratta degli stessi *item* che evidenziano un maggiore *gap* tra studenti italiani e stranieri, come ci si può attendere se si tiene presente che la comprensione del testo svolge un ruolo importante per la risoluzione dei quesiti matematici.

Le evidenze qui proposte e quelle che potranno emergere da un ampliamento della ricerca consentiranno da un lato di fornire elementi di riflessione sul rapporto tra competenze linguistiche e altre competenze, dall'altro di riflettere sugli stimoli utili a una formulazione e concezione linguistica delle prove più attenta alle specifiche esigenze degli studenti non madrelingua.

Il problema di come adeguare le prove standardizzate a un apprendente straniero è di difficile soluzione, specie per l'italiano – in particolare per la prova di lettura – in cui la complessità linguistica è parte integrante della competenza indagata e non può essere ridotta senza perdere la capacità di dar conto di tutti i livelli in gioco. In questo caso, l'aspetto veramente cruciale è che il *gap*, fisiologico nei primi tempi, decresca con l'avanzare dell'esperienza scolastica e l'inserimento nel contesto sociale dello studente straniero: il dato preoccupante rilevato in questo contributo non è tanto che vi sia un *gap*, quanto che la scuola sembri non riuscire a scalfirlo, anche nel lungo periodo.

Diverso, invece, il caso delle prove di matematica. Per quanto sia innegabilmente opportuno che gli studenti esercitino le loro competenze matematiche anche a partire da testi complessi, le prove standardizzate non rappresentano il momento migliore per porre lo studente (straniero in particolare) di fronte a complessità linguistiche che impattano con i risultati, come recenti studi hanno mostrato (Branchetti, Viale, 2015). Per questa ragione l'INVALSI negli ultimi tempi ha messo in atto azioni specifiche per evitare complicazioni linguistiche inutili nel testo dei quesiti di matematica.

Il lavoro sul testo matematico in tutte le sue forme e a diversi livelli di difficoltà deve semmai caratterizzare l'attività didattica quotidiana e rappresentare un terreno di lavoro comune tra insegnante di italiano e di matematica, come sperimentato in contesti di formazione insegnanti (Bolondi, Viale, in stampa).

Al di là di questo, l'analisi a vari livelli del *gap* tra studenti italiani e stranieri non deve essere intesa come una mera fotografia di una situazione alla quale rassegnarsi. La consapevolezza dei tipi di quesiti che creano maggiore difficoltà agli studenti dovrebbe semmai additare la strada per un'efficace azione didattica, tesa a lavorare sui problemi legati ai livelli di competenza più problematici e, forse proprio per questo, più trascurati. In questo senso, i risultati di maggiore interesse e utilità della ricerca proposta riguardano le implicazioni didattiche, per individuare aspetti della competenza linguistica su cui lavorare in classe nelle varie discipline.

L'obiettivo della scuola è portare tutti gli studenti – italiani e nuovi italiani – a elevati livelli di competenza linguistica, cruciali per una piena cittadinanza. Per dare – riprendendo una celebre affermazione di Gianni Rodari in *Grammatica della fantasia* (1973) – «tutti gli usi della parola a tutti […] Non perché tutti siano artisti, ma perché nessuno sia schiavo».

Riferimenti bibliografici

C. Barbaranelli, E. Natali, *I test psicologici. Teorie e modelli psicometrici*, Roma, Carrocci, 2005.

G. Bolondi, M. Viale, *Abilità linguistiche e discipline scientifiche: un'esperienza di formazione del corpo insegnante nel Polo dell'Emilia-Romagna del progetto «I Lincei per una nuova didattica nella scuola»*, in stampa negli atti XVIII Convegno nazionale GISCEL *Educazione linguistica e apprendimento / insegnamento delle discipline matematico-scientifiche* (Roma, 27-29 marzo 2014).

L. Branchetti, M. Viale, *Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico*, in *Quale didattica per l'italiano?*, a cura di

Marcello Ostinelli, Scuola Universitaria Professionale della Svizzera italiana, Dipartimento formazione e apprendimento, Locarno, 2015, pp. 139-148.

M. Castoldi, *Capire le prove INVALSI. Una guida intelligente*, Roma, CarocciFaber, 2014.

C. Giberti, A. Zivelonghi, G. Bolondi, *Gender differences and didactic contract: analysis of two tasks on power properties*, atti del 40th PME – Conference of the International Group for the Psychology of Mathematics Education (Szeged, Hungary, 3-7 agosto 2016).

M. G. Lo Duca, *Le prove di grammatica INVALSI e gli apprendenti immigrati*, «Working Paper», n. 22, 2014, Roma, Istituto nazionale per la valutazione del sistema educativo di istruzione e formazione.

G. Rasch, *Probabilistic models for some intelligence and attainment tests*, Chicago, University of Chicago Press, 1980.

4.4 Differenze di genere

Il numero di donne impegnate in materie STEM (acronimo di: Science, Technology, Engineering and Mathematics) negli ultimi anni risulta essere in aumento; nonostante ciò gli uomini continuano a essere la stragrande maggioranza in questi settori sia come studenti all'università, sia nel mondo del lavoro (Hill, Corbett & St Rose, 2010; OECD, 2015a).

Alcuni fattori che hanno portato a questa disparità di numero tra uomini e donne impegnati nelle materie scientifiche possono essere ricercati, in primo luogo, nella storia del costume e della società. Storicamente, infatti, lo studio della matematica era riservato quasi esclusivamente agli uomini e quasi nessun nome femminile viene ricordato nella storia della matematica prima del secolo scorso (MacKinnon, 1990; Leder & Forgasz, 2008). Questo fatto è da imputare anche alla presenza di fortissimi stereotipi (Leder & Forgasz, 2008) che impedivano alle donne di accedere alla cultura scientifica nei secoli passati: nel 1800, ad esempio, molti studiosi ritenevano che le donne avessero capacità intellettive inferiori agli uomini a causa delle minori dimensioni del cervello (Romanes, 1886).

Dall'inizio del 1900, sono stati fatti grandi passi avanti per permettere alle donne uguali diritti di accesso all'istruzione e al mondo del lavoro e per combattere gli stereotipi legati al genere. Dagli anni '70, in molti paesi sono state introdotte politiche educative finalizzate a incoraggiare le ragazze a studiare le scienze e la matematica e a garantire loro le stesse possibilità degli uomini anche nel mondo del lavoro (Leder & Forgasz, 2008).

Nel settore delle materie STEM la matematica risulta avere un ruolo fondamentale per lo sviluppo e la conoscenza anche delle altre discipline (Hill et al., 2010) e, per questo motivo, molti studi si sono focalizzati proprio sulle differenze di genere in matematica.

Nonostante l'introduzione di politiche educative finalizzate alle pari opportunità, differenze di genere nelle materie STEM sussistono ancora oggi (Hill et al., 2010): la cultura e gli stereotipi che vedono le ragazze "meno brave" dei ragazzi in matematica e nelle scienze in generale continuano a esistere, influendo negativamente sulle reali attitudini degli studenti, ostacolando le studentesse a raggiungere gli stessi risultati dei compagni maschi.

Negli ultimi anni, il tema delle differenze di genere è stato oggetto di numerosi studi condotti da diversi punti di vista. Sono numerose infatti anche le teorie che non rientrano nella ricerca in didattica ma che appartengono piuttosto ad altri settori (Leder, 1992; Byrnes, 2005; Pajares, 2005).

La psicologia anche negli ultimi anni ha sviluppato un ampio settore di ricerca proprio legato alle difficoltà di apprendimento della matematica da parte delle ragazze legando i minori risultati a fattori di tipo affettivo e psicologico come l'ansia e la sicurezza in se stessi (e.g. Primi, Busdraghi, Tomasetto, Morsanyi & Chiesi, 2014; Lindberg, Hyde, Petersen & Linn, 2010).

Nei prossimi paragrafi si presenterà lo stato dell'arte delle ricerche sulle differenze di genere in matematica, saranno prese in considerazione soprattutto le ricerche provenienti dal settore della didattica della matematica che saranno però affiancate dai principali studi sul tema provenienti da altri settori.

4.4.1 Studio delle differenze di genere attraverso le prove standardizzate

Lo studio delle differenze di genere in matematica è un tema che è stato ampiamente discusso anche in passato e che, negli ultimi anni, ha avuto un notevole impulso grazie anche al crescente impiego di prove standardizzate per valutare gli apprendimenti, sia a livello nazionale, sia a livello internazionale.

Leder e Forgasz nel 2008 forniscono una panoramica degli studi relativi alle differenze di genere nell'ambito della didattica della matematica e sottolineano l'importanza dell'uso delle prove standardizzate in questa direzione:

The results of large scale international testings, including the Programme for International Student Assessment (PISA), have attracted widespread attention from the general mathematics education research community as well as from those with a particular interest in gender differences in mathematics learning. (Leder & Forgasz, 2008)

Dal 2008 ad oggi, l'uso delle prove PISA per la ricerca in didattica della matematica è aumentato anche se le potenzialità che tali prove potrebbero avere, non sono ancora pienamente sfruttate (Maffia & Giberti, 2016). L'analisi delle differenze di genere in matematica risulta essere uno dei temi per i quali si fa maggiormente uso delle prove standardizzate, sia a livello generale sia focalizzando le analisi sugli esiti di una singola nazione (Owens, 2013; Maffia & Giberti, 2016).

I risultati delle ultime rilevazioni internazionali PISA 2015 e TIMSS 2015 confermano quanto osservato nelle precedenti relativamente al gender gap: gli studenti maschi ottengono risultati migliori in matematica rispetto alle studentesse in tutti i livelli scolastici e nella maggior parte delle nazioni (OECD, 2016a; Mullis et al., 2016). Dai risultati di entrambi i test, inoltre, emerge che l'Italia è una delle nazioni in cui il gender gap è maggiormente marcato e il divario di genere in matematica viene anche confermato dai test INVALSI effettuati a livello nazionale (INVALSI, 2016a; Di Tommaso et al., 2016).

Risultati PISA 2015

Nella tabella seguente è riportato il gender gap nella prova PISA 2015 di matematica in termini di differenza tra il punteggio medio ottenuto dai quindicenni maschi e quello ottenuto dalle coetanee. In blu scuro sono riportate le differenze statisticamente significative.

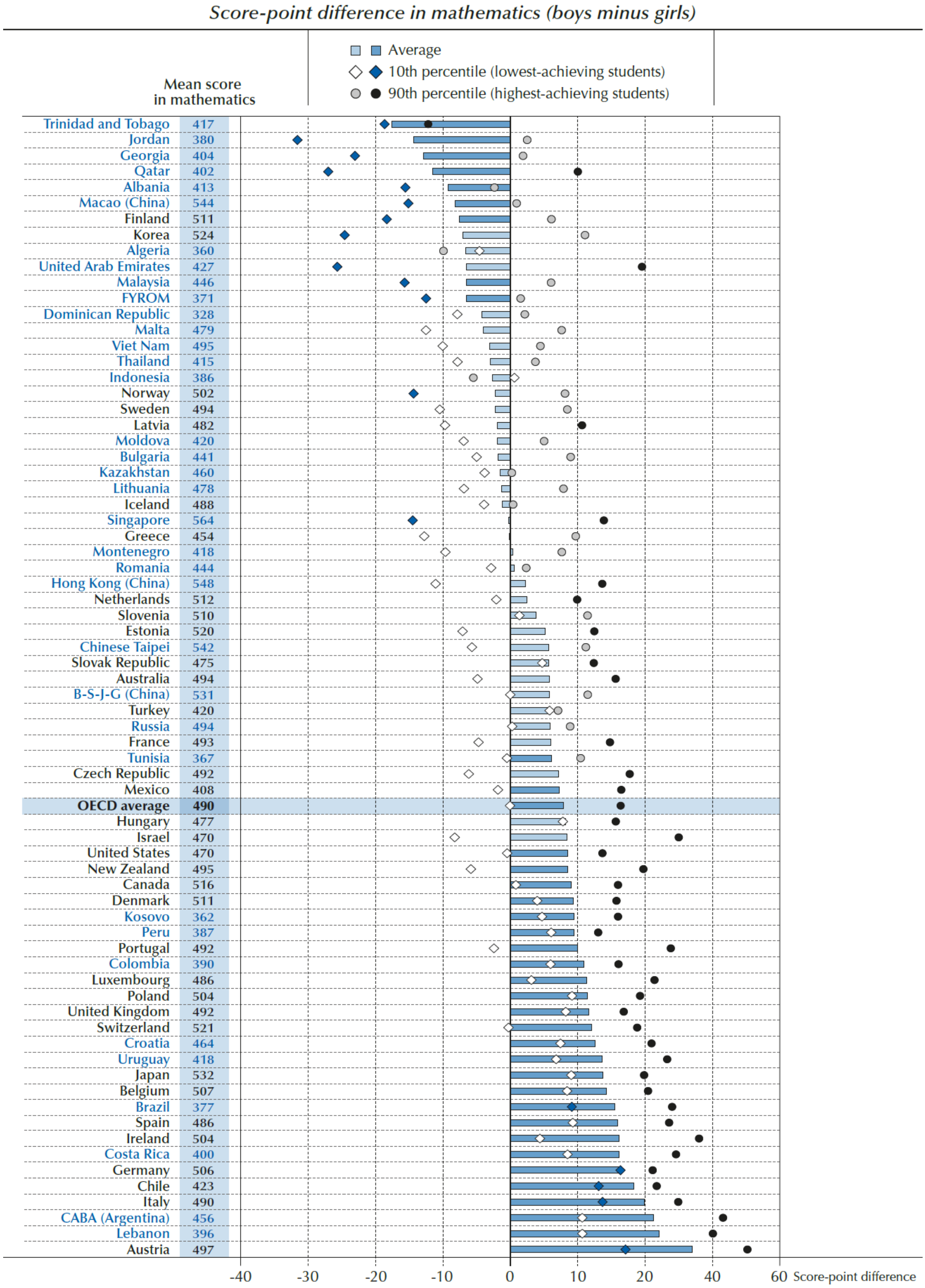


Figura 4.5: Differenze di performance in base al genere nelle prove PISA 2015 di matematica (OECD, 2016a). Fonte: *PISA 2015 Results*, *Excellence and equity in education* (Vol. I)

Si osserva che le differenze di genere in matematica non si manifestano in modo uniforme in tutti i paesi OECD: quasi in metà delle nazioni il gender gap a favore dei maschi è statisticamente significativo, in quasi altrettanti paesi il gap non è statisticamente significativo e addirittura in alcuni paesi i risultati delle ragazze superano quelli dei ragazzi.

Il divario medio all'interno dei paesi OECD è di 8 punti, ma il divario tra i "top-performers" risulta essere raddoppiato: confrontando il 10% migliore dei ragazzi con il 10% migliore delle ragazze il gap è di 16 punti (OECD, 2016a).

I risultati degli studenti italiani evidenziano che, pur ottenendo un punteggio medio complessivo pari alla media OECD, le differenze di genere sono molto marcate e pari 20 punti (OECD, 2016b). L'Italia è quindi uno dei paesi OECD in cui le ragazze sono maggiormente svantaggiate in matematica e dal 2012 al 2015 il gap risulta essere invariato se non leggermente aumentato (come si nota nel grafico riportato nella pagina seguente) mentre nella maggior parte dei paesi il gap negli stessi anni è invariato o leggermente diminuito.

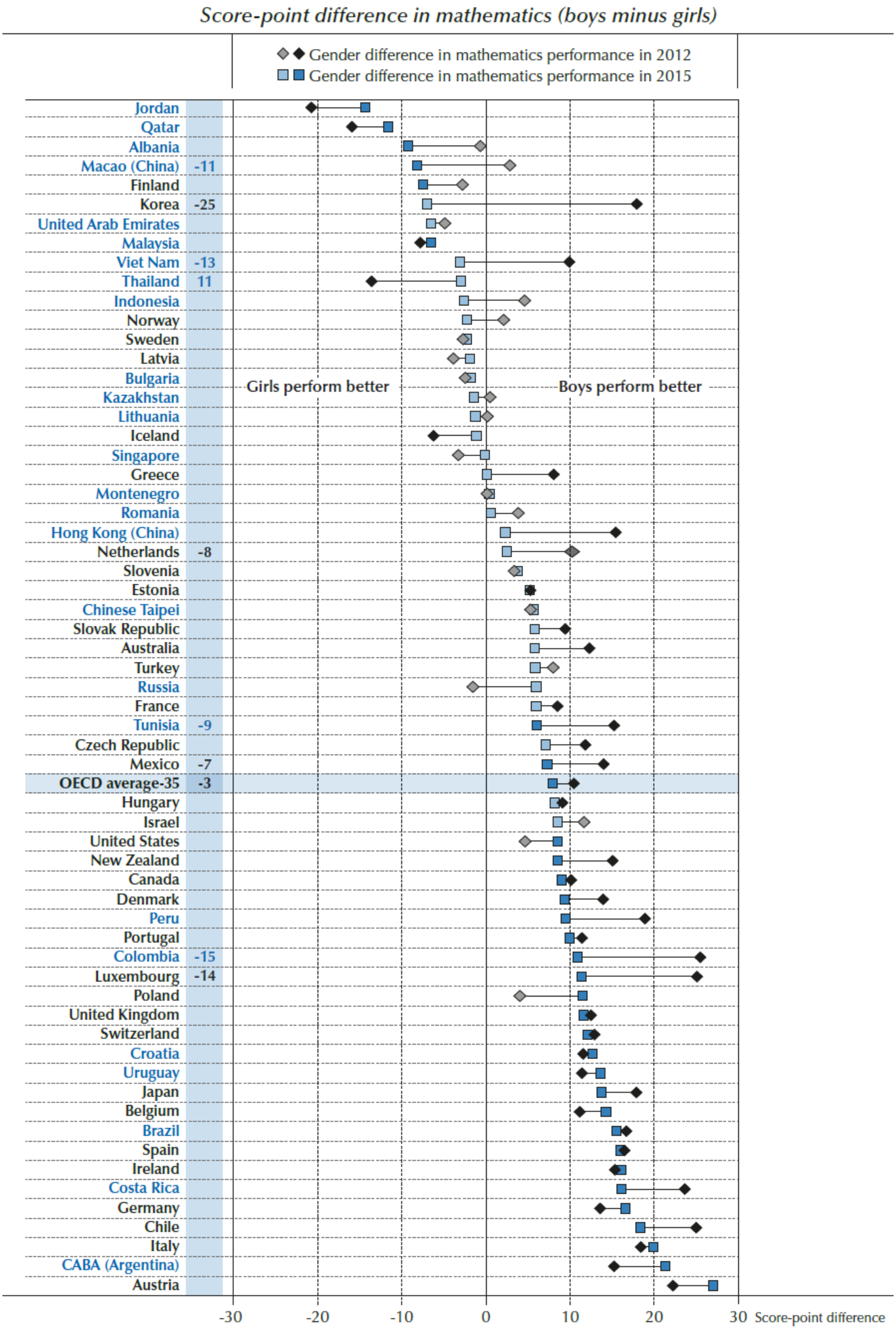


Figura 4.6: Evoluzione del gap di performance in base al genere tra le prove PISA 2012 e PISA 2015 di matematica (OECD, 2016a). Fonte: *PISA 2015 Results*, *Excellence and equity in education* (Vol. I)

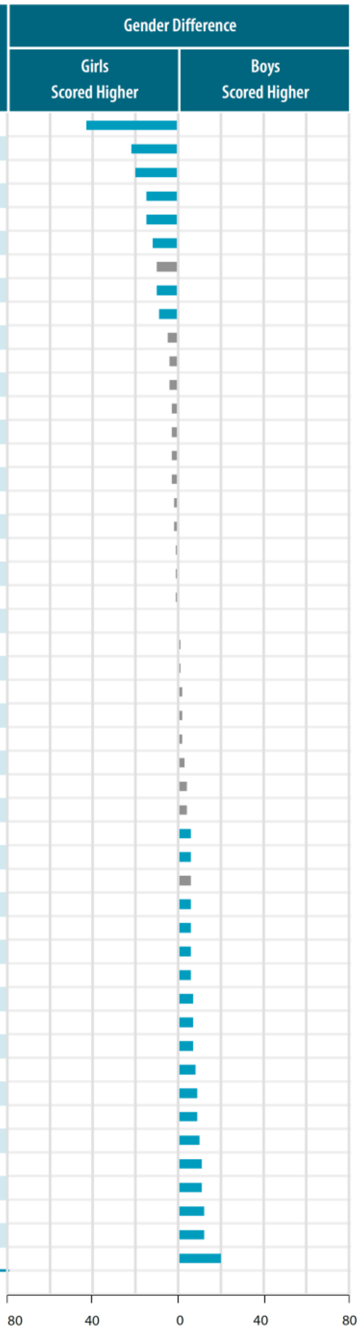
Risultati TIMSS 2015

Le osservazioni appena fatte vengono in buona parte confermate anche dai risultati delle prove TIMSS 2015 che riportano i risultati per i livelli scolastici 4 e 8.

2015 4th Grade

Exhibit 1.10: Average Mathematics Achievement by Gender

Country	Girls		Boys		Difference (Absolute Value)
	Percent of Students	Average Scale Score	Percent of Students	Average Scale Score	
ψ Saudi Arabia	49 (1.0)	405 (4.4)	51 (1.0)	363 (6.5)	43 (7.7)
Oman	50 (0.7)	436 (3.0)	50 (0.7)	415 (2.8)	22 (2.9)
Jordan	46 (2.4)	399 (3.3)	54 (2.4)	379 (4.9)	20 (5.8)
South Africa (5)	48 (0.8)	384 (3.8)	52 (0.8)	368 (4.4)	15 (4.2)
2 Bahrain	50 (0.7)	459 (1.7)	50 (0.7)	443 (2.3)	15 (2.5)
ψ Kuwait	51 (2.0)	359 (5.4)	49 (2.0)	347 (5.6)	12 (6.2)
Iran, Islamic Rep. of	50 (0.9)	437 (4.5)	50 (0.9)	426 (4.5)	10 (6.3)
Indonesia	48 (0.6)	403 (4.0)	52 (0.6)	393 (3.9)	10 (2.7)
Finland	48 (0.8)	540 (2.3)	52 (0.8)	531 (2.6)	9 (2.9)
Bulgaria	49 (0.8)	527 (5.7)	51 (0.8)	522 (5.1)	5 (2.9)
Norway (5)	49 (0.9)	551 (2.6)	51 (0.9)	547 (3.1)	4 (2.9)
2 Singapore	48 (0.5)	620 (3.9)	52 (0.5)	616 (4.3)	4 (3.0)
United Arab Emirates	48 (2.2)	453 (3.9)	52 (2.2)	450 (3.4)	3 (5.4)
1 Georgia	49 (0.9)	465 (3.9)	51 (0.9)	461 (4.4)	3 (4.0)
3 Serbia	48 (0.8)	520 (3.7)	52 (0.8)	517 (4.7)	3 (4.7)
Qatar	51 (2.5)	440 (4.1)	49 (2.5)	438 (4.9)	3 (5.9)
2 Lithuania	50 (0.9)	537 (2.8)	50 (0.9)	534 (3.1)	2 (3.3)
Kazakhstan	49 (0.8)	546 (4.6)	51 (0.8)	543 (4.8)	2 (2.8)
Morocco	48 (0.7)	378 (3.5)	52 (0.7)	377 (3.9)	1 (2.8)
2 Sweden	49 (1.0)	519 (3.2)	51 (1.0)	518 (3.2)	1 (3.0)
Russian Federation	49 (0.9)	564 (3.7)	51 (0.9)	564 (3.7)	1 (2.8)
Japan	50 (0.5)	593 (2.0)	50 (0.5)	593 (2.5)	0 (2.3)
Chile	49 (1.7)	458 (2.8)	51 (1.7)	459 (3.0)	1 (3.2)
Poland	50 (0.8)	534 (2.3)	50 (0.8)	536 (2.7)	1 (2.5)
Turkey	49 (0.6)	482 (3.2)	51 (0.6)	484 (3.5)	2 (2.7)
‡ Northern Ireland	50 (1.1)	569 (3.8)	50 (1.1)	571 (3.1)	2 (3.8)
New Zealand	49 (0.7)	489 (2.8)	51 (0.7)	492 (2.6)	2 (2.8)
Germany	48 (0.7)	520 (2.4)	52 (0.7)	523 (2.3)	3 (2.3)
Ireland	47 (1.5)	545 (2.6)	53 (1.5)	549 (2.9)	4 (3.4)
Slovenia	49 (0.8)	518 (2.1)	51 (0.8)	522 (2.4)	4 (2.6)
Chinese Taipei	49 (0.6)	594 (2.2)	51 (0.6)	599 (2.3)	6 (2.5)
† Belgium (Flemish)	50 (0.9)	543 (2.4)	50 (0.9)	549 (2.4)	6 (2.4)
Hungary	49 (0.9)	526 (3.4)	51 (0.9)	532 (3.8)	6 (3.4)
France	49 (0.7)	485 (3.2)	51 (0.7)	491 (3.2)	6 (2.8)
2 † Denmark	49 (0.8)	536 (3.1)	51 (0.8)	542 (3.0)	6 (2.8)
England	51 (0.7)	543 (3.0)	49 (0.7)	549 (3.3)	6 (2.9)
Cyprus	49 (0.7)	520 (2.9)	51 (0.7)	526 (3.1)	6 (2.7)
2 † United States	51 (0.6)	536 (2.3)	49 (0.6)	543 (2.6)	7 (1.9)
Czech Republic	49 (0.9)	525 (3.0)	51 (0.9)	532 (2.5)	7 (3.2)
Korea, Rep. of	48 (0.5)	604 (2.3)	52 (0.5)	612 (2.5)	7 (1.9)
† Netherlands	50 (0.9)	526 (1.8)	50 (0.9)	534 (2.2)	8 (2.2)
Australia	49 (1.0)	513 (3.1)	51 (1.0)	522 (3.9)	9 (3.5)
1 2 † Canada	49 (0.5)	506 (2.5)	51 (0.5)	515 (2.6)	9 (2.1)
† Hong Kong SAR	46 (1.5)	609 (3.8)	54 (1.5)	619 (2.8)	10 (3.3)
2 Portugal	49 (0.8)	536 (2.4)	51 (0.8)	547 (2.5)	11 (2.2)
Slovak Republic	48 (0.9)	493 (3.0)	52 (0.9)	504 (2.6)	11 (2.6)
2 Spain	49 (0.9)	499 (2.7)	51 (0.9)	511 (2.7)	12 (2.4)
Croatia	49 (0.8)	496 (2.1)	51 (0.8)	508 (2.3)	12 (2.7)
2 Italy	49 (0.7)	497 (2.7)	51 (0.7)	517 (3.0)	20 (2.7)
International Avg.	49 (0.2)	505 (0.5)	51 (0.2)	505 (0.5)	



SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Figura 4.7: Differenze di performance in base al genere nelle prove TIMSS 2015 di matematica – livello 4. Fonte: *TIMSS 2015 International Results in Mathematics* (Mullis et al., 2016).

Exhibit 1.11: Average Mathematics Achievement by Gender

2015 8th Grade

Country	Girls: Percent of Students	Girls: Average Scale Score	Boys: Percent of Students	Boys: Average Scale Score	Difference (Absolute Value)
ψ Oman	48 (1.7)	420 (2.9)	52 (1.7)	388 (3.5)	32 (4.6)
ψ Botswana (9)	51 (0.6)	400 (2.5)	49 (0.6)	381 (2.5)	19 (2.9)
ж Jordan	50 (2.6)	395 (4.0)	50 (2.6)	376 (5.4)	19 (7.0)
Thailand	54 (1.5)	440 (5.2)	46 (1.5)	422 (5.7)	18 (5.5)
Bahrain	48 (0.9)	462 (2.4)	52 (0.9)	446 (2.2)	16 (3.6)
ж Saudi Arabia	51 (1.6)	375 (5.1)	49 (1.6)	360 (7.1)	14 (8.2)
United Arab Emirates	50 (2.5)	471 (3.5)	50 (2.5)	459 (4.0)	12 (6.4)
Malaysia	50 (1.8)	470 (3.8)	50 (1.8)	461 (3.8)	9 (2.8)
2 Singapore	49 (0.6)	626 (3.4)	51 (0.6)	616 (3.8)	9 (3.5)
ψ Egypt	53 (2.3)	397 (5.5)	47 (2.3)	387 (5.1)	9 (6.7)
ж South Africa (9)	51 (1.1)	376 (5.3)	49 (1.1)	369 (4.6)	7 (4.1)
ψ Kuwait	50 (2.5)	396 (4.6)	50 (2.5)	389 (7.1)	7 (7.5)
ψ Qatar	50 (3.0)	440 (3.2)	50 (3.0)	434 (4.5)	7 (4.9)
Turkey	48 (0.8)	461 (4.8)	52 (0.8)	455 (5.3)	6 (3.6)
Kazakhstan	49 (0.9)	531 (5.8)	51 (0.9)	525 (5.3)	6 (3.7)
ψ Iran, Islamic Rep. of	48 (0.9)	438 (5.0)	52 (0.9)	435 (7.5)	3 (8.9)
England	51 (1.6)	520 (5.2)	49 (1.6)	517 (4.8)	3 (5.6)
Malta	49 (0.3)	495 (1.8)	51 (0.3)	492 (1.6)	3 (2.8)
† New Zealand	51 (2.0)	494 (3.2)	49 (2.0)	491 (4.6)	3 (4.2)
Japan	51 (1.0)	588 (3.1)	49 (1.0)	585 (3.0)	2 (4.2)
ж Morocco	46 (0.7)	385 (2.3)	54 (0.7)	384 (2.6)	2 (2.0)
1 2 Georgia	47 (0.9)	454 (3.9)	53 (0.9)	453 (4.0)	1 (4.0)
Chinese Taipei	49 (0.8)	599 (2.6)	51 (0.8)	599 (3.0)	0 (2.8)
Korea, Rep. of	47 (0.5)	605 (2.6)	53 (0.5)	606 (3.1)	1 (2.7)
Norway (9)	50 (0.7)	511 (2.5)	50 (0.7)	512 (2.7)	1 (2.6)
† United States	50 (0.6)	517 (3.3)	50 (0.6)	519 (3.2)	2 (2.0)
Australia	51 (1.6)	504 (3.8)	49 (1.6)	506 (3.5)	2 (4.0)
3 Israel	49 (1.2)	510 (4.3)	51 (1.2)	512 (4.8)	2 (3.9)
Slovenia	48 (0.7)	515 (2.4)	52 (0.7)	518 (2.5)	2 (2.4)
Lebanon	53 (1.6)	441 (3.7)	47 (1.6)	444 (4.5)	3 (3.9)
2 Lithuania	50 (0.8)	510 (3.4)	50 (0.8)	513 (3.1)	3 (3.4)
1 † Canada	51 (1.0)	525 (2.0)	49 (1.0)	530 (2.7)	4 (2.0)
Ireland	50 (1.1)	521 (2.6)	50 (1.1)	526 (4.0)	5 (3.9)
Hong Kong SAR	47 (2.1)	591 (4.7)	53 (2.1)	597 (6.0)	5 (5.7)
2 Italy	49 (0.8)	491 (3.0)	51 (0.8)	498 (2.8)	7 (2.8)
Sweden	48 (1.0)	497 (3.3)	52 (1.0)	504 (3.1)	7 (3.2)
Hungary	50 (0.9)	510 (4.3)	50 (0.9)	519 (4.0)	9 (3.4)
Russian Federation	49 (0.9)	533 (5.1)	51 (0.9)	543 (4.6)	9 (2.9)
ψ Chile	48 (1.8)	418 (3.7)	52 (1.8)	436 (4.2)	18 (4.9)
International Avg.	50 (0.2)	483 (0.6)	50 (0.2)	480 (0.7)	

Gender Difference: Girls Scored Higher / Boys Scored Higher

SOURCE: IEA's Trends in International Mathematics and Science Study – TIMSS 2015

Figura 4.8: Differenze di performance in base al genere nelle prove TIMSS 2015 di matematica – livello 8. Fonte: *TIMSS 2015 International Results in Mathematics* (Mullis et al., 2016).

In entrambi i livelli infatti il gender gap non risulta uniformemente distribuito sulle diverse nazioni.

Le nazioni partecipanti nel 2015 per il livello 4 sono 49 e in 18 di queste è presente una differenza di genere statisticamente significativa in matematica a favore dei maschi, in 23 paesi non si ha alcuna differenza significativa e solo in 8 le femmine ottengono risultati migliori. Nella prova del livello 8, invece, le differenze di genere non risultano così marcate e in 26 nazioni su 39 non si ha alcun divario statisticamente significativo.

Risulta interessante però notare come, anche in queste prove, la situazione dell'Italia sia ben delineata. I risultati medi complessivi degli studenti italiani risultano in entrambe le prove essere vicini alla media dei paesi coinvolti nelle prove TIMSS, ma si può notare che sia al livello 8 e in particolar modo nella prova di livello 4, l'Italia mostra un forte divario di genere, uno dei più marcati in assoluto.

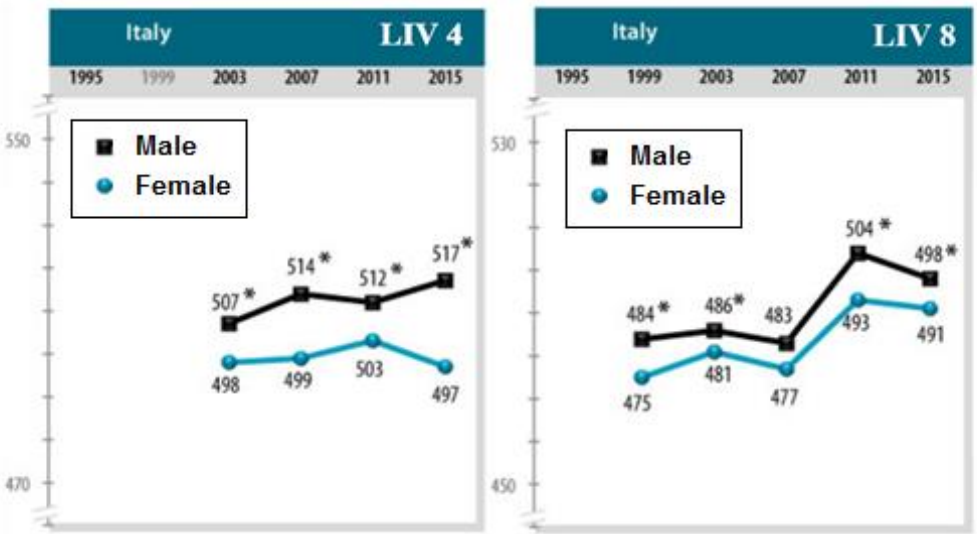


Figura 4.9: Evoluzione del gap di performance in base al genere nelle prove TIMSS di matematica dal 2003 al 2015. Fonte: *TIMSS 2015 International Results in Mathematics* (Mullis et al., 2016).

Dai grafici riportati si osserva inoltre che, se al livello 8 il gap di genere risulta tendenzialmente costante negli anni, questo non accade per il livello 4 in cui notiamo un incremento notevole del divario causato da un miglioramento delle performance dei ragazzi contro un peggioramento delle performance delle ragazze.

Dalle rilevazioni internazionali emergono quindi importanti informazioni utili per analizzare le caratteristiche delle differenze di genere in matematica. Molti studi infatti fanno uso dei dati estratti dalle prove PISA e TIMSS, ad esempio confrontandoli con i risultati di rilevazioni a carattere nazionale. Le prove standardizzate offrono la possibilità di approfondire questo tema su grandi popolazioni di studenti e analizzarne alcuni aspetti anche da un punto di vista quantitativo: ad esempio confrontando il gap tra diverse nazioni oppure studiando il gender gap in funzione del background sociale, economico e culturale.

4.4.2 Stato dell'arte su gender gap

Lo studio delle differenze di genere in matematica è un tema molto dibattuto negli ultimi decenni e sono numerose le ricerche che hanno cercato di indagare le caratteristiche e le cause del gender gap in matematica a favore dei maschi (Leder & Forgasz, 2008; Forgasz, 2010; Forgasz & Rivera, 2012; Di Tommaso et al., 2016).

Come si è potuto notare anche dall'analisi dei risultati PISA e TIMSS fin ora presentati, il gender gap in matematica è un problema estremamente complesso. Sono molteplici infatti i fattori che ne influenzano lo sviluppo e, inoltre, si tratta di un fenomeno che risente notevolmente delle caratteristiche degli studenti, del tipo di scuola e delle peculiarità della famiglia e della società.

4.4.2.1 Evoluzione del gender gap negli anni

In base a quanto riportano diverse ricerche, le differenze di genere in matematica non sono già presenti in età pre-scolare ma emergono durante i primi anni di scuola. Al termine della scuola primaria il gap nelle performance di maschi e femmine è evidente e continua ad aumentare alle scuole superiori (Fryer & Levitt, 2010; Hyde & Mertz, 2009; Hyde, Fennema, & Lamon, 1990; Spelke, 2005).

Fryer e Levitt (2010) hanno analizzato le performance in matematica di bambini e bambine seguendoli dalla scuola materna fino al termine della nostra scuola primaria (livello 5). In questo modo hanno evidenziato che il gap inizia nei primi anni di scuola e aumenta con il passare degli anni:

> *The results we obtain using these new data are informative and, in some cases, quite surprising. Consistent with the prior literature, when children enter kindergarten, girls and boys are observationally equivalent in both math and reading. By the end of fifth grade, however, girls have fallen more than 0.2 standard deviations behind their male counterparts in math. The math gap is equivalent to 2.5 months of schooling. Girls are losing ground in math in every region of the country, every racial group, all levels of the socio-economic distribution, every family structure, and in both public and private schools. By the end of the sample, girls do significantly worse than boys on every math skill tested.* (Fryer & Levitt, 2010)

Diversi degli studi sopra citati, tra cui anche quest'ultimo, si sono sviluppati negli Stati Uniti e sono basati sui test standardizzati che ogni stato deve svolgere ogni anno dall'introduzione della legge "No children Left Behind" del 2001. Negli Stati Uniti in più di 15 anni è stata così raccolta una enorme quantità di informazioni riguardanti le performance cognitive degli studenti in diversi livelli scolastici (dal livello 2 al 5) che hanno permesso di sviluppare numerosi studi longitudinali anche sull'apprendimento della matematica. Ad oggi, dopo diversi anni dalla loro introduzione, anche le prove PISA e le prove nazionali INVALSI possono essere utilizzate per studi simili. In un recente studio basato sulle prove INVALSI, Di Tommaso et al. (2016) osservano l'evoluzione del gender gap in matematica anche nel contesto italiano e notano un complessivo ampliamento del divario a favore dei maschi dal livello 2 al livello 10; l'aumento del gap risulta però essere concentrato negli anni della scuola primaria e nel passaggio dal livello 8 al livello 10, rimanendo pressoché stabile negli anni della scuola secondaria di primo grado.

4.4.2.2 Distribuzione del gender gap rispetto ai livelli di abilità

Le prove standardizzate e l'uso di modelli statistici per l'analisi dei dati permette anche di mettere in evidenza facilmente se un gap è particolarmente marcato per determinati livelli di abilità degli studenti. Nel caso delle differenze di genere risulta confermato da molti studi il fatto che il gap sia maggiore per gli studenti top-performers, ovvero quelli che sull'intera prova ottengono i risultati migliori.

Infatti, sempre dallo studio di Fryer e Levitt (2010), emerge che il divario tra maschi e femmine a favore dei primi risulta essere evidente in termini di punteggio medio ottenuto nella prova e soprattutto se si confrontano i risultati di studenti con alti livelli di abilità in matematica. Se alla scuola materna le bambine occupano il 45% dei 5 percentili più alti, dopo cinque anni di scuola solo il 28% dei 5 percentili più alti è formato da femmine.

Anche dall'analisi dei dati PISA 2009, emerge che nella maggior parte dei paesi partecipanti vi è un significativo gender gap in matematica a favore dei ragazzi ed è più marcato proprio tra gli studenti con alti livelli di abilità (González de San Román & De La Rica, 2012). Questo dato viene confermato anche dai risultati dei test PISA 2012 e PISA 2015: pur osservando una notevole variabilità tra le nazioni coinvolte nell'indagine, emerge come le ragazze risultino sottorappresentate tra gli "highest-achieving students" (OECD, 2016a; OECD, 2012b). Nel 2015 il gap medio tra le performance di maschi e femmine nei paesi OECD è di

8 punti ma, considerando solamente il 10% migliore dei maschi e il 10% migliore delle ragazze il divario raddoppia (OECD, 2016a).
Dalle rilevazioni PISA, l'Italia è uno dei paesi in cui questo fenomeno risulta più evidente e nel 2015 solo l'8% delle ragazze raggiunge il livello maggiore nella scala PISA contro il 13% dei ragazzi (OECD, 2016a). Di Tommaso et al. (2016) confermano questi risultati anche analizzando i dati INVALSI e, attraverso uno studio longitudinale delle prove dal livello 2 al livello 10, notano che al livello 2 il gap di genere non sussiste per i livelli più bassi ed è comunque poco marcato anche per i livelli medi e alti; con il passare degli anni questo gap compare anche per i livelli più bassi della distribuzione ma la crescita risulta molto più marcata per i livelli di abilità alti.

I risultati di molte ricerche svolte negli ultimi decenni e delle principali prove standardizzate nazionali e internazionali mostrano la presenza di forti differenze di genere in matematica e mettono in luce le caratteristiche principali di questo gap.

Nei prossimi paragrafi, analizzando i principali fattori che determinano il gender gap, si noterà quanto il fenomeno sia legato alle realtà culturali e sociali in cui si sviluppa e, per questo motivo, in alcuni paesi risulti molto più marcato. La non omogeneità della distribuzione geografica del gender gap risulta essere uno dei principali motivi per cui alcuni studi non rilevano un gap statisticamente significativo. Inoltre, l'evoluzione del gap non è sempre la stessa analizzando i dati relativi alle prove PISA e TIMSS in quanto gli stati coinvolti sono diversi e questo chiaramente influisce anche sui risultati complessivi (Winkelmann et al., 2008; Fryer & Levitt, 2010).

La ricerca sul tema del gender gap è molto controverso e un articolo che ha suscitato notevole scalpore è stato quello pubblicato da Hyde, Lindberg, Linn, Ellis, e Williams nel 2008 sulla rivista *Science*. Hyde et al., analizzando i dati di prove standardizzate degli Stati Uniti, concludono che "*In contrast to earlier findings, these very current data provide no evidence of a gender difference favoring males emerging in the high school years*".
Nello stesso anno, Leder e Forgasz (2008) pubblicano un articolo sulla rivista ZDM in cui mettono in luce le possibili cause che hanno portato Hyde a risultati discordanti con la maggior parte delle ricerche nel settore. Tra le motivazioni riportate, la principale è quella per cui le differenze di genere si evidenziano per livelli alti di abilità. Leder e Forgasz spiegano che l'uso, da parte di Hyde, di test formati soprattutto da item di livello cognitivo

medio e basso, può essere la causa per cui non emerge gender gap.

Conoscere le peculiarità di questo fenomeno permette quindi anche una riflessione approfondita su quale sia uno strumento adatto per rilevarlo al meglio e la costruzione dei test dovrebbe sempre tenere conto anche di questi fattori.

4.4.2.3 Fattori che influiscono sul gender-gap

Come si è già osservato nei paragrafi precedenti il tema del gender gap in matematica risulta essere ampiamente discusso anche a causa della complessità del fenomeno stesso che presenta una molteplicità di possibili cause e si manifesta in modo non uniforme.
Anche per quanto riguarda l'individuazione dei fattori che ne sono alla base, la letteratura è molto ampia e dibattuta: si possono trovare cause di natura biologica, sociale, culturale e psicologica che devono essere considerate congiuntamente per comprendere a pieno questo fenomeno.
Sono molteplici gli studi che analizzando le differenze di genere in matematica proprio al fine di individuarne le cause (Forgasz, 2010). Riprendendo i lavori di Leder (1992), Gallagher & Kaufmann (2005), a cui si riferisce anche l'articolo di Winkelmann et al. (2008), è possibile identificare *fattori interni* e *fattori esterni* che possono essere causa delle differenze di genere nell'apprendimento della matematica.

Fattori interni

Tra i fattori interni, che dipendono direttamente dall'individuo, alcuni studi considerano anche differenze di tipo biologico (Baron-Cohen & Wheelwright, 2004; Gallagher & Kaufman, 2004) come, ad esempio, una diversa struttura del cervello che permetterebbe ai maschi di sviluppare meglio alcune abilità legate all'apprendimento della matematica. Alla luce dei risultati delle rilevazioni internazionali, però, questa ipotesi deve essere superata o, quantomeno, affiancata a fattori di altra natura che risultano predominanti. Se le cause di natura biologica fossero il principale motivo delle differenze di genere in matematica, il gap dovrebbe presentarsi in maniera uniforme in tutte le nazioni, mentre, come abbiamo osservato nel paragrafo precedente, il gender gap varia notevolmente da un paese all'altro (Di Tommaso et al., 2016; OECD, 2016a; Hill et al., 2010).

Le differenze da un punto di vista delle principali abilità cognitive non sembrano essere tali da spiegare il gender gap in matematica. Molti studi, infatti, sostengono che non esistano differenze significative in termini di abilità cognitive generali (Ruffing, Wach, Spinath, Brünken, & Karbach, 2015; Halpern, Beninger, & Straight, 2011). Solo nel caso delle abilità visuo-spaziali, alcune ricerche riscontrano migliori risultati da parte degli studenti maschi (Lawton & Hatcher, 2005). Se questo dato fosse confermato si è visto però che tali abilità possono essere facilmente sviluppate attraverso un training mirato e di breve durata: il gap in questo settore potrebbe essere quindi facilmente colmato (Hill et al., 2010).

Tra i fattori strettamente collegati all'individuo che possono concorrere all'emergere delle differenze di genere non troviamo però solamente motivi di natura biologica e cognitiva, giocano un ruolo molto importante fattori psico-sociali legati alle motivazioni, alle convinzioni degli studenti e alla fiducia nelle proprie capacità (Winkelmann et al., 2008).
Già nella scuola primaria, si osserva una minore fiducia nelle proprie abilità in matematica da parte delle femmine e, anche quando i risultati tra i due generi sono simili, le femmine tendono a sottostimare le proprie capacità rispetto ai coetanei maschi (OECD, 2015a; Fredericks and Eccles, 2002; Herbert and Stipek, 2005).
In questa direzione, l'OECD analizza principalmente tre costrutti legati alla matematica: *math self-concept, math self-efficacy* e *math anxiety.*

Le definizioni fornite dall'OECD degli indici utilizzati per analizzare questi costrutti sono le seguenti (OECD, 2013b):

- *Mathematics self-efficacy:* index based on students' responses about their perceived ability to solve a range of pure and applied mathematics problems
- *Mathematics self-concept:* index based on students' responses about their perceived competence in mathematics
- *Mathematics anxiety*: index based on students' responses about feelings of stress and helplessness when dealing with mathematics

I primi due costrutti rientrano all'interno dei *self-beliefs* e sono strettamente interconnessi pur riflettendo aspetti diversi riguardanti la sfera delle convinzioni, delle motivazioni e delle emozioni degli studenti in matematica. Infatti, la *self-efficacy* riguarda le sensazioni e le convinzioni degli studenti nel momento in cui devono risolvere uno specifico quesito (ad esempio risolvere un'equazione) mentre nel caso del *math self-concept* le domande rivolte agli studenti sono più generali e legate alla disciplina nel suo complesso, non a uno specifico task (Pajares & Miller, 1994).

Il ruolo della *self-efficacy* e del *self-concept* risulta essere fondamentale nell'apprendimento di qualsiasi disciplina e in particolare nel caso della matematica (OECD, 2015a; Mash & O'Mara, 2008): uno studente con scarsa fiducia nei propri mezzi, infatti, sarà meno motivato davanti a un compito e, nel caso in cui si trovi in difficoltà, sarà meno propenso a perseverare per raggiungere l'obiettivo. Dall'indagine OECD-PISA 2012 emerge che, anche nella scuola secondaria, permane una notevole differenza tra ragazzi e ragazze sia in termini di *self-efficacy* sia in termini di *self-concept* in matematica.

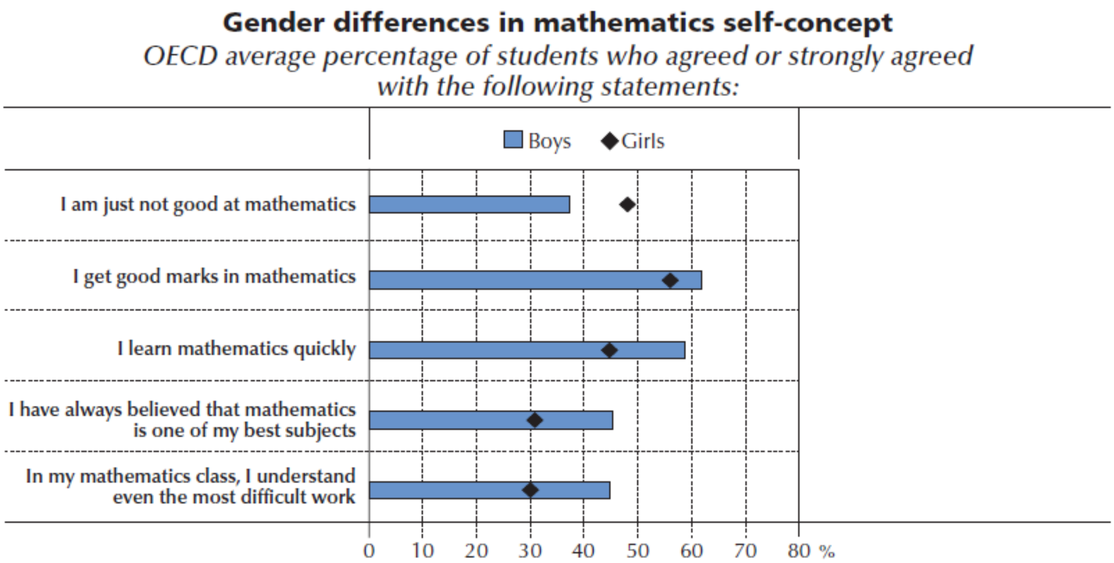


Figura 4.10: *Self-concept* in matematica in relazione al genere, dati relativi all'indagine OECD-PISA 2012. Le differenze statisticamente significative sono evidenziate con colori più scuri. Fonte: *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. (OECD, 2015a).

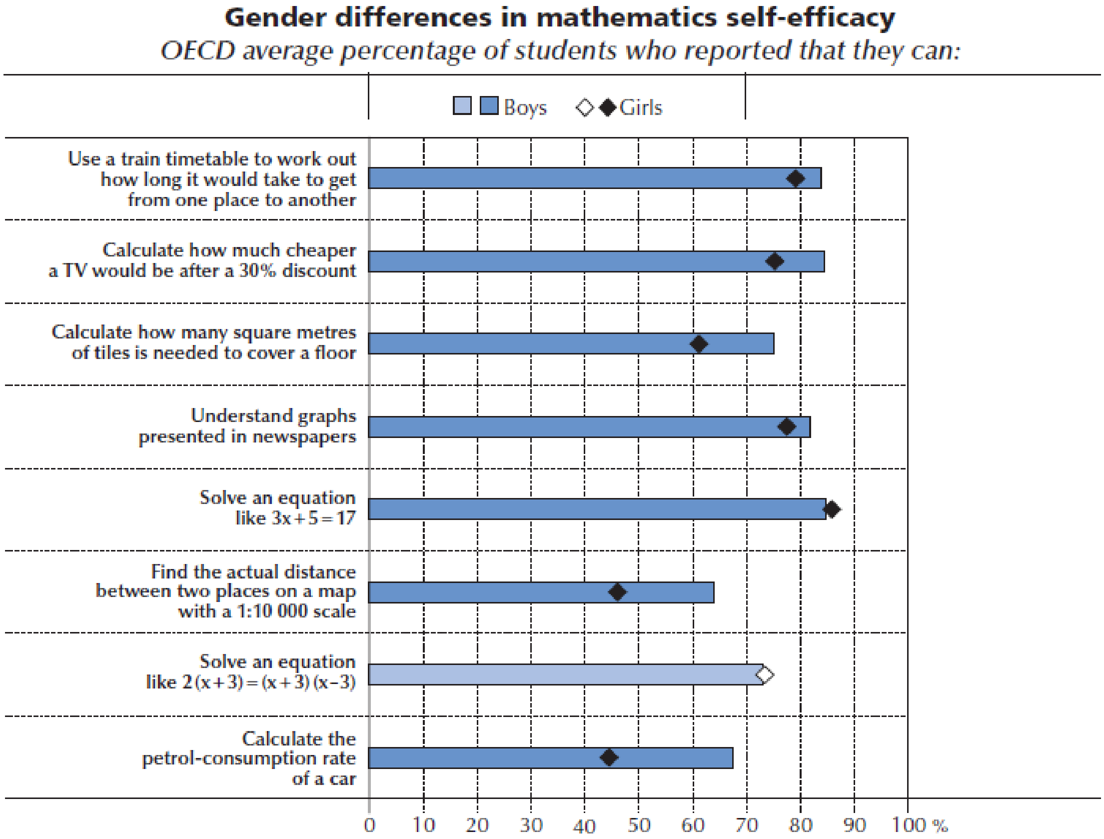


Figura 4.11: *Self-efficacy* in matematica in relazione al genere, dati relativi all'indagine OECD-PISA 2012. Le differenze statisticamente significative sono evidenziate con colori più scuri. Fonte: *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. (OECD, 2015a).

Come si osserva nei grafici sopra relativi all'indagine PISA 2012, in generale le ragazze mostrano livelli minori di fiducia nelle proprie capacità in matematica.

Le studentesse si considerano meno brave in matematica rispetto agli studenti maschi, pensano di non essere veloci nell'apprendimento della disciplina e dicono di trovarsi in difficoltà specialmente davanti agli argomenti più complessi (fig. 4.10). Sempre dal primo grafico, relativo al *math self-concept*, si evince che le risposte di maschi e femmine sono quantitativamente più vicine solo nel caso in cui si chiede di rispondere basandosi su un riferimento esterno, come i voti scolastici. Dalle indagini OECD si evince inoltre che la differenza tra maschi e femmine relativamente al *math self-concept* è presente anche a parità di abilità in matematica o a parità di risultati nella prova e questi risultati sono in accordo con letteratura precedente relativa a *self beliefs* e gender gap in matematica (OECD, 2015a; Jacobs, Lanza, Osgood, Eccles & Wigfield, 2002; Hill et al., 2010).

Per quanto riguarda la *math self-efficacy* è interessante notare (fig. 4.11) che gli unici due task in cui la differenza non è a favore dei maschi risultano essere quelli in cui si richiede di risolvere un'equazione: si tratta quindi in questo caso di esercizi risolvibili attraverso l'applicazione di procedure di routine, già svolte in classe che, a quanto pare, permettono alle ragazze di avere una maggiore fiducia nelle proprie capacità di risoluzione.

Un altro costrutto molto importante per analizzare i diversi risultati in matematica di maschi e femmine è l'ansia matematica (math anxiety), definita come "*an adverse emotional reaction to math or the prospect of doing math*" (Hembree, 1990). Numerosi studi hanno mostrato quanto l'essere ansiosi, spaventati e tesi nell'affrontare un compito di matematica porti gli studenti ad ottenere risultati inferiori rispetto alle proprie abilità (Hembree, 1990; Ma, 1999; Ma & Kishor, 1997; Dowker, Sarkar & Looi, 2016; Primi et al., 2014; OECD, 2015a).

Dai risultati OCSE relativi alla prova PISA del 2012 emerge una netta differenza nei livelli di ansia tra ragazzi e ragazze (fig. 4.12) confermata anche da numerosi altri studi sia a livello internazionale (Dowker et al., 2016; Hembree, 1990; Devine et al., 2012), sia a livello nazionale (Primi et al., 2014; Cargnelutti et al., 2016).

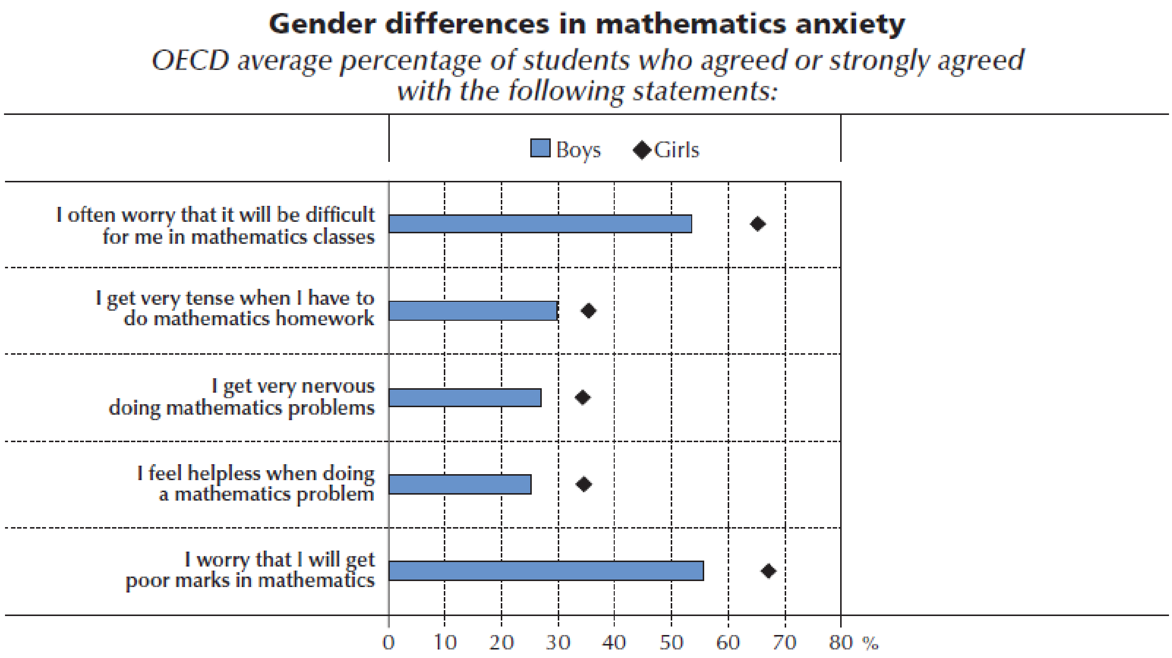


Figura 4.12: *Math anxiety* in relazione al genere, dati relativi all'indagine OECD-PISA 2012. Le differenze statisticamente significative sono evidenziate con colori più scuri. Fonte: *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. (OECD, 2015a).

I fattori appena descritti hanno una notevole influenza sui risultati degli studenti in generale e, come si è potuto osservare, incidono notevolmente anche sulle differenze di genere in matematica. Le ragazze infatti, mostrando una minore fiducia nelle proprie abilità in matematica e una maggiore ansia nell'approccio a questa disciplina, ottengono risultati inferiori e questo anche se si considerano studenti molto bravi (i cosiddetti *high achievers*). Se però si considerano studenti e studentesse con pari livelli di *math anxiety* e *math self-beliefs* allora si può notare che il gap nei risultati dei test di matematica scompare (OECD, 2015a).

Un discorso a parte merita, poi, l'analisi dei risultati in matematica in relazione al modello della *self regulation*, definita come:

> *the ability to control, direct, and plan one's thinking, emotions and behaviours*
> (Schunk & Zimmerman, 1997)

Dal punto di vista della disciplina, del rispetto delle regole, della partecipazione e dell'autocontrollo le ragazze risultano in generale essere migliori dei ragazzi (OECD, 2015a; Matthews, Ponitz, & Morrison, 2009) e questo le favorisce anche nel rendimento scolastico nelle diverse discipline, soprattutto per quanto riguarda le valutazioni date dall'insegnante durante il corso dell'anno. Un recente studio (Weis, Heikamp & Trommsdorff, 2013) mostra, ad esempio, che il controllo delle emozioni e del comportamento (*self-regulation*) sono strettamente legate alle performance scolastiche degli studenti (come già osservato anche da Calkins, 2007 e McClelland & Cameron, 2011) e che il fatto che le ragazze mostrino una migliore *self-regulation* possa essere il motivo per cui in alcuni studi non si riscontra un gender gap significativo in matematica.

Fattori esterni

Sembra quindi che, per spiegare le differenze di genere in matematica, i fattori interni, di natura biologica e psicologica, debbano essere accompagnati da altri fattori, questa volta esterni all'individuo e legati al contesto sociale e culturale in cui lo studente vive.
In questa prospettiva, molti ricercatori hanno evidenziato quanto cause di natura sociale e culturale influiscano sul differente rendimento in matematica di maschi e femmine; in particolare, sono diverse le ricerche che hanno messo in relazione il gender gap in matematica

con i principali indici di equità di genere utilizzati in campo economico e sociale. Queste ricerche hanno mostrato che nelle società in cui vi è una maggiore parità rispetto al genere, il gap tende a scomparire (Guiso et al., 2008; OECD, 2015a; Cascella, 2017).

Uno studio particolarmente significativo in questo senso è quello di Guiso et al. (2008) basato sull'analisi della rilevazione OECD PISA del 2003. Gli autori mettono in relazione il gender gap in matematica e in italiano con uno dei principali indici utilizzati per evidenziare il gap di genere nella società: GGI – World Economic Forum's Gender Gap Index. Questo indice riflette, per ogni nazione, il divario di genere in base alle condizioni economiche, politiche, educative e di salute.

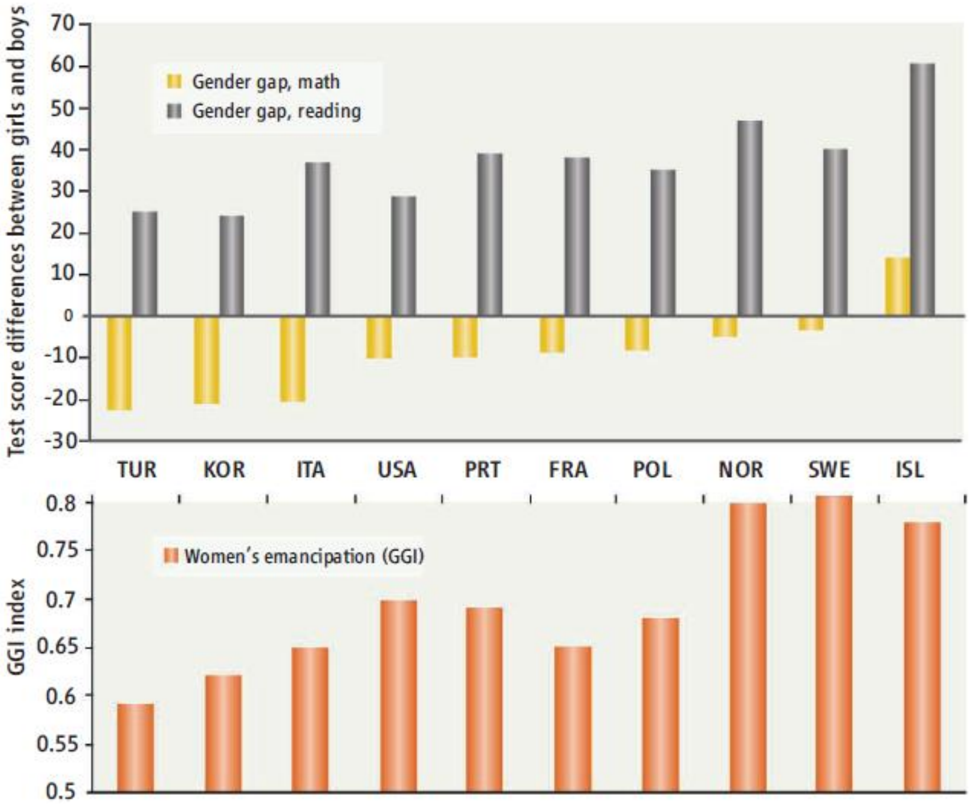


Figura 4.13: Analisi dei risultati del test PISA 2003 di comprensione del testo e matematica in relazione al livello di parità dei generi nella società (Fonte: Guiso et al., 2008).

Dai grafici sopra si può notare che nei paesi in cui il ruolo della donna nella società ha raggiunto alti livelli di emancipazione (GGI maggiore), il gender gap in matematica tende ad annullarsi e il gap nella comprensione del testo a favore delle ragazze incrementa. Questi risultati vengono confermati, nello stesso articolo, anche attraverso l'uso di altri indici statistici basati anche su fattori di natura culturale.

Dal 2006 al 2016, il Global Gender Gap Report mostra come l'Italia si trovi in una situazione piuttosto critica in fatto di parità di generi: nel 2016 l'Italia è risultata 50 esima su 144 paesi, risultato di gran lunga peggiore rispetto alla maggior parte dei paesi europei.

Le analisi di Guiso et al. sono state replicate anche da Fryer and Levitt (2010) e i risultati relativamente all'indagine OECD PISA sono stati confermati. Le stesse analisi, però, sono state ripetute anche utilizzando i risultati dell'indagine IEA TIMSS 2003 e, in questo caso, non si riscontra la stessa correlazione tra la riduzione del gender gap in matematica e l'equità di genere nella società. Fryer and Levitt spiegano che queste differenze sono legate al fatto che sono diverse le nazioni che partecipano alle due indagini: restringendo i risultati del TIMSS alle nazioni che partecipano anche all'indagine PISA, la correlazione tra GGI e la diminuzione del divario in matematica emerge nuovamente. Le nazioni che partecipano esclusivamente alle prove TIMSS sono principalmente paesi del Medio Oriente, in cui, stranamente, il ruolo della donna risulta essere molto svantaggiato nella società ma i risultati in matematica non sembra che ne risentano (Fryer and Levitt, 2010).

Questi studi mostrano quanto sia fondamentale il ruolo della cultura e della società nell'educazione degli studenti e quanto raggiungere una reale parità sociale dei generi influisca positivamente sulle possibilità delle generazioni future. Alcune ricerche si sono occupate delle conseguenze economiche e sociali causate dalle disparità di genere nell'istruzione e, ad esempio, hanno riscontrato una minore crescita economica nei paesi in cui le disuguaglianze sono maggiori (Klasen, 2002). Inoltre, politiche educative che mirano al raggiungimento di pari opportunità dei generi nel campo dell'istruzione vanno a vantaggio delle generazioni future in quanto si è visto che se le donne e le madri raggiungono ruoli di maggiore emancipazione nella società, questo influisce positivamente sull'istruzione e sulla salute dei figli (Schultz, 2002; Doepke, Tertilt, & Voena, 2011; Farré & Vella, 2007).

L'apprendimento della matematica e la disparità nei risultati in questa disciplina sono sicuramente influenzati dal contesto, dalle caratteristiche sociali-economiche e culturali dei singoli paesi e dal ruolo della donna all'interno della società. Esistono anche studi basati su indagini più recenti che hanno confermato questi risultati e indagato approfonditamente il rapporto tra emancipazione della donna e gender gap in matematica.
Nella ricerca di González de San Román e De La Rica (2012), attraverso l'analisi dei risultati

PISA 2009, viene confermata la forte influenza delle norme sociali e culturali del paese sulle differenze di genere riscontrate nei test:

> *We find that differences in culture and social norms across countries and across regions within the same country are crucial determinants in understanding gender differences in PISA 2009 test scores: girls perform relatively better in both maths and reading in societies where gender equality is enhanced, and the effect varies over the distribution of scores.*

Si considerano rilevanti, nel determinare le differenze di genere osservate, anche altri fattori legati maggiormente alle convinzioni e alla sfera psico-sociale ma di natura esterna all'individuo. Giocano infatti un ruolo fondamentale anche le convinzioni di insegnanti e genitori riguardo alle diverse abilità di maschi e femmine in matematica e gli stereotipi che vedono i ragazzi come più portati per le materie scientifiche mentre le ragazze favorite nelle materie letterarie. Si è riscontrato che tutto ciò incide notevolmente sulla percezione degli studenti riguardo alle proprie possibilità e influisce quindi sulle loro reali performance (Jacobs & Bleeker, 2004; Riegle-Crumb, 2005; Fryer & Levitt, 2010). Sono molteplici gli studi che sottolineano l'importanza del ruolo che la madre assume all'interno della famiglia e della società e come questo influenzi le performance in matematica dei figli e in particolare delle figlie femmine (Fryer & Levitt, 2010; Jacobs & Eccles, 1992b; González de San Román & De La Rica, 2012).

Gli stereotipi legati al genere che affondano le loro radici nella cultura del nostro paese e di molti altri, portano avanti l'idea che i maschi siano "naturalmente" più portati per le materie scientifiche e influenzano anche il modo in cui i genitori si rivolgono ai figli fin dai primi anni di vita (Jacobs & Bleeker, 2004; Tomasetto, 2013; Tomasetto, Mirisola, Galdi & Cadinu, 2015). Questi hanno comportamenti e aspettative diversi nei confronti di un figlio maschio o di una figlia femmina e una percezione diversa relativamente alle loro abilità e ai successi in matematica (Jacobs & Eccles, 1992a; Tiedemann, 2000; Tomasetto, 2013). Tutto ciò si riflette sulla percezione che ha lo studente delle proprie abilità, andando quindi a sfavorire le ragazze nelle materie scientifiche e in particolare in matematica (Jacobs & Eccles, 1992; Spinath & Spinath, 2005; Tomasetto, 2013).

Risulta molto interessante anche il fatto che siano diverse, in base al sesso del figlio, le motivazioni che i genitori adducono per giustificare un successo o un insuccesso in matematica: i successi dei figli maschi vengono spesso associati ad una predisposizione naturale per la disciplina mentre per le femmine un successo è ritenuto più spesso frutto di impegno e costanza; per quanto riguarda gli insuccessi invece i genitori tendono a motivarli con una mancanza di impegno dei figli maschi e a una scarsa abilità delle figlie femmine (Eccles, Jacobs, & Harold, 1990; Yee & Eccles, 1988; Tomasetto, 2013).

Gli stereotipi non riguardano però solamente i genitori, anche gli insegnanti infatti mostrano di essere influenzati dalle medesime convinzioni e tendono a considerare i maschi con maggiori abilità in matematica rispetto alle ragazze (Helwig, Anderson & Tindal, 2001; Li, 1999).

L'impatto che le convinzioni di insegnanti e genitori hanno, relativamente alle abilità degli studenti, è rilevante e il fatto che le studentesse femmine vengano ritenute meno dotate in matematica rispetto ai coetanei maschi fa sì che esse stesse abbiano meno fiducia nelle proprie abilità e ottengano quindi risultati effettivamente minori (Lindberg et al., 2010).

Infine, anche fattori strettamente collegati al contesto scolastico e alle pratiche didattiche possono fornire una possibile chiave di lettura delle differenze di genere in matematica. In un articolo del 1992, Gilah Leder evidenzia tra le possibili cause delle differenze di genere in matematica anche variabili legate ai curricula specifici della disciplina: gli argomenti trattati, la tipologia di quesiti, i metodi di valutazione e insegnamento hanno un ruolo fondamentale nell'emergere delle differenze di genere in matematica. Studi più recenti hanno confermato questa ipotesi e hanno mostrato che, oltre a variabili connesse ai curricula, anche i metodi di insegnamento e di valutazione, le pratiche didattiche e le norme socio-matematiche che si instaurano nella classe hanno una notevole influenza sul gender gap in matematica (Leder & Forgasz, 2008; OECD, 2016; Giberti, Zivelonghi & Bolondi, 2016).

Leder e Forgasz (2008) forniscono una panoramica delle ricerche riguardanti le differenze di genere e concludono sottolineando proprio l'importanza di tali variabili

> *The evidence is clear that groups of girls continue to be disadvantaged by previously identified dimensions—the mathematics curriculum, classroom practices, and assessment practices—as well as with respect to newer aspects including the adoption of technology into mathematics pedagogy.*

Learning abilities e learning strategies

Come detto, tra i fattori che influenzano il gender gap in matematica ci sono anche variabili legate ai contenuti specifici della matematica. Risulta quindi necessario analizzare il gender gap non solo da un punto di vista dell'intera disciplina, ma declinando l'abilità matematica nelle sue molteplici componenti e considerando i diversi processi cognitivi che gli studenti devono attivare nella risoluzione di un task matematico.

Le studentesse mostrano una maggiore difficoltà rispetto ai compagni maschi nell'approccio ad attività di problem solving, mentre mostrano capacità equivalenti o leggermente superiori in compiti che richiedono principalmente abilità di calcolo (Hyde et al., 1990; Byrnes & Takahira, 1993). Altri studi hanno dimostrato come maschi e femmine abbiano un modo diverso di affrontare attività di problem solving e adottino quindi strategie differenti: le ragazze tendono maggiormente ad ripetere procedure di routine, algoritmi già utilizzati e conosciuti e strategie convenzionali, mentre i ragazzi mostrano meno paura di sbagliare e tendono ad applicare anche nuovi metodi e approcci non convenzionali (Bell & Norwood, 2007; Gallagher De Lisi, Holst, McGillicuddy-De Lisi, Morely & Cahalan, 2000; Gould, 1996; Fennema, Carpenter, Jacobs, Franke & Levi, 1998).

Queste evidenze, oggi confermate da recenti studi anche in ambito psicologico, erano già state presentate nel 1998 da Fennema e Carpenter, attraverso uno studio longitudinale riguardante i primi anni della scuola primaria. I ricercatori non riscontrano particolari differenze di genere se non nel caso di una tipologia di problemi (extension problems) ed esclusivamente a livello 3, al contempo però

> *significant differences in problem-solving strategies were found in all grades. Girls tended to use concrete solution strategies like modeling and counting, and boys tended to use more abstract solution strategies that reflected conceptual understanding.*

Le differenti strategie di problem solving non sembrano strettamente legate a caratteristiche biologiche diverse tra maschi e femmine ma piuttosto a fattori di tipo socio-culturale e legati alla sfera delle convinzioni. Tra questi ritroviamo sicuramente la minore fiducia nelle proprie abilità da parte delle ragazze, stereotipi largamente diffusi per cui, ad esempio, "le brave ragazze seguono le regole" ('good girls follow the rules' - Langer, 1997) e fattori legati alla didattica, al sistema scolastico e alle pratiche d'aula (Boaler, 1997).

In questa direzione risulta particolarmente interessante il confronto con altre ricerche legate alle strategie di apprendimento applicate dagli studenti nello studio delle diverse materie. Non si riscontrano differenze in termini di abilità cognitive (Halpern et al., 2011), si hanno però differenze di genere significative per quanto riguarda le strategie d'apprendimento (Kesici, Sahin & Akturk, 2009; Marrs & Sigler, 2012; Virtanen & Nevgi, 2010).

La letteratura mostra quanto abilità cognitive e strategie d'apprendimento abbiano una forte influenza sulle performance degli studenti; questo è vero anche quando si vuole predire il successo accademico (Rohde & Thompson, 2007; Richardson, Abraham, & Bond, 2012). Un recente articolo "Learning strategies and general cognitive ability as predictors of gender-specific academic achievement" (Ruffing et al., 2015) tratta proprio le differenze di genere e le strategie di apprendimento in funzione del loro impatto sulle performance accademiche. Da questo articolo emerge che non vi sono particolari differenze in termini di abilità cognitive, ma che maschi e femmine mostrano un diverso approccio rispetto alle strategie di apprendimento prese in considerazione. In particolare, vengono considerate le seguenti strategie d'apprendimento:

- *effort*
- *attention*
- *organization*
- *Relationship*
- *Rehearsal*
- *Critical evaluation*
- *Time management*
- *Learning environment*
- *Learning with fellow students*
- *Literature*
- *Meta-cognition*

Fra queste ritroviamo solo due strategie che risultano essere maggiormente utilizzate dai maschi rispetto alle femmine: Reasoning e Critical Evaluation. Le ragazze, invece, mostrano un maggiore uso di molte delle rimanenti strategie: Effort, Organization, Reheasal, Time Management e Meta-Cognition (Ruffing et al., 2015).

Le differenze nelle "learning strategies" possono chiaramente essere alla base di un apprendimento con caratteristiche molto diverse per maschi e femmine. Questi risultati potrebbero essere una possibile spiegazione di alcuni fenomeni osservati nell'apprendimento della matematica: le ragazze, concentrandosi soprattutto sull'impegno, sull'organizzazione dello studio e sulla ripetizione dei concetti, potrebbero essere maggiormente legate alle procedure già viste in classe o durante lo studio a casa ed essere meno pronte a rispondere a stimoli nuovi e che richiedono un apprendimento più basato sulla comprensione profonda dei concetti al fine di una rielaborazione.

In conclusione, la letteratura è concorde e mostra che non è possibile individuare una unica causa del gender gap in matematica, ma che piuttosto vi sono molteplici e differenti fattori da considerare.
All'interno di questa tesi e degli articoli che verranno presentati in seguito, si considerano tali fattori e, in particolare, si mostrerà come risultino importanti non solamente quelli di natura sociale e culturale a livello generale, ma anche fattori micro-sociali, legati alle abitudini del contesto classe, al rapporto con l'insegnante e alle pratiche didattiche.
Le differenze evidenziate in psicologia tra maschi e femmine, come una maggiore ansia matematica e una minore fiducia nelle proprie capacità per le ragazze, accompagnate però a un maggiore controllo e disciplina, possono avere una notevole influenza sull'atteggiamento degli studenti in classe e sul rapporto che si instaura con l'insegnante. È possibile che queste differenze portino le ragazze ad essere in maggior misura dipendenti da ciò che si è svolto in classe e dal giudizio dell'insegnante; questo può condurre a incorrere in errori legati agli effetti del contratto didattico e, più in generale, a faticare a distaccarsi dalla routine d'aula.
L'incidenza di fattori micro-sociali, legati alla didattica e alle pratiche d'aula, è supportata anche dal fatto che, come osservato nei paragrafi precedenti, le differenze di genere emergono durante i primi anni di scuola e aumentano con il passare degli anni.

4.4.3 Differenze di genere nei test INVALSI

L'uso delle prove standardizzate per l'analisi delle differenze di genere è molto diffuso, soprattutto all'estero. In Germania, ad esempio, per monitorare gli apprendimenti in matematica esistono diversi test a livello nazionale che vengono spesso confrontati con i risultati delle prove OECD-PISA.

Le prove INVALSI sono somministrate nelle scuole italiane da quasi un decennio e offrono una enorme mole di dati relativi agli apprendimenti in italiano e in matematica degli studenti dalla scuola primaria alla scuola secondaria di secondo grado. Nonostante ciò, l'uso di queste prove nel campo della ricerca in didattica della matematica risulta ancora limitato (Maffia & Giberti, 2016) e, per quanto riguarda lo studio delle differenze di genere in matematica, questi dati vengono utilizzati quasi esclusivamente sul piano statistico, manca una interpretazione in chiave didattica del fenomeno a livello nazionale.

Lo scopo di questa tesi è mostrare le potenzialità dell'uso dei dati di prove su larga scala per la ricerca in didattica della matematica e, in particolare, buona parte del lavoro è stata proprio finalizzata a indagare le differenze di genere in matematica attraverso l'uso dei dati INVALSI.

Il gender gap viene solitamente presentato in termini di differenza tra il punteggio medio ottenuto dai maschi e quello ottenuto dalle femmine all'interno di una prova standardizzata (OECD-PISA o INVALSI). Le informazioni ricavate in questo modo e presentate nei principali rapporti sugli apprendimenti, nazionali e internazionali, sono sicuramente importanti perché forniscono una indicazione complessiva sulla presenza o meno di un gap di genere che è direttamente collegata alle performance degli studenti nel test.

Per analizzare il fenomeno in maniera più approfondita, per indagarne le cause e poter agire con interventi mirati, risulta però necessario analizzare i singoli item, individuare possibili differenze all'interno dei quesiti e approfondire l'analisi di quelli che creano maggiormente un divario.

Il modo più semplice per individuare un gap su un quesito è considerare la differenza tra la percentuale di risposte corrette dei maschi e la percentuale di risposte corrette delle femmine. Un 10% di differenza su una domanda molto semplice (che quindi ottiene percentuali di risposta molto alte) però dovrebbe avere un peso minore rispetto alla stessa differenza su un quesito più complesso.

Per questo motivo è stato creato un apposito indice (presentato nel paragrafo 4.2.1) che permette di analizzare il gender gap sui singoli item tenendo conto di questo fattore:

$$I_{GG_k} = \frac{M_k - F_k}{P_k} \qquad (4.1)$$

Dove:

- M_k è la percentuale di risposte corrette dei maschi all'item k
- F_k è la percentuale di risposte corrette delle femmine all'item k
- P_k è la percentuale di risposte corrette dell'intera popolazione all'item k.

Si è visto che lo studio delle differenze di genere in matematica risulta molto complesso e, per indagarne le cause, è necessario tenere conto di una notevole varietà di fattori. Una analisi che miri a individuare differenze sulle singole domande, in modo poi da collegarle a specifici apprendimenti, può essere un primo passo per avviare una riflessione mirata attraverso le lenti della didattica.

Le prime domande emerse in questa direzione sono perciò:

- Esistono item in cui il gender gap è più marcato rispetto ad altri?
- Esistono item in cui gender gap è a favore delle femmine?

Se rispondendo a queste domande dovessero emergere quesiti particolarmente significativi in termini di gender gap, allora si avrebbe la possibilità di focalizzare la ricerca proprio su questi item.

L'analisi dei distractor plot ottenuti separando gli studenti in base al genere permette poi di rispondere ad altre domande di ricerca più specifiche:

- Le differenze di genere in un determinato item, sono maggiormente marcate per uno specifico livello di abilità degli studenti?
- Il comportamento di maschi e femmine sui distrattori è il medesimo o si può attribuire il gender gap a una diversa scelta dei distrattori?

- È possibile individuare caratteristiche simili tra le domande che mostrano un maggiore gender gap a favore dei maschi, sia da un punto di vista misuratorio/psicometrico, sia da un punto di vista dei contenuti/didattico?

Le risposte a queste domande dovrebbero chiaramente sempre essere accompagnate da una riflessione e da una analisi dei quesiti attraverso le lenti fornite dalla didattica della matematica.

L'analisi a priori dei quesiti in base ai contenuti richiesti, ai processi cognitivi che lo studente deve attivare nel rispondere, alle motivazioni dei possibili errori, insieme allo studio dei protocolli forniti dagli studenti e a eventuali interviste, possono fornire un quadro molto interessante di particolari differenze di genere in matematica. Queste informazioni sono fondamentali per capire a pieno questo fenomeno e studiare interventi da proporre in classe che garantiscano una maggiore equità nell'educazione matematica.

Al fine di approfondire al meglio questa tematica, le analisi relative al gender gap effettuate durante il dottorato e quelle riportate in questa tesi sono state fatte adoperando la seguente metodologia:

- Analisi a livello dell'intero test (o di più test)
    - Studio del gap sull'intera prova in termini di punteggio medio di Rasch dei due gruppi di studenti
    - Analisi della distribuzione del gap sui singoli item
    - Individuazione degli item maggiormente significativi e di possibili caratteristiche in comune tra tali item (ambito, tipologia, difficoltà, …)
- Analisi a livello di un singolo item
    - Analisi a priori del quesito da un punto di vista didattico (*question intent*, legame con le Indicazioni Nazionali per il curricolo, processi cognitivi da attivare nella risoluzione, possibili errori legati a determinati costrutti didattici, …)
    - Analisi dei risultati complessivi del quesito (percentuali di scelta di ogni opzione, distractor plot, discriminatività, fit, …)
    - Analisi del comportamento differenziale dell'item suddividendo la popolazione in base al genere (percentuali di risposta di maschi e femmine, grafici relativi

all'andamento della risposta corretta in base al genere, distractor plot per genere)

- Eventuali interviste agli studenti per interpretare meglio le motivazioni alla base delle risposte fornite.

Nel prossimo paragrafo saranno riportate le analisi effettuate a livello dell'intero test per quattro prove INVALSI di differenti livelli, mentre nel paragrafo successivo si scenderà nel merito dei singoli item di queste prove, fornendo alcuni esempi di studio del gender gap in particolari quesiti.

4.4.3.1 Analisi delle prove INVALSI

Durante il corso del dottorato sono state analizzate, al fine di evidenziare differenze di genere, quasi tutte le prove INVALSI somministrate dal 2009 al 2015 nei diversi livelli scolastici.

Da tutte le prove di matematica è emersa una differenza di genere a favore dei maschi sull'intero test confermando ciò che è evidenziato anche nei rapporti tecnici dei singoli anni (ad es. INVALSI, 2016a).

Analizzando le differenze di genere sui singoli item, si è potuto osservare che il divario sull'intera prova non è dovuto a una differenza nelle performance distribuita su tutte le domande: vi sono domande che danno un gender gap molto marcato a favore dei maschi, parte delle domande che non mostrano differenze significative e alcune domande in ogni prova evidenziano una migliore performance delle femmine.

In questa tesi si è scelto di riportare le analisi relative al gender gap approfondite su una coorte di studenti che ha eseguito negli anni le prove:

- Livello 2 del 2009 (II primaria)
- Livello 5 del 2012 (V primaria)
- Livello 6 del 2013 (I secondaria di primo grado)
- Livello 8 del 2015 (Prova Nazionale – III secondaria di primo grado)

Nel 2017 la stessa coorte di studenti ha svolto la prova INVALSI al livello 10 (II secondaria di secondo grado) e, una volta resi disponibili i risultati, sarà possibile ampliare lo studio seguendo gli studenti in 5 prove, dalla primaria fino alla secondaria di secondo grado. Gli studenti che costituiscono il campione (quindi i dati analizzati) non sono gli stessi nelle diverse prove, ma per ogni test il campione è rappresentativo della medesima popolazione di studenti e quindi si può considerare questo studio come l'analisi longitudinale di una popolazione di studenti a cui sono state somministrate 5 prove INVALSI all'interno del percorso scolastico.

Applicando il modello di Rasch alle singole prove è possibile stimare un punteggio di abilità per ogni studente (compreso tra -4 e +4). Maschi e femmine possono essere quindi confrontati in base all'abilità complessiva sull'intero test.

I grafici relativi alla distribuzione percentuale di studenti e studentesse in funzione del punteggio di Rasch ottenuto nelle prove di matematica analizzate per la coorte A, sono già riportati nel paragrafo 4.2.1. Dai grafici non si osservano differenze tra maschi e femmine per i livelli di abilità bassi, in tutte le prove però si nota un gap rilevante per i livelli di abilità medio alti e alti e questo è dovuto al fatto che molte ragazze riescono a raggiungere solo punteggi intermedi e non riescono a inserirsi nella fascia degli high-achievers.

Passando quindi a una analisi più approfondita delle prove è stata calcolato l'indice di gender gap (utilizzando la formula 4.1) per tutti gli item dei test analizzati e i risultati sono riportati nelle tabelle seguenti. In ogni tabella sono riportati gli item di un test in ordine decrescente in base al gap, insieme alle principali caratteristiche di ogni quesito.

L02_2009				
Numero item	**INDICE GENDER-GAP (I_{GG})**	**Ambito**	**Delta**	**% risposte corrette**
M10	30%	NUMERI	0.79	33%
M20	18%	NUMERI	-0.55	62%
M14	13%	NUMERI	-0.35	57%
M15	10%	NUMERI	-0.26	55%
M2	9%	NUMERI	-0.19	54%
M3	7%	NUMERI	-0.33	57%
M22	7%	NUMERI	-0.69	64%
M16	7%	NUMERI	1.00	30%
M18	6%	SPAZIO E FIGURE	0.82	33%
M23	6%	NUMERI	-1.09	72%
M13	5%	SPAZIO E FIGURE	-1.38	78%
M17	4%	NUMERI	-1.46	78%
M21	2%	SPAZIO E FIGURE	0.58	38%
M5	1%	SPAZIO E FIGURE	-0.29	56%
M9	1%	NUMERI	0.97	30%
M6b	0%	DATI E PREVISIONI	-1.65	81%
M1	-1%	NUMERI	-1.38	77%
M12	-1%	DATI E PREVISIONI	-1.96	85%
M6a	-1%	DATI E PREVISIONI	-1.93	84%
M11	-2%	DATI E PREVISIONI	0.15	47%
M4	-2%	NUMERI	0.52	39%
M8	-4%	NUMERI	0.27	44%
M19	-9%	SPAZIO E FIGURE	1.48	22%
M7	-9%	NUMERI	-0.20	54%

Tabella 4.2: Item della prova INVALSI di matematica di II primaria del 2008/2009. Gli item sono riportati in ordine decrescente in base all'indice di gender gap. La percentuale di risposte corrette è riferita all'intera popolazione.

L05_2012				
Numero item	**INDICE GENDER-GAP (I_{GG})**	**Ambito**	**Delta**	**% risposte corrette**
D21_a	28%	NUMERI	0.63	37%
D30	27%	NUMERI	0.75	35%
D7_a	26%	NUMERI	0.34	43%
D18	25%	RELAZIONI E FUNZIONI	0.16	47%
D32	24%	NUMERI	0.30	44%
D7_b	23%	NUMERI	-0.01	50%
D8_a	23%	RELAZIONI E FUNZIONI	1.62	21%
D19_c	18%	DATI E PREVISIONI	0.74	35%
D9	18%	DATI E PREVISIONI	0.53	39%
D8_b	16%	RELAZIONI E FUNZIONI	1.88	18%
D4	16%	RELAZIONI E FUNZIONI	0.59	38%
D23	15%	RELAZIONI E FUNZIONI	0.77	35%
D25_b	13%	NUMERI	1.06	30%
D19_d	13%	DATI E PREVISIONI	-0.19	54%
D13	12%	RELAZIONI E FUNZIONI	-0.19	54%
D21_b	11%	NUMERI	-0.88	66%
D19_b	11%	DATI E PREVISIONI	-0.17	53%
D27	10%	SPAZIO E FIGURE	-0.29	55%
D19_a	9%	DATI E PREVISIONI	-0.32	56%
D11	8%	NUMERI	0.44	41%
D24	8%	NUMERI	-0.06	51%
D1	7%	NUMERI	-1.52	77%
D29	7%	RELAZIONI E FUNZIONI	1.35	25%
D12_b	6%	DATI E PREVISIONI	-0.97	68%
D5_a	6%	SPAZIO E FIGURE	-0.36	57%
D22	5%	SPAZIO E FIGURE	0.54	39%
D28	5%	NUMERI	0.66	37%
D33	4%	DATI E PREVISIONI	-0.81	66%
D6	4%	DATI E PREVISIONI	-0.26	55%
D12_a	4%	DATI E PREVISIONI	0.54	39%
D15	4%	NUMERI	-1.58	78%
D31	3%	SPAZIO E FIGURE	-0.47	59%
D3	3%	NUMERI	-1.00	69%
D5_b	3%	SPAZIO E FIGURE	0.34	43%
D16	3%	SPAZIO E FIGURE	0.08	48%
D14	2%	SPAZIO E FIGURE	0.89	67%
D20	1%	DATI E PREVISIONI	-1.48	77%
D17_b	0%	RELAZIONI E FUNZIONI	-0.62	62%
D2	0%	DATI E PREVISIONI	0.89	33%
D17_c	0%	RELAZIONI E FUNZIONI	0.25	45%
D17_a	0%	RELAZIONI E FUNZIONI	-0.31	56%
D26	-2%	SPAZIO E FIGURE	1.08	29%
D10	-2%	SPAZIO E FIGURE	-0.06	51%
D25_a	-5%	NUMERI	-0.07	51%

Tabella 4.3: Item della prova INVALSI di matematica di V primaria del 2011/2012. Gli item sono riportati in ordine decrescente in base all'indice di gender gap. La percentuale di risposte corrette è riferita all'intera popolazione.

L06_2013				
Numero item	**INDICE GENDER-GAP (I_{GG})**	**Ambito**	**Delta**	**% risposte corrette**
D17_b	47%	RELAZIONI E FUNZIONI	1.00	29%
D9_b	32%	RELAZIONI E FUNZIONI	0.70	35%
D12	29%	NUMERI	0.75	34%
D2_c	29%	RELAZIONI E FUNZIONI	2.17	13%
D27	29%	NUMERI	0.69	35%
D11	25%	SPAZIO E FIGURE	2.03	14%
D25_b	25%	NUMERI	-0.18	54%
D26_b	21%	RELAZIONI E FUNZIONI	0.87	32%
D26_a	20%	RELAZIONI E FUNZIONI	0.23	45%
D20_a	18%	SPAZIO E FIGURE	0.12	47%
D18	17%	SPAZIO E FIGURE	1.67	19%
D7_b	16%	NUMERI	0.31	43%
D29	15%	DATI E PREVISIONI	0.40	41%
D7_a	15%	NUMERI	0.91	31%
D21_a	14%	SPAZIO E FIGURE	0.94	31%
D16	14%	NUMERI	1.16	27%
D10_c	13%	DATI E PREVISIONI	1.78	17%
D23	12%	NUMERI	-0.34	57%
D2_b	11%	RELAZIONI E FUNZIONI	0.67	36%
D9_c	9%	RELAZIONI E FUNZIONI	1.24	25%
D4	8%	RELAZIONI E FUNZIONI	-0.98	70%
D17_a	8%	RELAZIONI E FUNZIONI	-1.21	74%
D14	7%	SPAZIO E FIGURE	1.17	26%
D24	5%	RELAZIONI E FUNZIONI	-1.07	72%
D10_a	5%	DATI E PREVISIONI	2.30	12%
D2_a	3%	RELAZIONI E FUNZIONI	0.89	31%
D20_b	3%	SPAZIO E FIGURE	0.88	32%
D6_c	2%	DATI E PREVISIONI	0.83	33%
D1_c1	2%	DATI E PREVISIONI	0.63	37%
D1_a	2%	DATI E PREVISIONI	-1.91	84%
D25_a	2%	NUMERI	-1.93	85%
D30	1%	NUMERI	0.28	44%
D1_b	0%	DATI E PREVISIONI	-1.84	84%
D15	-1%	SPAZIO E FIGURE	0.72	35%
D13	-1%	SPAZIO E FIGURE	-0.20	54%
D3	-1%	NUMERI	0.30	43%
D9_a	-1%	RELAZIONI E FUNZIONI	0.48	39%
D28	-1%	DATI E PREVISIONI	-0.61	63%
D21_b	-2%	RELAZIONI E FUNZIONI	0.21	45%
D6_a	-3%	DATI E PREVISIONI	-0.80	67%
D8_a	-3%	SPAZIO E FIGURE	-0.09	52%
D8_b	-3%	SPAZIO E FIGURE	0.71	35%
D10_b	-4%	DATI E PREVISIONI	-1.58	80%
D5	-5%	SPAZIO E FIGURE	-0.75	66%
D22	-8%	NUMERI	0.27	44%
D6_b	-10%	DATI E PREVISIONI	-0.17	54%
D19	-12%	NUMERI	0.14	47%

Tabella 4.4: Item della prova INVALSI di matematica di I secondaria di 1° grado del 2012/2013. Gli item sono riportati in ordine decrescente in base all'indice di gender gap. La percentuale di risposte corrette è riferita all'intera popolazione.

L08_2015				
Numero item	**INDICE GENDER-GAP (I_{GG})**	**Ambito**	**Delta**	**% risposte corrette**
D17	26%	SPAZIO E FIGURE	2.81	8
D16_b	18%	NUMERI	0.36	42
D18	14%	NUMERI	0.60	38
D3	12%	SPAZIO E FIGURE	-0.75	65
D15_b	12%	RELAZIONI E FUNZIONI	-0.93	69
D10	11%	RELAZIONI E FUNZIONI	-1.04	71
D24	11%	SPAZIO E FIGURE	-0.35	57
D6	10%	DATI E PREVISIONI	-0.22	54
D22	9%	NUMERI	-0.42	59
D14	8%	DATI E PREVISIONI	-0.93	69
D21_b	8%	RELAZIONI E FUNZIONI	0.04	49
D1_c	7%	RELAZIONI E FUNZIONI	-1.26	74
D19	6%	NUMERI	0.30	44
D2_a	6%	DATI E PREVISIONI	-1.53	79
D2_b	5%	DATI E PREVISIONI	0.07	48
D13	5%	SPAZIO E FIGURE	1.66	20
D27	4%	DATI E PREVISIONI	-0.22	54
D16_a	4%	NUMERI	-0.19	54
D12	4%	DATI E PREVISIONI	-0.85	67
D1_a	4%	RELAZIONI E FUNZIONI	-2.53	90
D21_a	3%	RELAZIONI E FUNZIONI	-0.90	68
D20	2%	RELAZIONI E FUNZIONI	-0.16	53
D4	2%	NUMERI	0.43	41
D7	2%	RELAZIONI E FUNZIONI	-1.05	71
D15_a	2%	DATI E PREVISIONI	-2.43	89
D8_b	1%	SPAZIO E FIGURE	-1.23	74
D5	0%	SPAZIO E FIGURE	0.36	42
D1_b	0%	RELAZIONI E FUNZIONI	-2.02	85
D11_a	0%	SPAZIO E FIGURE	-0.52	61
D25_a	0%	RELAZIONI E FUNZIONI	-0.92	68
D9	-1%	NUMERI	-0.63	63
D28	-2%	RELAZIONI E FUNZIONI	-0.41	58
D26	-2%	SPAZIO E FIGURE	-0.63	63
D8_a	-3%	SPAZIO E FIGURE	-1.64	80
D25_b	-5%	RELAZIONI E FUNZIONI	-0.33	57
D23	-5%	NUMERI	-0.64	63
D11_b	-12%	SPAZIO E FIGURE	0.59	38

Tabella 4.5: Item della prova INVALSI di matematica di III secondaria di 1°grado del 2014/2015. Gli item sono riportati in ordine decrescente in base all'indice di gender gap. La percentuale di risposte corrette è riferita all'intera popolazione.

Da questa prima analisi si può osservare che le domande che mostrano un maggiore gap a favore dei maschi (in termini di I_{GG}) appartengono a tutti gli ambiti di contenuto indagati dalle prove[22]. L'ambito Spazio e Figure non sembra creare maggiore gender gap rispetto agli altri ambiti, questo sebbene in letteratura gli unici riscontri di differenze nelle abilità cognitive siano relative alle abilità visuo-spaziali.

Una seconda osservazione riguarda la difficoltà: gli item che creano maggiore gender gap non sono solamente item difficili, ma fra gli item evidenziati se ne trovano anche con percentuali di risposte corrette abbastanza alte (superiori al 50%) e Delta negativo. Questo vuol dire che le studentesse non raggiungono alti livelli di abilità, non tanto perché falliscano negli item più complessi, ma anche perché sbagliano maggiormente alcuni item di difficoltà media. Può essere quindi interessante indagare cosa accade per queste domande.

Si può inoltre osservare che il gender gap non è distribuito equamente su tutte le domande: parte degli item non mostrano un divario significativo e, in ogni prova, per alcuni quesiti le ragazze ottengono una percentuale di risposte corrette maggiore rispetto ai ragazzi.
Volendo quindi approfondire maggiormente questa osservazione, si è deciso di suddividere le domande in cinque categorie in base al valore dell'indice, in particolare:

- $I_{GG} > 10$ ⟶ forte gender gap a favore dei maschi
- $3 < I_{GG} \leq 10$ ⟶ lieve gender gap a favore dei maschi
- $-3 \leq I_{GG} \leq 3$ ⟶ gender gap nullo o non significativo
- $-3 < I_{GG} \leq -10$ ⟶ lieve gender gap a favore delle femmine
- $I_{GG} < -10$ ⟶ forte gender gap a favore delle femmine

I grafici nelle pagine seguenti riportano il numero di item per ciascuna delle categorie appena descritte.

---

[22] È necessario tenere conto che una analisi più approfondita relativamente agli ambiti delle domande che creano maggiore gender gap dovrebbe tenere conto anche delle caratteristiche della prova stessa e quindi ad esempio della numero di item appartenenti a ogni ambito e della difficoltà dei quesiti di ogni ambito.

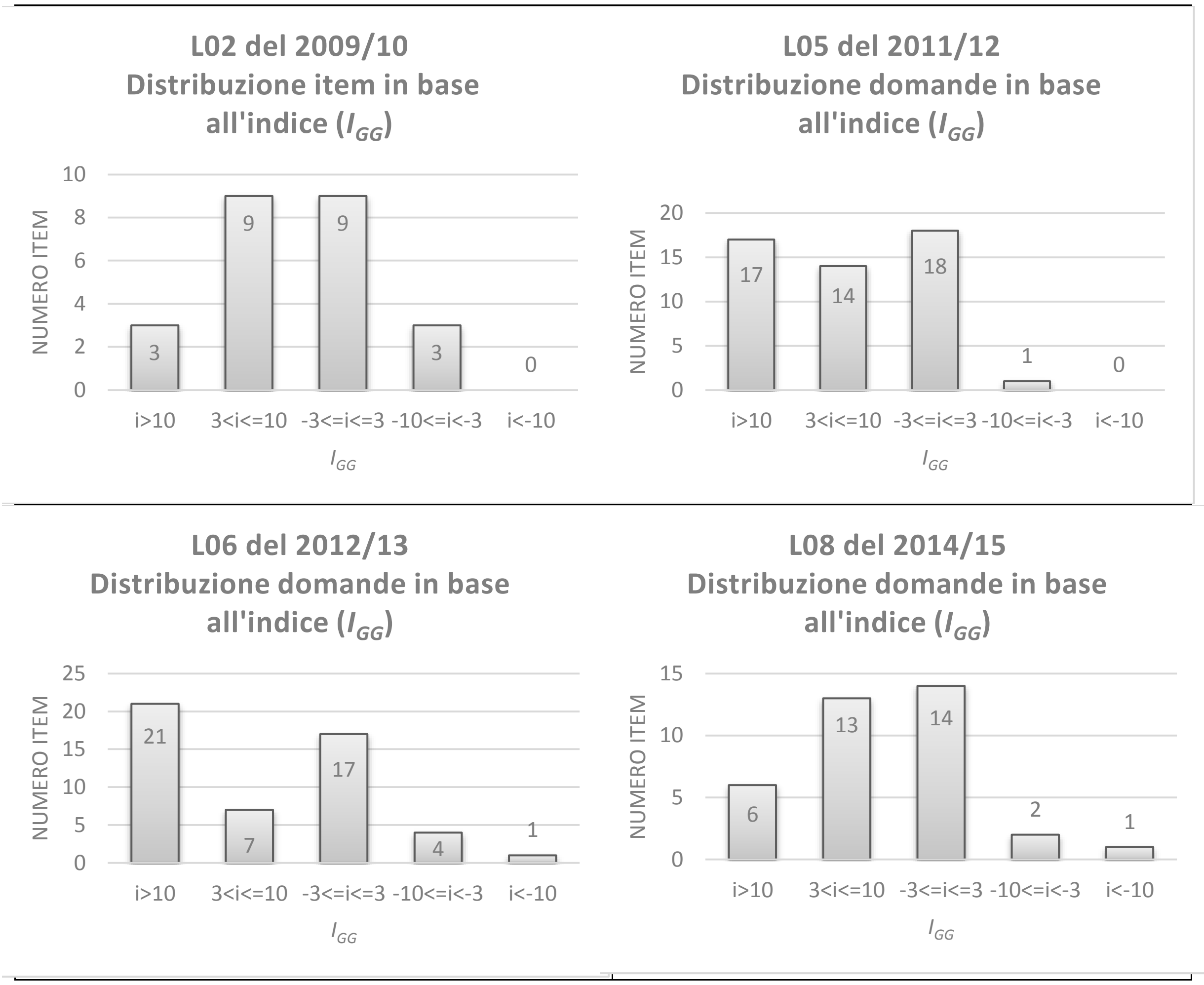


Figura 4.15: Distribuzione domande di ogni prova analizzata in funzione dell'indice di gender gap (I_{GG}).

Dai grafici e dalla tabella sottostante si evince che la distribuzione del gap non risulta in alcun modo uniforme. In ogni prova, circa un 35%-40% delle domande non mostra un gender gap significativo, le domande in cui i ragazzi rispondono meglio sono più del 50% del totale e in ogni test vi sono alcuni quesiti in cui le ragazze ottengono risultati superiori.

	Gender gap a favore dei maschi	**Domande neutre**	**Gender gap a favore delle femmine**
L02	50%	38%	13%
L05	62%	36%	2%
L06	56%	34%	10%
L08	53%	39%	8%

Tabella 4.6: Distribuzione percentuale degli item di ogni prova in funzione dell'indice di gender gap (I_{GG}). Sono considerate neutre le domande con indice compreso tra -3 e + 3 (compresi), a favore dei maschi quelle con indice superiore a 3 e a favore delle femmine quelle con indice inferiore a -3.

4.4.3.2 Analisi di alcuni quesiti

Alla luce di quanto detto, risulta interessante focalizzare l'attenzione sugli item più significativi: quelli che creano un divario molto marcato, quelli di difficoltà media che evidenziano comunque un divario notevole e i quesiti in cui le ragazze ottengono risultati migliori.

In questo paragrafo saranno esaminati brevemente alcuni dei quesiti tratti dalle prove analizzate nelle pagine precedenti e saranno messi a confronto con altri tratti sempre da prove INVALSI. Per alcuni di questi item si trova una analisi più approfondita negli articoli riportati in questo capitolo.

In particolare si è scelto di rintracciare, all'interno delle prove analizzate, cluster di quesiti che mostrassero un forte gender gap su un particolare contenuto matematico (confronto tra numeri decimali, stima di una misura, operazioni con le percentuali). L'analisi di uno stesso contenuto che, nei diversi livelli scolastici, ha mostrato sempre una maggiore difficoltà da parte delle studentesse può fornire importanti indicazioni relativamente agli ostacoli incontrati nell'apprendimento e, al contempo, mettere in luce possibili strategie di insegnamento che possano aiutare le studentesse a chiudere il gap.

D'altra parte, risulta anche importante osservare quali quesiti non evidenziano alcun gender gap o, addirittura, mostrano una percentuale di risposte corrette maggiore per le femmine. Sempre con riferimento alle prove della coorte presentata, verranno presentati alcuni quesiti che non hanno evidenziato gender gap e, in particolare, si mostrerà che tra questi ritroviamo anche domande relative alle abilità visuo-spaziali che rimanevano le uniche abilità cognitive in cui si riscontrava un gap in letteratura.

Gender gap nel confronto tra numeri decimali

L'analisi della prova di V primaria del 2012 ha fatto emergere una differenza di performance tra maschi e femmine, a favore dei primi, relativamente alla comprensione dei numeri decimali. All'interno della stessa prova, infatti è stato possibile confrontare i risultati di diversi quesiti riguardanti questo argomento (*D1 – D7a – D7b – D15 – D21a – D21b)* e in tutti l'indice relativo al gender gap segnala una maggiore difficoltà per le studentesse.

L'item D21 risulta particolarmente interessante perché mette in luce una misconcezione legata al confronto di numeri decimali, già ampiamente studiata in letteratura. Si tratta di una delle misconcezioni che emergono nel momento in cui lo studente che ha imparato a svolgere determinate operazioni in un insieme numerico (in questo caso il confronto tra numeri nell'insieme dei numeri naturali), si ritrova a svolgere le medesime operazioni in un ampliamento di questo insieme (numeri razionali) e tende a considerare valide le proprietà del primo insieme numerico anche nel nuovo insieme numerico.

In questo caso quindi lo studente, abituato a operare un confronto tra numeri naturali, è portato a confrontare le parti decimali di due numeri come se fossero numeri naturali:

> *'se si tratta di mettere in ordine 1,2 e 1,15, è noto che la competenza acquisita sui naturali può dare problemi interpretativi; la letteratura segnala casi in cui lo studente afferma: **«A parità di parte intera, siccome 15>2, allora 1,15>1,2».** Non sempre si rivela naturale scrivere 1,3 nella forma 1,30; ad impedire la naturalezza di questo passaggio sta anche una regola acquisita precedentemente, in base alla quale aggiungendo uno 0 "in fondo" ad un numero lo si moltiplica per 10; anche in questo caso, una regola valida in N viene erroneamente ed impropriamente estesa ai numeri razionali' (Sbaragli, 2012)*

Nel primo item del quesito D21, per identificare la nazione prima classificata, gli studenti devono arrivare a confrontare Italia e Stati Uniti che hanno rispettivamente i punteggi 80,12 e 80,2. La misconcezione appena descritta può portare gli studenti a sbagliare pensando che, visto che 12>2, allora 80,12>80,2.

Le percentuali di risposta corretta sono abbastanza basse e il valore del *Delta* conferma la difficoltà della domanda.

Nel primo item si evidenzia inoltre un forte gender gap a favore dei maschi e i grafici relativi al DIF mostrano che il divario è complessivamente uniforme per quasi tutti i livelli di abilità ad esclusione dei più bassi.

L'item 21b, a prima vista, propone la stessa richiesta del primo item ma risulta essere molto meno complesso. Per rispondere, infatti, non è necessario confrontare i numeri considerando anche la parte decimale, ma basta confrontare le parti intere e, quindi, non si incorre nella misconcezione sopra descritta.

Anche se le differenze di genere persistono, in questo caso risultano essere meno marcate ed è particolarmente interessante il confronto tra l'andamento della risposta corretta per maschi e femmine evidenziato dal grafico del DIF. In questo caso, infatti, sembra esserci un DIF solamente per livelli di abilità bassi e medi, mentre non vi è alcuna differenza per i livelli di abilità alti (probabilmente proprio perché da un certo livello di abilità in poi gli studenti si rendono conto della possibilità di confrontare solamente le parti intere dei punteggi e la misconcezione non ha più effetto).

Per una analisi più approfondita di questa domanda, da un punto di vista del funzionamento differenziale e del fit dell'item con il modello teorico si rimanda agli articoli riportati nei paragrafi 4.4.5 e 4.4.6.

D21 Livello 5 del 2012

D21. **Nella semifinale di una gara internazionale di ginnastica artistica i punteggi ottenuti complessivamente dalle atlete delle diverse nazioni sono i seguenti:**

Nazione	Punteggio
Austria	68,8
Croazia	71,8
Finlandia	72,0
Giappone	68,08
Grecia	60,8
Inghilterra	69,8
Italia	80,12
Stati Uniti	80,2
Svezia	70,2
Svizzera	78,1

a. **Quale nazione si è classificata prima?**

Risposta: ....

b. **Quale nazione si è classificata quarta?**

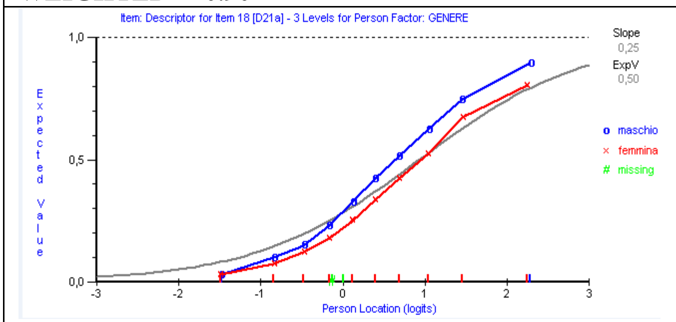
Risposta: ....

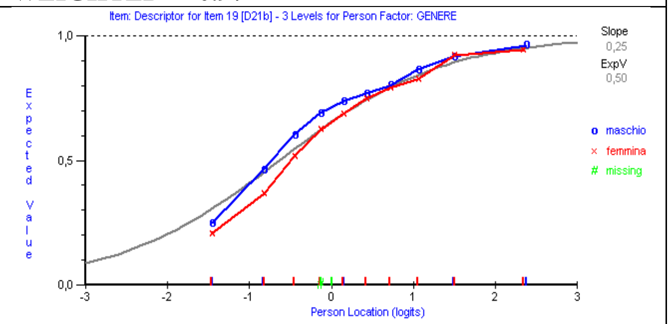
D21_a	TUTTI	FEMMINE	MASCHI
CORRETTA	37%	32%	43%
ERRATA	61%	66%	56%
MANCANTE	2%	2%	2%
INDICE	*28%*		

D21_b	TUTTI	FEMMINE	MASCHI
CORRETTA	67%	63%	70%
ERRATA	31%	35%	28%
MANCANTE	2%	2%	2%
INDICE	*11%*		

DELTA = 0.63
DISCRIMINATION = 0.55
WEIGHTED = 0.90

DELTA = -0.88
DISCRIMINATION = 0.46
WEIGHTED = 0.97





Riferimento alle Indicazioni Nazionali

- Traguardi IN - TP-V Ricerca dati per ricavare informazioni e costruisce rappresentazioni (tabelle e grafici). Ricava informazioni anche da dati rappresentati in tabelle e grafici.
- Obiettivi IN - Ob5-15 Leggere, scrivere, confrontare numeri decimali.
- Obiettivi IN - Ob5-35 Rappresentare relazioni e dati e, in situazioni significative, utilizzare le rappresentazioni per ricavare informazioni, formulare giudizi e prendere decisioni.

Figura 4.16: Analisi Item D21 prova INVALSI di livello 5 del 2012.

Questi risultati potrebbero essere in seguito approfonditi mediante interviste agli studenti, al fine di comprendere se i processi cognitivi attivati nel rispondere ai due item e gli errori siano effettivamente legati alla misconcezione descritta.

Sempre dall'analisi della stessa coorte di studenti e quindi delle prove analizzate, sono emersi altri quesiti legati all'ordinamento e al confronto di numeri decimali che hanno confermato una evidente differenza di genere.

Nella stessa prova da cui è stato tratto l'item precedente ritroviamo, ad esempio, un altro quesito relativo ai numeri decimali e anche in questo caso si osserva un notevole gender gap a favore dei maschi. I grafici relativi al DIF mostrano che il divario è relativo a tutti i livelli di abilità degli studenti e si attenua leggermente per i livelli molto alti e molto bassi in particolare per l'item b.

D7 Livello 5 del 2012

D7. Fai una crocetta sul numero che si avvicina di più a quello scritto a parole:

a. un decimo

A. ☐ 10
B. ☐ 0,09
C. ☐ 0,01
D. ☐ 0,15

b. sette centesimi

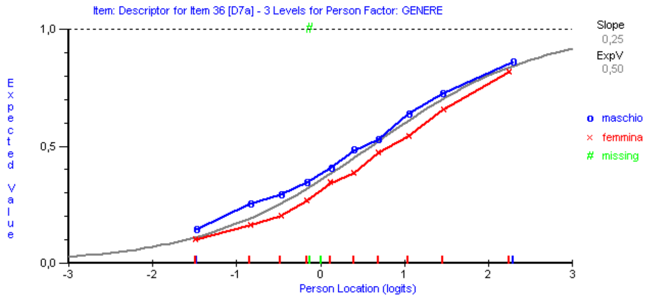
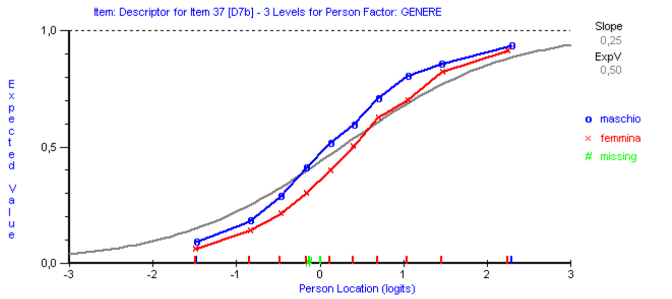
A. ☐ 700
B. ☐ 6,07
C. ☐ 0,08
D. ☐ 7

D7_a	TUTTI	FEMMINE	MASCHI
A	22%	25%	19%
B	**43%**	**38%**	**49%**
C	18%	20%	16%
D	16%	17%	16%
M	1%	1%	1%
INDICE	*26%*		

DELTA = 0.34
DISCRIMINATION = 0.45
WEIGHTED = 1.01

D7_b	TUTTI	FEMMINE	MASCHI
A	18%	20%	15%
B	20%	22%	18%
C	**50%**	**44%**	**56%**
D	11%	12%	10%
M	1%	2%	1%
INDICE	*23%*		

DELTA = -0.01
DISCRIMINATION = 0.56
WEIGHTED = 0.90

Grafico DIF (D7a) *risposta corretta per maschi e femmine*	*Grafico DIF (D7a)* *risposta corretta per maschi e femmine*
	

Riferimenti alle Indicazioni Nazionali

- Traguardi IN - TP-X Riconosce e utilizza rappresentazioni diverse di oggetti matematici (numeri decimali, frazioni, percentuali, scale di riduzione, ...).
- Obiettivi IN - Ob3-05 Leggere, scrivere, confrontare numeri decimali, rappresentarli sulla retta ed eseguire semplici addizioni e sottrazioni, anche con riferimento alle monete o ai risultati di semplici misure.
- Obiettivi IN - Ob5-15 Leggere, scrivere, confrontare numeri decimali.

Figura 4.17: Analisi Item D21 prova INVALSI di livello 5 del 2012.

Inoltre, sembrerebbe che le maggiori difficoltà delle ragazze nel confronto dei numeri decimali non si risolvano nel passaggio alla scuola secondaria: nella prova di livello 6 del 2013, a cui ha risposto quindi la medesima coorte di studenti, ritroviamo un altro quesito su questo tema.

Anche in questo caso si riscontra una differenza di genere significativa, messa in luce sia dalle percentuali di risposta, sia dai grafici del DIF che mostrano un divario marcato soprattutto per livelli di abilità alti. Risulta interessante inoltre osservare in che modo maschi e femmine hanno scelto le diverse opzioni di risposta; buona parte del gap percentuale si riversa nella scelta del distrattore B, probabilmente sempre per effetto della misconcezione descritta precedentemente. Questa risposta, infatti, può essere scelta da chi si concentra solo sulla parte decimale dei primi numeri di ogni sequenza e, visto che 5 > 28 > 124, allora conclude che 3,5 > 3,28 > 3, 124 (ignorando l'ultimo numero della sequenza).

Per approfondire la scelta dei distrattori si possono anche confrontare i distractor plot suddivisi in base al genere: le differenze nella scelta dell'opzione B è concentrata soprattutto sui livelli medio-bassi, ma il distrattore rimane attrattivo, per le ragazze, anche nei livelli alti di abilità; l'andamento del distrattore A è il medesimo, mentre il distrattore C risulta essere quasi costante nel grafico delle studentesse e decrescente per i maschi, che lo scelgono in percentuale maggiore alle femmine nei livelli bassi.

D23 Livello 6 del 2013

D23. **In quale dei seguenti gruppi i numeri sono disposti in ordine crescente?**

A. ☐ 3,5 ; 3,043 ; 3,28 ; 3,124

B. ☐ 3,5 ; 3,28 ; 3,124 ; 3,043

C. ☐ 3,043 ; 3,5 ; 3,124 ; 3,28

D. ☐ 3,043 ; 3,124 ; 3,28 ; 3,5

D23	TUTTI	FEMMINE	MASCHI
A	14%	15%	12%
B	17%	20%	15%
C	8%	8%	9%
D	**57%**	**54%**	**61%**
M	4%	4%	3%
INDICE	*12%*		

DELTA = -0.34
DISCRIMINATION = 0.56
WEIGHTED = 0.90

Distractor Plot D23 - Maschi

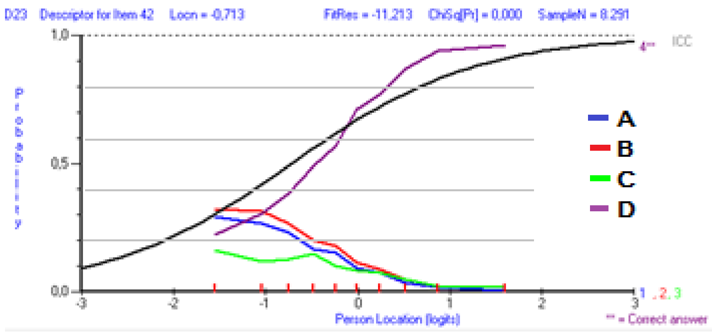
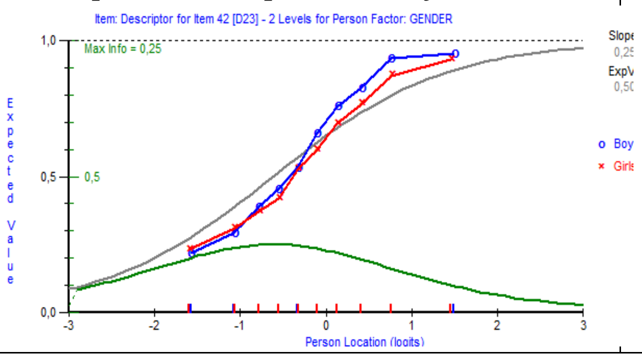
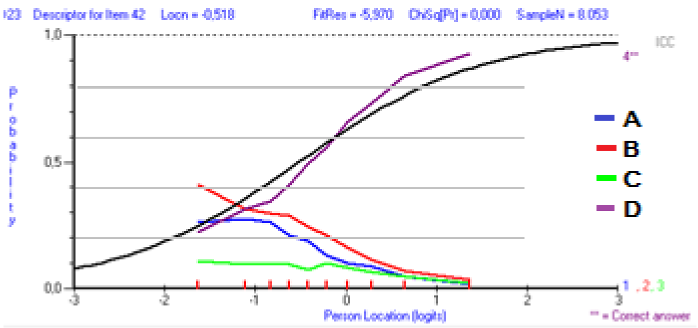


Grafico DIF
risposta corretta per maschi e femmine



Distractor Plot D23 - Femmine



Riferimenti alle Indicazioni Nazionali

- Traguardi IN - TS-XII Si muove con sicurezza nel calcolo anche con i numeri razionali, ne padroneggia le diverse rappresentazioni e stima la grandezza di un numero e il risultato di operazioni.
- Obiettivi IN - Ob5-15 Leggere, scrivere, confrontare numeri decimali.
- Obiettivi IN - Ob8-42 Eseguire addizioni, sottrazioni, moltiplicazioni, divisioni, ordinamenti e confronti tra numeri conosciuti (numeri naturali, numeri interi, frazioni e numeri decimali), quando possibile a mente oppure utilizzando gli usuali algoritmi scritti, le calcolatrici e i fogli di calcolo e valutando quale strumento può essere più opportuno.

Figura 4.18: Analisi Item D23 prova INVALSI di livello 6 del 2013.

Gender gap e abilità di stima di una misura

Dall'analisi delle stesse prove, è stato possibile collegare diverse domande con il medesimo *question intent* che hanno evidenziato una maggiore difficoltà delle ragazze nello stimare la misura di un oggetto reale.

Un quesito che ha mostrato gender gap in questa direzione è il seguente, tratto dalla prova di livello 6 del 2013.

D7b Livello 6 del 2013

b. Il treno su cui Nina sale è composto dalla locomotiva e da 9 vagoni.

Quanto è lungo all'incirca il treno di Nina?

A. ☐ Circa 10 m

B. ☐ Circa 50 m

C. ☐ Circa 250 m

D. ☐ Circa 1000 m

D7b	TUTTI	FEMMINE	MASCHI
A	6%	8%	5%
B	23%	21%	25%
C	**43%**	**40%**	**47%**
D	23%	26%	20%
M	5%	6%	4%
INDICE	*16%*		

DELTA = 0.31
DISCRIMINATION = 0.14
WEIGHTED = 1.20

Distractor Plot D23 – Maschi

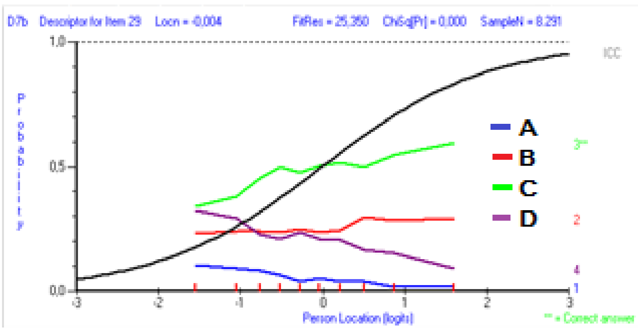
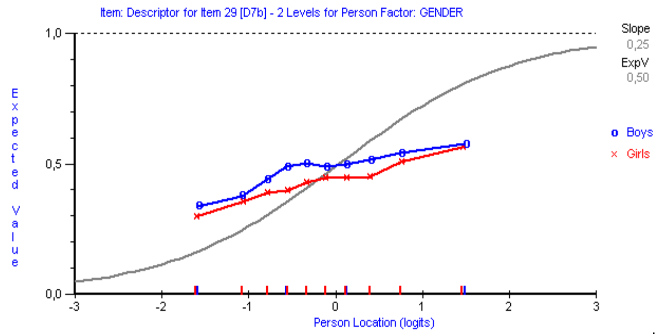
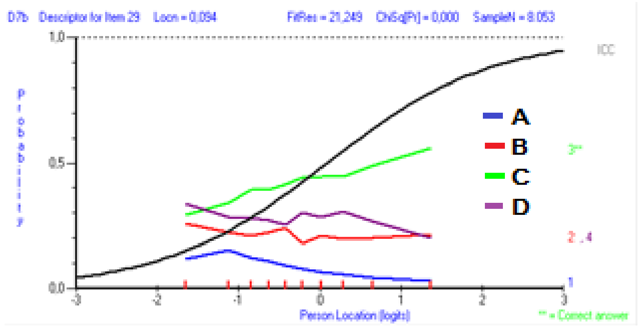


Grafico DIF
risposta corretta per maschi e femmine



Distractor Plot D23 - Femmine



Riferimenti alle Indicazioni Nazionali

- Traguardi IN - TS-XIV Analizza e interpreta rappresentazioni di dati per ricavarne misure di variabilità e prendere decisioni.
- Obiettivi IN - Ob5-38 Utilizzare le principali unità di misura per lunghezze, angoli, aree, volumi/capacità, intervalli temporali, masse, pesi per effettuare misure e stime.

Figura 4.19: Analisi Item D7b prova INVALSI di livello 6 del 2013.

In questo item, analizzato anche negli articoli riportati nei paragrafi seguenti, si chiede di effettuare la stima della lunghezza di un treno conoscendo il numero di vagoni da cui è composto. La percentuale di risposte corrette è in generale abbastanza bassa e vi è una differenza consistente tra quella dei maschi (47%) e delle femmine (40%).

La domanda è poco discriminativa e il fit con il modello risulta essere inadeguato. Osservando il grafico del DIF, infatti, si nota che il modello sottostima i rispondenti con bassi livelli di abilità e sovrastima quelli con livelli alti di abilità. La risposta corretta infatti mostra un andamento solo lievemente crescente e quasi lineare: non vi è una grande differenza nell'individuare la risposta corretta tra gli studenti meno bravi e quelli più bravi. L'ipotesi fatta è che questa domanda in particolare sia poco legata alla pratica didattica e cioè che a scuola il tema della stima venga affrontato in modo marginale. Questo spiegherebbe perché questo tipo di domanda metta in difficoltà tutti gli studenti, quasi indipendentemente dal loro livello di abilità.

Il divario nella scelta della risposta corretta è presente su tutti i livelli di abilità e particolarmente marcato nei livelli medi. Si è ipotizzato che il gap sia dovuto proprio al fatto che la domanda sia poco "scolastica": gli studenti non hanno lavorato molto su di un metodo per affrontare i problemi di stima e per questo motivo è necessario mettere in gioco strategie risolutive nuove e, come si è visto nella letteratura analizzata, questo risulta più facile per gli studenti maschi.

Il distrattore D (sovrastima del risultato) è decrescente con l'abilità nel caso dei maschi mentre risulta essere molto attrattivo anche per le femmine più brave e per questo motivo nel loro caso mostra un andamento solo lievemente decrescente.
L'opzione B (sottostima) è la più attrattiva per i maschi e nel relativo distractor plot risulta addirittura lievemente crescente. Il distrattore A (sottostima completamente fuori bersaglio) è decrescente sia per i maschi sia per le femmine, anche se risulta essere leggermente più scelto da queste ultime. Il fatto che una parte di studenti, seppur ridotta, scelga questa risposta può essere attribuito a un fenomeno di contratto didattico (10 è il risultato dell'addizione degli unici due numeri presenti: 1 locomotiva + 9 vagoni).

Per rispondere a questa domanda gli studenti potrebbero eseguire la stima in due modi diversi: stimando la lunghezza complessiva del treno, oppure stimando la lunghezza di un elemento e moltiplicando poi per 10.

Nelle risposte a questa domanda non abbiamo indicazioni relative alla strategia seguita quindi si è deciso di riproporre questa domanda all'interno del progetto *Variazioni 2* (paragrafo 5.3), proprio per approfondire gli aspetti emersi da questa prima analisi.

La conferma di quanto evidenziato dal quesito precedente è stata fornita dall'analisi di un altro quesito sottoposto due anni dopo, sempre alla stessa coorte di studenti. Il quesito in questione è il D3 della prova di livello 8 del 2015 e richiede di stimare l'altezza di un palazzo.

D3 Livello 8 del 2015

D3. **Osserva l'edificio nella foto.**

Quanto può essere alto l'edificio?

A. ☐ meno di 10 metri

B. ☐ tra 15 e 20 metri

C. ☐ tra 25 e 30 metri

D. ☐ più di 35 metri

D23	TUTTI	FEMMINE	MASCHI
A	8%	8%	8%
B	65%	61%	69%
C	22%	24%	19%
D	5%	6%	3%
M	0%	1%	0%
INDICE	*12%*		

DELTA = -0.75
DISCRIMINATION = 0.29
WEIGHTED = 1.11

Distractor Plot D23 – Maschi

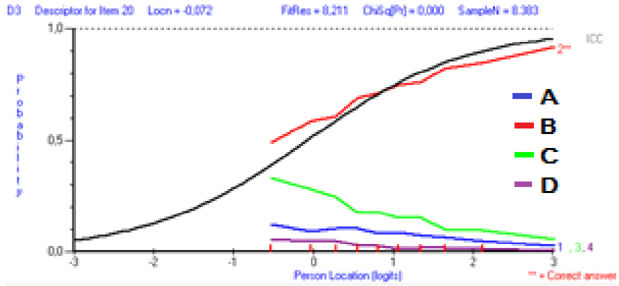
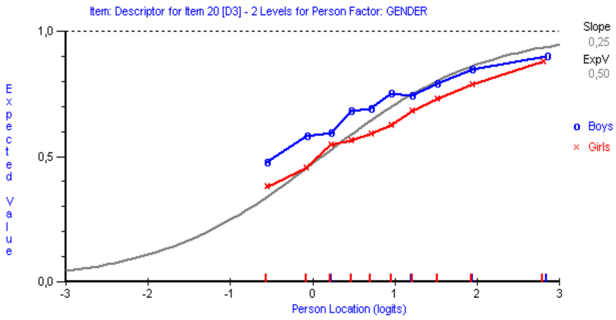
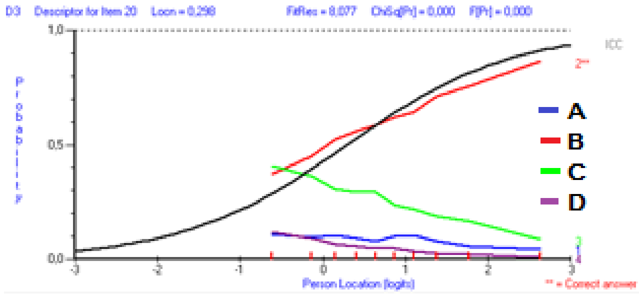


Grafico DIF
risposta corretta per maschi e femmine



Distractor Plot D23 - Femmine



Riferimenti alle Indicazioni Nazionali

- Traguardi IN - TS-XV Riconosce e risolve problemi in contesti diversi valutando le informazioni e la loro coerenza.
- Obiettivi IN - Ob8-75 Calcolare l'area e il volume delle figure solide più comuni e dare stime di oggetti della vita quotidiana.

Figura 4.20: Analisi Item D3 prova INVALSI di livello 5 del 2015.

In questa domanda la percentuale di risposte corrette è maggiore, probabilmente anche perché è stata somministrata a studenti di due anni più grandi rispetto alla precedente e perché la lunghezza da stimare è minore.

Nonostante il fit con il modello risulti migliorato, anche in questo caso si ha una sottostima dei livelli di abilità più bassi e una sovrastima dei livelli più alti. Inoltre, la crescita della risposta corretta è pressoché lineare; queste caratteristiche dell'item possono essere ricondotte alle stesse motivazioni fornite per il quesito precedente.

Il gender gap è marcato e presente su tutti i livelli di abilità (ad esclusione del decile più alto), confermando quindi una maggiore difficoltà delle studentesse nello stimare la misura di un oggetto reale. Le percentuali di risposte mancanti sono quasi nulle, il che porta a pensare che gli studenti siano abbastanza sicuri delle proprie risposte. Infine, anche in questo caso le ragazze tendono a sovrastimare il risultato e vengono attratte maggiormente dal distrattore C, per i livelli di abilità più bassi per le ragazze questo distrattore risulta più attrattivo anche rispetto alla risposta corretta, cosa che non avviene invece per i ragazzi.

Gender gap nell'operare con le percentuali

Dall'analisi delle altre prove, non relative alla coorte presentata in questo capitolo, è emerso un altro argomento che in più item ha mostrato una marcata differenza di genere, sempre a favore dei maschi: le percentuali. Da una analisi analoga a quella compiuta sulle prove precedenti ma focalizzata sulle prove INVALSI di livello 10, risulta infatti che le ragazze hanno maggiori difficoltà quando si tratta di lavorare con le percentuali.

Per gli item presentati in questo paragrafo non sono state ancora compiute analisi approfondite del DIF e le differenze di genere non sono ancora state studiate in funzione dei livelli di abilità. Le evidenze però rendono molto interessante l'analisi di questi quesiti relativamente al gender gap e analisi più approfondite saranno sicuramente portate avanti nei prossimi anni.

Il primo quesito che ha fatto emergere questo fenomeno è particolarmente interessante perché, a una prima lettura, potrebbe essere collegato a stereotipi.

D25 Livello 10 del 2012

D25. **In un negozio un abito è messo in vendita con uno sconto del 30% sul prezzo originario. Durante la stagione dei saldi il prezzo già scontato viene ancora abbassato del 10%. Qual è la percentuale complessiva di sconto sul prezzo originario dell'abito?**

A. ☐ 20%

B. ☐ 33%

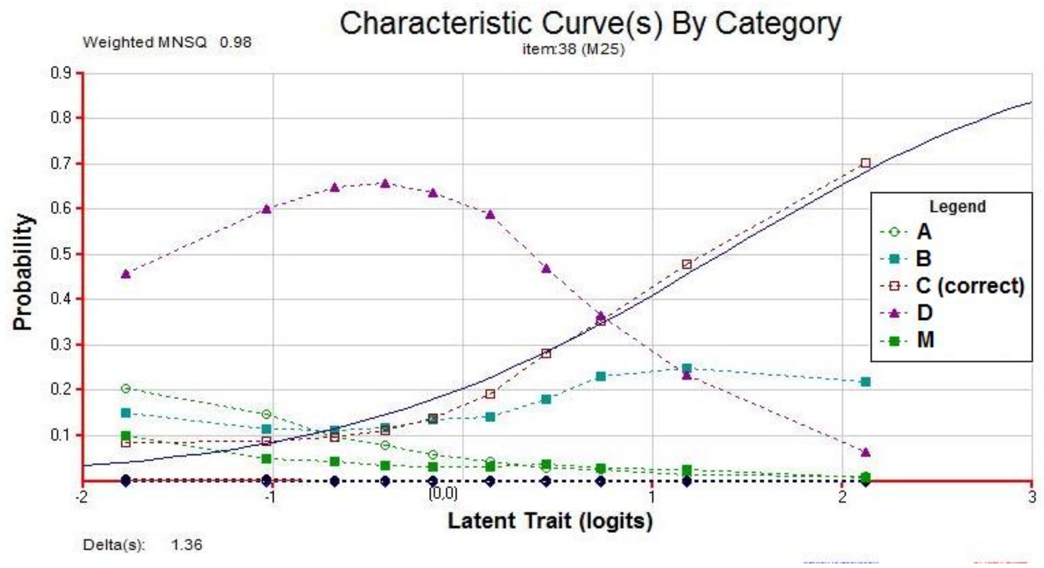
C. ☐ 37%

D. ☐ 40%

D25	TUTTI	FEMMINE	MASCHI
A	7%	9%	6%
B	16%	13%	19%
C	**24%**	**18%**	**30%**
D	48%	56%	41%
M	4%	4%	4%
INDICE	***54%***		

DELTA = 1.36
DISCRIMINATION = 0.43
WEIGHTED = 0.98

Distractor Plot



Riferimenti alle Indicazioni Nazionali

- Competenze AC - AC-C1 Utilizzare le tecniche e le procedure del calcolo aritmetico ed algebrico, rappresentandole anche in forma grafica
- Obiettivi AC - Impostare uguaglianze di rapporti per risolvere problemi di proporzionalità e percentuale.
- Obiettivi LG-IN - Le operazioni con i numeri interi e razionali. Operare con i numeri interi e razionali e valutare l’ordine di grandezza dei risultati. Calcolo (mentale, con carta e penna, mediante strumenti) con i numeri interi, con i numeri razionali sia nella scrittura come frazione che nella rappresentazione decimale.
- Obiettivi LG-IN - Rapporti e percentuali. Proporzionalità diretta e inversa.

Figura 4.21: Analisi Item D25 prova INVALSI di livello 10 del 2012.

Il quesito è stato somministrato a studenti frequentanti la seconda secondaria e mostra notevoli difficoltà nell’operare con le percentuali. Il gender gap risulta molto marcato e solo il 18% delle ragazze risponde correttamente, a fronte di un 30% dei ragazzi. In questo caso quindi si può dire che non è la tipologia di contesto o la vicinanza del contesto alle esperienze quotidiane ad influenzare il divario di genere perché, in tal caso, le ragazze avrebbero dovuto ottenere risultati pari o superiori ai compagni.

L’opzione D sembra essere la principale causa del gender gap: le femmine vengono attratte notevolmente da questo distrattore che viene indicato nel 56% dei casi, mentre per quanto riguarda i maschi, pur essendo il distrattore preferito, questa percentuale si riduce al 41%. Gli studenti che scelgono questo distrattore sono probabilmente guidati da una misconcezione molto simile a quella vista in precedenza per i numeri decimali: chi risponde

40% tende ad operare con le percentuali seguendo le stesse procedure utilizzate nei decimali e per trovare lo sconto totale somma i due sconti successivi (pensando che 30%+10%=40%).

Si potrebbe quindi ipotizzare anche in questo caso che sia una diversa influenza della misconcezione a creare il divario. Inoltre dal distractor plot si nota un comportamento molto particolare di questo distrattore che risulta attrattivo particolarmente per livelli di abilità medio-bassi ma che viene scelto da una buona parte dei rispondenti anche per livelli di abilità medio alti.

Un altro stimolo molto interessante relativo a questa domanda, che meriterebbe un ulteriore approfondimento, deriva dall'analisi suddivisa in base alla tipologia di scuola. Dalla tabella seguente, si osserva che il quesito nei licei, pur risultando leggermente più semplice, crea un gender gap maggiore. Sempre nei licei è minore la percentuale di studenti che scelgono il distrattore D e quindi sul totale degli studenti sembra essere meno forte la misconcezione. In realtà il divario tra maschi e femmine sul distrattore D è più marcato rispetto alle altre scuole e arriva a 23 punti percentuali. Quindi la misconcezione sembra avere una particolare influenza sul gender gap nei licei.

	LICEO			**ISTITUTO TECNICO**			**ISTITUTO PROFESSIONALE**		
	FEMMINE	MASCHI	TOTALE	FEMMINE	MASCHI	TOTALE	FEMMINE	MASCHI	TOTALE
A	7%	4%	6%	9%	6%	7%	15%	12%	13%
B	15%	22%	18%	11%	17%	15%	9%	18%	14%
C	**20%**	**40%**	**28%**	**18%**	**27%**	**24%**	**11%**	**18%**	**15%**
D	54%	31%	45%	58%	46%	50%	59%	48%	53%
M	4%	4%	4%	4%	3%	4%	6%	5%	5%
	INDICE = 71%			INDICE = 38%			INDICE = 47%		

Tabella 4.7: Analisi Item D25 prova INVALSI di livello 10 del 2012 per tipologia di scuola.

Le differenze di genere relativamente a questo argomento possono essere riscontrare anche in numerosi altri quesiti che vengono riportati di seguito con una breve descrizione.

Il quesito D15 della stessa prova, ad esempio, conferma la difficoltà degli studenti nel calcolare la percentuale di una percentuale. Anche in questo caso si osserva un forte divario tra i risultati delle ragazze e dei ragazzi ed entrambi i distrattori B e C risultano più attrattivi per le studentesse. Anche in questo caso le risposte potrebbero essere influenzate dalla stessa misconcezione osservata nel quesito precedente: chi sceglie il distrattore B infatti potrebbe aver sottratto le due percentuali come fossero numeri naturali (70%-20%=50%).

D15 Livello 10 del 2012

D15. **Nelle ultime elezioni svoltesi in un paese europeo è andato a votare il 70% degli aventi diritto al voto. Di questi il 20% ha votato per il partito *A*. Quale percentuale di aventi diritto al voto ha votato per il partito *A*?**

A. ☐ 60%

B. ☐ 50%

C. ☐ 20%

D. ☐ 14%

D15	TUTTI	FEMMINE	MASCHI
A	2%	2%	2%
B	24%	28%	20%
C	35%	37%	33%
D	**37%**	**31%**	**42%**
M	2%	3%	2%
INDICE	*30%*		

DELTA = 0.45
DISCRIMINATION = 0.46
WEIGHTED = 0.99

Distractor Plot

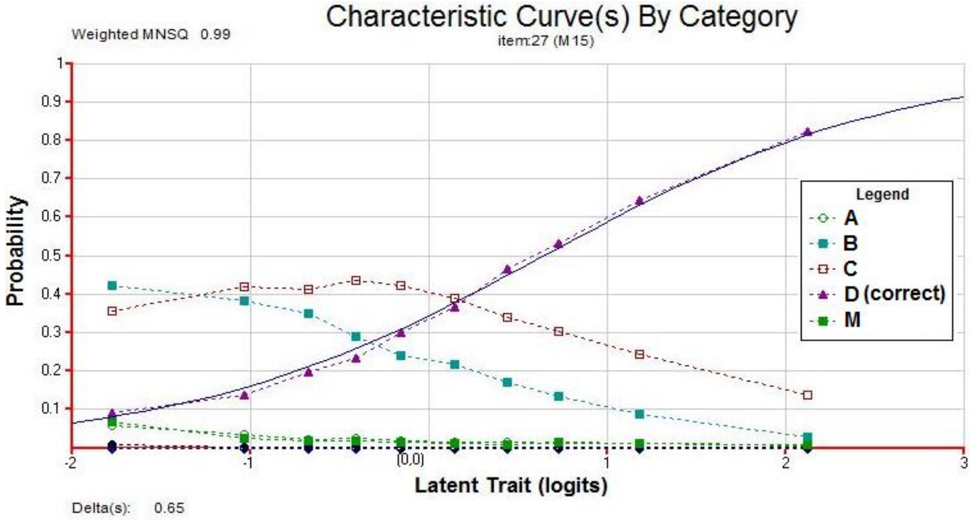


Figura 4.22: Analisi Item D15 prova INVALSI di livello 10 del 2012.

La prova INVALSI da cui sono stati tratti i quesiti precedenti non è però l'unica a confermare le osservazioni fatte relativamente al gender gap sulle percentuali.

Nei prossimi due quesiti le richieste sono molto simili a quelle precedenti ma si fornisce anche agli studenti il dato iniziale di cui si vuole calcolare la percentuale. In questo modo le domande dovrebbero risultare più semplici, poiché non richiedono di calcolare la percentuale di una quantità incognita.

D11 Livello 10 del 2016

D11. A novembre il prezzo di un paio di scarpe è di 100 €. A dicembre il prezzo aumenta del 10%. A gennaio, nel periodo dei saldi, il prezzo viene scontato del 10% rispetto al prezzo di dicembre. Qual è il prezzo delle scarpe in saldo?

A. ☐ 100 €

B. ☐ 99 €

C. ☐ 90 €

D. ☐ 91 €

D11	TUTTI	FEMMINE	MASCHI
A	49%	54%	45%
B	**35%**	**30%**	**40%**
C	10%	11%	9%
D	4%	4%	4%
M	2%	2%	2%
INDICE	***29%***		

DELTA = 0.75
DISCRIMINATION = 0.51
WEIGHTED = 0.98

Distractor Plot

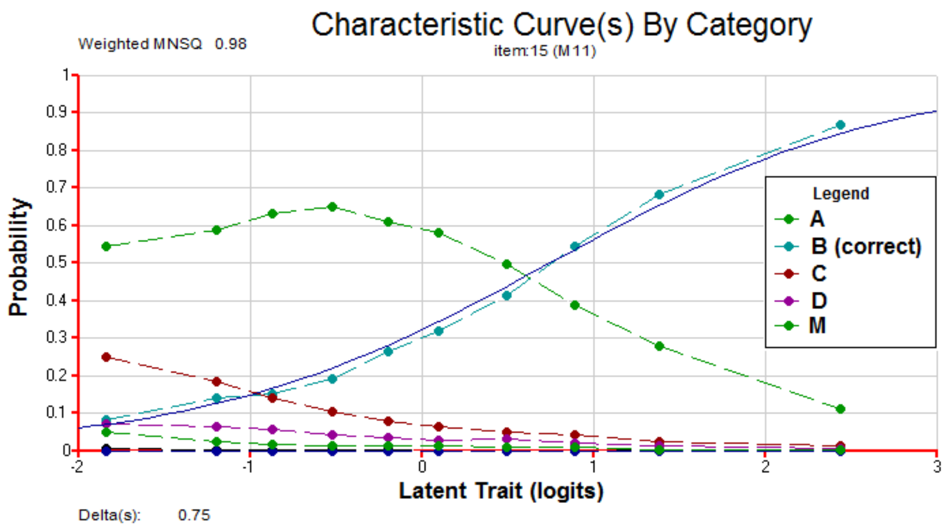


Figura 4.23: Analisi Item D11 prova INVALSI di livello 10 del 2016.

Nonostante nel quesito D11 il prezzo iniziale delle scarpe sia dato, il quesito continua a evidenziare notevoli difficoltà da parte di tutti gli studenti e in particolare delle ragazze che raggiungono solamente il 30% di risposte corrette. Il distrattore A è il più forte ed è alla base di quasi tutto il divario di genere riscontrato sulla risposta corretta. Si tratta anche in questo caso del distrattore legato alla misconcezione già discussa e dal distractor plot si nota che anche in questo caso è particolarmente attrattivo per livelli di abilità medio-bassi ma viene scelto anche da molti studenti che presentano livelli alti di abilità.

Infine si può osservare che queste difficoltà si riscontrano anche al termine del primo ciclo di istruzione. In questo caso l'item E15 è tratto dalla prova di livello 8 del 2012, nonostante l'item risulti comunque medio-difficile, si può notare come la percentuale di risposte corrette sia maggiore rispetto all'item precedente, che è stato somministrato a studenti di due anni più grandi.

E15 Livello 8 del 2012

E15. **Una grande azienda nel 2009 aveva 100 impiegati. Nell'anno 2010 il numero degli impiegati è diminuito del 20% rispetto al 2009 mentre nel 2011 è aumentato del 20% rispetto al 2010. Al termine dei due anni gli impiegati dell'azienda sono**

A. ☐ diminuiti del 4%

B. ☐ diminuiti del 10%

C. ☐ aumentati del 4%

D. ☐ aumentati del 10%

D11	TUTTI	FEMMINE	MASCHI
A	47%	43%	50%
B	15%	15%	14%
C	17%	18%	16%
D	19%	20%	17%
M	3%	4%	3%
INDICE	*15%*		

DELTA = 0.36
DISCRIMINATION = 0.49
WEIGHTED = 0.92

Distractor Plot

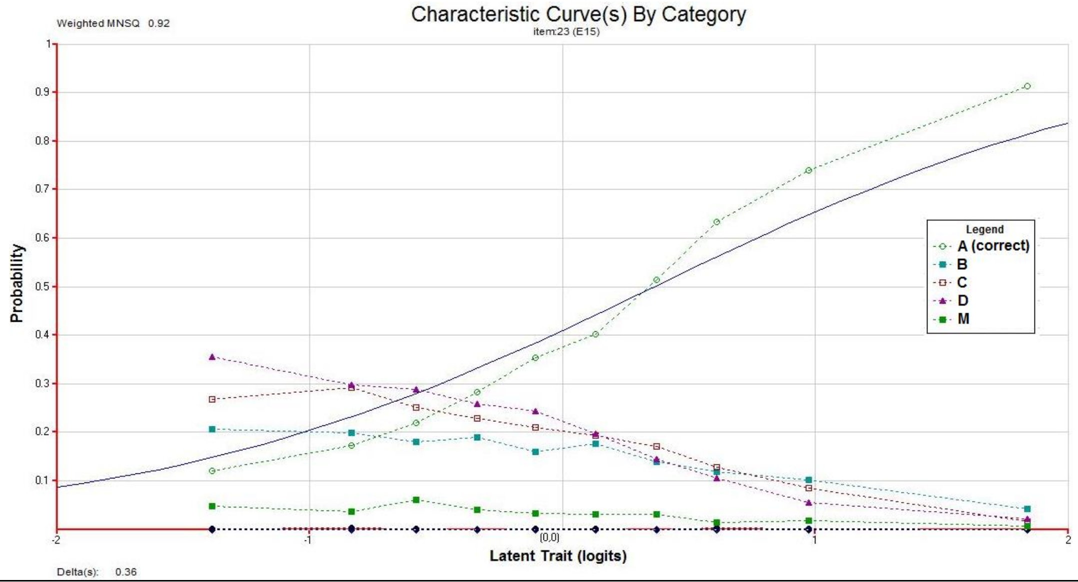


Figura 4.24: Analisi Item E15 prova INVALSI di livello 8 del 2012.

La minore difficoltà di questo item è probabilmente imputabile al fatto che nessuno dei distrattori prende in causa la misconcezione relativa alle percentuali: se uno dei distrattori fosse stato 0% oppure "rimasti invariati" probabilmente i risultati sarebbero stati più vicini a quelli dell'item precedente. Si osserva infatti che tra i distrattori non ve ne è uno notevolmente più attrattivo rispetto agli altri.

Anche il gender gap risulta minore rispetto agli item precedenti ma sempre presente. Inoltre anche per quanto riguarda la scelta dei distrattori non si nota una opzione che risulti particolarmente più attrattiva per le studentesse, il gender gap non è quindi dovuto a una particolare opzione che mostra un comportamento differenziale.

Questo item è stato anche inserito all'interno del progetto *Variazioni 2* (cap. 5), per osservare eventuali cambiamenti nelle risposte degli studenti a seguito di variazioni di contesto e tipologia di distrattori.

Gender gap assente o a favore delle studentesse

L'analisi preliminare delle prove INVALSI relative alla coorte di studenti su cui sono state concentrate le ricerche, ha messo in evidenza che non tutti i quesiti mostrano una differenza nelle performance di maschi e femmine.

Ci sono infatti domande che non mostrano alcun gender gap o altre, più rare, in cui le ragazze ottengono risultati migliori rispetto ai ragazzi. Dalle stesse prove (livello 2 del 2009, livello 5 del 2012, livello 6 del 2013 e livello 8 del 2015) si è quindi scelto di riportare alcuni esempi di domande di questo tipo.

L'item D19 del livello 6 del 2015 ha una difficoltà media e sono molto alte le percentuali di risposte mancanti anche perché si tratta di un quesito a risposta aperta. In questo caso le ragazze ottengono risultati superiori rispetto ai ragazzi e la percentuale di risposte mancanti risulta essere maggiore per questi ultimi. Osservando alcuni protocolli sembra che buona parte degli studenti, per rispondere a questo item, svolgano l'operazione e confrontino il loro risultato con quello proposto. Il quesito, se affrontato in questo modo, risulta abbastanza standard e vicino alle pratiche scolastiche e potrebbe essere questo il motivo del divario invertito.

D19 Livello 6 del 2013

D19. Andrea svolge sul quaderno questa divisione:

1632 : 4 = 48

Il risultato ottenuto da Andrea è sbagliato. Quale errore può aver fatto?

Risposta: ....

....

D19	TUTTI	FEMMINE	MASCHI
Corretta	47%	50%	44%
Errata	39%	38%	40%
M	14%	12%	16%
INDICE	*-13%*		

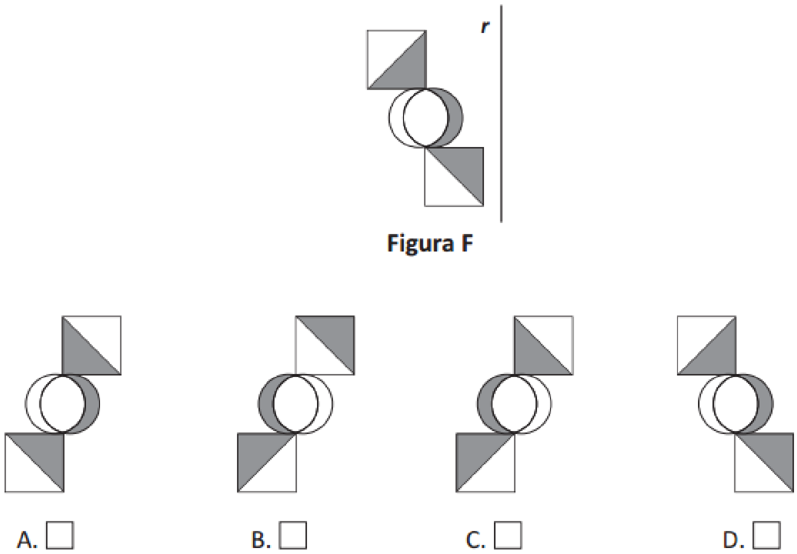
DELTA = 0.14
DISCRIMINATION = 0.40
WEIGHTED = 1.00

Figura 4.25: Analisi Item D19 prova INVALSI di livello 6 del 2013.

Dall'analisi delle prove del livello 5 del 2012 e del livello 8 del 2015, si è scelto di analizzare due quesiti che evidenziano come il gender gap non sia un fenomeno strettamente legato alla geometria e che i quesiti che richiedono abilità visuo-spaziali non per forza mettono maggiormente in difficoltà le ragazze.

D10 Livello 5 del 2012

D10. Quale tra le seguenti figure è simmetrica alla figura *F* rispetto all'asse di simmetria *r* ?



D10	TUTTI	FEMMINE	MASCHI
A	7%	7%	6%
B	10%	10%	10%
C	**51%**	**52%**	**50%**
D	31%	30%	32%
M	1%	1%	1%
INDICE	*0%*		

DELTA = -0.06
DISCRIMINATION = 0.45
WEIGHTED = 1.01

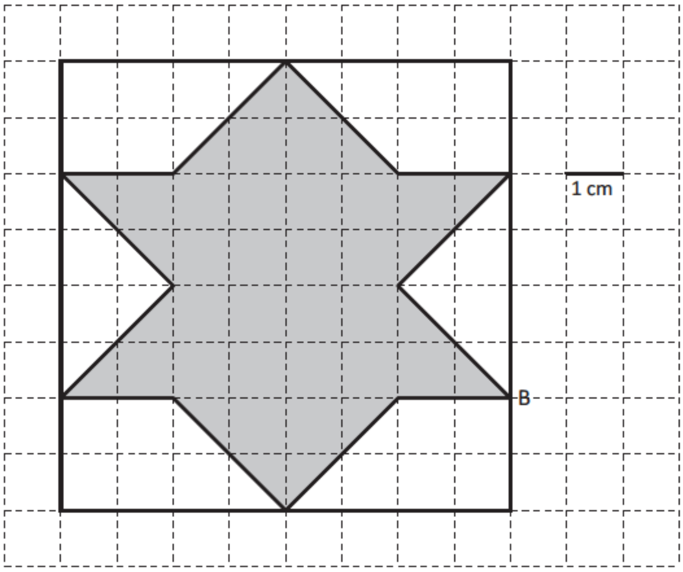
Figura 4.26: Analisi Item D10 prova INVALSI di livello 5 del 2012.

Nella prima domanda si richiede di individuare, rispetto ad una figura complessa data, la sua simmetrica. Per questa domanda le risposte corrette delle femmine sono addirittura superiori a quelle dei maschi.

La seconda domanda (Fig. 4.25) è composta da due item e chiede di ragionare sull'area e sugli assi di simmetria di un poligono particolare. Nel primo item le percentuali delle risposte corrette sono le medesime tra maschi e femmine, mentre nel secondo item le ragazze ottengono risultati migliori dei compagni.

D11 Livello 8 del 2015

D11. Osserva la seguente figura formata da un quadrato al cui interno è disegnato un poligono di colore grigio.



a. Qual è l'area del poligono grigio?

Risposta: .... cm^2

b. Disegna una diagonale del quadrato. La diagonale è asse di simmetria del poligono grigio?

A. ☐ Sì, perché la diagonale divide il poligono grigio in due parti uguali e simmetriche

B. ☐ Sì, perché la diagonale è asse di simmetria del quadrato

C. ☐ No, perché il poligono grigio non ha assi di simmetria

D. ☐ No, perché il simmetrico di B rispetto alla diagonale non è un vertice del poligono grigio

D11a	TUTTI	FEMMINE	MASCHI
Corretta	**61%**	**61%**	**61%**
Errata	34%	34%	34%
Mancante	5%	5%	6%
INDICE	*0%*		

DELTA = -0.52
DISCRIMINATION = 0.46
WEIGHTED = 0.97

D11b	TUTTI	FEMMINE	MASCHI
A	46%	44%	48%
B	7%	7%	8%
C	6%	6%	7%
D	**38%**	**40%**	**36%**
M	2%	2%	2%
INDICE	*-10%*		

DELTA = 0.59
DISCRIMINATION = 0.41
WEIGHTED = 1.00

Figura 4.27: Analisi Item D11 prova INVALSI di livello 8 del 2015.

Conclusioni

Le analisi presentate in questo capitolo e negli articoli riportati nei prossimi paragrafi, mostrano come il gender gap in matematica sia un fenomeno molto complesso. Nonostante ciò, si è mostrato che analisi quantitative di test su larga scala possono fornire importanti indicazioni per approfondirne le cause, le peculiarità legate alla disciplina e intervenire con azioni didattiche mirate.

Abbiamo infatti potuto osservare che il gap sull'intera prova non è determinato da un divario uniforme e presente in tutte le domande che la compongono, ma si presenta più marcato su alcuni quesiti e, all'interno di un test, si ritrovano anche alcuni quesiti in cui le ragazze ottengono risultati migliori dei ragazzi. Questo rende ancora più affascinante lo studio di questo fenomeno e, anche nell'analisi dei singoli item, gli strumenti statistici forniti dalla Teoria Classica dei Test e dalla Teoria della Risposta all'Item sono strumenti preziosi per compiere una analisi approfondita di tale fenomeno se integrati con una interpretazione didattica.

In particolare è stato possibile individuare cluster di item legati a un medesimo contenuto che, anche in prove di diversi livelli scolastici, mostravano un marcato gap a favore dei ragazzi. Questo significa che alcune difficoltà riscontrate maggiormente dalle femmine nei primi anni di scuola, non vengono poi colmate durante il percorso scolastico. Dalle prime analisi si è visto come, ad esempio, le studentesse siano maggiormente influenzate da particolari misconcezioni legate al passaggio tra l'insieme dei numeri naturali e l'insieme dei numeri razionali. Queste misconcezioni portate avanti negli anni diventano veri e propri ostacoli all'apprendimento se non vi è una consapevolezza e un lavoro mirato da parte dell'insegnante per evidenziarle e superarle.

L'analisi quantitativa basata sul modello di Rasch e sul DIF ha permesso di studiare tali domande e il gap evidenziato, anche in funzione del livello di abilità degli studenti. L'intreccio tra l'analisi statistica e l'interpretazione didattica ha permesso quindi di giungere a una comprensione più approfondita del fenomeno e di legare le evidenze statistiche ai processi cognitivi messi in atto dagli studenti.

Anche i quesiti che richiedono di stimare la misura di un oggetto reale hanno evidenziato, in diversi livelli scolastici, un notevole divario a favore dei maschi. Il gap in questo caso risulta uniforme rispetto al livello di abilità degli studenti e perciò per ogni livello di abilità la capacità di stimare una misura risulta più agevole per i maschi e ostacola maggiormente le femmine. Inoltre, l'analisi del fit mostra come questa sia una abilità che risulta "sghemba" rispetto al tratto latente considerato dai test e, in generale, non è molto marcata la differenza tra rispondenti con alti livelli di abilità e rispondenti con bassi livelli di abilità.

Queste osservazioni ottenute dall'analisi quantitativa, viste attraverso le lenti della didattica, permettono di ipotizzare che il comportamento dell'item sia dovuto al fatto che si tenda, generalmente, a lavorare poco in classe sulla stima di misure. La lontananza di questo tema dalla pratica didattica potrebbe essere inoltre una possibile spiegazione del gender gap presente in questi quesiti.

Le ultime analisi mostrano invece come, nelle prove INVALSI analizzate, le differenze di genere non si riscontrano su quesiti legati alle abilità visuo-spaziali che, in letteratura, vengono indicate come possibile causa del gender gap in matematica. Dalle analisi compiute tra i diversi ambiti non ve ne è uno in cui si concentrano maggiormente gli item con maggiore gap e l'ambito Spazio e Figure non sembra avere una particolare incidenza sulle differenze di genere.

Le ricerche riportate in questa tesi spingono a ricercare le cause del gender gap in problemi legati alla pratica didattica, al contesto sociale e alla sfera psico-sociale dell'individuo piuttosto che a specifiche abilità cognitive. Si è visto quanto aspetti legati alla pratica didattica e a fattori micro-sociali, possano essere una delle possibili spiegazioni delle differenze di genere riscontrate in matematica. I quesiti analizzati suggeriscono che le dinamiche d'aula intrecciate con le convinzioni degli studenti potrebbero agire diversamente su maschi e femmine, portando quindi a una differente influenza di fenomeni didattici come misconcezioni ed effetti del contratto didattico sui due gruppi di studenti.

Nei prossimi paragrafi sono riportati due articoli presentati in due convegni internazionali e pubblicati nei relativi atti. Il terzo articolo vuole essere un approfondimento dell'articolo presentato al convegno SEMT. Le ricerche presentate si basano su alcuni dei quesiti già analizzati nelle pagine precedenti e mostrano come queste analisi possano essere approfondite da un punto di vista statistico e didattico.

Possibili sviluppi

Le analisi compiute in questa tesi relativamente alle differenze di genere hanno mostrato le potenzialità dell'analisi quantitativa di prove standardizzate applicata anche a livello delle singole domande.

Analisi di altre domande, anche a partire da altre coorti di studenti, potranno mettere in luce altri cluster di domande che mostrano un particolare gender gap e sarà quindi possibile proseguire con le analisi sia da un punto di vista dei contenuti, sia relativamente a particolari fenomeni didattici (come misconcezioni ed effetti del contratto didattico) che potrebbero essere alla base delle differenze di genere in matematica.

L'analisi dei singoli quesiti, basata sull'analisi quantitativa dei risultati e l'interpretazione attraverso i costrutti della didattica, può essere ulteriormente approfondita attraverso una indagine qualitativa a posteriori basata su interviste, seguendo lo schema di ricerca utilizzato nell'articolo "Gender differences and didactic contract: analysis of two INVALSI tasks on power properties" (paragrafo 4.4.3). Interviste mirate sui singoli quesiti potrebbero infatti confermare o smentire le ipotesi fatte attraverso l'interpretazione didattica e mettere in luce i processi risolutivi adottati dagli studenti.

Come sottolineato anche in precedenza, gli studi sul gender gap possono intrecciarsi con gli studi sulle variazioni della formulazione dei quesiti. Risulta infatti interessante introdurre all'interno della metodologia proposta in questa tesi per lo studio dell'impatto di una variazione (Capitolo 5), anche quesiti già analizzati che hanno mostrato un forte gender gap che potrebbe essere legato a una particolare formulazione del quesito. In questo modo, è possibile studiare come una variazione di quel quesito possa influenzare il gender gap e su quali livelli di abilità questa variazione abbia influito maggiormente. Questo tipo di analisi può portare a una migliore comprensione delle differenze di genere ma anche a una maggiore consapevolezza nella fase di costruzione dei quesiti.

Infine le ricerche effettuate relativamente agli item di matematica potrebbero essere affiancate da indagini relative ad altri fattori, già esaminati in letteratura, che incidono sul gender gap in matematica. Gli aspetti legati alla sfera psico-sociale, gli affect e i beliefs strettamente legati alla matematica, se indagati nel questionario studente, potrebbero fornire un quadro più completo di questo fenomeno e sarebbe possibile studiare le correlazioni tra questi costrutti e le risposte degli studenti a singoli item o a cluster di item.

In questa direzione, sempre nella sperimentazione *Variazioni 2* (paragrafo 5.3), è stato inserito al termine del test di matematica un questionario mirato per la misurazione dell'ansia matematica.

Inoltre dal confronto del gap di genere e il gap di cittadinanza potrebbero nascere interessanti ricerche. Come già accennato, analisi analoghe a quelle compite per studiare il gender gap, effettuate suddividendo gli studenti in base alla cittadinanza, hanno portato a risultati molto diversi.

Per quanto riguarda la distribuzione degli studenti in funzione del punteggio di Rasch (abilità sull'intera prova) la curva degli italiani e quella degli stranieri sono traslate una rispetto all'altra, evidenziando quindi un gap relativo a tutti i livelli di abilità (come specificato nel paragrafo 4.2.1). Al contrario di quanto osservato per il gap di genere che si concentra soprattutto sui livelli medi e alti.

Inoltre, utilizzando l'indice anche per studiare il gap tra italiani e stranieri in matematica, sulla stessa coorte di studenti sottoposti alle medesime prove, emerge che gli studenti stranieri ottengono risultati inferiori nella quasi totalità degli item e il gap di cittadinanza in matematica è distribuito in maniera più uniforme su tutti gli item.

Questo sta a indicare che le differenze di genere e quelle di cittadinanza mostrano caratteristiche molto diverse, sicuramente a causa della diversa natura del gap e delle differenti cause che ne sono alla base. Il gap in matematica tra italiani e stranieri è sicuramente fortemente influenzato dalle difficoltà di comprensione della lingua italiana da parte degli studenti stranieri che, in maniera più o meno marcata, incide su tutti i quesiti di un test; il gap di genere, invece, è dovuto a una molteplicità di fattori diversi e si presenta in forma disomogenea, con particolarità che necessitano uno studio più approfondito per giungere a una interpretazione del fenomeno.

4.4.4 Gender differences and didactic contract: analysis of two INVALSI tasks on powers properties [23]

Chiara Giberti, Università di Trento
Alessia Zivelonghi, Università di Trento
Giorgio Bolondi, Università di Bologna

The results of standardized tests such as PISA and the Italian INVALSI, point out the existence of a gender gap in mathematics. This gap is deeply studied in mathematics educations literature. In this paper we analyse two INVALSI items of grade 10 in which male and female answers have distinctly different behaviour. Our aim is to observe if this different trend of male and female answers is influenced in particular by effects of didactic contract. In this analysis we integrate quantitative and qualitative methods. The quantitative analysis is based on IRT models and it allow us to highlight the trend of the correct and wrong answers, distinguishing between male and female. The qualitative analysis involves interviews to students and confirm that the choice of a particular response is influenced by didactic contract effects.

INTRODUCTION

The INVALSI tests are national standardized tests administered every year in different grades of primary and secondary schools in order to have systematic checks on students' knowledge and skills in maths and Italian. The increasing importance given to standardize tests such as INVALSI and PISA, provides new opportunities not only in the evaluation of educational systems' performances, but also in the educational field. If 10 years ago the usage of PISA results in mathematics education was still limited (Sfard, 2005), in the recent years, many

[23] Research Report presentato al convegno 40th PME – Conference of the International Group for the Psychology of Mathematics Education. Pubblicato nei proceedings del convegno *PME - International Group for the Psychology of Mathematics Education*, p. 275.

researchers began to use standardize assessments for their studies. For instance, the results of PISA and INVALSI tests showed the existence of a gender gap in mathematics in favour of male and gave the opportunity of study this issue in large populations using also specific statistic tools. In our analysis, we observe that the gender gap is not uniformly distributed on all the items of a test: only some of the tasks present a marked gender gap (in terms of percentage of correct answer of male and female). Moreover, according to recent studies on INVALSI tests (Cascella, 2015), the psychometrical analysis of the item functioning reveals that some items present different performances for male and female.

In this paper, we focus our attention on two INVALSI items of grade 10 in which male and female populations have a strictly different behaviour. We select these two items also because we suppose that two wrong choices (we are dealing with multiple choice tests) are related with didactic contract effects. Our purpose is to investigate if, in this particular case, the gender gap and the different behaviour revealed in the quantitative analysis, can be influenced also by a different response to didactic contract for male and female. The first part of our study is a quantitative analysis of the two items based on Item Response Theory models and evidences the different trends related to gender. The second part is purely qualitative and consists in interviews with the purpose of understanding the processes that led students to choose a particular distractor.

THEORETICAL LENSES

In the recent years, national (INVALSI) and international (such as PISA) assessments pointed out that in mathematics male and female have different performances: boys outperform girls at all school levels and in almost all the countries. This issue has been debated for several years and a large number of studies has focused on the determinants of gender-gap (Forgasz et al., 2010). Standardized assessment, and in particular PISA studies, have given an increasing importance to investigation on gender differences in academic achievement. Various studies highlight the importance of social factors to explain gender-gap in mathematics, evidencing that in more gender-equal cultures this gap disappears (Guiso et al., 2008). This hypothesis is also supported by the fact that this gap is not present at the early stage of school but it raises during the school years. In this research we endorse this idea and, in particular, we assume that education is the main cause of gender differences. A recent study based on PISA 2009 results confirmed this hypothesis and reveals also that "In addition, gender role attitudes within the family environment […] is found to affect girls' performance positively." (González de San Román & De La Rica, 2012).

This research hence fits into a constructivist perspective in which cognitive functions are formed according to the context, and are described as products of social interactions. The learning process cannot be separated from this interactive context defined on the bases of three components: student, teacher and knowledge (Chevallard, 1985). In this paper, we use the idea of didactic contract defined by Guy Brousseau as

behaviour of the teacher expected from the pupil and the behaviour of the pupil expected from the teacher constitute the didactic contract. (Brousseau, 1980).

The didactic contract imposes rules of behaviour and it is the key to analyse the students response to the items analysed in this paper. The relations established between students and teacher, within the *milieu*, could be also studied in detail through other theoretical constructs, such as the concepts of *coutume didactique* and sociomathematical norms (for a comparison of these concepts and an example of how they can network, see Ferretti et al., ; Ferretti, 2015). Moreover our aim is to study if didactic contract, as a product of social context, have a different influence on male and female and, therefore, on gender gap in maths. At first, it's interesting to notice that, as we have already seen for gender gap in mathematics, also didactic contract seems to be not present in pre-scholar pupils (Baldisserri et al., 1993) but it origins in primary school.

METHODOLOGY

In this research, we use both quantitative and qualitative analysis. The first part of the analysis is purely quantitative and give us the opportunity to observe the behaviour of the items in a large-scale assessment and to make assumption about that behaviour. The second part of the analysis has the purpose of validate these assumptions through interviews to a restricted group of students.

Quantitative analysis

The two items investigated in this study belong to 2011 and 2012 INVALSI tests of grade 10. For each test, the data analysed are those of INVALSI's sample. This sample consisted of approximately 40.000 students and it is representative of the population of Italian grade 10 students. The INVALSI team proved the consistency of both tests by using the Classical Test Theory tools and made a first analysis using IRT models and in particular the Rasch Model (INVALSI, 2012b; Rasch, 1960). The Rasch model is a simple logistic model and it is useful to analyse a standardized test such as INVALSI because it allows joint estimation of two kind of parameters: a difficulty parameter for each item and an ability parameter for each student. More specifically, this model express the probability of choosing the correct answer in an item as a function of the item's difficulty and the ability of the students in the whole test and this function is called *Item Characteristic Curve*. In this way, it is possible to use Rasch parameters to represent also the empirical data and, in particular, we can represent the trend of each possible response as a function of the students' ability. Those specific graphs are named *Distractor Plots*.

In the INVALSI National Annual Report 2011, gender differences in math tests are identified on the basis of total medium score observed for male and female (INVALSI, 2011). This gap is perceived in both of the tests analysed and it is statistically significant (INVALSI, 2011; INVALSI, 2012a).

Starting from these INVALSI results and the same dataset, we compare percentages of male and female answers, then we use the Rasch Model to study distractor plots for male and female separately. Distractor plots allow us to study gender differences in relation with the ability level of the students and, in particular, we can observe if there are differences not merely in choosing the right answer but also in trends of the incorrect ones.

Qualitative analysis

The second part of our study is purely qualitative and consists in interviews of a restricted group of students about the items analysed in this paper. For this purpose, we administer in two classes of the same high school a brief questionnaire, consisting in 5 mathematical items including one of the two items studied in this research (Fig. 3). The other items are designed to contextualize the studied task into a mathematical test and to evidence if a student face up the test seriously. Just after the correction of the questionnaire, we select 22 of the 49 students for the interview. We select them on the basis of their response to the task studied and their maths score provided by the teacher. In particular, we choose to interview principally students good at maths (school mark > 6.5/10) who didn't answer correctly to the item. The interviews are semi-structured, task based and in couples. We decide to interview together students that had selected different options and ask them to explain to the classmate the reasons of their decision. At a later stage, we present them the other item and ask to compare it with the first one. Each interview takes about 20 minutes and is audio taped. At last, we transcribe the interviews and analyse the transcriptions.

ITEMS ANALISYS

In this research, we focus our attention on two similar items: the question intent is the same, both concern the same content (powers properties) and the answers are analogous. Both tasks are multiple-choice questions with only one correct answer but one is set into an algebraic context (Fig. 1) and the other into an arithmetical context (Fig. 3). Moreover, in both items we register a remarkable difference in male and female performances.

The expression $a^{37} + a^{38}$ is also equal to

A. ☐ $2a^{75}$

B. ☐ a^{75}

C. ☐ $a^{37}(a+1)$

D. ☐ $a^{37\cdot 38}$

Figure 1: Item from the grade 10 INVALSI test administered in 2012 [1]

The correct answer is C and it is chosen only by 35% of students. Option A, in which the base is the sum of the basis and the exponent is the sum of exponents, is chosen by 19% of students. Option B and D are similar because the resulting power has the same base of the original ones but the resulting exponent is the sum of two exponent in option B and the product in option D. The 26% of the respondents select answer B, which is the most attractive wrong answer and option D is chosen by 16% of students. In addition, only 3% of students do not respond to this question and this may mean that students are fairly confident about their answers. In the table below, we can also observe that male responded better than female: 38% of male give the right answer compared with 31% of female.

	Total	Male	Female
A	19 %	19 %	20 %
B	26 %	27 %	26 %
C	**35 %**	**38 %**	**31 %**
D	16 %	14 %	19 %
Missing	3 %	3 %	3 %

Table 1: Results of item (Fig. 1) from INVALSI test administered in 2012.

Observing the table (Tab. 1) of percentage, we can also see that the response D is more attractive for girls. These and others particularities of these item responses are more visible using the results of Rasch analysis to graph distractor plots.

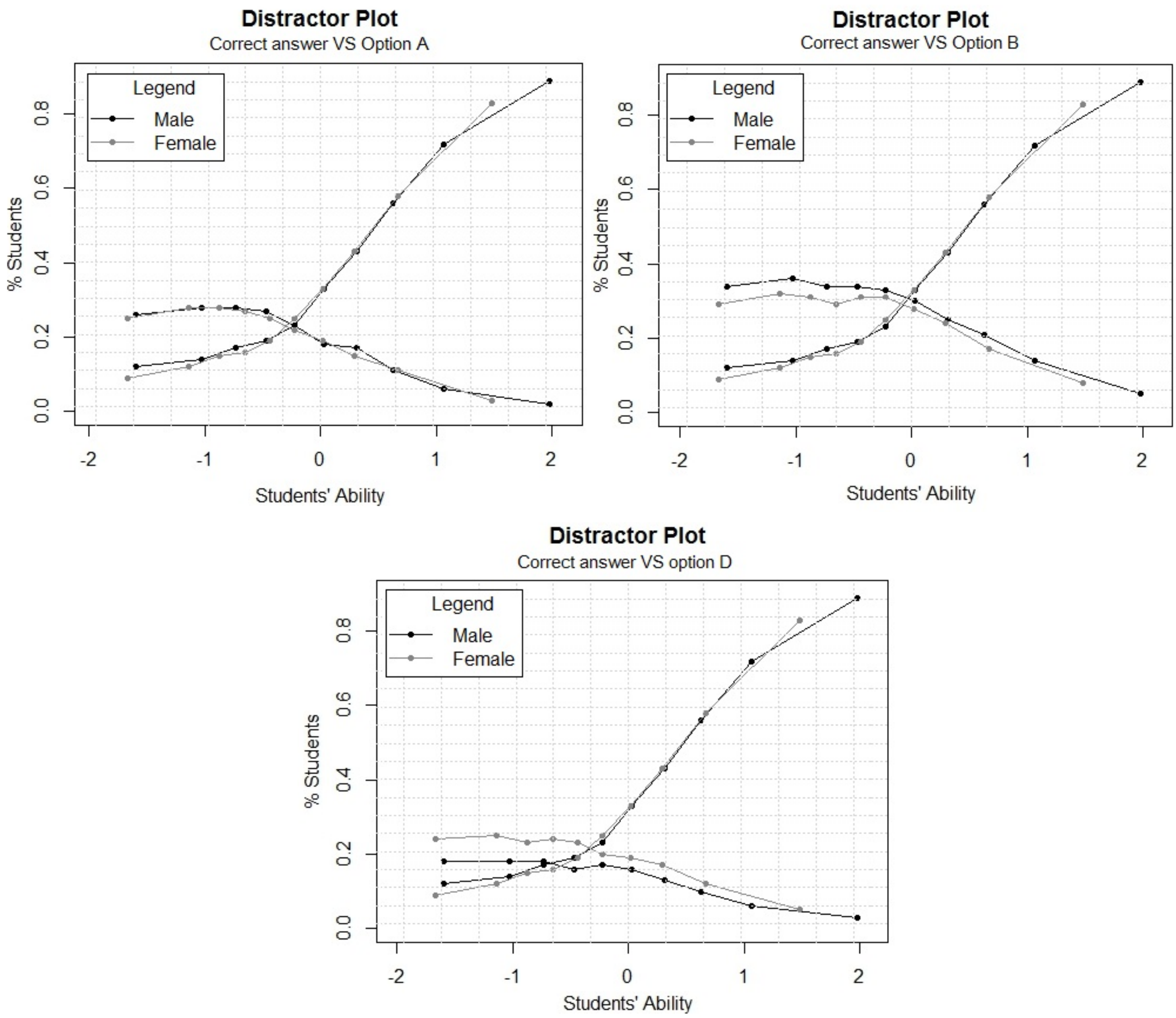


Figure 2: Distractor plots of the first item [1]: comparison between the right answer and the other options (the right answer is the one with increasing trend).

Distractor plots reveal that the trend of the correct answer is almost the same for male and female, although the different percentage seen before. This means that girls and boys with the same ability level choose the correct response with the same percentage, the gender gap observed before in the correct answer (Tab. 1) arises from the fact that female reaching highest ability levels are fewer than male. In Figure 2 we can notice that also the trend of the answer A is almost the same for both male and female but the differences are evident in the behaviour of the other two options. As we observed before, answer D is more attractive for female at all ability levels and, obviously, especially for the lower ones in which it is chosen by a higher percentage of students. Furthermore, boys at all levels of competency prefers the response B compared to girls.

These results become more relevant compared with those of the second item analysed. Indeed, the evidence observed before are all confirmed analysing the second item (Fig. 3) in which the sum of power is analogous but given in an arithmetical context.

The expression $10^{37} + 10^{38}$ is also equal to

☐ A. 20^{75}

☐ B. 10^{75}

☐ C. $11 \cdot 10^{37}$

☐ D. $10^{37 \cdot 38}$

Figure 3: Item from the grade 10 INVALSI test administered in 2011.

This second item results more difficult than the first one and the percentage of correct answers is only 22%. Furthermore, the gender-gap is still present and more relevant than before: only 18% of female choose the correct answer in comparison with the 27% of male. It is very interesting to notice that the trends observed in the distractor plots of the first item, are the same that we can notice in this question, despite the fact that the first one has an higher percentage of correct answers.

STUDENT'S INTERVIEWS

The interviews reveal that most of the students facing with these items are immediately led to identify and apply some kind of rule, in particular those who choose a wrong option. This is a typical didactic contract behaviour. Many students say clearly that, when they see two powers with the same bases they directly think to have to apply powers properties and the reason of this behaviour is inherent into didactical practice:

1 S1: I had in mind powers rules, I was confused. I didn't even think to option C.

2 I: So when you see powers you immediately think...

3 S1: I think to rules. I think: "It will be something with rules".

4 I: Why?

5 S1: I don't know, it is what we've always done in the exercises.

Students who choose options B and D remember more or less correctly power properties and for this reason, they expect that the result have the same basis. Then they exclude option A because of the base. Moreover, we observe that options B and D are attractive also for students who know properly powers' properties. In the interview below, for example, the student remember the properties for the product and therefore he excludes option B, but even more he is led to find some other rules to solve the exercise.

1 I: Why students choose $10^{37 \cdot 38}$?

2 S2: Maybe there was a rule according to which to sum powers with the same basis you have to multiply the exponents.

3 I: And this one? Why did they select 10^{75}?

4 S2: Also for this one because there was a rule that says to sum the exponents, but this (rule) is when there is a product! This (answer) is not right because this rule is valid when there is a product!

Students who choose options B and D, often explain their decision on the basis of some kind of rules that derive from their school experience and, in particular, from their relation with the teacher and the milieu habits. Those responses can be related with didactic contract: students choose answers B and D believing that when they solve an exercise that include powers with the same bases they have to apply powers' rules. This behaviour is also observed in students who know powers properties. Moreover, we observed that this attitude belongs to the classroom habits and routine, therefore we can refer this phenomenon to didactic costume seen as the habits picked up in didactical practice during mathematical lessons.

CONCLUSIONS AND FUTURE PERSPECTIVES

This paper presents a study of gender-gap in mathematics from a two-fold point of view: we study the behaviour of male and female facing a mathematical task using specific statistic tool and, in particular, distractor plots. The results of standardized tests analysed using Rasch Model enable us to observe gender differences not only in the whole test, but also focusing on a particular item, objet of our study. Moreover, distractor plots evidence the different behaviour of male and female in choosing each possible answer related their ability in the whole test. The two items analysed in this paper presented the same interesting evidences comparing male and female performances. The gender-gap (in terms of percentage of correct answer) is remarkable in both the tasks. Moreover, distractor plots reveals that, in both the items, Option A have the same trend for male and female, male prefer Option B and female prefer option C at all levels of competency, also for the highest ones. The interviews allow us to interpret these results on the base of students responses and underline that answer B and D can be explained using the construct of the didactic contract. Indeed, the students interviewed always refer to classroom practice, relation with the teacher and the milieu habits. Integrating all these information, we notice that the gender gap, in these particular tasks, is influenced by didactic contract effects: indeed, Option A and the missing percentages are the same for male and female in both the items and therefore the gap in the correct answer percentages, is due to options B and C which are related to didactic contract. We also assume that, in this tasks, male and female are influenced in a different way by didactical contract because, even though both options B and D are related with this construct, the first is preferred by male and the second by female. This different behaviour of male and female could be analyse deeper in future studies, including more interviews.

Finally, the structure of this research could be used also to analyse other items or to study differences not only between male and female but also for other groups of students. Indeed, the Rasch analysis and, in particular, the study of distractor plots shall can provide numerous others evidences that can be interpreted using qualitative analysis and interviews.

References

Baldisserri, F., D'Amore, B., Fascinelli, E., Fiori, M., Gastaldelli, B., & Golinelli, P. (1993). I palloncini di Greta. *La matematica e la sua didattica*, 4, 444-451.

Brousseau, G. (1980a). Les échecs électifs dans l'enseignement des mathé-matiques à l'école élémentaire. *Revue de laringologie otologie rinologie*, 3, 107-131.

Cascella, C. (2015). Male and Female Performance in Mathematics: Empirical Evidence from Italy. *The International Journal of Interdisciplinary Educational Studies*, 9(3-4), 1-9.

Chevallard Y. (1985). *La transposition didactique. Du savoir savant au savoir enseigné.* Grenoble: La Pensée Sauvage.

Ferretti, F, Gambini, A., & Bolondi, G. (in press). The age of the earth effect: a situation of didactic contract. *13th International Congress on Mathematical Education* (ICME-13), Hamburg, Germany.

Ferretti, F. (2015). *L'effetto "età della Terra". Contratto didattico e principi regolativi dell'azione degli studenti in matematica* (Doctoral Dissertation, University of Bologna, Italy). Retrieved Genuary, 2016, from http://amsdottorato.unibo.it/7213.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *SCIENCE-NEW YORK THEN WASHINGTON*, 320(5880), 1164.

González de San Román, A., & De La Rica, S. (2012). Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes. *IZA Discussion Paper,* 6338*, Institute for the Study of Labor.*

INVALSI, (2011). *Rilevazione nazionale degli apprendimenti 2010-2011. Le rilevazioni degli apprendimenti.* Retrieved Genuary, 2016, from http://www.invalsi.it/snv1011/documenti/ Rapporto_SNV%202010-11_e_Prova_nazionale_2011.pdf .

INVALSI, (2012a). *Rilevazione nazionale degli apprendimenti 2011-2012. Quadro di sistema.* Retrieved Genuary, 2016, from http://www.invalsi.it/snv2012/documenti/ Rapporti/Rapporto_rilevazione_apprendimenti_2012.pdf .

INVALSI, (2012b). *Rilevazione nazionale degli apprendimenti 2011-2012. Rapporto tecnico.* Retrieved Genuary, 2016, from http://www.invalsi.it/areadati/concidee/ Rapporto_Tecnico_Prove.pdf .

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Denmarks Paedagogiske Institut.

Sfard, A. (2005). What Could be More Practical than Good Research?. *Educational Studies in Mathematics*, 58(3), 393-413.

Forgasz, H., Becker, J., Lee, K., & Steinthorsdottir, O. (2010). International perspectives on gender and mathematics education. *Charlotte, NC: Information Age Publishing.*

4.4.5 Highlights on gender gap from Italian standardized assessment in Mathematics[24]

Bolondi Giorgio – Università di Bolzano

Cascella Clelia – INVALSI

Giberti Chiara – Università di Trento

Abstract

In this paper we present a work in progress study on math gender-gap analysing Italian standardized tests. The methodology explained is designed to study gender gap in Maths not only in the whole test but more deeply on specific items. The analysis of Item Characteristic Curves (ICCs) plotted both in male and female group on the basis of Rasch estimates, is useful to examine test of various grades and also to compare similar items of different grades. In this paper we focus on primary students and we provide two examples from grade 5 and 6 of items analysis integrating statistical enquiry and didactical interpretation of students' responses, for the purpose of revealing the potentiality of this approach in studying gender differences in math.

Introduction

Gender gap in math

The issue of gender gap in math education is deeply studied since the last century and, in recent years, the increasingly important role of national and international assessments have given new emphasis on this theme (Leder & Forgasz, 2008). The results of international investigation such as PISA 2015 and TIMSS 2015 confirm that in mathematics male and female have different performances: boys outperform girls at all school levels and in almost

[24] Research Paper presentato al convegno 20th SEMT – International Symposium Elementary Mathematics Teaching (Agosto 2017). Pubblicato nei proceedings del convegno.

all the countries (OECD, 2016; Mullis et al., 2016). In both this tests administered in 2015, Italy is one of the country in which the gender gap is more remarkable and this gap is also confirmed by results of national standardized test, called INVALSI tests (INVALSI, 2016; Di Tommaso et al., 2016).

Gender difference in math performances have been debated in several studies and plenty of researches has focused on the determinants of gender-gap (Forgasz, 2010). As reported by Winkelmann et al. (2008), we identify both external and internal factors that can be a cause of gender differences in maths achievement. Internal factors include biological variables considered responsible for gender gap in math in some studies, but this hypothesis needs to be overcome because international assessment have revealed that gender gap in math differs enormously across countries (Di Tommaso et al., 2016; OECD, 2016). Possible, yet-to-be-confirmed, biological, internal factors must be accompanied with other explanations connected to social and cultural factors, related to the context in which the students live. In this perspective, many researchers highlight the importance of social and cultural reasons, evidencing that in more gender-equal cultures this gap disappears (Guiso et al., 2008; OECD, 2015; Cascella, 2017). Furthermore, beliefs of teachers and parents about boys and girls math abilities, parental expectations and gender stereotypes play an important role in students' self-perception and then have a huge influence on their performances (Jacobs & Bleeker, 2004; Riegle-Crumb, 2005; Freyer & Levitt, 2010). For instance, many studies found that the role of the mother in the family and in the society is strictly related with female performances in math tests (Freyer & Levitt, 2010; Jacobs & Eccles, 1992; González de San Román & De La Rica, 2012).

Several studies have also shown differences between boys and girls in metacognitive aspects related to maths that have a negative impact on their performances: for instance, girls tend to be more influenced by math anxiety and display less math self-efficacy (Cargnelutti et al., 2016; OECD, 2016; Pajares, 2005).

At last, also factors strictly related to the school context seems to be a possible explanation of the gender gap in math. Leder (1992), in order to explain gender differences reflects also on "curriculum variables, like content areas of mathematics, types of the items and method of assessment and instruction" and more recent studies have shown that not only mathematics curriculum but also classroom practices and assessment practices, educational methods have a huge impact on gender gap in math (Leder & Forgasz, 2008; OECD, 2016; Giberti et al., 2016). In this perspective, many studies have shown that male and female use different strategies in problem solving activities: girls usually prefer routine procedures, well known algorithm and conventional strategies, whereas boys are less afraid of making mistake and try new methods and unconventional approaches (Bell & Norwood, 2007; Gallagher et al., 2000; Gould, 1996; Fennema & Carpenter, 1998). Also in this case, the reasons of differences in problem solving strategies seems to be not so much related to biological differences between male and female but mostly to stereotypes, such as 'good girls follow the rules' (Langer, 1997) and to pedagogy, school system and classroom practice (Boaler, 1997).

As we have shown there is not only one possible explanation of gender gap in mathematics, but there are numerous factors that influence this topic. In this paper, we endorse the idea that gender gap in math achievement is particularly influenced by cultural and social causes and also micro-social factors related to the mileu habits and classroom practice. This hypothesis, in opposition to biological causes, is also supported by the fact that this gap is not present at the early stage of school but it raises during the school years.

Math gender gap and large scale assessment

The issue of gender gap in maths is deeply studied in mathematics education from both a qualitative and a quantitative point of view. Many studies focus on small group of students and try to understand gender differences on specific tasks thought interviews but most of the recent researches on gender gap in maths are based on national and international standardize tests, such as PISA and TIMSS. Indeed, using the results of these tests, it is possible to analyse gender gap in several ways, for example many studies compare the relevance of the gap in different countries and evidence where this gap is more remarkable. Standardized assessments give the possibility to study this issue on a large population of students and from a statistical point of view. In this paper we study gender differences in Italian standardized tests (INVALSI tests) using particular statistical tools to evidence gender-gap not only in the whole test but also evidencing the items in which this gap is more remarkable and focusing the analysis on these specific items. In the INVALSI reports (INVALSI, 2016) we observe gender differences on the basis of total medium score on the whole test and in the same way the gender-gap is studied in PISA reports (OECD, 2016). This information is important to signal the existence of a gap between male and female performances in maths and this is also useful to make comparison between different geographical areas and to analyse the relation between gender gap and social and cultural factors. Nevertheless, if we want to make assumption on the origins of this gap and analyse it with the lenses of the maths education theories, it is essential to analyse this gap on the single items of the test.

Are male better than female in all mathematical items? Are there some items in which the gender gap is more remarkable than in the others? In this case, it could be interesting to focus on the items in which the gender gap is more evident and make a deeper analysis from both a statistical and a didactical perspective. This kind of analysis seems to be able to provide at least two main research products: 1) an insight into the classrooms in order to ascertain possible links between didactical protocols and skills development in males and females; and, 2) individuation of recurring items' features that cause item bias in sub-groups of students clustered by gender. Moreover, the latter is undoubtedly be useful in order to construct *standardized* achievement test that, as well known, in accordance to Item Response Theory approach frequently used in education research, must guarantee measurement invariance. Although undoubtedly a lot of different factors that might affect gender gap exist, exploring the relationship between gender gap in Math and/or in STEM and socio-cultural and economic environment within which students grow up is an ongoing issue timely topic in EU agenda but it is outside the scope of this work. Nevertheless, we claim that the observation of gender gap for different ability levels and the exploration of possible links between didactical protocols and differences in male and female test scores can be absolutely useful in order to formulate new research hypotheses and to provide new answers to the questions that our results might get emergent.

Methodology

Data

Our analysis is conducted using answers given by Italian students to Math achievement test developed by the Italian National Institute for the evaluation of Educational Systems. We analysed several of the 14000 items of that make up the INVALSI database but in this paper we give two examples of item analysis. Questions in Mathematics test cover the subareas of "space and shape" (roughly geometry), "change and relationship" (algebra), "quantity" (arithmetic) and "uncertainty" (probability), in a range of difficulty from those that require simple mathematical operations to those that require complex thinking. The mathematics scores have been estimated by using the Rasch model and have been scaled in an empirical range equal to [-4; +4], i.e. the latent trait along which both items and persons are scaled depending on their difficulty and their ability respectively.

Measurement bias and differential item functioning in psychometric literature: an overview of our methodological procedure

The wide use of standardized achievement tests in Social and Human Sciences has put the spotlight of psychometrical and educational literature on the treatment of item bias. The term bias, when used to describe mental tests, has a specific technical meaning: it is defined as a systematic error in measurement process that can alter the expected items' psychometrical properties and cause a consistent distortion of a statistic, i.e. for achievement tests, unfair person and items estimates. In other words, item bias refers to differences in the way a test item functions across sub-groups of persons matched on the attribute measured by the test (e.g. ability) and stratified by gender, socio-cultural background and/or any other variables not explicitly hypothesized by the statistical model used to analyse data (Osterlind, 1983; Camilli 2006; Camilli and Shepard 1994; Holland and Wainer 1993; Embretson & Reise, 2000; Penfield and Camilli 2007; Osterlind 2009). Therefore, when the same item shows different behaviour in different sub-groups of persons, it is biased and it is a grave violation of Item Response Theory (IRT) models, frequently used to assess students' ability. In particular, the Italian National Institute for the Evaluation of Educational System employs the Rasch model, according to which the probability of a correct answer depends on students' relative ability, i.e. his/her ability compared to item's difficulty: no other variables (such as for example students' socio-demographical characteristics) can affect this probability. Therefore, according to the mathematical mechanism underlying the model, an achievement test developed to be analysed by the Rasch model has to be constructed to avoid any possible effects exerted by each other possible variable.

In the sectorial literature, term bias refers to the whole student's test scores and thus it has a more general meaning compered to *Differential Item Functioning* (DIF). DIF refers to each single item and to item behaviour in sub-group of students *matched on ability*. In order to control item functionality, a lot of different techniques to detect DIF have been proposed, such as, for example, Mantel-Haenszel procedure and its follow-up consisting in the implementation of a revised version of chi square (Osterlind, 2009, pp. 29-37). Some other examples are: 1. the "areas measure (Rudner, 1977; Rudner, Getson, & Knight, 1980; Linn & Harnisch, 1981; Shepard, Camilli, & Williams, 1985); 2. the delta plot method (Angoff & Ford, 1973); 3. the Draba statistic (of Wright, Mead, & Draba, 1976); 4. the Wald statistics (Lord, 1980); the likelihood ratio methods (Thissen, Steinberg, & Wainer, 1988; Thissen, Steinberg, & Gerrard, 1986); and so on. Many studies carried out on both real and simulated data have showed the efficacy of these techniques to detect DIF in small or medium sample. For larger dataset, unfortunately, the number of useful techniques decreases significantly. In the present study, we carried out a DIF analysis within Item Response Theory (IRT) framework, based on the comparison of the *Items' Response Function* (IRF) that links the probability of a correct answer to student's ability. In our study, we have two IRFs, one for males and one for females. Comparing them, we can compare item functionality in both male and female sub-group of students scaled along the same latent trait, i.e. in each sub-group, we have the same number of students with the same ability level. Figure 1 provides an example of this: 1. The first graph represents

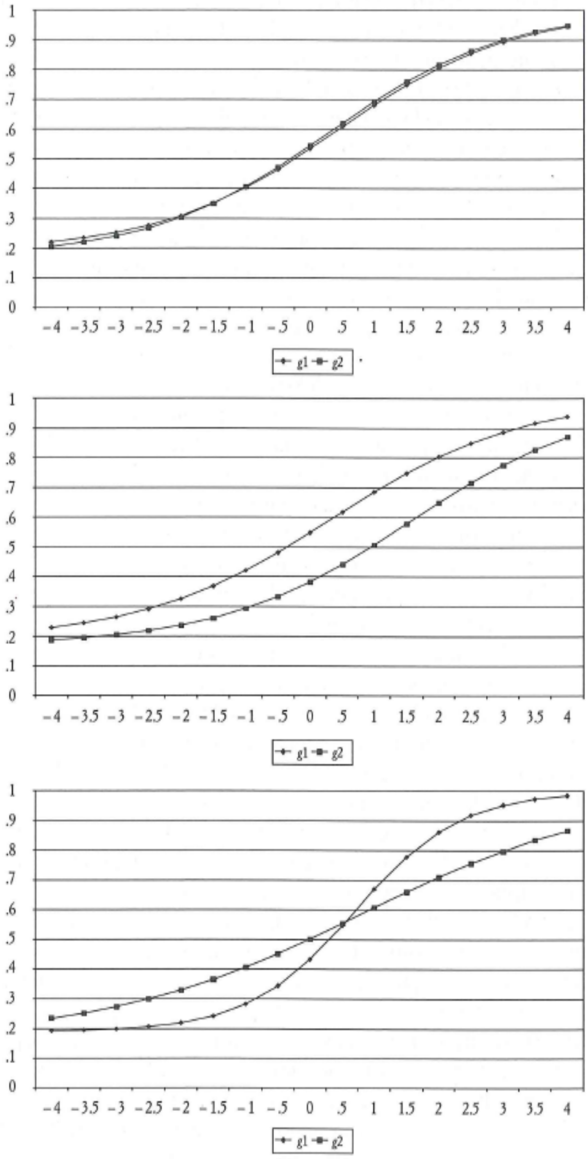


Figure 1 - Some examples of items without bias (at the top), with uniform bias (in the middle) and with not uniform bias (ad the bottom) (Barbaranelli, Natali, 2005, 175)

a not biased item (IRFs in both group have the same behaviour); 2. The second graph shows a uniform DIF between students in group g1 e in group g2 (item functionality is different in these two groups because the probability of a correct answer is higher in the first group compared to the second one, in each ability level); and, finally, 3. The third graph represents a not uniform DIF (the probability of correct answer is higher in one group relative to the other one only for specific ability level).

The DIF analysis based on the comparison of IRFs provides very useful information: 1. It identifies advantaged versus disadvantaged sub-group of students; and, 2. It locates along the latent trait students' probability of a correct answer in each sub-group. Unfortunately, many statistical packages do not plot also distractors' behaviour in each sub-group of students. Distractor plot is a graph plotted only for multiple choice item that shows the functionality of an items' distractors. For this reason, we carried out three analysis: first at all, we analysed male and female answers all together (DIF analysis); then we analysed them separately and compared distractor graphs plotted for males and females.

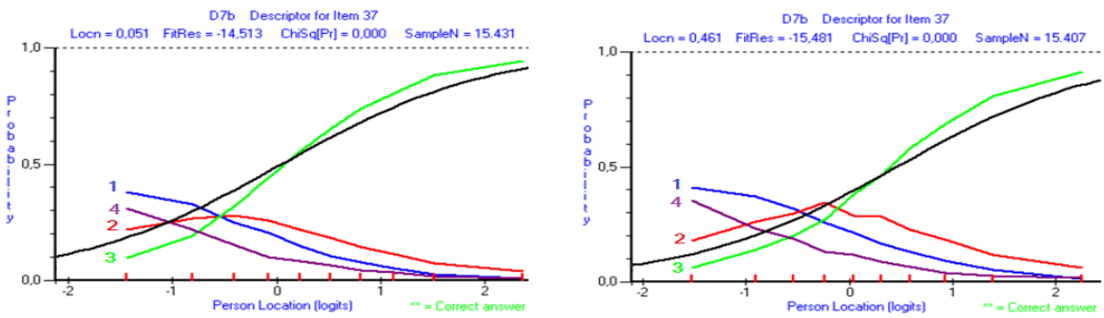


Figure 2 – An example of how compare two distractor plots. Analysis carried out on data collected by INVALSI in 2015.

As shown in figure 2, distractor graphs plotted for males and females show the characteristic curve of each item's distractor, using the same Rasch scale: By comparing them, we can immediately compare the probability that a male or a female, *with the same ability level*, has to choose a distractor. Each distractor has a different attraction power that is inversely proportional to the student's ability level. In fact, the mathematical mechanism underlying the Rasch model guarantees the cumulativity of each item: the attractivity of each wrong option decreases when student's ability increases. Therefore, the graphical inspection of distractors' characteristic curve might reveal students' misconception as well as specific their weaknesses that might be very useful to formulate specific research hypotheses as done in our research.

Gender gap index

As regards measurement bias, the first way to identify the entity of the gender gap on each item is to compute the difference between the percentage of correct answer of male and the percentage of correct answers of female. In this way we observe which items contribute most to creating the gender gap on the whole test: in fact if the measure of gender gap on the test is on the bases of total medium score, it derives directly from the raw scores and hence to the percentage of correct answers. The problem of considering the gender gap of an item in this way is that a difference of 10 point percentage on an easy item in which the percentage of correct answer is around 80% is the same of a difference of 10% on a very difficult item in which only the 30% of students choose the right answer. To take into account also the difficulty of the item, we create a specific index to evidence the size of gender gap on each item.

$$I_k = \frac{M_k - F_k}{P_k}$$

In which:

M_k is the percentage of correct answer for males to the item *k*

F_k is the percentage of correct answer for female to the item *k*

P_k is the percentage of the whole population to the item *k*.

In this way we can observe that the items with a positive value of the index are those in which male outperform female and the items with a negative index are those in which female results are better than male's ones. The index gives also an evaluation of the magnitude of the gap considering also the difficulty of the item analysed and items selected in this way can be than studied in deeper using DIF and the other procedures described above.

Results

In our research we analyse INVALSI tests using the statistical tools described in the previous section. In particular, we focus our attention on four INVALSI test belonging to different levels and different years: level 2 of 2009, level 5 of 2012, level 6 of 2013 and level 8 of 2015. These tests are particularly interesting because they were administered to different sample of students but representative of the same population, growing during the years. In this way it is possible also to observe the changing of a specific population of students at different stages of their schooling.

First, we used gender-gap-index to point out which items highlight mostly gender gap. In this way it is possible to notice that the gender gap is not distributed on all the items of a test, but it's most marked on some of the items and there are also some items with no gender differences in favour of male and few items in which female have better results. It is interesting to compare the results of this first analysis with the results of DIF analysis compute on the same tests. DIF analysis point out not merely which items create a higher gap, but also if this gap is more remarkable for particular ability levels of the students.

In the next section we focus on two tasks belonging to tests mentioned above. Both tasks refer to the same math content that is comparison between decimal numbers but the kind of task, the context and the level of the test are different.

Item analysis

The first task belongs to the INVALSI test of level 5 administered in 2012.

D21. In the final of an international competition of floor gymnastic, the scores obtained overall by the athletes are the following:

Nation	Score
Austria	68,8
Croatia	71,8
Finland	72,0
Japan	68,08
Greece	60,8
England	69,8
Italy	80,12
United States	80,2
Sweden	70,2
Switzerland	78,1

a. What nation ranked first? ....

b. What nation ranked fourth? ....

Figure 3 – Items 21a-21b from the grade 5 INVALSI test administered in 2012 and results of the items (9-10 years old students).

ITEM A	ALL	FEMALE	MALE
CORRECT	37%	32%	43%
WRONG	61%	66%	56%
MISSING	2%	2%	2%
INDEX	28%		

ITEM B	ALL	FEMALE	MALE
CORRECT	67%	63%	70%
WRONG	31%	35%	28%
MISSING	2%	2%	2%
INDEX	11%		

This question is composed by two items and in both of them we observe a significant gender-gap favouring boys in terms of percentage of correct answer. At a first glance, both items require to order decimal numbers but the first item is much more difficult than the second one and moreover the first item point out a wider gender gap. There are many studies concerning decimal numbers that evidence misconceptions related to this issue and, in particular, to comparison between decimal numbers. Misconceptions concerning decimal numbers are frequent and often arise during the transition between natural numbers and decimal numbers: many students tend to continue applying procedure and properties valid for natural numbers also when they operate with decimals (Sbaragli, 2012). To solve the first item, students have, in particular, to compare two decimals with the same integer part: the countries with better results are Italy (80,12 points) and United States (80,2 points). Many students in this case compare the decimal part of the two numbers and since 12>2, they concluded that 80,12>80,2. Observing the percentage for the first item we note that less than 40% of students give the right answer and this may confirm that this is a widespread misconception. It is also clear that girls are more influenced than male: only the 32% of female answer correctly compared to the 43% of male.

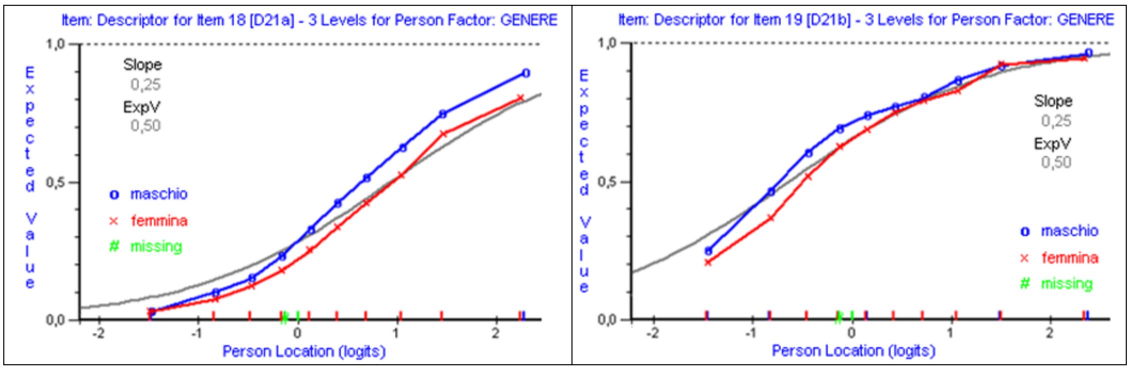


Figure 4 – Item D21a administered by INVALSI to pupils attending the 5th grade of primary school, in 2012 (9-10 years old).

Results of DIF analysis confirm the existence of a gender-gap in item 21a that is null in the lower decile but increases with the ability of students (Fig. 4). This means that the different influence of misconception between male and female is deeper for high ability levels: with increasing ability-levels male are better at getting over barriers given by misconceptions, while girls struggle most.

Even if the question of the second item seems similar to first one, we observe that the second item is less difficult (67% of correct answer, INDEX = 11%), the gender-gap in terms of percentage of correct answer still exist but we can see that the distribution of this gap is strictly different. In fact, DIF analysis shows that gender-gap is focused on medium and low ability levels and there is no gender differences for higher ability levels. This different behaviour of the two items analysed can be explained observing that both items ask to compare decimal numbers but, in the second case, it is possible to get around this difficulty and relative misconceptions: students of high ability levels do not consider decimal numbers but only their integer part, because they realize that it's enough to find the fourth ranked. In this way, the procedure applied to answer is changed and, since the misconception existing in the first item no longer influences, also the gender-gap is filled. Furthermore, analysing the fit of the first item, we notice that the model overestimates lower levels and underestimates higher ones and this misfit of the item might be related with the influence of the misconception. This hypothesis becomes even more interesting analysing the fit of the second item: we observe that for lower ability levels the model overestimates the results but for medium and higher levels, when we assume a change of procedure, the fit is optimal, the gender-gap disappears.

The other task analysed belongs to INVALSI test of level 6 administered in 2013 and, even there, students have to compare decimal numbers.

D23. In which of the following groups numbers are arranged in ascending order?

A. 3,5 ; 3,043 ; 3,28 ; 3,124

B. 3,5 ; 3,28 ; 3,124 ; 3,043

C. 3,043 ; 3,5 ; 3,124 ; 3,28

D. 3,043 ; 3,124 ; 3,28 ; 3,5

Figure 5 – Item 23

ITEM	ALL	FEMALE	MALE
A	14%	15%	12%
B	17%	20%	15%
C	8%	8%	9%
D	**57%**	**54%**	**61%**
MISSING	4%	4%	3%
INDEX	12%		

from the grade 6 INVALSI test administered in 2013 and results of the item.

The results point out that difficulty related to this content still exist also at level 6 and this item (also compared with other items belonging to other tests) give us an endorsement of the fact that girls tends to have more difficulty than male in comparing decimal numbers and are more influenced by the misconception explained before. Comparing the percentage of each answer we can observe that female are more attract than male by answer B: in this case the first three numbers are ordered following the misconception by which 3,5<3,28<3,124 because 5<28<124. Deeper informations not merely on the right answer but also on the trends of the incorrect ones can be sourced using DIF-distractor plots: we observe that the difference in selecting option B is particularly due to the lowest ability level in which female have a probability of choosing this answer slightly less than 50%. Moreover, also in this case, the model overestimates lower levels and overestimates the higher ones.

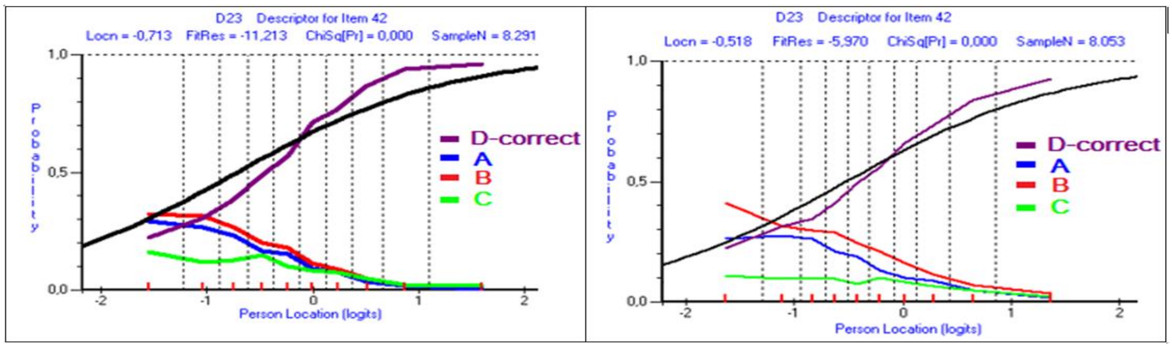


Figure 3 - Male (left plot) and female (right plot) distractor graphs plotted for the item D23 administered by INVALSI to students of grade 6 in 2013 (10-11 years old students)

Further perspectives

In this paper we have reported examples of the methodological strategy we are adopting to analyse around 1400 Math items administered from 2008 up to now. Focusing on two items of grade 5 and 6, we used this method to observe gender differences related to a specific misconception in comparing decimal numbers. We expect that the analysis carried out on the entire database will allow us to identify recurring features of Maths items influencing gender gap and will give us the possibility to produce a first mapping of these items' characteristics (type, mathematical contents and context). We believe that, by combining statistical analysis and didactical interpretation of students' cognitive processes, it will be also possible to point out some possible causes of gender differences in Maths.

References

Bell, K. N., & Norwood, K. (2007). Gender equity intersects with mathematics and technology: Problem-solving education for changing times. *Gender in the classroom*, 225-258.

Boaler, J. (1997). Reclaiming school mathematics: The girls fight back. *Gender and Education*, 9(3), 285-305.

Cargnelutti, E., Tomasetto, C., & Passolunghi, M. C. (2016). How is anxiety related to math performance in young students? A longitudinal study of Grade 2 to Grade 3 children. *Cognition and Emotion*, 1-10.

Cascella, C. (2017). "Male and female social roles in daily file, pupils' perceptions and gender gap in Math test scores. Some empirical evidences from Italian primary school. (Forthcoming)

Che, M., Wiegert, E., & Threlkeld, K. (2012). Problem solving strategies of girls and boys in single-sex mathematics classrooms. *Educational Studies in Mathematics*, 79(2), 311-326.

Di Tommaso, M. L., Mendolia, S., & Contini, D. (2016). The Gender Gap in Mathematics Achievement: Evidence from Italian Data. *IZA Discussion paper*, n.10053, Bonn.

Embretson, S. E. and Reise, S. P. (2000), Item Response Theory for Psychologists, Mahwah, NJ: Lawrence Erlbaum Associates.

Fennema, E., Carpenter, T. P. (1998). New perspectives on gender differences in mathematics: an introduction. Educational Researcher, 27(5), 4-5.

Forgasz, H. J. (2010). International perspectives on gender and mathematics education. IAP.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-240.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. Journal of experimental child psychology, 75(3), 165-190.

Giberti, C., Zivelonghi, A., & Bolondi, G. (2016). GENDER DIFFERENCES AND DIDACTIC CONTRACT: ANALYSIS OF TWO INVALSI TASKS ON POWERS PROPERTIES. 40th PME proceedings, 275.

González de San Román, A., & De La Rica, S. (2012). Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes. *IZA Discussion Paper*, 6338, *Institute for the Study of Labor.*

Gould, S. L. (1996). Strategies used by secondary school students in learning new concepts which require spatial visualization. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *SCIENCE-NEW YORK THEN WASHINGTON*, 320(5880), 1164.

INVALSI, (2016). *Rilevazione nazionale degli apprendimenti 2015-2016. Le rilevazioni degli apprendimenti.* Retrieved March, 2017, from http://www.invalsi.it/invalsi/doc_evidenza/2016/07_Rapporto_Prove_INVALSI_2016.pdf

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter?. *New directions for child and adolescent development*, 2004(106), 5-21.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers’ gender-role stereotypic beliefs on mothers’ and children`s ability perceptions. *Journal of personality and Social Psychology*, 63(6), 932-944.

Langer, E. J. (1997). The power of mindful learning.

Leder, G., & Forgasz, H. (2008). Mathematics education: new perspectives on gender. *ZDM – The International Journal on Mathematics Education*, 40(4), 513-518.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. *TIMSS & PIRLS International Study*.

OECD, (2015). The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. PISA, OECD Publishing.

OECD, (2016). PISA 2015 Results (Volume I): Excellence and Equity in Education. OECD Publishing, Paris.

Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. Gender differences in mathematics: An integrative psychological approach, 294-315.

Riegle-Crumb, C. (2005). The cross-national context of the gender gap in math and science. *The social organization of schooling*, 227-243.

Sbaragli, S. (2012). Il ruolo delle misconcezioni nella didattica della matematica. *Bolondi B., Fandiño Pinilla MI (2012). I quaderni della didattica. Metodi e strumenti per l’insegnamento e l’apprendimento della matematica*, 121-139.

Winkelmann, H., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: Results from a German large-scale study. *ZDM*, 40(4), 601-616.

4.4.6 Gender Gap in Mathematics and Misconceptions: a study based on large-scale results.[25]

Bolondi Giorgio – Università di Bolzano

Cascella Clelia – INVALSI

Giberti Chiara – Università di Trento

Introduction

When considering a mathematics test, are males better than females in all mathematical items included in the test? If yes, which are the features (content, involved processes, formulation…) of these items?

If we deal with this question, it could be interesting to focus on the items in which the gender gap is more evident and make a deeper analysis from both a statistical and a didactical perspective. This kind of analysis seems to be able to provide at least two main research products: 1) an insight into the classrooms in order to ascertain possible links between didactical protocols and skills development in males and females; and, 2) individuation of recurring items' features that cause item bias in sub-groups of students clustered by gender. Moreover, the latter is undoubtedly also useful in order to construct good *standardized* achievement tests that, as well known, in accordance to Item Response Theory approach frequently used in education research, must guarantee measurement invariance.

Moreover, although undoubtedly a lot of different factors that might affect gender gap exist, exploring the relationship between gender gap in math and/or in STEM and socio-cultural and economic environment within which students grow up is an ongoing issue timely topic in EU agenda but it is outside the scope of this work.

The issue of gender gap in mathematics education has been deeply studied from both a qualitative and a quantitative point of view. For instance, many studies focus on small groups of students and try to understand gender differences on specific tasks by means of task-based interviews: on the other hand, most of the recent researches on gender gap in maths are based

[25] Articolo di approfondimento dei contenuti presentati nel convegno SEMT 17. A seguito del convegno l'articolo sarà sottomesso per la rivista *Educational Studies*.

on large-scale surveys, national and international standardized tests, such as PISA and TIMMS. Indeed, by using the results of these tests, it is possible to analyse gender gap from more general perspectives: for example, many studies compare the relevance of the gap in different countries highlight where this gap is more remarkable. Standardized assessments give the possibility to study this issue on a large population of students and from a statistical point of view. In the INVALSI reports (INVALSI, 2016) we observe gender differences on the basis of total medium score on the whole test and in the same way the gender-gap is studied in PISA reports (OECD, 2016). This information is important to point out the existence of a gap between males and females performances in maths and this is useful to make comparisons between different geographical areas and to analyse the relation between gender gap and social and cultural factors. Nevertheless, if we want to make assumption on the origins of this gap and analyse it with the lenses of the maths education theories, it is essential to analyse this gap on the single items of the test.

The observation of gender gap for different ability levels and the exploration of possible links between didactical protocols and differences in males and females test scores can be absolutely useful in order to formulate new research hypotheses and to provide new answers to the questions that our results might get emergent.

In this paper, we study gender differences in Italian large-scale mathematics surveys (INVALSI tests) with a mixed method QUAN/QUAL -> QUAL approach. By using a specific statistical tool (DIF analysis), we point out evidences not only in the whole test but also highlighting the items in which this gap is more remarkable, then we focus the qualitative analysis on these specific items, by individuating common features of them. This allows to state conjectures about the didactics aspects of the phenomena observed.

State of art

Gender gap in Mathematics

Several studies in Mathematics education focused on the different performance of male and female in mathematics. The growing importance of national and international standardized assessments has given new emphasis on this issue (Leder & Forgasz, 2008). The results of international survey such as PISA and TIMSS and national ones as well, such as INVALSI, highlighted an important gap in favour of male in mathematics performances in almost all the countries (OECD, 2016; Mullis, Martin, Foy & Hooper, 2016). Considering both PISA and TIMSS latest results, the gap was not equally distributed in all the countries and Italy was one of the countries in which the gender gap was more remarkable. Furthermore, the results of national standardize tests in Italy evidenced the existence of this gap in all school grades (INVALSI, 2016; Di Tommaso, Mendolia & Contini , 2016).

Researches on this issue revealed that factors determining gender differences are several and varied (Forgasz, 2010). Winkelmann, van den Heuvel-Panhuizen & Robitzsch (2008) distinguished between internal and external factors. Internal factors include biological differences between male and female, but several recent studies put aside this hypothesis and gave more relevance to social and cultural factors connected with the context in which student's live (Di Tommaso et al., 2016; OECD, 2016).

In support of social and cultural determinants of gender gap in mathematics, many researches have shown that in countries with a more gender-equal culture this gap decreases or even disappears (Jacobs & Eccles, 1992; Guiso, Monte, Sapienza & Zingales, 2008; González de San Román & De La Rica, 2012; OECD, 2015; Cascella, 2017).

Moreover, the beliefs of teachers and parents about males and females math skills and gender stereotypes of society have a profound effect on students' self-perception and thereby affect their performances (Jacobs & Bleeker, 2004; Riegle-Crumb, 2005; Freyer & Levitt, 2010). In fact, differences in metacognitive aspects related to maths, such as, in the case of girls, higher levels of math anxiety and less math self-efficacy, play an important role in determining the different results of males and females in math (Pajares, 2005; Cargnelutti, Tomasetto & Passolunghi, 2016; OECD, 2016).

Studies considered also other factors related to the school context, such as curriculum variables (Leder, 1992), classroom practices, assessment methods and teaching methods (Leder & Forgasz, 2008; OECD, 2016; Giberti, Zivelonghi & Bolondi, 2016). In this perspective, researchers have also focused on the different strategies used in problem solving activities, pointing out that females apply routine procedures and well-known strategies more frequently compared to males. The explanations of this different behaviour are inherent to metacognitive aspects: males seem less worried of being wrong and more inclined to try new methods and alternative approaches (Gould, 1996; Fennema & Carpenter, 1998; Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, Morely, & Cahalan, 2000; Bell & Norwood, 2007).

In this research, we endorse the idea that the main reasons of the observed differences in the performance of males and females in mathematics are related to cultural and social context in which the students live. We also consider that micro-social factors related to the *mileu* habits, curriculum variables and classroom practice might be central keys to understanding the issue of gender gap.

How to measure gender gap in math test scores: a brief methodological review

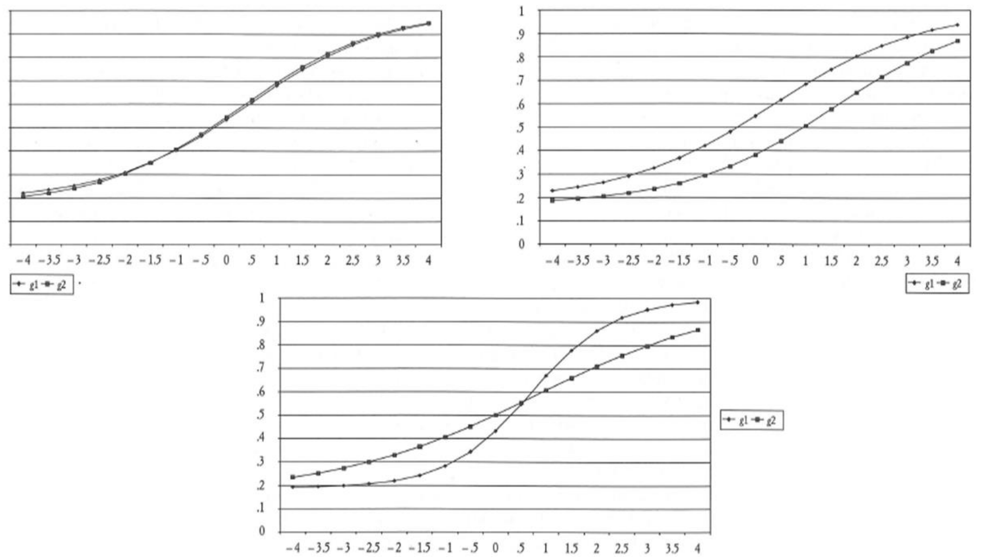
The wide use of standardized achievement tests in Social and Human Sciences has put the spotlight of psychometrical and educational literature on the treatment of item bias. From a measurement point of view, the term '*bias*', when used to describe mental tests, has a specific technical meaning: it is defined as a systematic error in measurement process that can alter the expected items' psychometrical properties and cause a consistent distortion of a statistic, i.e. for achievement tests, unfair person and items estimates. In other words, item bias refers to differences in the way a test item functions across sub-groups of persons matched on the attribute measured by the test (e.g. ability) and stratified by gender, socio-cultural background and/or any other variables not explicitly hypothesized by the statistical model used to analyse data (Osterlind, 1983; Holland & Wainer 1993; Camilli & Shepard 1994; Embretson & Reise, 2000; Camilli 2006; Penfield and Camilli 2007; Osterlind 2009). Therefore, when the same item shows different behaviours in different sub-groups of persons, it is biased and it is a severe violation of the Rasch model, frequently used to assess students' ability. In particular, INVALSI employs the Rasch model, according to which the probability of a correct answer depends on students' relative ability, i.e. his/her ability compared to item's difficulty. According to the Rasch model, each examinee responding to a test item possesses some amount of the underlying ability (e.g., ability in Mathematics) ideally represented as a latent trait along which both items and subject are scaled. Items are scaled depending on their difficulty (i.e. the amount of ability an examinee has to have to provide a correct answer). Subjects are scaled along the same latent trait depending on the amount of their own underlying ability. Thereby, more talented an examinee is, higher the probability of a correct answer is. An Item Characteristic Curve (ICC) exactly expresses this concept: it describes the relationship between the probability of correct response to an item and the ability scale.

The model assumes that the probability of a correct answer depends on student's ability exclusively and no other variables (such as for example students' socio-demographic characteristics) can affect directly this probability (of course they may have, and in fact they do have, an indirect impact). Therefore, according to the mathematical mechanism underlying the model, an achievement test developed to be analysed by the Rasch model has

to be constructed to avoid any possible effects exerted by each other variable.
In the sectorial literature, term bias refers to the whole student's test scores and thus it has a more general meaning compared to *Differential Item Functioning* (DIF). DIF refers to each single item and to item behaviour in sub-group of students *matched on ability*. In order to control item functionality, a lot of different techniques have been proposed, such as, for example, Mantel-Haenszel procedure and its follow-up consisting in the implementation of a revised version of chi square (Osterlind, 2009). Some other examples are: 1. the "areas measure" (Rudner, 1977; Rudner, Getson, & Knight, 1980; Linn & Harnisch, 1981; Shepard, Camilli, & Williams, 1985); 2. the delta plot method (Angoff & Ford, 1973); 3. the Draba statistic (of Wright, Mead, & Draba, 1976); 4. the Wald statistics (Lord, 1980); 5. the likelihood ratio methods (Thissen, Steinberg, & Gerrard, 1986; Thissen, Steinberg, & Wainer, 1988); and so on. Many studies carried out on both real and simulated data showed the efficacy of these techniques to detect DIF in small or medium sample. For larger dataset, unfortunately, the number of useful techniques decreases significantly. In the present study, we carried out a DIF analysis within Item Response Theory (IRT) framework, based on the comparison of the *Items' Response Function* (IRF) that links the probability of a correct answer to student's ability.

In our study, we have two IRFs, one for males and one for females. Comparing them, we can compare item functionality in both male and female sub-group of students scaled along the same latent trait, i.e. in each sub-group, we have the same number of students with the same ability level. Figure 1 provides an example of this: 1. The first graph represents a not biased item (IRFs in both group have the same behaviour); 2. The second graph shows a uniform DIF between students in group g1 e in group g2 (item functionality is different in these two groups because the probability of a correct answer is higher in the first group compared to the second one, in each ability level); and, finally, 3. The third graph represents a not uniform DIF (the probability of correct answer is higher in one group relative to the other one only for specific ability level).

Figure 4 - Some examples of items without bias (in the left), with uniform bias (on the right) and with not uniform bias (at the bottom)



Source: our adaptation from Barbaranelli and Natali (2005, p. 175)

The graphical inspection of each ICC provides also further information, such as for example item discrimination and guessing as well.

Discrimination describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location. This property essentially reflects the steepness of the item characteristic curve in its middle section: when an item is over-discriminating, the model over-estimates the probability of a correct answer for low-ability level students and vice versa for the other levels. However, an over-discrimination can represent a severe problem in terms of model fit and thus can cause unfair estimation.
In general, higher is the difference between observed probability (represented, for each ability group, by points in the figure shown on the right side of Table 7) and the estimated probability (the represented by the grey line reported in the same figure), less data fit model. Nevertheless, in our case and in general for all big dataset (table 1), model fit can be assessed within a tolerance interval instead of a punctual value as recommended for smaller samples. Fit analysis carried out for all items presented in this paper are internal to the tolerance intervals provided by the sectorial literature (INVALSI, 2016b; Wright et al., 1994).

Guessing expresses the probability of providing a correct answer by chance, independently on real student's ability. In this case, for lowest ability levels, the probability of encountering an item successfully is greater than what expected by the model: for example, in figure 1, the guessing effect equal to 0.2, i.e. for students' scaled at -4 logit along the latent trait the probability of a correct answer is equal to 20% instead of 0%.

Methodology

In this study, we analysed answers given by Italian students to Mathematics achievement tests developed by INVALSI. In accordance to the Italian national curricula, INVALSI math achievement test is aimed at revealing students' capacity in using mathematical concepts to interpret the surrounding environment and to solve daily life problems (INVALSI, 2016).

1) We considered a bank item of 1400 items were mathematics test (subministered by the INVALSI to the whole population) scores have been estimated and classified. In the bank item, estimation is performed by using the Rasch model and have been scaled in an empirical range equal to [-3; +3], i.e. the latent trait along which both items and persons are scaled depending on their difficulty and their ability respectively. Classification is performed according to the INVALSI theoretical framework (INVALSI, 2017), according to the dimensions of contents and the dimension of processes. We performed a supplementary DIF analysis of the items of the bank and we selected items were the difference of performance between male and female was quantitatively and statistically relevant, and the didactical issues involved were neat.

2) We focused our attention on five items, derived from INVALSI math achievement test administered in 2012 to pupils attending the 5th grade of primary school (i.e., D21a, D21b); in 2013 to boys and girls attending the 1st grade of lower secondary school (i.e., D23, D7b); and, in 2015 to students attending the 3rd grade of lower secondary school (i.e., D3).

These items are particularly interesting in order to highlight the following didactical issues: 1) misconceptions in comparing decimal numbers (i.e., D21a, D21b, and D23); and, 2) difficulty of students in estimating a measure (D7b and D3). These are well known problems in Mathematics education: for literature reviews see, for instance, Sbaragli, 2012.

The items selected comes from the following large scale surveys:

Table 1 – Numbers of students involved in INVALSI administration at the 5th grade of primary school, in 2012, and at the 1st and 3rd grade of lower intermediate school, in 2013 and 2015 respectively

	5th grade primary school (2012)	**1st grade lower intermediate school (2013)**	**3rd grade lower intermediate (2015)**
Population (N)	489.580 students	484.441 students	520.920 students
Sample (n)	30.843 students	16.383 students	28.494 students

Source: INVALSI data

All our analysis have been performed in the answers provided by the students of the sample. For each of these items we provide: the original item phrasing, some statistical descriptive, the item characteristic curve and the DIF analysis carried out by gender in comparison to the ANOVA.

All analyses were carried out by using RUMM2030. The item characteristic curves plotted by this package clearly highlight differences between expected and observed items' behaviour, both in male (focal) and female (reference) group. This kind of analysis provides information on the interaction between gender gap and student's ability level more than on the gap itself. This information was compared and discussed by using the results from the research in mathematics education.

In the following section, all the information gathered using statistical analysis is presented and connected to a qualitative didactical interpretation of students' cognitive processes, in order to reach a deeper understanding of gender differences related to specific contexts.

Results and discussion

As a first step of analysis, we calculated the percentage of males and s correct answers, item by item. These measures showed a remarkable difference between male and female responses in items related to misconceptions in comparing decimal numbers (D21a, D21b, and D23) and difficulty of students in estimating a measure (D7b and D3).

We compared items with similar content (decimal number/estimation) but different features (i.e., item response format and contextual situation).

Item D21

The first task belongs to the INVALSI test of level 5 of primary school administered in 2012.

Table 2 - Item D21 administered by INVALSI to pupils attending the 5th grade of primary school, in 2012 (9-10 years old).

D21. The following table contains final scores achieved by some gymnasts from different Countries in an international competition.

Nation	Score
Austria	**68.8**
Croatia	**71.8**
Finland	**72.0**
Japan	**68.08**
Greece	**60.8**
England	**69.8**
Italy	**80.12**
United States	**80.2**
Sweden	**70.2**
Switzerland	**78.1**

a) Which Nation was placed first? __________________________

b) Which Nation was placed fourth? __________________________

D21_a	ALL	FEMALE	MALE
Correct	**37%**	**32%**	**43%**
Wrong	61%	66%	56%
Missing	2%	2%	2%

D21_b	ALL	FEMALE	MALE
Correct	**67%**	**63%**	**70%**
Wrong	31%	35%	28%
Missing	2%	2%	2%

Percentage, by answer option

Both item a and b require comparison between decimal numbers reported in the table. For both items, percentage of correct answers showed a remarkable gap between males and females: males outperformed females with a gap of 11 points percentage in the first item and 7 points percentage in the second.

The two items require the same math competency, which is ordering decimal numbers, but students face greater difficulties in the first one and the percentage of correct answer in the second item is nearly twice the percentage of the first one.

Many researchers have studied difficulties of students in using decimal numbers and, in particular, the obstacles encountered in the transition between natural numbers and rational numbers. This is a critical step because it is possible that this transition contributes to the creation of misconceptions. In fact, students facing for the first time with decimal numbers tend to continue applying procedure and properties valid for natural numbers. If these beliefs are not discussed and abandoned, they may become misconceptions and impact on students´understanding and learning also in later school years (Steinle & Stacey, 2003; Steinle & Stacey, 2004; Roche & Clarke, 2004; Sbaragli, 2012).

Analysing item D21a we noticed that, at last, students have to compare two decimal number with the same integer part: Italy (80.12 points) and United States (80.2 points) are top ranked and the comparison between these two numbers allow students to identify winning country. A specific misconception, already studied in literature (Steinle & Stacey, 2003; Steinle &

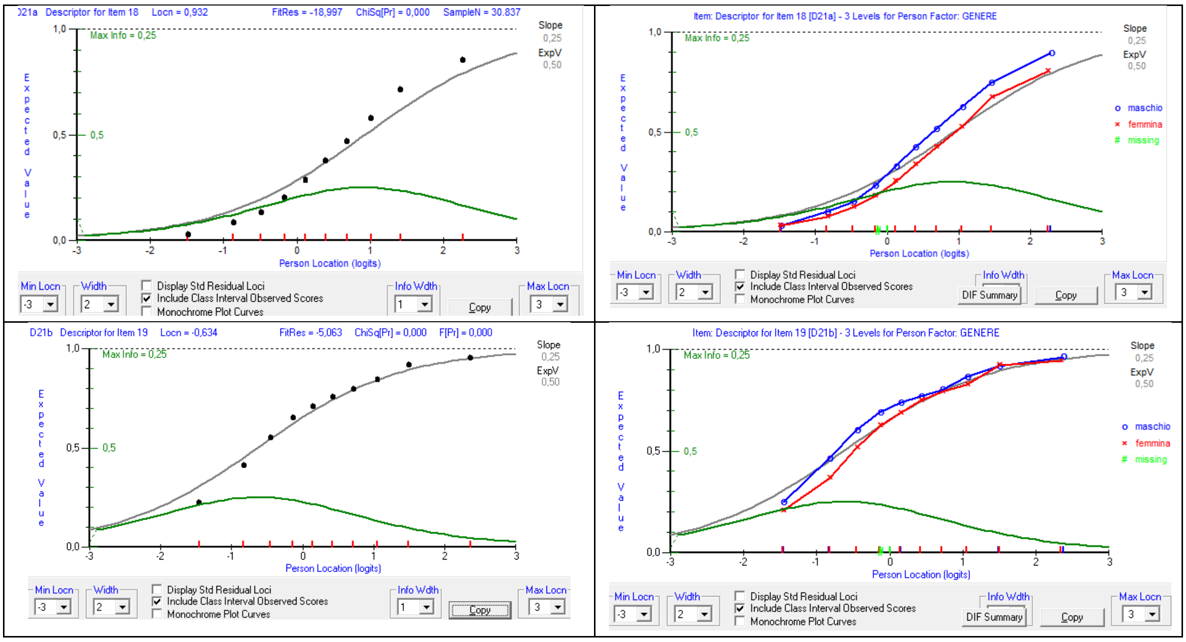
Stacey, 2004; Roche & Clarke, 2004; Sbaragli, 2012), is involved in comparing two decimals with the same integer part. Indeed, in this case, students are inclined to consider the decimal parts of the two numbers and compare them as naturals: since 12>2, they concluded that 80.12>80.2. According to other researches, this behaviour is widespread in middle years of schooling (Steinle & Stacey, 2004) and it is probably the cause of the low percentage of correct answer in the first item (less than 40%). Moreover, in this item we observed that the 43% of males gave the right answer in opposition to the 32% of females; this may mean that this misconception have a greater influence on females.

According to our analysis, both items showed a significant differential functioning but with a different shape for item a and item b, i.e. stressing different students' ability levels as shown in the following table 3.

In the following

Table *3*, we reported the Item Characteristic Curve (ICC – on the left side) and the traditional graphical DIF analysis provided by RUMM2030 (on the right side).

Table 3 - Item Characteristic Curve (on the left) and DIF graphical inspection (on the right) of the item D21 administered to students attending the 5th grade of primary school, in 2012



Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Results of DIF analysis confirmed the existence of a gender gap in item 21a and, when we compared item behaviour in male and female groups (respectively, blue and red lines in

graphs reported on the right side of
Table *3*), we observed some interesting differences, in particular for the highest ability level. In both groups, the probability of a correct answer was approximately equal both for males and for females for low ability levels (-1.50 ÷ 0.50) but we observed an increasing difference between the probabilities of a correct answer, in favour of males, in correspondence of higher ability level.

If we can assume that the above-mentioned misconception is the main reason for the difficulties of students in items like this, then our analysis suggest that a cause for the gender gap may be rooted also in the different sensibility of girls with respect to this misconception. The growth of gender gap with the ability of students may means that for males with medium and high ability-levels is easier to get over barriers given by misconceptions, while girls with the same ability struggle most.

These evidences became even more interesting if related with the results of the second item, which seems to be similar to the first one but reveals different results. Indeed, in the second item the percentage of correct answers grows to 67% instead 37% of the first one.

Furthermore, we noticed that also this item reveals a significant gender gap in terms of percentage of correct answers but this was less marked (63% of correct answer for females, 70% for males) and DIF graphs showed a different distribution of this gap (Tabel 3). In fact, observing the DIF plots the gender-gap is visible only for students with low and medium ability levels and considering higher ability levels this gap disappeared.

The different results in terms of gender gap evidenced by these two similar items might be explained focusing on the possible strategies adopted by the students to solve them. In fact, to solve the second item, high-level students probably realized that comparing decimal numbers is unnecessary: to pinpoint the fourth ranked Country, they could just compare the integer part of each number. When students employ this strategy, they are not yet challenged by the necessity of comparing decimal numbers, thereby they are not tricked by misconception and, as consequence, and percentage of correct answer grows up. Thus, if differences in male and female performances in the first item is related to the existence of the misconception, as we assumed, this change of procedure for higher ability levels is also an explanation of the disappearance of differences for higher levels: with the change of strategy, the influence of misconception no longer exists and then the gender gap is closed.

Furthermore, the *fit* analysis of the two items with the model supported this hypothesis. We observed that item D21a showed a little over-discrimination, i.e. item slope is higher than what expected by the model: the model overestimates lower ability levels and underestimates higher ones. It is possible that the observed misfit of the item is connected with the influence of the misconception. Moreover, also the fit of the second item (D21b) reveals that, for lower ability levels the model overestimates the results but for medium and higher levels, when we assume a change of procedure, the fit is optimal and, as already pointed out, the gender-gap disappears.

This evidence may confirm that in the second item the misconception influences students' responses only for medium and lower ability levels and that the disappearance of the gap for higher is because the gender gap is strictly related to the misconception.

Our interpretation was confirmed by other data coming from the DIF analysis carried out by RUMM2030. According to the two-way ANOVA (Table 4), both items have a significant gender main effect [p = 0.000] whereas the interaction effect is statistically significant for item D21b [p = 0.01] but not for item D21a [p = 0.99], as confirmed by the graphical inspection of ICC plotted for males and females. In fact, over-estimation is chiefly "explained" by male probability of correctly answering that is strongly higher than what expected by the Rasch model. In contrast, female answering behaviour is consistent to model expectation.

Table 4 -Two-way ANOVA for item D21a (figure on the left) and D21b (figure on the right) administered to students attending the 5th grade of primary school, in 2012.

Analysis of Variance for ITEM 18 [D21a:Descriptor for Item 18]

SOURCE	S.S	DF	M.S	F-RATIO	Prob
BETWEEN	769,268	20	38,463		
ANOVA-Fit[CInt]	604,227	9	67,136	83,398830	0,000000
DIF[gender]	188,728	2	94,364	117,222200	0,000000
gender-by-CInt	-23,687	9	-2,632	-3,269420	0.999999
TOTAL Item DIF	165,041	11	15,004	18,638150	0,000291
TOTAL Misfit	769,268	20	38,463	47,780460	0,000000
WITHIN	24417,340	30332	0,805		
TOTAL	5186,610	30352	0,830		

Note: Negative SS for interaction is due to unequal cell sizes.

Analysis of Variance for ITEM 19 [D21b:Descriptor for Item 19]

SOURCE	S.S	DF	M.S	F-RATIO	Prob
BETWEEN	292,395	20	14,620		
ANOVA-Fit[CInt]	192,081	9	21,342	23,297530	0,000000
DIF[gender]	80,634	2	40,317	44,010500	0,000000
gender-by-CInt	19,681	9	2,187	2,387061	0,010406
TOTAL Item DIF	100,314	11	9,119	9,954959	0,000000
TOTAL Misfit	292,395	20	14,620	15,959120	0,000688
WITHIN	27700,270	30238	0,916		
TOTAL	7992,670	30258	0,925		

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

In Table 5, mean differences between males and females and residuals analysis are reported. In particular, the first table shown in Table 5 (on both sides) reports mean of observed scores in each ability group levels. It provides a glance on differences in Math scores estimated for subgroups of males and females matched on the ability and show the average of persons' parameter within each group estimated on the basis of answers provided by each student to the entire achievement test. For item D21a, medium test score in male group is equal to 0.43 whereas it is equal to 0.33 in female group. This difference increases taking into account each class interval: a strong superiority of males relative to females emerges for each class interval, in particular for medium ability levels. On average, these differences can be revealed also for item D21b (right side of Table 5).

The second and the third tables (reported on both sides of Table 5) respectively show case numerosity in each class interval and estimations' residuals in each of them for both males and females.

Table 5 - DIF analysis carried out by RUMM2030 for item D21a (on the left) and D21b (on the right) administered to students attending the 5th grade of primary school, in 2012

Mean of Observed Scores for ITEM 18 [D21a:Descriptor for Item 18]

Level	Total [N]	Level Mean	CInt1 Mean	CInt2 Mean	CInt3 Mean	CInt4 Mean	CInt5 Mean	CInt6 Mean	CInt7 Mean	CInt8 Mean	CInt9 Mean	CInt1(Mean
males	[15.453]	0,43	0,03	0,10	0,15	0,23	0,32	0,42	0,52	0,62	0,75	0,89
females	[15.415]	0,33	0,03	0,07	0,12	0,18	0,25	0,34	0,42	0,53	0,67	0,80
missing	[1]	*****	*****	*****	*****	0,00	*****	*****	*****	*****	*****	*****
Overall		0,38	0,03	0,08	0,13	0,20	0,28	0,38	0,47	0,58	0,71	0,86

Mean LOCATIONS for Class Interval [CInt]

	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
male	-1,48	-0,85	-0,48	-0,17	0,12	0,39	0,68	1,04	1,45	2,29
female	-1,49	-0,84	-0,47	-0,17	0,11	0,39	0,68	1,04	1,45	2,24
missing	0,00	0,00	0,00	-0,13	0,00	0,00	0,00	0,00	0,00	0,00

Item Residual Analysis for ITEM 18 [D21a:Descriptor for Item 18]

FREQUENCIES [Total N]

gender-by-CInt		CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	[15.181]	1383	1327	1366	1444	1394	1486	1549	1576	1719	1937
females	[15.171]	1562	1579	1669	1589	1646	1549	1487	1350	1314	1426
missing	[1]	0	0	0	1	0	0	0	0	0	0
TOTAL	30.353	2945	2906	3035	3034	3040	3035	3036	2926	3033	3363

RESIDUALS for ITEM 18 [D21a:Descriptor for Item 18]

gender-by-CInt	Mean	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	0,07	-0,20	-0,13	-0,12	-0,04	0,04	0,11	0,16	0,21	0,26	0,27
females	-0,09	-0,20	-0,20	-0,19	-0,17	-0,12	-0,07	-0,03	0,01	0,12	0,07
missing	-0,59	*****	*****	*****	-0,59	*****	*****	*****	*****	*****	*****
TOTAL	-0,01	-0,20	-0,17	-0,16	-0,11	-0,05	0,02	0,06	0,12	0,20	0,19

Mean of Observed Scores for ITEM 19 [D21b:Descriptor for Item 19]

Level	Total [N]	Level Mean	CInt1 Mean	CInt2 Mean	CInt3 Mean	CInt4 Mean	CInt5 Mean	CInt6 Mean	CInt7 Mean	CInt8 Mean	CInt9 Mean	CInt10 Mean
males	[15.453]	0,72	0,25	0,46	0,60	0,69	0,74	0,77	0,80	0,86	0,92	0,96
females	[15.415]	0,64	0,20	0,37	0,51	0,62	0,69	0,75	0,79	0,83	0,92	0,94
missing	[1]	*****	*****	*****	*****	0,00	*****	*****	*****	*****	*****	*****
Overall		0,68	0,22	0,41	0,55	0,65	0,71	0,76	0,80	0,85	0,92	0,96

Mean LOCATIONS for Class Interval [CInt]

	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	-1,46	-0,83	-0,46	-0,14	0,15	0,42	0,71	1,05	1,49	2,38
females	-1,46	-0,83	-0,45	-0,14	0,14	0,42	0,71	1,05	1,50	2,34
missing	0,00	0,00	0,00	-0,13	0,00	0,00	0,00	0,00	0,00	0,00

Item Residual Analysis for ITEM 19 [D21b:Descriptor for Item 19]

FREQUENCIES [Total N]

gender-by-CInt		CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	[15.133]	1446	1397	1359	1451	1381	1505	1487	1702	1745	1660
females	[15.125]	1679	1629	1667	1574	1646	1521	1436	1458	1318	1197
missing	[1]	0	0	0	1	0	0	0	0	0	0
TOTAL	30.259	3125	3026	3026	3026	3027	3026	2923	3160	3063	2857

RESIDUALS for ITEM 19 [D21b:Descriptor for Item 19]

gender-x-CInt	Mean	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	0,05	-0,14	0,02	0,11	0,14	0,12	0,06	0,02	0,06	0,08	0,07
females	-0,05	-0,24	-0,17	-0,06	0,00	0,00	0,01	-0,01	-0,05	0,10	-0,01
missing	-1,28	*****	*****	*****	-1,28	*****	*****	*****	*****	*****	*****
TOTAL	0,00	-0,19	-0,08	0,02	0,07	0,05	0,04	0,01	0,01	0,09	0,04

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Item D23

The second task analysed belongs to INVALSI test of level 6 administered in 2013. In this case too, students have to compare decimal numbers.

Table 6 - Item D23 administered to students attending the 1st grade level of lower intermediate school in 2013 (on the left) and percentage of students' answers for each distractor (in the right)

D23. In which of the following groups are numbers arranged in ascending order?

- **A.** □ **3.5; 3.043; 3.28; 3.124**
- **B.** □ **3.5; 3.28; 3.124; 3.043**
- **C.** □ **3.043; 3.5; 3.124; 3.28**
- **D.** □ **3.043; 3.124; 3.28; 3.5**

D23	ALL	FEMALES	MALES
A	14%	15%	12%
B	17%	20%	15%
C	8%	8%	9%
D (correct)	**57%**	**54%**	**61%**
Missing	4%	4%	3%

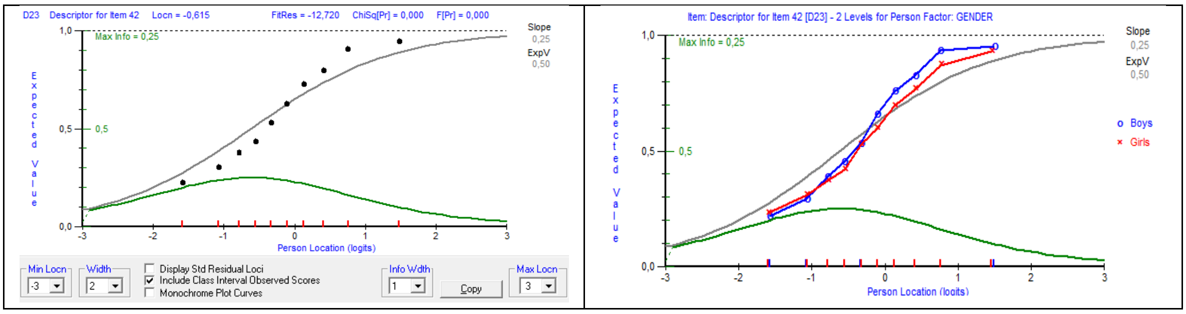
Percentage, by answer option

Even if this question was administered at level 6, the percentage of correct answers is less than 60%. Thus, this means that difficulties related to the comparison between decimals are still evident also at the end of the first year in lower secondary school.

Moreover, also this item presented a gender gap of 7 points percentage in favour of males and, similarly to other items administered at different grades, it evidenced specific difficulties for females in managing this mathematical competence (i.e., ordering decimal numbers).
In particular, according to DIF analysis reported in table 7, the gap is relevant for medium and high ability levels.

The analysis of the fit of this item revealed a behaviour similar to item D21a: it showed a little over-discrimination (on the left side of table 7) and this occurred, in particular for the highest ability level, both in male and female group. Also in this case, the model overestimates the lower levels and underestimates the higher ones, especially for what concerns males.

Table 7 - Item Characteristic Curve (on the left) and DIF graphical inspection (on the right) of the item D23 administered to students attending the 1st grade of lower intermediate school, in 2013

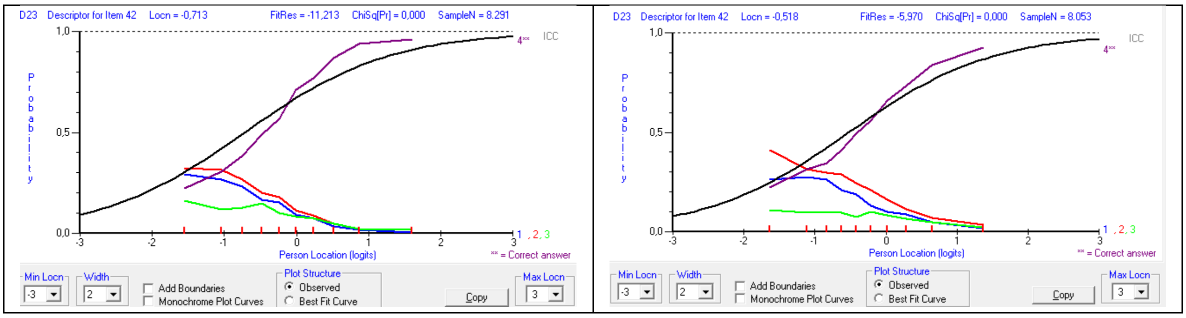


Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Since this is a multiple choice item, we draw its distractor plot in order to study also wrong options' functionality (Table 8). Females were more attracted than males by answer B and this supported our hypothesis about the role played by misconception related to decimal numbers (Sbaragli, 2012): the first three numbers were probably ordered following the wrong idea according to which 3.5<3.28<3.124 because 5<28<124.
The comparison by gender between correct and wrong answer options' shapes did not reveal any significant differences between males and females, with exception of option B.

Table 8 - Male (plot on the left) and female (plot on the right) distractor graphs drawn for the item D23 administered by INVALSI to students attending the 1st grade level of lower intermediate school in 2013

(10-11 years old students)



Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Figure 5 – Two-way ANOVA carried out by RUMM2030 for item D23 administered to students attending the 1st grade level of lower intermediate school in 2013

Analysis of Variance for ITEM 42 [D23:Descriptor for Item 42]

SOURCE	S.S	DF	M.S	F-RATIO	Prob
BETWEEN	430,029	19	22,633		
ANOVA-Fit[CInt]	394,794	9	43,866	52,109540	0,000000
DIF[Gender]	32,555	1	32,555	38,672790	0,000222
Gender-by-CInt	2,680	9	0,298	0,353711	0,956584
TOTAL Item DIF	35,235	10	3,523	4,185620	0,000000
TOTAL Misfit	430,029	19	22,633	26,886420	0,000000
WITHIN	13218,840	15703	0,842		
TOTAL	3648,860	15722	0,868		

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

The two-way ANOVA on the person factor 'gender' and class intervals showed a significant main effect but a not statistically significant interaction (respectively, p = 0.000 and p = 0.956).

The first table shown in Figure 3 reported mean of observed scores in each ability group levels. It provided a glance on differences in math scores estimated for subgroups of males and females matched on the ability and showed the average of persons' parameter within each group estimated on the basis of answers provided by each student to the entire achievement test. This table did not show relevant differences in ability between boys and girls. Finally, the second and the third tables showed case numerosity in each class interval and estimations' residuals in each of them for both males and females.

Figure 6 – DIF analysis carried out by RUMM2030 for item D23 administered to students

attending the 1st grade level of lower intermediate school in 2013

Mean of Observed Scores for ITEM 42 [D23:Descriptor for Item 42]

Level	Total [N]	Level Mean	CInt1 Mean	CInt2 Mean	CInt3 Mean	CInt4 Mean	CInt5 Mean	CInt6 Mean	CInt7 Mean	CInt8 Mean	CInt9 Mean	CInt10 Mean
males	[8.291]	0,62	0,22	0,29	0,39	0,45	0,53	0,66	0,76	0,82	0,94	0,95
females	[8.054]	0,55	0,23	0,31	0,37	0,42	0,53	0,60	0,70	0,77	0,87	0,93
Overall		0,59	0,22	0,30	0,38	0,43	0,53	0,63	0,73	0,80	0,91	0,95

Mean LOCATIONS for Class Interval [CInt]

males	-1,58	-1,07	-0,79	-0,56	-0,33	-0,10	0,13	0,41	0,76	1,49
females	-1,60	-1,07	-0,78	-0,55	-0,33	-0,11	0,13	0,41	0,76	1,46

Item Residual Analysis for ITEM 42 [D23:Descriptor for Item 42]

FREQUENCIES [Total N]

Gender-by-CInt		CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	[7.996]	749	764	747	718	750	790	797	842	888	951
females	[7.727]	822	825	827	856	825	781	775	729	687	600
TOTAL	15.723	1571	1589	1574	1574	1575	1571	1572	1571	1575	1551

RESIDUALS for ITEM 42 [D23:Descriptor for Item 42]

Gender-x-CInt	Mean	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	0,05	-0,14	-0,20	-0,14	-0,12	-0,08	0,07	0,17	0,20	0,35	0,22
females	-0,04	-0,10	-0,17	-0,17	-0,20	-0,09	-0,06	0,04	0,07	0,19	0,15
TOTAL	0,00	-0,12	-0,18	-0,16	-0,16	-0,08	0,01	0,11	0,14	0,28	0,19

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Item D7b

Let's now use the same methods to study a different math content: estimation of a measure. In particular, the items presented below ask students to make a length estimate of a real object: the first items refers to the length of a train while the second one concerns the height of a building.

The question intent of the first item consist in estimating the length of the train in the figure, which is composed by nine wagons and the locomotive. This item was administered in 2013 at level 6 and only 43% of the students choose the correct answer.

Table 9 - Item D7b administered to students attending the 1st grade level

of lower intermediate school, in 2013.

b. The train Nina went in on is composed by the locomotive and 9 coaches.

How long is the train approximately?

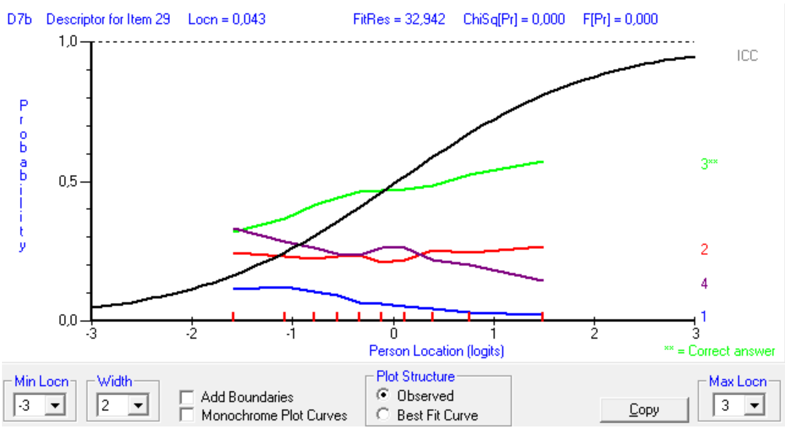
A. □ **Approximately 10 meters**
B. □ **Approximately 50 meters**
C. □ **Approximately 250 meters**
D. □ **Approximately 1000 meters**

D7b	ALL	FEMALES	MALES
A	6%	8%	5%
B	23%	21%	25%
C (correct)	**43%**	**40%**	**47%**
D	23%	26%	20%
Missing	5%	6%	4%

Percentage, by answer option

This item presents interesting behaviour not only observing DIF results but also studying distractor plot of the whole population.

Figure 4 – Distractor analysis carried out for the entire population of the item D7b administered to students attending the 1st grade of lower intermediate school



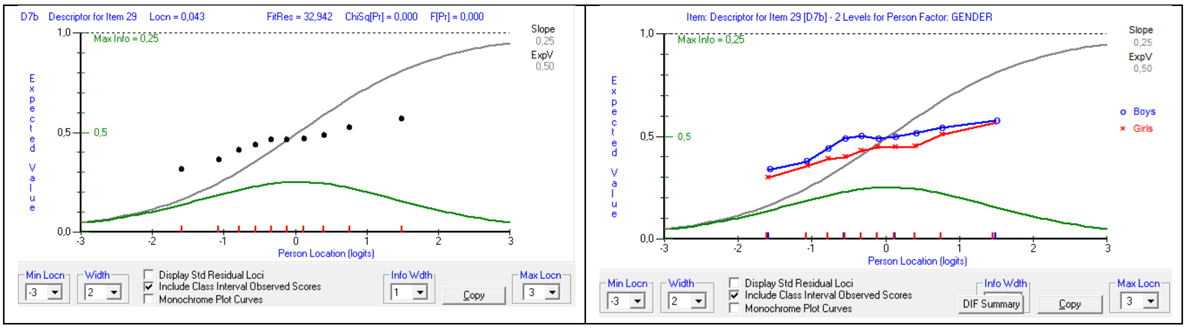
Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

This item showed high under-discrimination (**Errore. L'origine riferimento non è stata trovata.** and Fig. 4), i.e.the probability of a correct answer was overestimated for low-ability students, and vice versa for high-ability students. Under-discrimination is usually due to a violation of unidimensionality assumption, i.e. an item refers to a further dimension not covered by the other test items (Hambleton & Swaminathan, 1985), for example because it refers to a specific mathematical sub-domain not represented by other items within the achievement test or because it involves some abilities/competences not relevant to cope other items within the test.

Furthermore, also the functionality of the other options are unusual and, in particular, the percentage of students who choose option B is almost the same for all ability levels.

This particular item behaviour is interesting from a didactical point of view because support the idea according to which students' competence in estimating a measure is weakly related with the idea of mathematical competency (reported by Italian national curricula) investigated using this test and so with the latent trait: all students give almost the same answers independently from their ability level. Moreover, estimation is mentioned as an important goal in national curricula in Italy not only in secondary school, but also in primary school: students' ability in estimate dimensions of real objects should be developed since the first years of schools. These results highlighted that all students had difficulty in estimating a measure in all grades and this may mean that it is a competency not developed in classroom practice.

Table 10 – Item Characteristic Curve (on the left) and DIF graphical inspection (on the right) of the item D7b administered to students attending the 1st grade of lower intermediate school, in 2013



Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

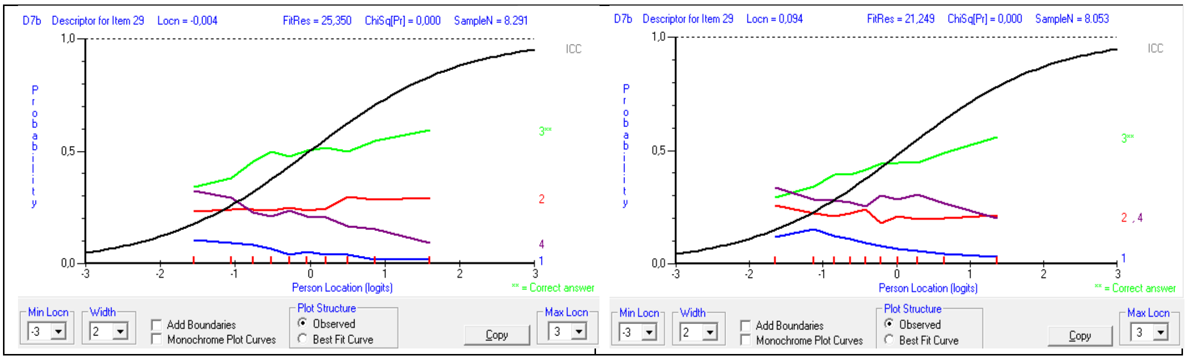
The percentage of each answer divided by gender pointed out that the gap between males and females was remarkable and approximately 7 percentage points. In fact, only the 40% of females indicated the right answer compared to the 47% of males. Furthermore, option B (underestimation of the length of the train) is more attractive for males and option D (overestimation of the length) is more attractive for females.

Moreover, DIF analysis pointed out the different trend of the correct answer comparing male and female results and we observed a strong superiority of males relative to females in each ability level with few exceptions for the highest and lowest ability levels (on the right side of **Errore. L'origine riferimento non è stata trovata.**).

Distractor analysis showed some interesting evidences about differences in male and female group. In particular, the distractor D (purple lines in both graphs below) was particularly attractive for females irrespectively of ability level in contrast to item *cumulativity* hypothesized by the Rasch model (Hambleton & Swaminathan, 1985). On the contrary, in male group, observed functionality was consistent to Rasch model expectation. Students choosing this distractor made an overestimation of the length of the train, and did not control the reasonableness of the length reconducting the length of the train to the length of a single wagon: in fact if the train would be 1000 metres long, each wagon would be 100 metres.

Some differences can be observed also for distractor A (blue lines in the graphs below) showing a minor attractivity power for males relative to females in each class intervals. Finally, distractor B (red lines) showed a clear anti-cumulativity behaviour: in male group, it had the same *attractivity* for low and medium ability levels and an higher *attractivity* for high ability level as well; in female group, it had approximately the same *attractivity* with few exceptions for medium ability levels (i.e. around 0.000 on the latent trait).

Table 11 - Distractor analysis carried out in male (figure on the left) and female group (figure on the right)



Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

In Figure 7 and in Figure 8, we reported RUMM2030 DIF analysis for item D7b. A significant main gender effect [p = 0.000] but a not significant interaction [p = 0.999] can be disclosed. Moreover, differences in medium persons' parameters confirms a superiority of males relative to females in each class intervals although person parameter estimated for males and females on the basis of the answer given by them to the entire test does not significantly differ depending on gender. Finally, similarly to the previous output, the second and the third tables reported in Figure 8 show case numerosity and analysis of residuals, respectively.

Figure 7 - Two-way ANOVA carried out by RUMM2030 for item D23 administered to students attending the 1st grade of lower intermediate school in 2013

Analysis of Variance for ITEM 37 [D7b:Descriptor for Item 37]

SOURCE	S.S	DF	M.S	F-RATIO	Prob
BETWEEN	905,762	20	45,288		
ANOVA-Fit[CInt]	706,636	9	78,515	98,369210	0,000586
DIF[gender]	223,437	2	111,718	139,968500	0,000868
gender-by-CInt	-24,311	9	-2,701	-3,384230	0.999999
TOTAL Item DIF	199,126	11	18,102	22,679910	0,000306
TOTAL Misfit	905,762	20	45,288	56,740090	0,000000
WITHIN	24235,570	30364	0,798		
TOTAL	5141,330	30384	0,827		

Note: Negative SS for interaction is due to unequal cell sizes.

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Figure 8 - DIF analysis carried out by RUMM2030 for item D7b administered to students attending the 1st grade of lower intermediate school in 2013

Mean of Observed Scores for ITEM 37 [D7b:Descriptor for Item 37]

Level	Total [N]	Level Mean	CInt1 Mean	CInt2 Mean	CInt3 Mean	CInt4 Mean	CInt5 Mean	CInt6 Mean	CInt7 Mean	CInt8 Mean	CInt9 Mean	CInt10 Mean
males	[15.453]	0,56	0,09	0,18	0,29	0,41	0,51	0,59	0,71	0,80	0,86	0,94
females	[15.415]	0,45	0,06	0,14	0,21	0,30	0,40	0,50	0,63	0,70	0,82	0,91
missing	[1]	*****	*****	*****	*****	0,00	*****	*****	*****	*****	*****	*****
Overall		0,51	0,07	0,16	0,24	0,35	0,45	0,55	0,67	0,76	0,84	0,93

Mean LOCATIONS for Class Interval [CInt]

	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	-1,48	-0,85	-0,48	-0,17	0,12	0,40	0,69	1,04	1,45	2,29
females	-1,49	-0,84	-0,47	-0,17	0,12	0,39	0,69	1,04	1,45	2,24
missing	0,00	0,00	0,00	-0,13	0,00	0,00	0,00	0,00	0,00	0,00

Item Residual Analysis for ITEM 37 [D7b:Descriptor for Item 37]

FREQUENCIES [Total N]

gender-by-CInt		CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	[15.228]	1422	1349	1374	1449	1401	1491	1550	1540	1730	1922
females	[15.156]	1589	1592	1664	1587	1643	1550	1488	1315	1308	1420
missing	[1]	0	0	0	1	0	0	0	0	0	0
TOTAL	30.385	3011	2941	3038	3037	3044	3041	3038	2855	3038	3342

RESIDUALS for ITEM 37 [D7b:Descriptor for Item 37]

gender-x-CInt	Mean	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	0,08	-0,18	-0,15	-0,08	0,03	0,10	0,12	0,21	0,26	0,23	0,19
females	-0,09	-0,26	-0,25	-0,24	-0,19	-0,14	-0,07	0,04	0,04	0,14	0,14
missing	-0,82	*****	*****	*****	-0,82	*****	*****	*****	*****	*****	*****
TOTAL	0,00	-0,22	-0,21	-0,17	-0,09	-0,03	0,02	0,12	0,16	0,19	0,17

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Item D3

It is interesting to compare the previous results with those of the item D3 administered in 2015 to students at grade 8.

Figure 9 - Item D3 administered to students attending the 3rd grade of lower intermediate school in 2015

D9. Look at the building below

How high can be the building?

A. □ **Less than 10 meters**
B. □ **Between 15 and 20 meters**
C. □ **Between 25 and 30 meters**
D. □ **More than 30 meters**

D3	ALL	FEMALES	MALES
A	8%	8%	8%
B (correct)	**65%**	**61%**	**69%**
C	22%	24%	19%
D	5%	6%	3%
Missing	0%	1%	0%

Percentage, by answer option

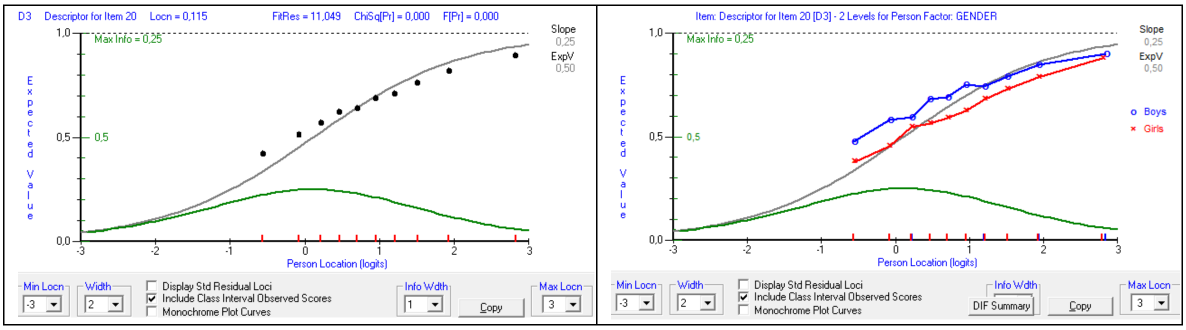
This item appeared less difficult than D7b, and the percentage of the correct answer was 65%. This was probably because the respondents were two years older and the measure they had to estimate was smaller and nearest to their everyday experience. However, also in this case, there was a marked gender-gap evidenced by both percentage and DIF analysis.

Observing the percentage of each answer, we noticed that the correct answer is selected by the 69% of males and the 61% of females, the gap is approximately of 8 points and we perceived a strong difference in choosing option C: the 19% of males choose this answer versus the 24% of females. In fact, also in this case, females tend to overestimate the measure. It is interesting to notice that almost all the students answered to this question and the missing answer were practically nil; this means that students were confident of their answers.

The item characteristic curve of item D3 did not show a very good discrimination, at global level (left side of Table 12) even if it was better than the discrimination of the previous item (D7b). In this case, the fit with the model was slightly better but we identified that the behaviour was similar to the first one: the characteristic curve obtained using Rasch model underestimated the responses for lower and medium ability levels and vice versa for highest ones. In fact, accordingly to item D7b,the trend of the correct answer was roughly linear for both males and females.

However, gender person factor plot reveals that a parallel structural difference is present between the boy and girl level, with a strong superiority of boys relative to girls, with higher differences in low and medium ability levels (right side of Table 12), as also detailed shown in the first table reported in Figure 10.

Table 12 – Item Characteristic Curve (on the left) and DIF graphical inspection (on the right) of the item D7b administered to students attending the 1st grade of lower intermediate school, in 2013



Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Figure 10 - DIF analysis carried out by RUMM2030 for item D7b administered to students attending the 1st grade of lower intermediate school in 2013

Mean of Observed Scores for ITEM 20 [D3:Descriptor for Item 20]

Level	Total [N]	Level Mean	CInt1 Mean	CInt2 Mean	CInt3 Mean	CInt4 Mean	CInt5 Mean	CInt6 Mean	CInt7 Mean	CInt8 Mean	CInt9 Mean	CInt10 Mean
males	[8.401]	0,72	0,48	0,58	0,60	0,68	0,69	0,75	0,74	0,79	0,85	0,90
females	[7.946]	0,61	0,38	0,46	0,55	0,56	0,59	0,62	0,68	0,73	0,79	0,88
Overall		0,66	0,42	0,52	0,57	0,62	0,64	0,69	0,71	0,76	0,82	0,89

Mean LOCATIONS for Class Interval [CInt]

	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	-0,56	-0,08	0,22	0,47	0,70	0,95	1,20	1,51	1,94	2,84
females	-0,56	-0,08	0,22	0,46	0,71	0,95	1,21	1,51	1,93	2,80

Item Residual Analysis for ITEM 20 [D3:Descriptor for Item 20]

FREQUENCIES [Total N]

Gender-by-CInt		CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	[8.356]	753	773	793	709	916	820	755	938	901	998
females	[7.904]	873	852	837	743	898	803	749	798	728	623
TOTAL	16.260	1626	1625	1630	1452	1814	1623	1504	1736	1629	1621

RESIDUALS for ITEM 20 [D3:Descriptor for Item 20]

Gender-x-CInt	Mean	CInt1	CInt2	CInt3	CInt4	CInt5	CInt6	CInt7	CInt8	CInt9	CInt10
males	0,08	0,29	0,26	0,14	0,19	0,10	0,12	-0,01	-0,03	-0,04	-0,14
females	-0,09	0,08	0,01	0,04	-0,05	-0,11	-0,16	-0,16	-0,18	-0,21	-0,19
TOTAL	0,00	0,18	0,13	0,09	0,07	0,00	-0,02	-0,09	-0,10	-0,11	-0,16

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

The ANOVA summary confirms the graphical evidence and report that this item has both a significant interaction [p = 0.000] and a significant gender main effect [p = 0.000] (Figure 11).

Figure 11 - Two-way ANOVA carried out by RUMM2030 for item D23 administered to students attending the 1st grade of lower intermediate school in 2013

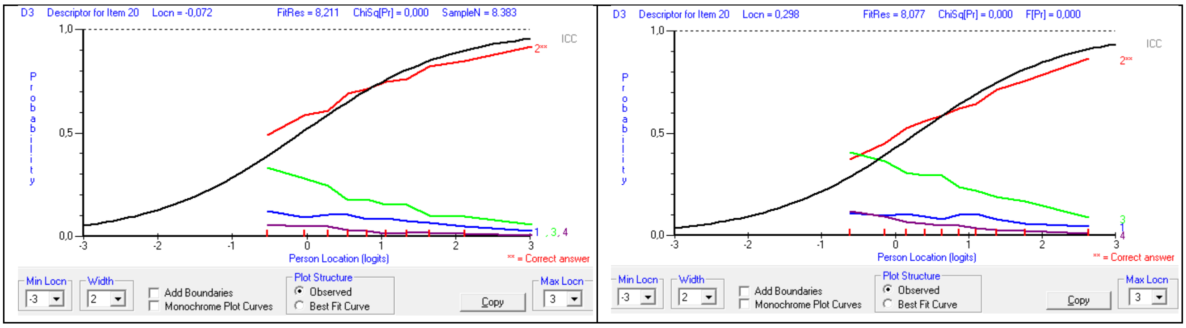
Analysis of Variance for ITEM 20 [D3:Descriptor for Item 20]

SOURCE	S.S	DF	M.S	F-RATIO	Prob
BETWEEN	340,409	19	17,916		
ANOVA-Fit[CInt]	189,650	9	21,072	18,997520	0,000000
DIF[Gender]	109,512	1	109,512	98,729840	0,000000
Gender-by-CInt	41,247	9	4,583	4,131815	0,000305
TOTAL Item DIF	150,759	10	15,076	13,591620	0,000183
TOTAL Misfit	340,409	19	17,916	16,152310	0,000093
WITHIN	18013,530	16240	1,109		
TOTAL	8353,940	16259	1,129		

Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Item D3 is multiple-choice. The distractor analysis reveals little differences between boys and girls with just one exception represented by option C (green line in the graphs below) that show a better functionality in female than in male group.

Table 13 - Distractor analysis carried out in male (figure on the left) and female group (figure on the right)



Source: our elaboration on INVALSI SNV data carried out by using RUMM2030

Similarly to the previous item, estimating height of a floor and then multiply it for their number was a possible item solution strategy.. This strategy allowed students to avoid immediately the first option because in this case the height of a story would be less than 2 meters, easily comparable with a person's height. Observing the trend of option A (blu line) in the distractor plots, we noticed that it was almost constant for all ability levels and for both males and females. Option D decreased for both group of students but, in the case of female group, for lower ability levels it had the same attractivity of the correct answer.

The comparison between these two items gave us the opportunity to point out that estimation is a very interesting issue because it seems to be a competency that is not strictly fitting with the other aspects of math ability, as measured by the INVALSI tests. Even so, ability to estimate a measure is one of the important goals of school education and is mentioned in national curricula of all school levels in Italy. The ability of the students in estimating the length of a real object seems not to fit with the model. Estimation is included also in curricula of many other nations and it is a relevant aspect of mathematical abilities. Nevertheless this issue is not extensively discussed in mathematics education literature and only few studies focus on the way estimation is developed in classrooms activities (Pizarro, Gorgorió & Albarracín, 2015).

In a recent paper, Pizarro et al. (2015) made a brief review of the literature on estimation of a measure and explained that, even it is considered an important ability in mathematics education, it is often considered as an issue distant from mathematics, discipline characterized by exactitude and rigour (Segovia & Castro, 2009).

Nevertheless, the persistence of gender-gap in both of the items and for almost all ability levels in favour of males, evidenced that estimation is an issue in which females struggle most and this is particularly interesting from a didactical point of view. In fact, this means that the processes involved in estimating a measure of a real object put females in a specific difficulty that cause a worst results also in the whole test. It is possible that girls are more influenced by didactical practice and this kind of issue give them more difficulty because they are not used to work on estimation in classrooms.

This view of estimation as something outer and different from the other aspects of mathematics and the scarcity of study regarding activity for teachers in this field, could be a possible explanation of our results and in particular may suggest conjecture about the didacrics issues involved in the origin of the gender gap observed.

Conclusions and further issues

In this paper, we have reported examples of the methodological strategy we are adopting to analyse around 1400 math items administered from 2008 up to now. Focusing on items from grade 5, 6 and 8 large scale assessments, we used this method to observe gender differences related to a specific misconception in comparing decimal numbers and to measure estimate. We linked the quantitative features of the differences observed between males and females to the qualitative features of the items, identifying possible didactical reasons for the origin of these differences. We expect that a systemic analysis carried out on the entire item bank will allow us to identify recurring features of maths items influencing gender gap and will give us the possibility to produce a first mapping of these items' characteristics (type, mathematical contents and context). We believe that, by combining statistical analysis and didactical interpretation of students' cognitive processes, it will be possible to point out some possible causes of gender differences in maths.

From another point of view, using DIF analysis applied to standardized tests, we displayed that misconceptions have a stronger influence on female and how they have different effects according to the ability level of the students. We think that this result may be the starting point for a qualitative empirical research.

References

Angoff, W. H., & Ford, S. F. (1973). Item-race interaction on a test of scholastic aptitude. *Journal of Educational Measurement*, *10*(2), 95-105.

Barbaranelli, C., & Natali, E. (2005). *I test psicologici: teorie e modelli psicometrici*. Carocci.

Bell, K. N., & Norwood, K. (2007). Gender equity intersects with Mathematics and technology: Problem-solving education for changing times. *Gender in the classroom*, 225-258.

Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., Vol. 4, pp. 221-256). Westport: American Council on Education & Praeger Publishers.

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*(Vol. 4). Sage.

Cargnelutti, E., Tomasetto, C., & Passolunghi, M. C. (2016). How is anxiety related to math performance in young students? A longitudinal study of Grade 2 to Grade 3 children. *Cognition and Emotion*, 1-10.

Cascella, C. (2017). Male and female social roles in daily file, pupils' perceptions and gender gap in Math test scores. Some empirical evidences from Italian primary school. (Forthcoming)

Di Tommaso, M. L., Mendolia, S., & Contini, D. (2016). The Gender Gap in Mathematics Achievement: Evidence from Italian Data. *IZA Discussion paper*, n.10053, Bonn.

Embretson, S. E. and Reise, S. P. (2000), Item Response Theory for Psychologists, Mahwah, NJ: Lawrence Erlbaum Associates.

Fennema, E., Carpenter, T. P. (1998). New perspectives on gender differences in mathematics: an introduction. *Educational Researcher*, 27(5), 4-5.

Forgasz, H. J. (2010). International perspectives on gender and mathematics education. IAP.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-240.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of experimental child psychology*, 75(3), 165-190.

Giberti, C., Zivelonghi, A., & Bolondi, G. (2016). Gender differences and didactis contract: analysis of two Inalsi tasks on power properties. *40th PME proceedings*, 275.

González de San Román, A., & De La Rica, S. (2012). Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes. *IZA Discussion Paper,* 6338*, Institute for the Study of Labor.*

Gould, S. L. (1996). Strategies used by secondary school students in learning new concepts which require spatial visualization. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. SCIENCE-NEW YORK THEN WASHINGTON, 320(5880), 1164.

Hambleton, R. K., & Swaminathan, H. (1985). A Look at Psychometrics in the Netherlands.

Holland, P. W., & Wainer, H. Differential item functioning. 1993. *Abingdon: Routledge Google Scholar*.

INVALSI (2016). Rilevazione nazionale degli apprendimenti 2015-2016. Le rilevazioni degli apprendimenti.

Retrieved March, 2017, from
http://www.invalsi.it/ invalsi/doc_evidenza/2016/07_Rapporto_Prove_INVALSI_2016.pdf

INVALSI (2016b). Rilevazione nazionale degli apprendimenti 2015-2016. Rapporto tecnico. Retrieved July 2017 from
http://www.invalsi.it/invalsi/doc_evidenza/2016/002_Rapporto_tecnico_2016.pdf

INVALSI (2017). Il Quadro di Riferimento delle Prove di Matematica del Sistema Nazionale di Valutazione.
Retrieved July, 2017, from
http://www.invalsi.it/invalsi/doc_evidenza/2017/QdR2017_190417.pdf

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter? New directions for child and adolescent development, 2004(106), 5-21.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children`s ability perceptions. Journal of personality and Social Psychology, 63(6), 932-944.

Leder, G. C. (1992). Mathematics and gender: Changing perspectives.

Leder, G., & Forgasz, H. (2008). Mathematics education: new perspectives on gender. *ZDM – The International Journal on Mathematics Education*, 40(4), 513-518.

Linn, R. L., & Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational measurement*, *18*(2), 109-118.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. *TIMSS & PIRLS International Study*.

OECD, (2015). The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. PISA, OECD Publishing.

OECD, (2016). PISA 2015 Results (Volume I): Excellence and Equity in Education. OECD Publishing, Paris.

Osterlind, S. J. (1983). *Test item bias* (Vol. 30). Sage.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (Vol. 161). Sage Publications.

Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. Gender differences in mathematics: An integrative psychological approach, 294-315.

Penfield, R. D., & Camilli, G. (2007). Differential item functioning and item bias. In C. R. Rao & S. Sinharay (Eds.), Handbook of Statistics Psychometrics (Vol. 26, pp. 125-167). Amsterdam: Elsevier.

Pizarro, N., Gorgorió, N., & Albarracín, L. (2015, February). Primary teacher'approach to measurement estimation activities. In *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3227-3233).

Riegle-Crumb, C. (2005). The cross-national context of the gender gap in math and science. *The social organization of schooling*, 227-243.

Roche, A., & Clarke, D. M. (2004). When does successful comparison of decimals reflect conceptual understanding. *Mathematics education for the third millennium: Towards 2010*, 486-493.

Rudner, L. M. (1977). An Approach to Biased Item Identification Using Latent Trait Measurement Theory.

Rudner, L. M., Getson, P. R., & Knight, D. L. (1980). Biased item detection techniques. *Journal of Educational Statistics*, 213-233.

Sbaragli, S. (2012). Il ruolo delle misconcezioni nella didattica della matematica (Sbaragli Silvia), In In: Bolondi B., Fandiño Pinilla M.I. (2012). I quaderni della didattica. Metodi e strumenti per l'insegnamento e l'apprendimento della matematica, Napoli: Edises, 2012.

Segovia, I., & Castro, E. (2009). Computational and measurement estimation: curriculum foundations and research carried out at the University of Granada, Mathematics Didactics Department.

Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, *22*(2), 77-105.

Steinle, V., & Stacey, K. (2003). Grade-Related Trends in the Prevalence and Persistence of Decimal Misconceptions. *International Group for the Psychology of Mathematics Education*, *4*, 259-266.

Steinle, V., & Stacey, K. (2004). A longitudinal study of students' understanding of decimal notation: An overview and refined results. In *Proceedings of the 27th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 541-548).

Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group-mean differences: The concept of item bias. *Psychological Bulletin*, *99*(1), 118.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines.

Winkelmann, H., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: Results from a German large-scale study. *ZDM*, 40(4), 601-616.

Wright, B. D., Mead, R., & Draba, R. (1976). Detecting and correcting test item bias with a logistic response model. *Research Memorandum*, *22*.

Capitolo 5

Impatto di una variazione nella formulazione di un task

5.1 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 richiesta.

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 nella 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 e per una revisione della letteratura si rimanda al primo articolo riportato di seguito (paragrafo 5.2.1). 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 formulazioni di una stessa domanda senza che la risposta alla seconda domanda somministrata sia influenzata dall'aver già risposto alla prima domanda.

Le ricerche riportate in questo capitolo si propongono di superare questo ostacolo attraverso l'uso di prove standardizzate. L'idea di base, proposta da Branchetti e Viale (2015), consiste nel predisporre diverse versioni di uno stesso test; tutti i fascicoli hanno una parte consistente di item in comune che rimangono invariati mentre, le domande oggetto di studio, compaiono in versioni diverse all'interno dei diversi fascicoli. In questo modo, gli studenti che rispondono alle diverse versioni di un item non sono gli stessi, ma le loro risposte alle diverse versioni possono essere confrontate grazie alla stima dell'abilità degli studenti operata sulla parte in comune del test.

Nella prima ricerca effettuata, *Variazioni 1,* è stata validata la metodologia statistica e si sono avuti i primi riscontri su particolari tipologie di variazioni.

La seconda ricerca è stata svolta in collaborazione con l'istituto INVALSI ed ha permesso di studiare l'impatto di determinate variazioni sia da un punto di vista di didattico, sia per quanto riguarda le proprietà psicometriche dell'item.

La prima ricerca sarà esposta brevemente e per gli approfondimenti relativi al quadro teorico e alla metodologia utilizzata si rimanda agli articoli riportati di seguito.

Il progetto *Variazioni 2* non è ancora concluso, la sperimentazione in classe è terminata ma i dati non sono ancora a disposizione. Per questo motivo nella tesi verranno descritta la struttura della ricerca, le finalità, la costruzione dei fascicoli e verranno analizzate a priori alcune delle domande oggetto di studio con le relative variazioni.

5.2 Progetto *VARIAZIONI 1*[26]

Il progetto *Variazioni 1* nasce con l'obiettivo di fare una prima sperimentazione quantitativa che possa validare la metodologia proposta da Branchetti e Viale (2015) per l'analisi dell'impatto di una variazione, nella formulazione di un quesito in matematica, sulle risposte degli studenti.

Le domande di ricerca relative al progetto *Variazioni 1* sono:

1) **Come si può misurare l'impatto di una variazione?**

   Considerando poi che una variazione potrebbe causare differenze significative nelle risposte degli studenti appartenenti all'intera popolazione oppure differenze sostanziali nelle performance di un particolare gruppo di studenti della popolazione emerge una seconda domanda di ricerca

2) **Come si può evidenziare, attraverso l'analisi dei dati, la rilevanza di queste differenze nell'intera popolazione o in un sottoinsieme della popolazione?**

Si è scelto di partire da una prova standardizzata che fosse già stata somministrata a livello nazionale ed è stata scelta in particolare la prova INVALSI somministrata a livello 6 del 2013. La possibilità di utilizzare un test già somministrato a livello nazionale ha dato la garanzia di avere una prova costituita da item già validati e che mostravano buone proprietà misuratorie. Inoltre questo ha permesso anche di operare un confronto tra i dati ottenuti nella nuova sperimentazione e i dati della rilevazione nazionale.

La scelta del livello della prova (livello 6) è stata dettata dall'esigenza che gli studenti non avessero già affrontato tale prova, nonostante fosse stata somministrata negli anni passati; nel 2016, anno della sperimentazione relativa al progetto, la prova INVALSI di livello 6 non veniva più svolta già da alcuni anni e questo ha fatto sì che gli studenti non avessero utilizzato le prove passate per cimentarsi nei test.

La prova scelta è composta da 48 item in totale (30 domande) e il campione nazionale della prova somministrata nel 2013 era composto da 27 504 studenti (1528 in Emilia Romagna). A partire da questo test sono stati creati due fascicoli distinti:

- FASCICOLO ORIGINALE – identico alla rilevazione nazionale anche a livello di impaginazione

[26] Il progetto Variazioni 1 è stato svolto in collaborazione con Giorgio Bolondi (Università di Bolzano), Laura Branchetti (Università di Palermo), Alice Lemmo (Università di Bologna) e Boninsegna Rebecca (Università di Bologn

- FASCICOLO VARIATO – in cui 41 item sono stati riproposti invariati rispetto alla versione originale e 7 item sono stati modificati applicando diversi tipi di variazioni.

Le variazioni sono state effettuate in diverse direzioni (variazioni nella sintassi, variazioni della tipologia di domanda, variazioni grafiche, …) perché lo scopo primario della sperimentazione era quello validare lo strumento statistico, mettendone in luce le potenzialità per l'analisi di qualsiasi tipologia di variazione.

La parte di item in comune ai due fascicoli, chiamata *core-test*, è stata scelta in modo da essere rappresentativa dell'intero fascicolo e ha lo scopo di permettere il confronto dell'abilità degli studenti. Per questo motivo, si è scelto di variare solo una piccola percentuale degli item presenti nella prova.

La struttura dei due fascicoli è riportata nella tabella seguente e i fascicoli completi sono inseriti in appendice.

FASCICOLO ORIGINALE	FASCICOLO VARIATO	TIPOLOGIA DI VARIAZIONE
D1a	D1a	
D1b	D1b	
D1c	D1c	
D2a	D2a	
D2b	D2b	
D2c	D2c	
D3	**D3_var**	**Variazione nella tipologia di quesito: da quesito a risposta univoca a quesito a risposta multipla.**
D4	D4	
D5	D5	
D6a	D6a	
D6b	D6b	
D6c	D6c	
D7a	D7a	
D7b	D7b	
D8a	D8a	
D8b	**D8b_var**	**Rovesciamento della richiesta**
D9a	D9a	
D9b	D9b	
D9c	D9c	
D10a	D10a	
D10b	D10b	
D10c	D10c	
D11	D11	
D12	D12	
D13	D13	
D14	**D14_var**	**Variazione nella disposizione del testo rispetto all'immagine**
D15	D15	
D16	**D16_var**	**Variazione della sintassi**
D17a	D17a	

D17b	D17b	
D18	**D18_var**	**Assenza della figura**
D19	D19	
D20a	D20a	
D20b	D20b	
D21a	D21a	
D21b	D21b	
D22	**D22_var**	**Variazione dei dati**
D23	D23	
D24	D24	
D25a	D25a	
D25b	D25b	
D26a	D26a	
D26b	D26b	
D27	**D27_var**	**Variazione della sintassi**
D28	D28	
D29	D29	
D30	D30	

Tabella 5.1: Struttura dei fascicoli (originale e variato) somministrati per il progetto *Variazioni 1*. In grigio sono riportati gli item oggetto di variazioni.

La sperimentazione ha coinvolto 40 classi prime di scuole secondarie di primo grado in tutta l'Emilia Romagna. In ogni classe, metà degli alunni ha risposto al fascicolo originale e l'altra metà al fascicolo variato. In questo modo sono stati somministrati in tutto 777 fascicoli: 380 originali e 397 variati.

L'analisi del *core-test*, operata separatamente sui due fascicoli, ha evidenziato una buona tenuta in termini di coerenza interna degli item (α di Cronbach >0.8) e l'applicazione del modello di Rasch alla parte di core-test ha permesso osservare una distribuzione degli item in termini di difficoltà, molto simile a quella ottenuta nella prova a livello nazionale.

Per procedere nello studio e confrontare i risultati delle due versione dei 7 item non appartenenti al core-test, si è scelto in primo luogo di applicare nuovamente il modello di Rasch ai risultati ottenuti dall'intera popolazione di studenti (777 individui) sui 41 item in comune. È stato così possibile stimare un parametro di abilità per ogni studente e porre i 777 studenti sulla medesima scala di abilità. In base a questi parametri è stato quindi possibile creare i distractor plot delle restanti 7 domande nelle due versioni in cui ciascuna compare e confrontare i grafici delle due varianti.

Inoltre, i 41 item appartenenti al *core-test*, hanno avuto anche la funzione di *àncora* per applicare ai due fascicoli una procedura di *Test Equating*. Attraverso una calibrazione congiunta applicata utilizzando il software JMetrik è stato possibile esprimere sulla stessa scala i punteggi (abilità studenti e difficoltà domande) ottenuti nei due test. I risultati della calibrazione, relativamente alla stima della difficoltà dei quesiti, sono riportati nella tabella sottostante e si può osservare una differenza statisticamente significativa per tutte le variazioni apportate.

ITEM	DELTA	ST. ERR.	ITEM	DELTA	ST. ERR.	ITEM	DELTA	ST. ERR.
D1a	-2,06	0,11	**D9a**	0,82	0,08	**D20a**	-0,5	0,08
D1b	-2,4	0,12	**D9b**	0,63	0,08	**D20b**	0,65	0,08
D1c1	-1,25	0,09	**D9c**	1,32	0,09	**D21a**	0,75	0,08
D1c2	0,38	0,08	**D10a**	1,84	0,1	**D21b**	0,5	0,08
D2a	0,77	0,08	**D10b**	-2,26	0,11	**D22o**	0,1	0,11
D2b	0,19	0,08	**D10c**	1,33	0,09	**D22v**	-0,51	0,11
D2c	1,56	0,1	**D11**	1,49	0,09	**D23**	-0,76	0,08
D3o	0,03	0,11	**D12**	0,09	0,08	**D24**	-1,18	0,09
D3v	-0,57	0,11	**D13**	-1,32	0,09	**D25a**	-1,63	0,09
D4	-1,79	0,1	**D14o**	0,52	0,11	**D25b**	-0,47	0,08
D5	-1,24	0,09	**D14v**	0,88	0,12	**D26a**	-0,25	0,08
D6a	-0,74	0,08	**D15**	-0,38	0,08	**D26b**	0,59	0,08
D6b	-0,82	0,08	**D16o**	0,31	0,11	**D27o**	0,55	0,12
D6c	1,53	0,1	**D16v**	0,8	0,12	**D27v**	0,37	0,11
D7a	0,12	0,08	**D17a**	-1,47	0,09	**D28**	-0,45	0,08
D7b	0,03	0,08	**D17b**	0,38	0,08	**D29**	0,06	0,08
D8a	-0,1	0,08	**D18o**	0,68	0,12	**D30**	0,3	0,08
D8bo	0,47	0,11	**D18v**	1,09	0,12			
D8bv	0,77	0,12	**D19**	0,24	0,08			

Tabella 5.2: Difficoltà degli item dopo la procedura di Test Equating.

La domanda D22 è stata analizzata approfonditamente negli articoli che seguono perché permette di illustrare a pieno le potenzialità di questo tipo di analisi. Si è infatti osservato che, per questa domanda, la variazione non solo ha determinato un calo significativo della difficoltà dell'item, ma ha avuto una influenza diversa in base al livello di abilità degli studenti evidenziato dai distractor plot. Inoltre è stato molto interessante osservare che questa variazione ha avuto un impatto diverso tra maschi e femmine: per i maschi l'item variato è molto più semplice rispetto a quello originale, mentre per le femmine questo cambiamento risulta molto più lieve.

5.2.1 Uno strumento per analizzare l'impatto di una variazione nella formulazione di una domanda matematica[27] *(Revised Version)*

Rebecca Boninsegna, Università di Bologna
Giorgio Bolondi, Università di Bologna
Laura Branchetti, Università di Palermo
Chiara Giberti, Università di Trento
Alice Lemmo, Università di Palermo

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

In this paper we present a new methodology that allows to measure and analyse the impact of a variation in the formulation of a math question on students' responses. There are many researches on how the formulation of a question influences students' performances in solving a task; analysing the impact of a single variation in the formulation of a task is very complex in terms of students' resolution processes because it is not possible to administer two similar

[27] Articolo presentato al seminario "I dati INVALSI: uno strumento per la ricerca" tenutosi a Roma il 20 settembre 2016. L'articolo è in fase di pubblicazione sul volume omonimo.

tasks to the same student without them affecting each other. The statistical tool presented allows to overcome this obstacle by means of standardized tests analysed through the Rasch model and the main statistical indices. Specifically, the article describes the statistical tool used and its validation plan, based on a study that involves about 800 students starting from an INVALSI test for grade 6 administered in 2012-13. Finally, we analyse an example of a task to highlight the potential of the methodology that we present. Such analysis is presented not only to analyse the impact of a variation in students' performance but also to obtain educational information through a qualitative approach.

1. Introduzione

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

2. Presentazione del problema

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

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

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

Non è facile indagare quantitativamente l'effetto che le variazioni hanno sulle prestazioni degli studenti poiché è difficile, se non impossibile, realizzare la situazione di osservazione ottimale. In tal caso, infatti, uno stesso studente, a distanza di pochi minuti, dovrebbe rispondere a due domande molto simili, senza che la risposta fornita alla prima interferisca e influenzi la risoluzione dell'altra. Un eventuale studio qualitativo a posteriori, condotto mediante una discussione in aula e riguardante le strategie utilizzate dagli studenti nella risoluzione di un quesito, potrebbe suggerire interpretazioni a posteriori delle difficoltà incontrate in due quesiti simili ma diversi dal punto di vista della formulazione. Anche in questo caso però risulta comunque complicato superare l'ostacolo della influenza reciproca tra i due quesiti (quello originale e quello variato).

Lo scopo di questa ricerca è quello di indagare sperimentalmente in che modo le variazioni di formulazione del testo di un quesito matematico influenzino le risposte degli studenti. Docenti e ricercatori coinvolti nella produzione e nell'analisi dei test standardizzati sono particolarmente interessati a questo tema, che è cruciale nel momento in cui bisogna scegliere, tra diverse formulazioni di uno stesso quesito, quale somministrare. In questo caso, la metodologia qualitativa basata su un approccio interattivo non può essere considerata adeguata.

Branchetti e Viale (2015) hanno proposto una metodologia basata sulla IRT (*Item Response Theory*) e sul modello di Rasch (Rasch 1960) per studiare tale problema. Gli autori hanno condotto uno studio pilota su una popolazione di circa 200 studenti di scuola secondaria di I grado (livello 6 e 7), nel quale sono stati indagati gli effetti delle variazioni linguistiche, soprattutto sintattiche, apportate ad alcuni quesiti della prova di Matematica di livello 6 somministrata dall'INVALSI nell'anno 2009-10. Focalizzandosi sulla parte non variata del test, gli autori hanno confrontato le risposte degli studenti raccolte nella loro popolazione con le risposte agli stessi quesiti presentati nel test del 2009-10. Grazie a questo confronto è stato possibile confrontare i risultati degli studenti negli item variati in funzione del punteggio ottenuto sulla pare di test non variata e confrontare tali risultati con quelli della rilevazione nazionale; questo ha permesso di individuare quale percentuale di studenti era stata potenzialmente influenzata dalla variazione. In seguito gli autori hanno eseguito un'analisi qualitativa dei risultati. Lo studio presentato in questo report intende migliorare la metodologia qui descritta facendo uso di tecniche statistiche più sofisticate e validandola su una popolazione più ampia.

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

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

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

3. Lo strumento statistico

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

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

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

La procedura che ci proponiamo di esporre e validare è la seguente. Partiamo da un test (T) composto da N domande già sottoposto a un campione di studenti. Nel nostro caso, questo campione è composto da circa 27.000 studenti che nel 2013 hanno svolto la prova INVALSI di livello 6 ed è quindi rappresentativo della popolazione degli studenti italiani frequentanti la classe prima della scuola secondaria di I grado in quell'anno. La robustezza del campione nazionale e le analisi statistiche effettuate dall'INVALSI su questi dati permettono quindi di partire da un test che mostra ottime caratteristiche misuratorie, sia in termini di singoli item sia in termini globali.

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

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

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

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

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

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

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

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

4. Piano di validazione

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

Siamo partiti da un test INVALSI somministrato su scala nazionale nel maggio del 2013 a 590.728 studenti frequentanti la classe prima della scuola secondaria di I grado (livello 6). Il test originale (T) era composto da N=48 domande. Le analisi statistiche dell'INVALSI sono state condotte su un campione rappresentativo di circa 27000 studenti di cui un sottogruppo di 1.528 formava il campione rappresentativo della regione Emilia Romagna.

Abbiamo scelto m=7 domande del test T e le abbiamo modificate secondo diversi criteri legati alle variabili redazionali descritte da Laborde (1995). Abbiamo somministrato il nuovo test T', contenente gli item variati, e il test originale T a 777 studenti della stessa età e della stessa regione (Emilia Romagna), assicurandoci che gli studenti non avessero già risposto alla prova del 2013. In ciascuna delle 40 classi coinvolte nella ricerca, metà degli studenti hanno svolto il nuovo test T' (per un totale di 397 studenti) ed il resto ha risposto alla prova originale T (per un totale di 380 alunni). Gli alunni di ogni classe sono stati suddivisi in modo casuale allo scopo di considerare paragonabili le due popolazioni così ottenute.

In primo luogo abbiamo confrontato i risultati globali dei nostri test con i risultati del campione nazionale e con quello dell'Emilia Romagna. Per fare ciò abbiamo applicato il modello di Rasch sia sui test interi T e T', sia sulle 41 (N-m) domande in comune del *coretest* CT. Questo ci ha permesso di confrontare i risultati ottenuti dalla somministrazione dei due fascicoli con la rilevazione a livello nazionale del 2013. Infatti, attraverso l'analisi delle mappe di Wright, è stato possibile confrontare le distribuzioni, relative alle diverse somministrazioni, delle 41 domande del CT in funzione dei parametri di difficoltà di Rasch e verificarne la corrispondenza. In aggiunta al modello di Rasch, abbiamo utilizzato indici specifici provenienti dalla teoria classica dei test: per esempio, l'alpha di Cronbach ha permesso di verificare che fosse rispettata la coerenza interna del test e le principali caratteristiche psicometriche degli item.

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

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

5. Esempio di analisi

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

D22. Quale dei seguenti numeri interi è più vicino al risultato di questa moltiplicazione?

4,82 x 9,95

A. ☐ 36

B. ☐ 42

C. ☐ 48

D. ☐ 50

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

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

D22. Quale dei seguenti numeri è più vicino al risultato di questa moltiplicazione?

482 x 995

A. ☐ 360.000

B. ☐ 420.000

C. ☐ 480.000

D. ☐ 500.000

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

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

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

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

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

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

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

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

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

Tab. 4 - Percentuali di risposta per l'item D22 (in forma originale e variata).
Risposta corretta: C.

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

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

A questo punto, può essere interessante analizzare i *distractor plot* per indagare se le differenze precedentemente identificate sono distribuite uniformemente su tutti gli studenti o se questi cambiamenti hanno influenzato maggiormente studenti con un certo livello di abilità. I *distractor plot* (fig. 2) sono stati realizzati come funzione dell'abilità degli studenti valutati sul CT.

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

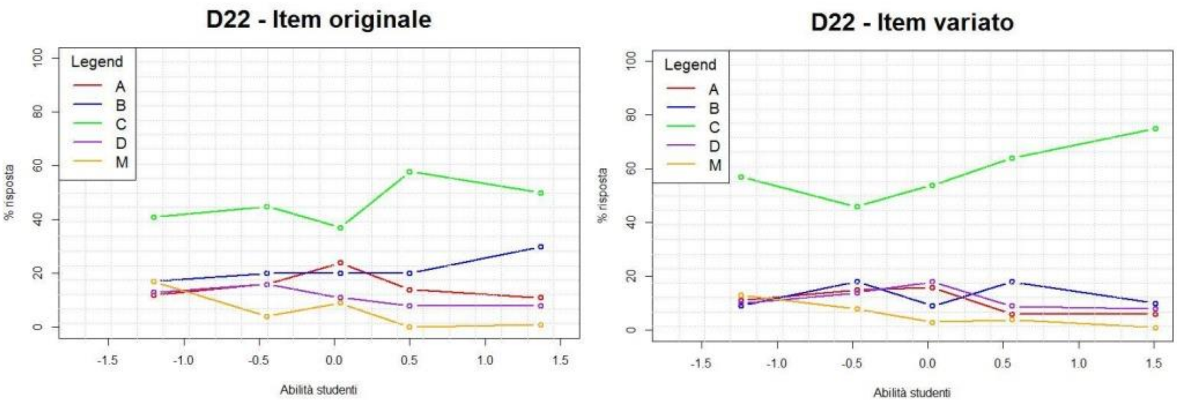


Fig. 2 - D22 – Distractor plot: item originale e variato.

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

Altri elementi interessanti emergono dall'analisi dell'andamento relativo alla scelta delle altre opzioni di risposta. Ad esempio, l'opzione B (in blu, fig. 2) mostra una variazione più sensibile della percentuale di risposta: viene scelta infatti dal 21% degli studenti nella versione originale e dal 13% degli studenti nella versione variata. Questo risultato può essere approfondito analizzando l'andamento della risposta nei *distractor plot*; si nota infatti che nella versione originale l'opzione B viene scelta maggiormente da studenti con un alto livello di abilità, cosa che non avviene per la versione variata. Dopo la modifica questo distrattore risulta molto meno appetibile per questi studenti, i quali optano invece per la risposta corretta, che in questo modo risulta crescente per livelli di abilità medi e alti.

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

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

Tab. 5 - Percentuali di risposta all'item D22 in base al genere.
Risposta corretta: C.

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

6. Conclusioni

La metodologia messa a punto per indagare l'impatto di una variazione nella formulazione di un quesito sulla distribuzione di frequenza di risposte degli studenti si è rivelata efficace in quanto ha fatto emergere alcune differenze nelle distribuzioni di risposte studiate in relazione al livello di abilità degli studenti manifestata nel *core test* e al genere. Questo strumento statistico permette di evidenziare se la variazione apportata al quesito, ha influito sulle risposte degli studenti e su quali livelli di abilità l'impatto è stato più significativo. Inoltre, questo approccio ha permesso di evidenziare differenze di performance tra diverse categorie di studenti e di indicare percorsi per ulteriori indagini sulle cause di queste differenze. Questa metodologia sembra adeguata per analizzare gli effetti di ulteriori categorie di variazioni. La metodologia quantitativa presentata potrebbe contribuire, in future ricerche, nel far emergere macro-fenomeni che possono successivamente essere investigati attraverso un'impostazione sperimentale qualitativa. Attraverso questa seconda fase è possibile verificare le ipotesi di un cambio di strategia indotto dal cambiamento nella formulazione delle domande. Un approccio di questo tipo all'analisi dell'impatto di una variazione in un quesito di matematica potrebbe quindi diventare un tassello fondamentale di una metodologia mista quantitativa e qualitativa (Johnson and Onwuegbuzie 2004) che consenta di indagare anche qualitativamente nuovi fenomeni partendo da evidenze quantitative.

Riferimenti bibliografici

Barbaranelli C. e Natali E. (2005), *I test psicologici: teorie e modelli psicometrici*, Carrocci, Roma.

Branchetti L. and Viale M. (2015), *Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico*, in Ostinelli M., a cura di, *Didattica dell'italiano. Problemi e prospettive*, , *Dipartimento formazione e apprendimento* - Scuola universitaria professionale della Svizzera italiana (*SUPSI*), Locarno: 138 - 148.

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

De Corte E. and Verschaffel L. (1985), "Beginning first graders' initial representation of arithmetic word problems", *The Journal of Mathematical Behavior*, 4: 3-21.

Daroczy G., Wolska M., Meurers W.D. and Nuerk HC. (2015), "Word problems: a review of linguistic and numerical factors contributing to their difficulty", *Frontiers in Psychology*, 6: 1-13.

Duval R. (1991), "Interaction des niveaux de représentation dans la compréhension de textes", *Annales de Didactique et de Sciences Cognitives,* 4: 163-196.

INVALSI (2013), *Rilevazioni nazionali sugli apprendimenti 2012-2013. Rapporto tecnico*, testo disponibile al sito: http://www.invalsi.it/snvpn2013/rapporti/Rapporto_tecnico_SNV2013_12.pdf, data di consultazione: 31/05/2017.

Johnson R.B. and Onwuegbuzie A.J. (2004), "Mixed methods research: A research paradigm whose time has come", *Educational Researcher*, 33, 7: 14-26.

Kolen M.J. and Brennan R.L. (1995), *Test Equating: Methods and Practices*, Springer, New York, NY.

Laborde C. (1995), "Occorre imparare a leggere e scrivere in matematica?", *La Matematica e la sua Didattica*, 2: 121-135.

Mayer R. (1982), *The psychology of mathematical problem solving*, in Lester F. and Garofalo L.J., eds., *Mathematical Problem Solving. Issues in Research*, The Franklin Institute Press, Philadelphia, PA.

Nesher P. (1982), *Levels of description in the analysis of addition and subtraction word problems*, in Moser J., ed., *Addition and Subtraction: A Cognitive Perspective*, Lawrence Erlbaum Associates, Hillsdale, NJ.

OECD (2013), *Pisa 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing, Paris.

Rasch G. (1960), *Probabilistic Models for Some Intelligence and Attainment Tests*, Danmarks Paedagogiske Institut, Copenhagen.

Thevenot C. and Oakhill J. (2005), "The strategic use of alternative representation in arithmetic word problem solving", *Quarterly Journal of Experimental Psychology*, 58: 1311–1323.

5.2.1 A tool for analysing the impact of the formulation on the performance of students answering to a mathematical item

Laura Branchetti, Università di Palermo
Chiara Giberti, Università di Trento
Giorgio Bolondi, Università di Bologna

We present a methodology for measuring and analysing the impact on students' performances of a change in the formulation of an item. We describe the statistical tools employed and the implemented validation plan.

STATEMENT OF THE PROBLEM

When students face a mathematical task, their performances are influenced by the formulation of the task; in particular, if we think to a student answering an item of a standardized test, the formulation is particularly important because the task is not proposed by the teacher who usually assesses the students. Analyzing the factors affecting the problem solving activities, Nesher (1982) lists three components that may vary in a *word problem*: logical (operations, lack or abundance of data, ...), syntactic (position of the question in the text, number of words, ...) and semantic (contextual relations, implicit suggestions, ...). Daroczy, Wolska, Meurers & Nuerk (2015) proposed a review of the factors affecting the difficulty of word problems, distinguishing between “three components of WP difficulty: (i) the linguistic complexity of the problem text itself, (ii) the numerical complexity of the arithmetic problem, and (iii) the relation between the linguistic and the numerical complexity of a problem”. Hence, modifications in the text, even small, may cause changes in the students' approaches to problem solving, as D'Amore (2000) highlighted. Duval (1991) named these modifications *redational variables* and Laborde redefined them in 1995 in order to include also not verbal modification, like introducing draws (Laborde, 2005). In the quoted studies, the effect of variations was studied from a qualitative point of view and in interaction with the students.

The effect of variations on the students' performances is not easy to investigate because the optimal situation to study is impossible to realize. A student should answer indeed two very similar questions, and should “forget” to have faced the first question while answering the second. Inevitably the first task would influence the second or the change should be so evident to transform deeply the nature of the question itself. While a qualitative interactive study of the evolution of meanings and strategies in group or classroom discussions could suggest indirectly interpretations *a posteriori* of some students' difficulties in a particular task, we want to exploring directly the questions: how the variations influence the students' answers during the test? What would the same students do if they have to answer independently the original and the modified question? Does a variation cause a significantly different behaviour in "real-time" while a student is answering the question during the test?

Teachers and researchers involved in standardized tests are particularly interested in these aspects, that are crucial when one of the two formulations of the same question has to be disregarded in favour of the other. In this case, qualitative methodology based on an interactive approach is not adequate to answer the question.

Branchetti & Viale (2015) proposed a methodology based on the IRT (*Item Response Theory*) and on the Rasch model (Rasch, 1960) in order to study such a problem. They carried out a pilot study with about 200 students concerning linguistic changes in the texts of some of problems of a standardized test, in particular varying the syntax used. The authors decided to compare the expected students' answers to the questions formulated in the original way with the real answers to the modified ones, forecasting the expected answers by means of the statistical data provided by the National Institute for Assessment and Evaluation of the Educational and Instructional system (INVALSI). They performed only a qualitative analysis of some cases, without the support of any software. In this study we go on improving their methodology using statistical techniques and grounding the validation of the methodology on a wider population's data analysis.

Our research questions are:

- How can we measure the impact of a variation?
- A variation may cause significant changes in the answers of a population or in the performances of a particular group of students. How can we make emerge from the data analysis the relevance of a variation on the whole population or on some subgroups?

We present here the research methodology and one example of analysis of the impact of a variation (numerical: numbers magnitude and kind of number) carried out in the frame of a wider research in which we analysed the effects of different kinds of variation on 777 students.

THE STATISTICAL TOOL

National and International surveys - such as INVALSI (the Italian National Institute for the Evaluation of Educational System) in Italy and OECD-PISA on a world scale - often use a Rasch Model to analyse the results of the students, especially when it is necessary a comparison between two different tests or the comparison between students (Barbaranelli & Natali, 2005; INVALSI, 2013; OECD, 2013). Rasch Model is a simple logistic model that belongs to the Item Response Theory (IRT) class; it estimates a difficulty parameter for each item of the test and an ability parameter for each student. The core concept of this approach to test results evaluation is the probability of a student of answering correctly to a task. The IRT models, and between these the Rasch Model, express the probability of choosing the correct answer in an item as a function of the item's difficulty and the ability of the student. In particular, the probability of choosing the correct answer of an item is function of the relative ability of the student, which is his inherent ability compared with the difficulty of the item. The relation between the students' ability and the probability of the correct answer is represented by a curve called *Item Characteristic Curve* (ICC). In a similar way it is possible to use Rasch parameters to represent also the empirical data and, in particular we can represent the trend of each possible response to an item as a function of the students' ability. This specific graphs are called *Distractor Plots* and give many information regarding how students respond to an item. The prediction is significant if the students' sample size is large enough and if the statistical parameters' values' constraints (p-value, Cronbach alpha, among others) are respected. If such a prediction is valuable, this information may be used as a blind information about the students' expected performances in respect of an item.

Let us now reconsider the main research question, i.e. by means of what methodology it is possible to evaluate the impact of a variation of the formulation on students' performances in standardized tests.

The procedure that we propose and validate is the following one. We start from a test (T) composed by N items administered to a sample of students (in our validation case, this sample is the sample composed of approximately 27.000 students chosen for a national INVALSI Survey, which was representative of the population of Italian grade 6 students). In this test (T), a core test (CT) - composed of N-m items of the task- which gives a statistically consistent (with respect to the statistical parameters) measurement of students' ability is fixed.

A Rasch analysis is then performed on this core task composed by N-m items. Let us denote by $A_1, A_2, \ldots, A_m$ the remaining items. The probability that a student of a given ability level *p* (measured on the Rasch scale) answers correctly to the item A_j can be computed, and the relationship between the ability of the student and the probability of the different choices can be visualized through a graph (the above-mentioned distractor plot). We note that, due to the characteristics of the Rasch model, this relationship includes also the information on how the choices of the students with a given score (roughly speaking, the percentage of correct answers to the whole test) on the test T are distributed, when answering to the item A_j (correct answer, missing, choice of a particular distractor....).

Each one of the items $A_1, A_2, \ldots, A_m$ is then modified, by performing on it a single, well-individuated variation, getting a new set of items $A'_1, A'_2, \ldots, A'_m$. A new test T' composed by the fixed N-m items of the core test CT and the varied m items is then assembled. A sample of classes is selected, and in each class the original test T is administered to half of the students (randomly chosen) and the varied test T' to the other half. The core test CT allows to anchor our population' results to the results of the original Invalsi's sample. Population. The Rasch model is then applied to the common core CT of items, included in T'. By comparing the Rasch score calculated using the common core of items CT for the two groups, and the characteristic curves and distractor plots of resp. A_j and A'_j, j= 1, …, m, we can visualize the first evidences of differences and similarities between the two variants of the item. Then we can represent how the students answered to the modified items as function of their ability, in other words we can make distractor plots of the two versions of an item, plotting the empirical data as functions of the ability parameter calculated on common part of the test (CT).

For a more rigorous description of the statistical procedure, in order to compare the scores of two students obtained from two different tests that measure the same latent trait, we have then to apply a specific procedure of *test equating*. This procedure has the task of expressing on the same scale the results of the two tests. In particular, we use a *concurrent calibration* procedure, which is considered to be more precise than a separated calibration (Kolen & Brennan, 1995), that permit to estimate a difficulty parameter for each item and an ability parameter for each student, considering the results of both of the tests at the same time. This method allows us to express all the parameters estimated on the same scale and, in particular we can compare the difficulty parameters of the items $A_1, A_2, \ldots, A_m$, respectively with the parameters of the items $A'_1, A'_2, \ldots, A'_m$. This kind of procedure used to link two different tests is often applied to compare results of students over time and it is called *anchoring technique*: two tests are administered to different groups of students, but the two tests contain a set of common items used to make an anchorage between the results. In this way we can first of all formulate quantitative-based conjectures about the effects of specific kinds of variations, both with top-down and with bottom-up approaches.

THE VALIDATION PLAN

We implemented a validation plan for our procedure as follows. Our test T consisted of the test administered to 590.728 Italian students of grade 6, during may 2013. The test was composed of N=48 items and the statistical analysis was performed on a representative sample of 27.504 students. 1.528 of them were a representative sample of the students from the Italian region of Emilia-Romagna. We choose m=7 items and we changed them along different directions, as described in the theoretical framework (by operating on the lexicon, the syntax, the use of figures, the registers of representation and so on). We administered the new test T' and the original test T to 777 students of the same age from the same region, who had not participated to the test T in 2013. In particular, in each of the 40 classes involved in the trial, half of the class randomly chosen responded to the new test T' and the rest of the students responded to the original test T. The validation plan is based on the analysis of the 777 tests, including 380 original test T and 397 varied test T', and the two population of students who responded to the two test are comparable because are sourced randomly in the same classes. First, we compared the global results of our tests and the results of the national and regional original INVALSI test. For this purpose, we used results of Rasch analysis on the whole test and on the Core Test, but we also observed specific indexes belonging to Classical Test Theory, such as the alpha-Cronbach that gives us the internal consistency of the tests. In particular, we observed that the distribution of the 41 items as function of the Rasch difficulty parameter is similar in the two tests of our trial and, even if there are minot local differences, this distribution is also globally similar to that of the national and regional survey. Then the statistical procedure of Rasch analysis and test equating has been applied to T and T', and this allowed to point out several quantitative effects of the variations performed.

We implemented also a qualitative analysis in order to confirm, from a different point of view, our methodology and to complete our research. On the basis of this analysis, a semi-structured interview has been prepared and 15 students out of the 777 who have been interviewed. The interviews have been transcribed and compared with the quantitative results of the tests.

AN EXAMPLE OF ANALYSIS

In this paper we propose and analyse one of the seven modified items. In this case the variation is numerical: numbers magnitude and kind of numbers. The item in the original test asks the student to estimate the result of a product of two decimal numbers.

D22. Which of the following integer numbers is closer to the result of this multiplication?

4,82 x 9,95

A. 36

B. 42

C. 48

D. 50

Figure 1: D22 – item in the original form (test T).

The *question intent* of the varied item is the same but the numbers are integer and their order of magnitude is different. It is important to notice that all the responses in the varied form are analogous to the ones in the original item.

D22. Which of the following integer numbers is closer to the result of this multiplication?

482 x 995

A. 360.000

B. 420.000

C. 480.000

D. 500.000

Figure 2: D22 – item in the varied form (test T').

Daroczy et al. (2015) stated that variations in problem solving strategies have been studied extensively and can depend on linguistic factors like wording, semantic categories and propositions. However, how individuals come up with mathematical solution strategies can also be influenced by numerical factors like number magnitude (Thevenot and Oakhill, 2005). Such variables, which are independent from the linguistic ones and make the *word problems* even harder, have rarely been studied.

We decided to analyse this kind of variation in order to explore the potentiality of our methodology. Our hypothesis was that the students' performances would have been different passing from decimal numbers to natural numbers; our research tools permitted to us to investigate how much the variation influenced the answers, to compare on a unique scale the students' answers and to observe more in depth the results differentiating between students with different Rasch scores, different gender, foreign or mother tongue. Thus we had the possibility to see which categories of students were more influenced by the variation and to hypothesize why they should have been influenced by the variation.

The question intent concerns the estimate of operations results and the students were explicitly asked to say which was to number closer to the product result. Nevertheless, even if the kind of numbers involved shouldn't change the nature of the problem, we may expect that the students are lead by the classroom habits to approximate or to compute with different procedures the result.

In this case an *a priori* analysis lead us to conjecture that changing the kind of numbers in a multiplication would have changed the students' strategies. The change is usually thought to be due to differences in the teaching practices in which operations with decimal and natural numbers are involved at school, but if it is the only reason why the students are driven to choose one distractor or the correct answer, the change should influence all the students in the same way. With our analysis we can explore instead possible differences, that may be due to a different stability and generality of the students' knowledge. Our analysis can thus provide information concerning the features of the students' knowledge about methods for estimating; the students may in fact be lead to round to multiples of 10 with natural numbers but may have never used this strategy with decimal numbers. As a results they should have a partial knowledge of methods for estimate, concerning only some kind of numbers. On the contrary we might observe that the students, once overpassed an intermediate Rasch score, are not conveyed by this kind of change since their knowledge is more complete.

Since the item was formulated as a multiple choice question, our *a priori* analysis can only concern the students' choices in a shortlist of four possibilities and we can try to conjecture why a student could choose a particular possibility. In the original version and in the modified one we proposed the same digits but a difference in the magnitude (the numbers in the modified version were 2 orders bigger), both in the text and in the answers.

A common strategy would lead to the same answer, with the only difference of the magnitude.

A significant difference in the item score and in the percentage of students' choices of a distractor in respect of another are thus signals of different strategies.

A priori, some of the possible strategies are:

- to round both number to the closer integers (or hundreds);
- to consider only the first digit;
- to approximated both numbers by excess or by defect;
- mixed strategies

The data analysis, performed as we have explained before, demonstrates that with this variation the item becomes much easier. Indeed the percentage of correct answers in the original item is 46% and in the varied one increases to 59%.

	% original item	% varied item
A	15%	11%
B	21%	13%
C	46%	59%
D	11%	12%
MISSING	6%	6%

Table 1: Answers percentages for item D22 (original form and varied form). Correct answer: C.

As we can see in the table before, the variation has mostly influenced the choice of distractor B. Whether the distractors A and D increase or decrease only of some percentage points, answer B loses about 8% in the varied item. The variation does not influence the percentage of students who don't answer to the item and this may mean that, despite the different difficulty of the two versions, almost all the students are confident enough to try to answer.

The *concurrent calibration technique* applied to both of the test simultaneously allows us to estimate the difficulty parameters of all the items (including the two version of the 7 items modified) and to consider them on the same scale. The comparison between difficulty parameters estimated in this way gives us an additional proof that the variation in this case had made the item easier. In fact, after using the anchoring procedure, the difficulty parameter of the original item is 0,10 and the difficulty of the varied item is -0.51, both with a standard error of 0.11, hence the difference is statistically significant.

At this point of our analysis, it is interesting to use Rasch analysis and in particular distractor plots to investigate if the differences identified before are evenly distributed on all the students or if these changes have influenced in particular students with a certain ability level. Using the Rasch Model to analyse the core test (CT) composed by 41 common item, we estimate an ability parameter for each of the 777 students. The distractor plots below represent the empirical data of the two versions of the item D22 as function of the ability of the students evaluated before using the core test.

The behaviour of the correct answer and the incorrect ones is different in the two items. In particular the trend of the correct answer is more regularly increasing in the varied item. In both of the two forms we observe that this question is not very discriminant. In other words this item doesn't distinguish well between students with high ability levels and students with lower ones. This information is also confirmed by the analysis of the whole tests: the discrimination parameter of this item is 0,15 for what concerns the original test and 0,22 in the varied test (hence slightly better). Also in the national survey this item has a discrimination above the threshold (0,20) but it is interesting to notice that the variation improved the statistical properties of the item. Indeed, the discrimination is higher in the varied form of the item and also comparing the weighted index of the two items, the varied one gives better results.

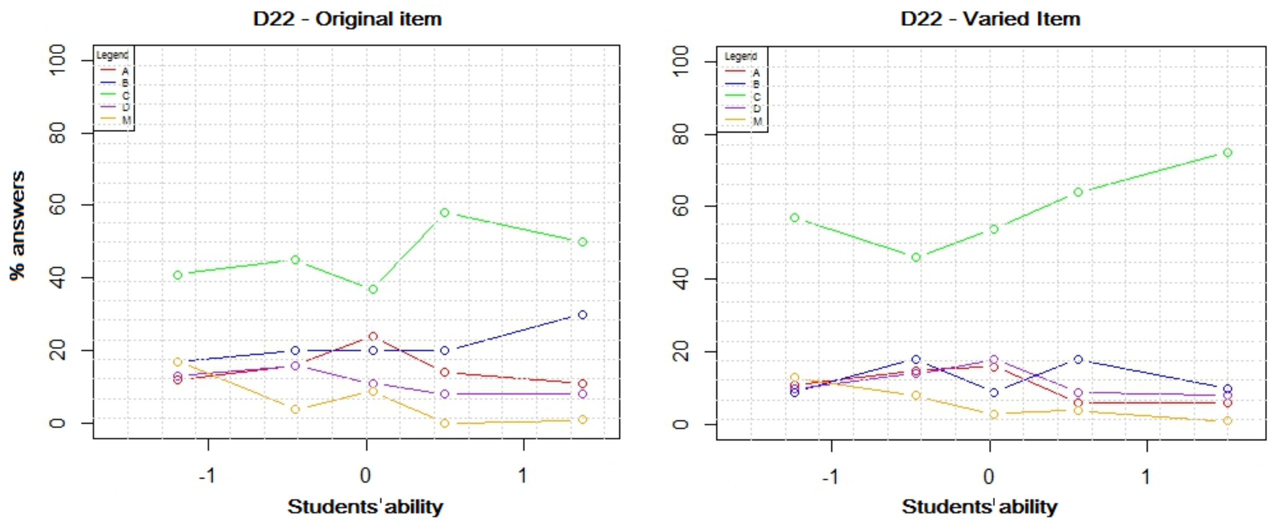


Figure 3: D22 – Distractor plots: original item and varied item.

In this paper we are interested in analysing the comparison between the two items and, in particular, many useful details raise from the comparison between distractor trends. For example, we can focus on the trend of the distractor B (blue) which is the one with the most change in the percentage seen before: comparing the two plots we can see that distractor B is often chosen by students with high ability level in the original item but the variation makes it much less attractive for those students.

Furthermore, the analysis of this item is very interesting also because the variation has a very different impact on males and females. In the table below the percentage of each answer for male and female and for both the two versions of the item are presented.

	MALE		**FEMALE**	
	% original item	% varied item	% original item	% varied item
A	16%	10%	12%	11%
B	23%	12%	20%	13%
C	**44%**	**62%**	**50%**	**56%**
D	11%	11%	10%	13%
MISSING	5%	4%	8%	7%

Table 3: Answers percentages for item D22 distinguishing male and female. Correct answer is C.

In the original item with the decimal numbers, correct answers are the 44% for males and 50% for females. The variation has a huge impact on males' performances and, answering to the varied item, 18% more of the students choose the correct answer. Instead, if we observe female percentages, we can see that the correct answer takes only the 6% more in the varied answer. This may mean that male and female apply different strategies to solve this item and the strategies used by female are almost the same even if the numbers are changed.

CONCLUSION

Our procedure has shown to be efficient in pointing out differences in the students' performances and in indicating paths for further investigations concerning the causes of these differences. The methodology resulted to be effective since it provided us information about the effects of variations on different categories of students. Moreover, it seems likely that it can allow to analyse the effects of further possible categories of variations in the formulation of an item, hypothesized *a priori* by means of theoretical dissertations or qualitative analyses carried out with small groups of students.

This statistical procedure may become part of a mixed method research (Johnson & Onwuegbuzie, 2004). In particular, we are going to adopt for our further researches a design

QUAN → QUAL. In the a posteriori qualitative part, that we are redefining after a pilot study and we are going to carry out in the future, we expect several interesting indicators to emerge from the activities of paraphrasing and explaining the task, which could confirm our hypotheses, based on the quantitative results, concerning a change of strategy induced by the change of formulation .

References

Barbaranelli, C., & Natali, E. (2005). *I test psicologici: teorie e modelli psicometrici.* Roma: Carrocci Editore.

Branchetti, L., & Viale, M. (2015). Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico. In Ostinelli M. (2015). *La didattica dell'italiano. Problemi e prospettive.* Proceedings of the conference "Quale didattica dell'italiano? Problemi e prospettive", Locarno, ottobre 2014.

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

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* 6:348.

Duval, R. (1991), Interaction des différents niveaux de représentation dans la compréhension de textes. *Annales de Didactique et de sciences cognitives*, pp. 136-193. Strasbourg : Edition de l′IREM

INVALSI (2013), Rilevazione nazionale degli apprendimenti 2012-2013. Rapporto tecnico. http://www.invalsi.it/invalsi/doc_evidenza/2013/034_Rapporto_Prove_INVALSI_2015.pdf

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come, *Educational Researcher, 33*(7), 14-26.

Kolen, M. J., & Brennan, R. L. (2013). *Test equating: Methods and Practices*. Springer Science & Business Media.

Laborde, C. (1995), Occorre imparare a leggere e scrivere in matematica?, *La matematica e la sua didattica,* vol. 2, pp. 121-135.

Nesher, P. (1982). Levels of description in the analysis of addition and subtraction word problems. In J. Moser (Ed.), *Addition and subtraction: A cognitive perspective.* Hillsdale, NJ: Lawrence Erlbaum.Associates

OECD (2013), Pisa 2012 Assessment and Analitical Frame- work: Mathematics, Reading, Science, Problem Solbing and Financial Literacy, OECD Publishing. http://dx.-doi.org/10.1787/9789264190511-en

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmarks Paedagogiske Institut.

Thevenot C., Oakhill J. (2005). The strategic use of alternative representation in arithmetic word problem solving. *Q. J. Exp. Psychol.* 58 1311–1323

5.2.2 A tool for analysing the impact of the formulation on the performance of students answering a mathematical item *(new version)*[28]

Laura Branchetti, Università di Palermo
Chiara Giberti, Università di Trento
Giorgio Bolondi, Università di Bolzano

1.Introduction

When facing a mathematical task, students are influenced by the formulation of the task itself. In particular, this may influence in a significant way their performance when dealing with an assessment task, for instance in the case of a word problem (D'Amore, 2014; D'Amore, 2000). This is a classical topic in educational research and for instance, a recent literature review for the case of arithmetical *word problems* is Daroczy, Wolska, Meurers & Nuerk (2015).

Understanding the relationships between formulation of a problem/task, students' approaches to problem solving and performances may have three kinds of impact:

- First, a theoretical one, in the direction of a better understanding of the results of the teaching-learning processes;
- Second, a practical one: it may help task-assessment designers (teachers, large-scale assessment authors, researchers….) both in well-defining the *question intent* and in monitoring different levels of difficulty. Hence it can contribute to better define the summative aspects of assessment;
- Third, a didactical one: it may help in interpreting students' behaviors when answering an assessment question. Hence, it can give a contribution to formative assessment.

In this paper, we propose our contribution to this general problem by designing and validating

[28] Revisione e approfondimento dell'articolo presentato al convegno ICME, in fase di sottomissione alla rivista *Frontiers in measurement and assessment*.

a quantitative methodology for measuring the impact of a variation in the formulation of an item on students’ performances. The methodology, elaborated to deal with this important issue in Mathematics education, is general enough to be useful also to investigate other questions in the field but also to go beyond it, when general constraints concerning the features of data are respected.

In the first part of the paper, we propose a brief literature review concerning studies in the field of *word problems* and text formulation within the field of Mathematics education, in order to show the relevance of the topic, to focus the attention on the results that we decided to investigate better and to explain the reasons why we investigated quantitatively the impact of variations in the formulation on students' performances. Since we are not trying to discuss here new findings, but we are just presenting a quantitative, statistically significant, methodological approach to the investigation of the classical topic in order to propose new insights inspired by distributions and graphics, we didn't pursue the goal of being exhaustive in this review but just to identify significant problems to which apply our method.

In the second part of the paper we present in dephts the methodology, comparing it with other methodologies used to investigae the same topic, and, in the last part, we provide two examples of possible further qualitative data analysis inspired by our results.

2.Backgrounds of the research and statement of the problem

2.1. Variables in the formulation of a problem

Considering the general problem of the comprehension of a text and the information retrieval (not necessarily bounded to a mathematical *word problem*), Duval (1991) introduced what he named *variables rédactionnelles*, stating that they influence the student´s cognitive processes. Laborde redefined them in 1995 including also not verbal modifications and different forms of representations (Laborde, 1995). She listed factors concerning editing, punctuation, syntactical complexity, word density, order of the information, spell out of intermediate objects needed for the solution. The research developed in this thread showed that modifications in the text, even small, may cause changes in the students' approaches to problem solving, as D'Amore (2000) highlighted, providing also a detailed list of previous researches and related results. As pointed out by Bagni (2005), the crucial point is not the fact that a formulation is necessarily better or worse than another one, but the fact that changes in the formulation actually change the problem.

Nowadays, we can say that an extensive research literature has been developed on the theme, allowing to classify from different perspective the factor that may impact on the students' performances. We list just some of them that we used as references to choose specific problems to investigate.

Analysing the factors affecting the problem solving activities, Nesher (1982), while categorizing the variations, listed three components that may vary in a *word problem*: logical (operations, lack or abundance of data, ...), syntactic (position of the question in the text, number of words, ...) and semantic (contextual relations, implicit suggestions, ...). How individuals come up with mathematical solution strategies can also be influenced by numerical factors like number magnitude (Thevenot & Oakhill, 2005). Daroczy et al. (2015) proposed a review of the factors affecting the difficulty of *word problems,* distinguishing between "three components of WP difficulty: (i) the linguistic complexity of the problem text itself, (ii) the numerical complexity of the arithmetic problem, and (iii) the relation between the linguistic and the numerical complexity of a problem" (p.1). Daroczy et al. (2015) also stated that variations in problem solving strategies were studied extensively and can depend on linguistic factors like wording, semantic categories and propositions.

Branchetti & Viale (2015) contributed to this part of the general research, focusing on linguistic variations and showing how far linguistic variations (in particular, in the syntactic structure of the sencences) can affect students´ performances. This study is relevant for us since it introduced a methodology to investigate the topic that is the starting point of our approach.

2.2. Research methodologies

The methodologies used in order to investigate the impact of these variations are almost quantitative and often consists in the display of different test containing original forms of the tasks and varied ones. In some studies, the same question is revised and reformulated in many versions and all the different forms of the task are administered to the same group of students (e.g. Lepik, 1990; Cummins, Kintsch, Reusser & Weimer, 1981; De Corte, Verschaffel & De Win., 1985; Thevenot, Devidal, Barrouillet, & Fayol, 2007). In this case, the ability of the students responding to the different versions of each task is the same but the main problem of this approach consist in the unavoidable influence of the first task administered on the resolution of the second one. In almost all of this researches, the way to partially overcome this obstacle consist in changing the order of the version proposed to the students or to allow time to pass between the first time the student faces the task and when he faces the other version (e.g. Vicente, Orrantia & Verschaffel, 2007). For example in the research of De Corte et al. (1985) the authors prepared two tests: in test A the problems compare in a form similar to those of textbooks problems and in test B the same problems were reformulated to be more clear to students. They administered both of the tests one week apart to a sample of 170 students but half of the students faced test A and one-week later test B and the other half face the two test in the opposite order.

Other quantitative approaches to this issue try to overcome to the obstacle explained before by using different populations of students, and this is the case for instance of Nesher research (1976): 4 different tests containing different version of the same problems were administered to 800 students in total, but each student answered only to one version of the test.

In some cases, we find also qualitative researches with the aim of analyze the impact of a variation based on interviews of the students and asking them to compare the two versions of a task or based on the analysis of protocols of the students (e.g. Spanos, Rhodes & Dale, 1988).

In general, the methodologies can be framed either in quantitative methods or in qualitative methods. Nevertheless, recent educational research is moving more and more towards a mixed method approach (Johnson & Onwuegbuzie, 2004). In this direction, we found the work of Abedi and Lord (2001) that combines two steps: the first one consists in interviews of students and the second one is quantitative and based on two version of the same test administered to a sample of 1174 students. Moreover, this research is particularly interesting for us because the two tests where composed by 20 varied problems (which compared in the original form in one of the two tests and in the varied form in the other) and 5 control items, unchanged in the two tests.

Branchetti & Viale (2015) presented another example of mixed method research on this issue. In particular, they proposed a methodology based on the IRT (*Item Response Theory*) and on the Rasch model (Rasch, 1960) in order to study the problem of the syntactic structure of the formulation of a task. They carried out a pilot study with about 200 students concerning linguistic changes in the texts of some of problems of a standardized test, in particular varying the syntax used. The authors decided to compare the expected students' answers to the questions formulated in the original way with the real answers to the modified ones, forecasting the expected answers by means of the statistical data provided by the National Institute for Assessment and Evaluation of the Educational and Instructional system (INVALSI). They then performed a qualitative analysis of some cases.

The innovative nature of this approach, respect to the previous quantitative researches, consist in applying the statistical tools generally used to analyse results of large scale assessments (Rasch model) also for the quantitative part of an empirical research in math education.

2.3. Statistical tools

National and International large-scale surveys - such as INVALSI (the Italian National Institute for the Evaluation of Educational System) in Italy and OECD-PISA on a world scale - often use the Rasch Model to analyse the results of the students, especially when it is necessary a comparison between two different tests or the comparison between students (Barbaranelli & Natali, 2005; INVALSI, 2013; OECD, 2013). Rasch Model is a simple logistic model that belongs to the Item Response Theory (IRT); it estimates a difficulty parameter for each item of the test and an ability parameter for each student. The IRT models, and between these the Rasch Model, express the probability of choosing the correct answer in an item as a function of the item's difficulty and the ability of the students measured over the entire test.

For each item of the test, the relation between the students' ability and the probability of the correct answer is represented by a curve called *Item Characteristic Curve* (ICC). In a similar way, it is possible to use Rasch parameters to represent also the empirical data and, in particular, we can represent the trend of each possible response to an item as a function of the students' ability. These specific graphs are called *Distractor Plots* and give many information regarding how students respond to an item.

The information gathered using the Rasch model are significant and predictive, in case of new administration of the same test, if the students' sample size is large enough and if the statistical parameters' values' constraints (p-value, Cronbach alpha, among others) are respected. If such a prediction is valuable, this information is used as a blind information about the students' expected performances in respect of an item. It is for these reasons that our statistical tool, which is described in detail in the next paragraph, is based on the Rasch model.

Furthermore, using the Rasch model, we are also able to apply a specific statistical procedure of test equating based on this model. In this research, we will compare the results of students obtained from two different tests that measure the same latent trait and have a common group of items. This procedure has the task of expressing on the same scale the results of the two tests. In particular, we use a *concurrent calibration* procedure, which is considered more precise than a separated calibration (Kolen & Brennan, 1995) and allows to estimate a difficulty parameter for each item and an ability parameter for each student, considering the results of both of the tests at the same time. This kind of procedure used to link two different tests is often applied to compare results of students over time and it is called *anchoring technique*: two tests are administered to different groups of students, but the two tests contain a set of common items used to make an anchorage between the results.

2.4. Statement of the problem

In the literature review concerning the impact of variations of formulation on the performance of students, we may observe that the different methodologies and results actually show that this impact exists and it is relevant, but there is no way for "measuring" this impact, and a fortiori for comparing evidences arising from different studies. Furthermore, in general, it is difficult to analyse this impact in specific subgroups of students, whilst this would be important in the perspective of equity in education (for instance, if one is interested in measuring the impact of linguistic variations in mixed-languages situations).

Indeed, the effect of variations on the students' performances is not easy to investigate because the optimal situation to study is impossible to realize. A student involved in the research should answer indeed to two very similar questions in the same test, and should "forget" to have faced the first question while answering the second one. Inevitably, the first task would influence the second one or the change should be so evident to transform deeply the nature of the question itself. While a qualitative interactive study of the evolution of meanings and strategies in groups or classroom discussions could suggest indirectly interpretations *a posteriori* of some students' difficulties in a particular task, we want to explore directly the questions: how much the variations influence the students' performances during the test? What would the same students do if they have to answer independently the original or to the modified question? Does a variation cause a significantly different behaviour in "real-time", while a student is answering the question during the test?

Hence, our research problem is to design a quantitative methodology which integrates with the existing research approaches in order to address the two points of measuring and differentiating the impact of a variation in the formulation of a problem, on the students' performances.

In particular our research questions are:

- Is it possible to design a quantitative tool based on the Rasch model and anchoring techniques that allows us to measure the impact of a variation on the performance of the population?

A variation may cause significant changes in the answers of the whole population or in the performances of a particular group of students.

- In particular, could such a tool point out from data analysis the relevance of a variation on specific sub-groups of the population?

3. A tool for measuring the impact of variations in the formulation

In this paper we explain how a quantitative approach such as the one proposed by Branchetti and Viale (2015), integrated with anchoring techniques, can highlight interesting evidences, which cannot be observed in other ways. Moreover, this approach give also the opportunity to "measure" the impact of a variation, in relation to a well defined scale. We present in this paragraph our methodology, based on a test linking, equating techniques and the use of the Rasch model.

3.1. Design of the tool

The procedure that we propose and validate is the following one. We start with a core-test (CT) composed by *n* items that measure the same latent trait. The core-test must give a statistically consistent (with respect to the statistical parameters) measurement of students' ability. This core test is then considered as part of a whole test (T) composed by *m* items assessing the same latent trait. Let us denote by $A_1, A_2, \ldots, A_{m\text{-}n}$ the *m-n* items of the test (T) that do not belong to the core-test. Each one of the items $A_1, A_2, \ldots, A_{m\text{-}n}$ is then modified, by performing on it a single, well-individuated variation, getting a new set of items A'_1, $A'_2, \ldots, A'_{m\text{-}n}$. A new test T' composed by the fixed *n* items of the core test CT and the varied *m-n* items is then assembled.

We select a sample of classes, and in each class we administered the original test T to half of the students (randomly chosen) and the varied test T' to the other half. Let us denote with P1 the population to which T is administered, and P2 the population to which T' is administered. P is then the union of P1 and P2.

The first analysis concern only the common items (CT) of the tests T and T'. We apply the Rasch model to the CT on P, on P1 and on P2; we also apply other specific statistical tools, as for example the alpha-Cronbach index, to P, P1 and P2, in order to measure the internal consistency of the CT and its statistical validity. In this way, we have the first information about the comparability of the two sample of students who have answered to the two tests, and on the fact that adding items to the CT does not affect significantly the latent trait measured by the CT itself.

Furthermore, if the test CT has been previously administered to a particular population (for instance, to a statistical sample of a school population), it is possible to get statistical data which enables to establish comparisons between the case studied and some benchmark population. Once we have compared the results of the students on the common part of the tests, we pursue with the analysis of the other items that occur in different forms in the two tests.

The second step is performing the same analysis (Rasch model and other standard statistical tools) on the test T on P1, and T' on P2, in order to measure the internal consistency of T and T' and their statistical validity.

In the third step, we reconsider the results of Rasch analysis considering only the *m-n* common items of the core-test. The ability of each student is then measured using the same items that constitute the CT, independently to the test administered. The probability that a student of a given ability level *p* (measured on the Rasch scale based on CT) answers correctly to the item A_j can be computed, and the relationship between the ability of the student and the probability of the different choices can be visualized through a graph (the above-mentioned distractor plot). We note that, due to the characteristics of the Rasch model, this relationship includes also the information on how the choices of the students with a given ability on the test T are distributed, when answering to the item A_j (correct answer, missing, choice of a particular distractor....).

At this stage, we are interested in studying the different impact of the two formulations of an item (A_1 vs A'_1, A_2 vs A'_2, and so on) on the two groups P1 and P2 of students and, for this purpose, we use the Rasch ability of the students measured on the core-test. In particular, we represent how the students answered to the modified items as function of their ability; in other words, we make distractor plots of the two versions of an item (A_i and A'_i), plotting the empirical data as functions of the ability parameter calculated on common part of the test (CT). In this way, it is possible to observe the trend of each possible answer in the two versions of an item and compare them analysing the different behaviour of the students. It is also possible to observe if this variation has a particular impact on a specific ability level and, using deeper analysis always based on distractor plots, it is possible to notice if the variation has a greater impact on a subgroup of the population (for example male or female).

Moreover, we analyse the impact of the variations also using test equating to confirm the results obtained using our procedure. We decide to use the concurrent calibration applied to the results obtained by the new administration of the two tests T and T' and this allows us to express all the parameters estimated (student's ability and item's difficulty) on the same scale. In particular, we can compare the difficulty parameters of the items A_1, A_2, …, A_m, respectively with the parameters of the items A'_1, A'_2, …, A'_m and we can notice if these differences are statistically significant.

3.2. Output of the tool

For each pair of varied items (A_1 vs A'_1, A_2 vs A'_2, and so on) the procedure will give:

1) An index of difficulty for each version, placed on a common scale, anchored by the CT;

2) A distractor plot for each version, where the x-axis measure the same ability.

Roughly speaking, our tool will measure how the two variations differ as difficulty, and how the variation of the formulation impacts on the performances, as a function of the ability of the student.

3.3. Utilization criteria

This approach needs specific controls and checks after the testing and only after this checking one can use a quantitative tool like our, based on statistical indicators.

UC1) First, it is necessary to verify the internal consistency of the varied tests: the Cronbach alpha of CT on P, P1 and P2 must have acceptable and not statistically different values. The same must happen for the Cronbach alpha of T on P1 and of T' in P2.

UC2) Second, it is necessary to verify that the three test (CT, T and T') are related to the same latent trait- in our case, the mathematical ability. In order to verify this, it is needed a comparison of the results of the two tests in terms of distribution of students and distribution of the items in relation to Rasch parameters. This is of course very delicate with a purely statistical approach; in most research situations, it is helpful and easier to apply a qualitative analysis observing Wright maps. We first compare the Wright maps of both of the tests obtained applying the Rasch model exclusively to the core test, and then we do the same thing comparing maps of the whole tests. If the variations have not influenced too much the core test and the latent trait measured, the distribution of the core test items must be similar in the two maps.

UC3) Third, at items' level, we must highlight if each variation has worked well and, at the same time, it is important to see if the varied items maintained good psychometrical features in relation to the core test and in relation to the whole test. To confirm that, we use the Rasch model and specific indexes of the classical test theory to analyse fit and discrimination of each item which must fulfil the standard validation criteria.

4. Validation plan

Our validation plan consists in

- Starting with a test T for which there are already available solid data from a large scale assessment;
- Individuating among the items of T a core test CT, which can be assumed as a good test for measuring the mathematical ability;
- Considering variation of items testing contents on which there are important threads of research in mathematics education;
- Testing the tool on a large population, comparable as characteristics to the population of the large scale assessment;
- Verifying the utilization criteria and the comparability of the results with the results of the large-scale assessment;
- Verifying the statistical coherence and the didactic relevance of the data and the related evidences;
- Comparing these quantitative evidences with the results of the previous researches.

4.1. Design and administration of the Tests

4.1.1. The starting test

Our test T consisted of the INVALSI test administered to 590.728 Italian students of grade 6, during May 2013. We decide to use an INVALSI test as basis for our validation plan because in this way we start from a test was previously administered to a large population of student and analysed from the INVALSI statistical team. Thereby we are sure to start from a test with good values concerning statistical reliability and coherence and that gives a statistically consistent measure of the latent trait that we can identify with mathematical competence. Furthermore, we have also the possibility to compare our results with the results of the national survey, understand if our sample is representative of the national INVALSI sample and therefore of the whole Italian grade 6 students population. The INVALSI test was composed of m=48 items and the statistical analysis were performed on a representative sample of 27.504 students. 1.528 of them were a representative sample of the students from the Italian region of Emilia-Romagna.

4.1.2. The variations

We chose n=7 items and we changed them along different directions, as described in the theoretical framework (by operating on the lexicon, the syntax, the use of figures, the registers of representation and so on). Five items are in the domain of numbers (arithmetic and estimations) and two are with geometrical content (Euclidean and analytical geometry).

Two arithmetic items are without context. In the first case, a multiplication between natural numbers without result is reported and students are asked to choose the quantity of digits of the result of the operation among 4 option. The item is a cloze open question and the variation concerned the type of item: the varied item is a *multiple-choice* one. This kind of variation has been studied by many authors.

In the second item, students are asked to choose among 4 alternatives in *multiple-choice* item, the right estimation of the result of a multiplication between rational numbers in the decimal representation. The variation concerned the kind of numbers and their magnitude: the numbers of the varied item were obtained multiplying the numbers by 100 and making them natural.

Two other items in the domain of numbers are arithmetic *multiple-choice word problems* in which data were natural numbers and the question refers to a context described verbally. In both the cases, the variation acted on the syntax of the sentences. The last variation concerned the editing of the item.

All the original versions of the geometrical items are *word problems* with a mixed text i.e. a text characterized by an integration between data in the verbal and in the graphic form. Two aspects of geometry are explored: a) representation of points by means of coordinates in a Cartesian plan; b) measure of lengths and areas.

In the first case, in the verbal text a path from a point (whose coordinated are reported in the text) to another is described, referring to a graphic representation. The students are asked to write the coordinates of the second point. In the varied version, the graphics representation is the same, but the coordinates of the final one are given and the students are asked to identify the coordinates of the starting point. Hence we have here a substantial variation of the relationship between the stimulus and the task.

The second item a problem concerning area and perimeter of composed polygons with a question in the text and the situation represented graphically. The variation concerns the level of representation of the situation, that is transformed in a pure verbal one by removing the graphic representation.

4.1.3. The population

We administered the new test T' and the original test T to 777 students of the same age from the same region (Emilia Romagna); of course they had not participated to the test T in 2013. In particular, in each of the 40 classes involved in the trial, half of the class randomly chosen responded to the new test T' and the rest of the students responded to the original test T. The validation plan is based on the analysis of the 777 tests, including 380 original test T and 397 varied test T'.

4.2. Data analysis- UC and comparison with the INVALSI large-scale results

In this section, we verify that our trial satisfies the utilization criteria, and at the same time we verify that the quantitative results that we obtained are coherent (along the three lines of the UC) with the results of the large-scale assessment from which T is derived- hence giving more strengthness to our validation.

4.2.1. UC1 and comparison with the large-scale INVALSI assessment

First, we compared the global results of our tests and the results of the national and regional original INVALSI test. For this purpose, we used results of Rasch analysis on the whole test and on the Core Test, also observing specific indexes belonging to Classical Test Theory, such as the alpha-Cronbach that gives us the internal consistency of the tests.

First of all, we verify the internal consistency of the test in our trial using the alpha-Cronbach index.

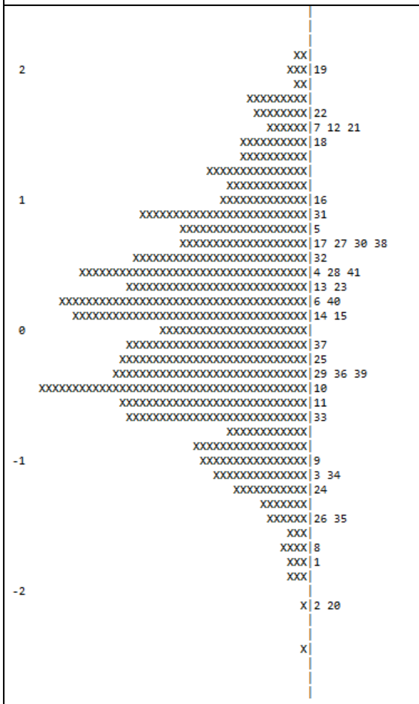
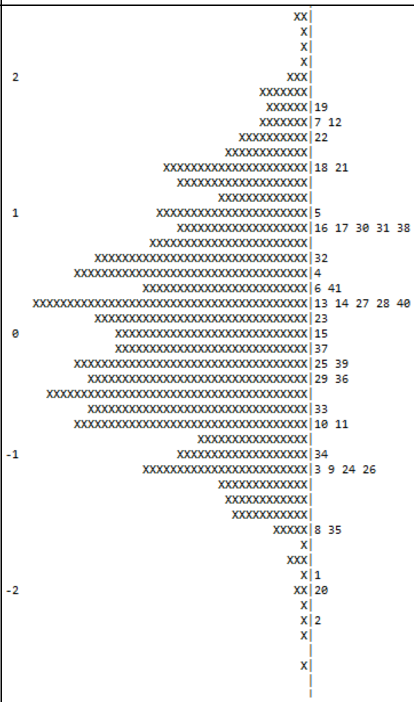
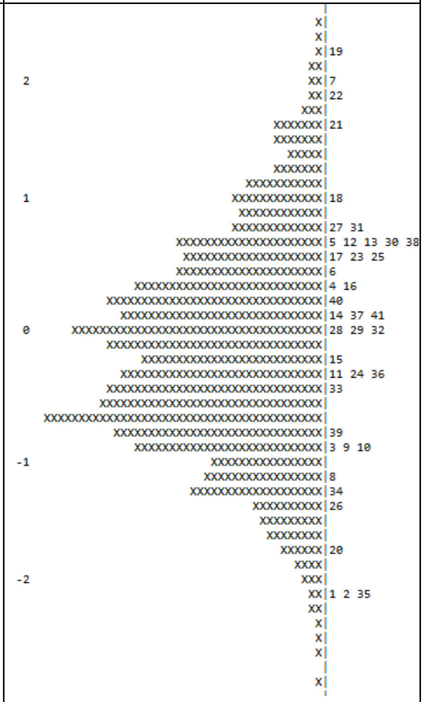
	Sample size	**α-Cronbach of the whole test**	**α-Cronbach of the core test (CT)**
National INVALSI survey	27.504 students	0.86	0.85
Regional restriction of the national INVALSI survey	1.528 students	0.86	0.84
Populations P, P1 and P2	• Population P: 777 students □ P1 (original test): 380 students □ P2 (varied test): 397 students	○ original test: 0.86 ○ varied test: 0.87	0.85 (considering the two tests together)

Table 1: Comparison between the internal consistency of the test administered to the populations P, P1, P2 and the same test administered in National INVALSI survey.

As reported in Table 1, alpha-Cronbach values are acceptable in each test, and the internal consistency is verified both considering the whole test and analysing only the core test.

4.2.2. UC2 and comparison with the large-scale INVALSI assessment

Moreover, as explained in the validation criteria, we observed that the distribution of the 41 items of the CT as function of the Rasch difficulty parameters is similar in the two tests of our trial and, even if there are minor local differences, this distribution is also globally similar to that of the national and regional survey.

Wright map of the Core Test in the Original version of the test	Wright map of the Core Test in the Varied version of the test	Wright map of the Core Test in the National version of the test
		

Item names key:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
D1_a	D1_b	D1_c1	D1_c2	D2_a	D2_b	D2_c	D4	D5	D6_a	D6_b	D6_c	D7_a	D7_b	D8_a	D9_a	D9_b	D9_c	D10_a	D10_b	D10_c
22	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	
D11	D12	D13	D15	D17_a	D17_b	D19	D20_a	D20_b	D21_a	D21_b	D23	D24	D25_a	D25_b	D26_a	D26_b	D28	D29	D30	

Table 2: Wright map of the CT relative to our trial and INVALSI National survey; distribution of items (right side) as function of Rasch difficulty parameters and distribution of students (left side) as function of Rasch ability parameters.

Also the comparison between the Wright maps obtained analysing tests (original and varied) as a whole, highlights that the distributions of students and items in relation to the Rasch parameter is similar and then the variations did not influence the tests taken as a whole.

4.2.3. UC3 and comparison with the statistical indexes of the original items

At last, we consider the statistical parameters of each one of the 7 items varied for the trial and we verify that their statistical parameters are still acceptable.

item	Population P1 – original test		Population P2 – varied test		National INVALSI survey	
	weighted	discrimination	weighted	discrimination	weighted	discrimination
D3	1.04	0.33	1.01	0.35	0.99	0.40
D8b	1.04	0.30	1.01	0.39	1.07	0.31
D14	1.02	0.35	1.04	0.35	1.06	0.29
D16	1.05	0.31	1.12	0.25	1.09	0.25
D18	0.87	0.56	0.80	0.60	0.87	0.52
D22	1.18	0.15	1.15	0.22	1.12	0.25
D27	0.95	0.40	1.10	0.28	1.00	0.38

Table 3: Statistical parameters (weighted and discrimination) of the varied item.

In the previous table we observe that items' parameters obtained in our trial are similar, for population P1, to those calculated on the original INVALSI test. Generally, the parameters are acceptable also in our trial and, in cases of minor anomalies (such as the too low weighted value for D18), they were already present in the National Survey.

4.3. Data analysis- Outputs and qualitative analysis

We present here the outputs and possible suggestions for further qualitative analyses, comparing our data with the research framework and the results coming from previous researches, for three of the variations that we implemented. We recall that the outputs of our tool are the indexes of difficulty of the variations, anchored by means of a concurrent calibration technique, and the distractor plots relating the probability of answering correctly (to one or to the other variation) in function of the ability, measured on the same scale.

The complete result of this calibration in shown in Table 4, where the second column reports the calibrated index of difficulty, and the third one the standard error. This gives a measurement of the impact of the variation (given by the difference between the index of item Dn_o and Dn_v) which in most cases is statistically significant.

Item	Delta	St. err.	Item	Delta	St. err.	Item	Delta	St. err.	Item	Delta	St. err.
D1-a	-2,06	0,11	D7-a	0,12	0,08	D14_o	0,52	0,11	D22_o	0,10	0,11
D1-b	-2,40	0,12	D7-b	0,03	0,08	D14_v	0,88	0,12	D22_v	-0,51	0,11
D1-c1	-1,25	0,09	D8-a	-0,10	0,08	D15	-0,38	0,08	D23	-0,76	0,08
D1-c2	0,38	0,08	D8-b_o	0,47	0,11	D16_o	0,31	0,11	D24	-1,18	0,09
D2-a	0,77	0,08	D8-b_v	0,77	0,12	D16_v	0,80	0,12	D25-a	-1,63	0,09
D2-b	0,19	0,08	D9-a	0,82	0,08	D17-a	-1,47	0,09	D25-b	-0,47	0,08
D2-c	1,56	0,10	D9-b	0,63	0,08	D17-b	0,38	0,08	D26-a	-0,25	0,08
D3_o	0,03	0,11	D9-c	1,32	0,09	D18_o	0,68	0,12	D26-b	0,59	0,08
D3_v	-0,57	0,11	D10-a	1,84	0,1	D18_v	1,09	0,12	D27_o	0,55	0,12
D4	-1,79	0,1	D10-b	-2,26	0,11	D19	0,24	0,08	D27_v	0,37	0,11
D5	-1,24	0,09	D10-c	1,33	0,09	D20-a	-0,50	0,08	D28	-0,45	0,08
D6-a	-0,74	0,08	D11	1,49	0,09	D20-b	0,65	0,08	D29	0,06	0,08
D6-b	-0,82	0,08	D12	0,09	0,08	D21-a	0,75	0,08	D30	0,3	0,08
D6-c	1,53	0,10	D13	-1,32	0,09	D21-b	0,50	0,08			

Table 4: Test equating results, comparison between the difficulty parameters of the two forms (in respect to the standard error).

4.3.1. Examples of analysis 1: a number size variation (item D22)

In this case the variation is numerical: numbers magnitude and kind of numbers. The item in the original test asks the student to estimate the result of a product of two decimal numbers. This is a well known problem in mathematics education (see, f.i., De Corte et al., 1985)

D22. Which of the following integer numbers is closer to the result of this multiplication?

4,82 x 9,95

- **A. 36**
- **B. 42**
- **C. 48**
- **D. 50**

Figure 1: D22 – item in the original form (test T).

The *question intent* of the varied item is the same but the numbers are integer and their order of magnitude is different. It is important to notice that all the responses in the varied form are analogous to the ones in the original item: simply, each factor is multiplied by 100 and the option are multiplied by 100 000. The *type* and the *size* of the number involved is completely changed.

D22. Which of the following integer numbers is closer to the result of this multiplication?

482 x 995

A. 360.000

B. 420.000

C. 480.000

D. 500.000

Figure 2: D22 – item in the varied form (test T').

We decided to analyse this kind of variation in order to explore the potentiality of our methodology. Our starting hypothesis was that this variation would change drastically the strategies of the students, even in principle a parallel, simply scaled strategy would be possible. This is a good validation case for exploring the functioning of our tool. In particular, this case may allow to verify if our research tool is useful for observing more in depth the results differentiating between students with different Rasch scores, different gender, foreign or mother tongue. Thus, we have the possibility to see which categories of students were more influenced by the variation and to conjecture how and why they should have been influenced by the variation.

The question intent concerns the estimate of operations results and the students were explicitly asked to say which was the number closer to the product result. Nevertheless, even if the kind of numbers involved shouldn't change the nature of the problem, we may expect that the students are led by the classroom habits to approximate or to compute with different procedures the result.

In this case an *a priori* analysis lead us to conjecture that changing the kind of numbers in a multiplication would have changed the students' strategies. The change is usually ascribed to differences in the teaching practices in which operations with decimal and natural numbers are involved at school, but if it is the only reason why the students are driven to choose one distractor or the correct answer, the change should influence all the students in the same way. With our analysis, we can explore instead possible differences, that may be due to a different stability and generality of the students' knowledge. Our analysis can thus provide information concerning the features of the students' knowledge about methods for estimating; the students may in fact be lead to round to multiples of 10 with natural numbers but may have never used this strategy with decimal numbers. As a result, they should have a partial knowledge of methods for estimate, concerning only some kind of numbers. On the contrary, we might observe that the students, once overpassed an intermediate Rasch score, are not conveyed by this kind of change since their knowledge is more complete.

Since the item was formulated as a *multiple-choice* question, our *a priori* analysis can only concern the students' choices in a shortlist of four possibilities and we can try to conjecture why a student could choose a particular possibility. In the original version and in the modified one we proposed the same digits but a difference in the magnitude (the numbers in the modified version were 2 orders bigger), both in the text and in the answers.

A common strategy would lead to the same answer, with the only difference of the magnitude. A significant difference in the item score and in the percentage of students' choices of a distractor in respect of another are thus signals of different strategies.

A priori, the possible strategies are:

1) to round both number to the closer integers (or hundreds);
2) to consider only the first digit;
3) to approximated both numbers by excess or by defect;
4) mixed strategies
5) …...

The data analysis, performed as we have explained before, demonstrates that with this variation the item becomes much easier. Indeed the percentage of correct answers in the original item is 46% and in the varied one increases to 59%.

	% original item	% varied item
A	15%	11%
B	21%	13%
C	**46%**	**59%**
D	11%	12%
MISSING	6%	6%

Table 5: Answers percentages for item D22 (original form and varied form).
Correct answer: C.

As we can see in the table before, the variation has mostly influenced the choice of distractor B. Whilst the distractors A and D increase or decrease only of some percentage points, answer B loses about 8% in the varied item. The variation does not influence the percentage of students who don't answer to the item and this may mean that, despite the different difficulty of the two versions, almost all the students are confident enough to try to answer.

The *concurrent calibration technique* applied to both of the test simultaneously allows us to estimate the difficulty parameters of all the items (including the two version of the 7 items modified) and to consider them on the same scale. The comparison between difficulty parameters estimated in this way gives us an additional proof that the variation in this case had made the item easier. In fact, after using the anchoring procedure, the difficulty parameter of the original item is 0,10 and the difficulty of the varied item is -0.51, both with a standard error of 0.11, hence the difference is statistically significant.

At this point of our analysis, it is interesting to use Rasch analysis and, in particular, distractor plots to investigate if the differences identified before are evenly distributed on all the students or if these changes have influenced in particular students with a certain ability level. Using the Rasch Model to analyse the core test (CT) composed by 41 common item, we estimate an ability parameter for each of the 777 students. The distractor plots below represent the empirical data of the two versions of the item D22 as function of the ability of the students evaluated using the core test.

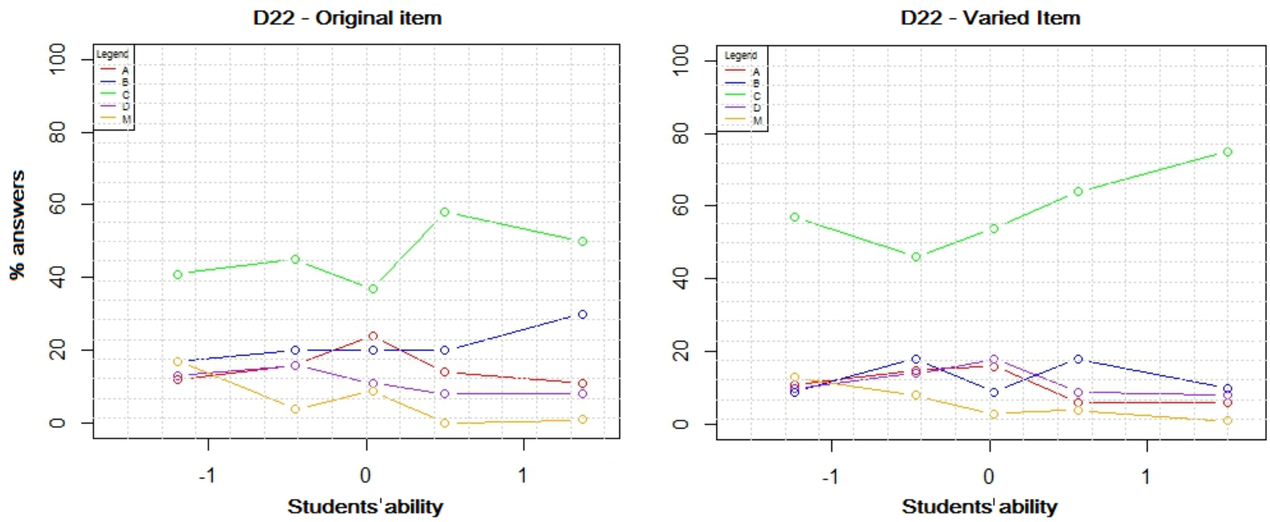


Figure 3: D22 – Distractor plots: original item and varied item.

The behaviour of the correct answer and the incorrect ones is different in the two items. In particular, the trend of the correct answer is more regularly increasing in the varied item. In both of the two forms, we observe that this question is not very discriminant. In other words this item doesn't distinguish well between students with high ability levels and students with lower ones. This information is also confirmed by the analysis of the whole tests: the discrimination parameter of this item is 0,15 for what concerns the original test and 0,22 in the varied test (hence slightly better). Also in the national survey this item has a discrimination above the threshold (0,20) but it is interesting to notice that the variation improved the statistical properties of the item. Indeed, the discrimination is higher in the varied form of the item and also comparing the weighted index of the two items, the varied one gives better results.

In this paper we are interested in analysing the comparison between the two items and, in particular, many useful details raise from the comparison between distractor trends. For example, we can focus on the trend of the distractor B (blue) which is the one with the most relevant variation in the percentage seen before: comparing the two plots we can see that distractor B is often chosen by students with high ability level in the original item but the variation makes it much less attractive for those students.

Furthermore, the analysis of this item is very interesting also because the variation has a very different impact on males and females. In the table below the percentage of each answer for male and female and for both the two versions of the item are presented. In the original item with the decimal numbers, correct answers are the 44% for males and 50% for females. The variation has a huge impact on males' performances and, answering to the varied item, 18% more of the students choose the correct answer. Instead, if we observe female percentages, we can see that the correct answer takes only the 6% more in the varied answer.

	MALE		FEMALE	
	% original item	% varied item	% original item	% varied item
A	16%	10%	12%	11%
B	23%	12%	20%	13%
C	**44%**	**62%**	**50%**	**56%**
D	11%	11%	10%	13%
MISSING	5%	4%	8%	7%

Table 6: Answers percentages for item D22 distinguishing male and female.
Correct answer is C.

This may mean that male and female, in general, apply different strategies to solve this item and the strategies used by female are almost the same even if the numbers are changed. As a natural issue, we can investigate how this gender gap difference is related to classroom practices (see f.i. Bolondi, Cascella and Giberti, 2017).

4.3.2. Example of analysis 2: a linguistic variation (items D16 and D27)

In both the cases the items are arithmetic *multiple-choices word problem* and the variation concern a complication of the periods syntax. In the first item, the variation concerns the syntactic level and is a complication of the period's structure. A linguistic analysis of the original text showed that it is composed by two periods and that, while the first has just a principal sentence, the second one has a principal sentence including the question and a relative subordinate sentence. In the modified item, the periods were joint. The principal contains the question and there are subordinate sentences depending on it that are a relative and a conditional one, that depends on a declarative subordinate (see figure 6). We reproduce here, for obvious reasons, the Italian text.

D16. Una scatola di cioccolatini contiene 15 cioccolatini al latte e 25 cioccolatini fondenti. Con 100 cioccolatini al latte e 180 fondenti, qual è il numero massimo di scatole con la stessa composizione della precedente che si possono riempire?

A. ☐ 5

B. ☐ 6

C. ☐ 7

D. ☐ 8

Figure 4: D16 – item in the original form (test T).

D16. Qual è il numero massimo di scatole di cioccolatini che si possono riempire con 100 cioccolatini al latte e 180 fondenti, sapendo che ogni scatola deve contenere 15 cioccolatini al latte e 25 fondenti?

A. ☐ 5

B. ☐ 6

C. ☐ 7

D. ☐ 8

Figure 5: D16 – item in the varied form (test T').

1°periodo principale

Una scatola di cioccolatini contiene 15 cioccolatini al latte e 25 cioccolatini fondenti. // Con 100

reggente

cioccolatini al latte e 180 fondenti, / qual è il numero massimo di scatole con la stessa

sub.relativa

composizione della precedente / che si possono riempire?

reggente sub. relativa

Qual è il numero massimo di scatole di cioccolatini / che si possono riempire con

sub. condizionale

100 cioccolatini al latte e 180 fondenti, / sapendo / che ogni scatola deve

sub. dichiarativa

contenere 15 cioccolatini al latte e 25 fondenti?

Figure 6. Syntactical analysis

Also the second item was modified at the syntactic level, but in a different way. The variation concerns again the syntactic level and is a reduction of the number of sentences and an inversion of order between the question and the data. This variation is thus syntactic but impact also the logic of resolution, since the data are once before (original), once after (varied) the question.

D27. **Nello zaino di Chiara ci sono il libro di scienze, che pesa mezzo chilo, il libro di matematica, che pesa 980 g, e due quaderni uguali. Libri e quaderni pesano in tutto due chilogrammi. Quanto pesa ciascun quaderno?**

A. ☐ 150 g

B. ☐ 260 g

C. ☐ 510 g

D. ☐ 520 g

Figure 7: D27 – item in the original form (test T).

D27. **Nello zaino di Chiara ci sono il libro di scienze, il libro di matematica e due quaderni uguali. Quanto pesa ciascun quaderno, sapendo che il libro di scienze pesa mezzo chilo, il libro di matematica pesa 980 g, e che libri e quaderni pesano in tutto due chilogrammi?**

A. ☐ 150 g

B. ☐ 260 g

C. ☐ 510 g

D. ☐ 520 g

Figure 8: D27 – item in the varied form (test T').

The original text was composed by three periods. The first was composed by a principal sentence and three relative subordinates at the same level. The other two are composed by principal sentences, and the second contains the question. The varied item has one period less (just two) and the last is a complex period with a principal, that contains the question, and three relatives, containing the data.

Looking at the global parameters of the items, some general trends emerged. The parameter of difficulty of the items changed with the variation of the formulations in both the cases, as much as the percentage of right answers. In the first case, the value of the parameter of difficulty passed from 0.37 to 0.82 and the percentage of right answers decreased from 41,58% to 33,00%. In the second case, we observed an opposite trend: even if the variation was expected to make the sentence more difficult, the value of the parameter of difficulty passed from 0,59 to 0,44 and the percentage of right answers increased, from 36,84% to 40,55%. In Table 7 and 8 the global results for every option in the two cases are reported.

Item Delta(s): 0.37

Label	Score	Count	% of tot
A	0.00	49	12.89
B	1.00	158	41.58
C	0.00	108	28.42
D	0.00	54	14.21
M	0.00	11	2.89

Item Delta(s): 0.82

Label	Score	Count	% of tot
A	0.00	43	10.83
B	1.00	131	33.00
C	0.00	132	33.25
D	0.00	62	15.62
M	0.00	29	7.30

Table 7: Parameter of difficulty and percentages of the original and varied versions of D16

Item Delta(s): 0.59

Label	Score	Count	% of tot
A	0.00	87	22.89
B	1.00	140	36.84
C	0.00	48	12.63
D	0.00	49	12.89
M	0.00	56	14.74

Item Delta(s): 0.44

Label	Score	Count	% of tot
A	0.00	45	11.34
B	1.00	161	40.55
C	0.00	56	14.11
D	0.00	73	18.39
M	0.00	62	15.62

Table 8: Parameter of difficulty and percentages of the original and varied versions of D27

We propose here a comparison between the distractor plots obtained categorizing the students by levels of competences (computed on the CT), in order to show what our method made emerge and to interpret this anomaly and to deepen the data analysis.

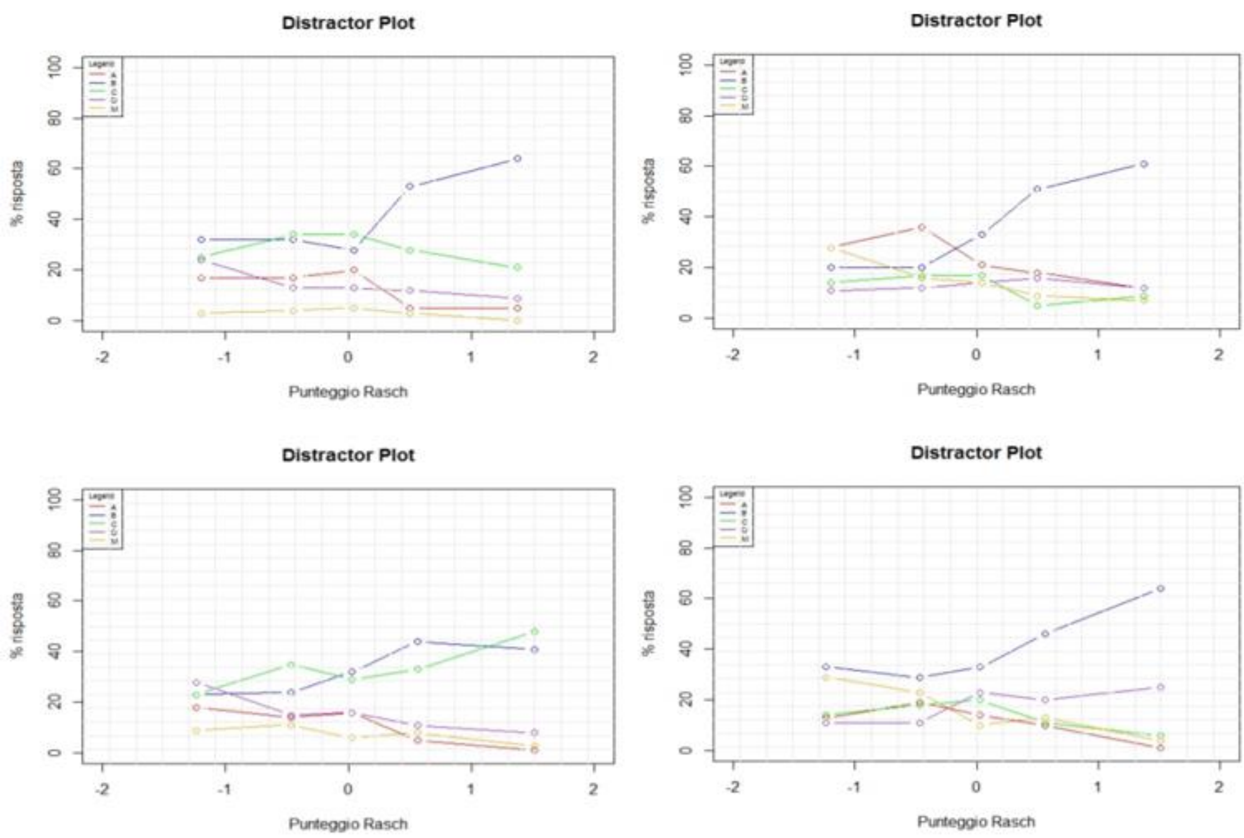


Figure 9 Distractor plots of the D16 (left) and the D27 (right), original (up) and varied (down) items

The correct answer is represented by a blue line. The yellow ones represent the missing answers, while the other represent the other options.

Comparing the graphs of the original versions, we observed first that the graphs of the original versions are very different for students with the lowest performances, but with significant resemblances for students with medium-high performances. The graphs corresponding to the varied items show that syntactic variations had in general different impacts on students with the same level of competence in the test. Furthermore, the different variations affected in the same category of students in different ways. Let’s analyse some emergent features of the graphs.

The students with the highest performances had the 60% of right answers to both the original items, but had significantly different performances in the varied items (around 40% in the first case, 60% in the second). In particular, in the first case there is an evident difference between the original and the varied items, while in the second case the percentage is almost the same in the original and varied version for what concerns students with high performances, so the variation didn't influence their performances. On the contrary in the case of students with an average competence no significant effects have been found. The students with lower performances in the whole process had a more complex behaviour. In the first question, the effect of the variation was to reduce the percentage of right answers and to make the answers distribution over the 4 options almost casual, since the options have very similar percentages. In the second case, the percentage of right answers increases a lot in the varied item, so the effect is the opposite of the one we observed with students with the best global performances.

Since the statistical parameters are good, we consider such an anomaly interesting and maybe due to relevant factors that could be interesting from a didactical point of view. This result suggests and encourages us to investigate better this phenomenon, in order to clarify the relationships between formulation, the mathematical competence, linguistic skills and assessment. Further investigations should be carried out with qualitative methods, like case studies, interviews and focus groups analysis with students, analysis of textbooks and teaching practices concerning arithmetic *word problems*.

Conclusions

The analysis performed in sections 4.2 and 4.3 allows us to state that our tool gave back data on the impact of different kinds of variations in the formulation of a problem that are consistent with the main findings reported in the literature that we chose as a reference and statistically significant. The hypothesized impact have not only been checked, but also measured and analysed in depth on specific sub-groups of students. The data analysis allowed to highlight phenomena already known in the literature, and to point out new aspects of these phenomena, like their different relevance for different subgroups of students (male and female, students with high or low performances in the test). Indeed, the analysis of answers distributions, provided interesting and provoking questions concerning the reasons why variations in the formulation of items should affect more some categories of students than others. Another important issue concerns the question intent: since the distribution of the answers may be very different in the varied case, it seems that students are demanded to activate different strategies and resources to answer the question.

This hypotheses could become a starting point for a further and more detailed in which this quantitative method, in order to give useful information, is integrated with qualitative analyses.

The limitations of the method concern its quantitative nature. Such a method needs integration to allow interpretation of the data. Moreover, the statistical requests for the validity of the tool are quite important: the tool requires a large number of students in order to be effective, especially if one wants to use it for studying subgroups.

References

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219-234.

Bagni, G. T., & D'Amore, B. (2005). Epistemologia, sociologia, semiotica: la prospettiva socio-culturale. *La matematica e la sua didattica*, *1*, 73-89.

Barbaranelli, C., & Natali, E. (2005). *I test psicologici: teorie e modelli psicometrici.* Roma: Carrocci Editore.

Branchetti, L., & Viale, M. (2015). Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico. In Ostinelli M. (2015). *La didattica dell'italiano. Problemi e prospettive.* Proceedings of the conference "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), 405-438.

D'Amore, B. (2000). Lingua, matematica e didattica. *La matematica e la sua didattica*, *1*, 28-47.

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

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* 6:348.

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), 460.

Duval, R. (1991), Interaction des différents niveaux de représentation dans la compréhension de textes. *Annales de Didactique et de sciences cognitives*, pp. 136-193. Strasbourg : Edition de l'IREM

INVALSI (2013), Rilevazione nazionale degli apprendimenti 2012-2013. Rapporto tecnico. http://www.invalsi.it/invalsi/doc_evidenza/2013/034_Rapporto_Prove_INVALSI_2015.pdf

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come, *Educational Researcher, 33*(7), 14-26.

Kolen, M. J., & Brennan, R. L. (2013). *Test equating: Methods and Practices*. Springer Science & Business Media.

Laborde, C. (1995), Occorre imparare a leggere e scrivere in matematica?, *La matematica e la sua didattica,* vol. 2, pp. 121-135.

Lepik, M. (1990). Algebraic word problems: Role of linguistic and structural variables. *Educational Studies in Mathematics*, *21*(1), 83-90.

Nesher, P. (1982). Levels of description in the analysis of addition and subtraction word problems. In J. Moser (Ed.), *Addition and subtraction: A cognitive perspective.* Hillsdale, NJ: Lawrence Erlbaum.Associates

OECD (2013), Pisa 2012 Assessment and Analitical Frame- work: Mathematics, Reading, Science, Problem Solbing and Financial Literacy, OECD Publishing. http://dx.doi.org/10.1787/9789264190511-en

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmarks Paedagogiske Institut.

Spanos, G., Rhodes, N. C., & Dale, T. C. (1988). of Mathematical problem solving: Insights and applications. *Linguistic and cultural influences on learning mathematics*, *221*.

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), 43-56.

Thevenot C., Oakhill J. (2005). The strategic use of alternative representation in arithmetic word problem solving. *Q. J. Exp. Psychol.* 58 1311–1323.

Vicente, S., Orrantia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British journal of educational psychology*, *77*(4), 829-848.

5.3 Progetto *VARIAZIONI 2*[29]

Realizzando il primo progetto è stato possibile mettere a punto la nuova metodologia di indagine per studiare l'impatto di una variazione nella formulazione di un task matematico. Questo tema di ricerca interessa non solo la ricerca in didattica della matematica ma è molto importante anche per chi crea i test e per chi ne analizza i risultati.

Il progetto *Variazioni 2* è nato in collaborazione con l'INVALSI e in particolare con una ricercatrice dell'Istituto, Clelia Cascella, che aveva già portato avanti un lavoro su questo tema, focalizzandosi però principalmente sulle caratteristiche psicometriche di un item e dello stesso item a cui sono state applicate variazioni.

Il livello scolastico a cui è indirizzato questo progetto è il terzo anno della scuola secondaria di primo grado; si tratta infatti di un livello particolarmente delicato e interessante, in quanto la prova INVALSI rientra all'interno delle prove dell'Esame di Stato. Le variazioni e i fascicoli sono stati messi a punto nel corso dei primi mesi del 2017 anche grazie a un primo *pilot study* che ha coinvolto 5 classi di una stessa scuola e che ha permesso di evidenziare refusi e criticità nei fascicoli. La fase di somministrazione dei fascicoli è stata gestita dall'Istituto INVALSI e si è svolta nei mesi di aprile, maggio e giugno del 2017; l'inserimento dei dati risulta quindi ancora parziale e i risultati definitivi si avranno solamente nel corso dei prossimi mesi. Essendo il progetto in fase di conclusione, in questa tesi sarà spiegata la struttura della ricerca, gli obiettivi che intende perseguire e i primi risultati, ricavati dall'analisi di circa il 10% dei fascicoli.

Obiettivi della ricerca

Se il primo progetto aveva come scopo principale quello di validare la metodologia di ricerca, il progetto *Variazioni 2* ha permesso di far confluire necessità di ricerca legate sia alla didattica sia alla psicometria ed è quindi caratterizzato da molteplici finalità:

1) Studiare come una determinata variazione della formulazione di un quesito di matematica possa incidere sulle risposte degli studenti e sui processi cognitivi messi in

[29] Il progetto è frutto della collaborazione con Giorgio Bolondi (Università di Bolzano) e Clelia Cascella (INVALSI). Alla prima fase del progetto ha collaborato Chiara Lotti (Università di Trento) mettendo a punto la griglia di correzione delle prove.

gioco. 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.

2) Capire se una determinata variazione potesse avere una maggiore influenza su uno specifico sottogruppo della popolazione. In particolare, 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 domanda.
3) Approfondire l'analisi dell'impatto di una variazione linguistica di un quesito di matematica.
4) Studiare l'impatto di una variazione sulla funzionalità di un item da un punto di vista psicometrico. A tal fine, sono stati scelti quesiti INVALSI che, nelle rilevazioni nazionali, avevano mostrato problemi relativi al fit con il modello e sono stati modificati per individuare i fattori alla base dello scarso adattamento dei dati al modello.
5) Sperimentare nuovi formati di risposta e riflettere sul formato di risposta più opportuno in funzione dello scopo e della natura della domanda.

Sono stati quindi intrecciati gli interessi legati alla didattica e quelli legati all'analisi dei quesiti relativamente al loro funzionamento psicometrico. Inoltre, per studiare approfonditamente come la variazione di un quesito modifichi le risposte degli studenti e, di conseguenza, il funzionamento del quesito stesso, si è scelto in questo caso di lavorare anche con più variazioni dello stesso quesito.

Costruzione dei fascicoli

Per la costruzione del *core-test* si è deciso, anche in questo caso, di riferirsi principalmente a una prova INVALSI già somministrata negli anni passati, la Prova Nazionale (livello 8) del 2010/2011. A partire da questa prova sono stati selezionati i quesiti che risultavano più significativi da un punto di vista misuratorio e maggiormente rappresentativi della prova dal punto di vista degli ambiti di contenuto. Le 8 domande così selezionate costituiscono la parte di *core-test* chiamata *àncora interna*, ovvero quesiti (invariati) che sono stati inseriti all'interno dei fascicoli alternandoli alle domande variate. Le procedure di ancoraggio in questo caso si sono basate su strumenti statistici più sofisticati che hanno permesso di ridurre il numero di item del *core-test* e aumentare il numero di domande con variazioni. Inoltre la procedura di *test equating* si è basata su doppio ancoraggio. Fanno parte del *core-test* infatti anche 6 quesiti invariati posti all'inizio di ogni fascicolo; questi ulteriori quesiti costituiscono l'*àncora esterna*, sono tratti da altre prove dello stesso livello scolastico e mostrano ottime proprietà misuratorie.

Il *core-test* risulta così costituito complessivamente da 14 quesiti (15 item) che mostrano tutti ottime caratteristiche misuratorie (fit ottimale e alta discriminatività); inoltre risultano rappresentativi dell'intera prova in termini di difficoltà degli item e ambiti di contenuto, in modo da permettere una stima dell'abilità degli studenti coerente (Kolen & Brennan, 2004).

Per poter somministrare più variazioni di una stessa domanda si è scelto di preparare quattro fascicoli distinti. Le domande individuate per rientrare nei fascicoli con variazioni sono state tratte da passate rilevazioni INVALSI (principalmente di livello 8), ricavate da ricerche effettuate nel campo della didattica oppure inventate ex novo per indagare un determinato problema. Per ciascuna di queste è stata progettata una variazione oppure tre variazioni distinte, in questo modo le domande non appartenenti al *core-test* si presentano:

- In due forme (forma originale e forma variata) ciascuna inserita in due dei quattro fascicoli
- In quattro forme (forma originale, forma variata1, forma variata2, forma variata3) ciascuna inserita in uno dei quattro fascicoli.

Le diverse forme sono state distribuite all'interno dei fascicoli in base alla presunta difficoltà dell'item, in modo da creare fascicoli equilibrati nonostante la presenza delle domande variate.

Anche da un punto di vista grafico, nei quattro fascicoli è stata mantenuta la posizione degli item nelle pagine e la struttura delle pagine stesse, per evitare ulteriori cambiamenti che potessero confondere i risultati relativi alle variazioni.

Nella tabella sotto è riepilogata la struttura dei quattro fascicoli somministrati.

	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_originale	D3_L062012_v3	D3_L062012_v2	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

Tabella 5.3: Struttura dei fascicoli somministrati per il progetto *Variazioni 2*.
Gli item evidenziati in giallo costituiscono l'àncora esterna, quelli in arancione l'àncora interna e gli altri item, in grigio, sono quelli presenti in diverse forme nei fascicoli.

Nei fascicoli le domande sono inserite con l'etichetta riportata nella prima colonna e, per risalire velocemente alla domanda originale, sono stati inserite in grigio delle etichette che riportano il riferimento alla prova originale.

L'ultima parte di ogni fascicolo è dedicata a un questionario con domande di contesto e ha lo scopo di mettere in luce le convinzioni e atteggiamenti degli studenti relativamente alla matematica, il livello di *math anxiety* e altre informazioni che potranno fornire importanti indicazioni se correlate con i risultati del test.

5.3.1 Esempi di Variazioni

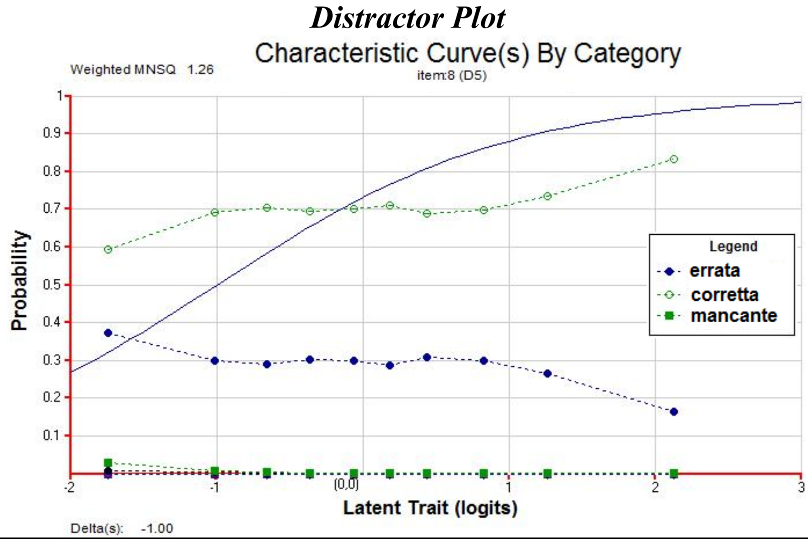
Viste le molteplici finalità del progetto, sono stati diversi anche i criteri con cui sono state individuate le domande da variare.

Alcune delle domande sono state selezionate tra quelle di prove passate che mostravano problemi da un punto di vista misuratorio. Può succedere infatti che una domanda non sia completamente coerente con il modello per alcuni livelli di abilità e che presenti quindi un fit non ottimale. L'analisi di alcune di queste domande attraverso il progetto *Variazioni 2* permetterà di indagare le motivazioni alla base di questo *misfit*, ad esempio:

- la formulazione della domanda potrebbe essere poco chiara e gli studenti sbaglierebbero 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 indipendentemente dall'abilità matematica e che farebbero sbagliare anche studenti molto bravi.

Un esempio paradigmatico è la domanda seguente, tratta dalla Prova Nazionale (livello 8) del 2011.

D5 Livello 8 del 2011
D5. Giovanni e Caterina si stanno allenando in piscina. Nuotano entrambi alla stessa velocità ma Giovanni ha cominciato più tardi ad allenarsi. Quando Giovanni ha fatto 10 vasche, Caterina ne ha fatte 30. Al termine dell'allenamento Giovanni ha fatto 50 vasche; quante ne ha fatte Caterina? Risposta: ....



DELTA = -1.00
DISCRIMINATION = 0.10
WEIGHTED = 1.26

Figura 5.1: Analisi Item D5 prova INVALSI di livello 8 del 2011 (item D12 nei fascicoli del progetto *Variazioni 2*).

Questo item, pur risultando semplice e raggiungendo quasi il 70% delle risposte corrette, mostra un pessimo comportamento da un punto di vista psicometrico. Il *misfit* della curva relativa alle risposte corrette rispetto al modello teorico è evidente sia osservando il distractor plot (il modello sottostima gli studenti con bassi livelli di abilità e sovrastima quelli medio-alti), sia in termini di *weighted* che risulta non accettabile. Inoltre la domanda ha una discriminazione molto bassa, anche questa al di sotto della soglia limite. Dal distractor plot si nota che la risposta corretta è crescente solo nel passaggio tra il primo e il secondo decile e nei decili più alti, per il resto è costante e l'andamento della risposta errata è speculare. Questo significa che gli studenti hanno la stessa probabilità di rispondere correttamente, indipendentemente dal loro livello di abilità e la ridottissima percentuale di risposte mancanti indica che gli studenti sono convinti della risposta fornita.

La domanda risulta ben formulata e non sembrano esserci motivi per cui si potrebbero creare fraintendimenti. L'ipotesi elaborata per spiegare questo risultato è piuttosto legata a fattori di natura didattica: probabilmente la tipologia di problema e il contesto portano molti studenti ad attivare un ragionamento di tipo proporzionale (10:30=50:x) e i protocolli confermano questa tendenza. Si è deciso quindi di mantenere inalterata la struttura del problema e i dati forniti ma di modificarne il contesto per capire se è il contesto stesso a portare gli studenti a questo tipo di ragionamento.

Le versioni in cui il quesito viene proposto nei diversi fascicoli sono le seguenti:

FASCICOLO 1	*D5_PN2011_originale* **D12. Giovanni e Caterina si stanno allenando in piscina. Nuotano entrambi alla stessa velocità ma Giovanni ha cominciato più tardi ad allenarsi. Quando Giovanni ha fatto 10 vasche, Caterina ne ha fatte 30. Al termine dell'allenamento Giovanni ha fatto 50 vasche; quante ne ha fatte Caterina?** **Risposta: ....** vasche
FASCICOLO 2	*D5_PN2011_v2* **D12. Giovanni e Caterina mettono la paghetta settimanale ognuno nel proprio salvadanaio. Ricevono la stessa paghetta, ma Giovanni ha iniziato più tardi a metterla nel suo salvadanaio. Quando Giovanni ha 10 euro, Caterina ne ha 30. A Natale, quando rompono i salvadanai, Giovanni ha 50 euro; quanti euro ha Caterina?** **Risposta: ....** euro
FASCICOLO 3	*D5_PN2011_v4* **D12. Giovanni e Caterina contano le macchine che passano davanti al cancello di casa. Giovanni ha cominciato più tardi a contare. Quando Giovanni ha contato 10 macchine, Caterina ne ha contate 30. Quando rientrano in casa, Giovanni ha contato 50 macchine; quante ne ha contate Caterina?** **Risposta: ....** macchine
FASCICOLO 4	*D5_PN2011_v3* **D12. Due cronometri A e B segnano il tempo. Il cronometro B è partito dopo il cronometro A. Quando B segna 10 secondi, A ne segna 30. Quando vengono fermati, B segna 50 secondi; quanti secondi segna A in quell'istante?** **Risposta: ....** secondi

Figura 5.2: Item D12 nella versione originale (D5 livello 8 del 2011) e nelle quattro versioni variate.

Nell'inserimento dei dati si terrà anche nota degli studenti che rispondono 150, applicando quindi un ragionamento proporzionale.

Altre domande sono state invece selezionate tra quelle che mostravano un gender gap particolarmente marcato, le variazioni sono state pensate per capire cosa fosse alla base di questo divario e se fosse possibile ridurlo senza incidere sul *question intent*.

Un esempio di queste domande è quella relativa alla stima della lunghezza di un treno, già analizzata anche nel paragrafo 4.4.3. In questo caso si è scelto di modificare l'immagine per incentivare gli studenti a ragionare sulla stima della lunghezza di un vagone e poi moltiplicare questa misura per 10 per ottenere la lunghezza del treno. A tal fine l'immagine inserita nella versione variata rappresenta un unico vagone e si fornisce un riferimento (il ragazzo) che possa aiutare nella stima.

Le versioni in cui il quesito viene proposto nei diversi fascicoli sono le seguenti:

<table>
<tr><th colspan="2">D7b Livello 6 del 2013</th></tr>
<tr><td>FASCICOLO 2
e
FASCICOLO 3</td><td>D7b_L062013_originale
D8. Nino sale su un treno composto dalla locomotiva e da 9 vagoni:

Quanto è lungo all'incirca il treno di Nino?
A. □ Circa 10 m
B. □ Circa 50 m
C. □ Circa 250 m
D. □ Circa 1000 m</td></tr>
</table>

FASCICOLO 1 e FASCICOLO 4	*D7b_L062013_v1new* **D8. Nino sale su un treno composto dalla locomotiva e da 9 vagoni:**  **Quanto è lungo all'incirca il treno di Nino?** A. □ **Circa 10 m** B. □ **Circa 50 m** C. □ **Circa 250 m** D. □ **Circa 1000 m**

Figura 5.3: Item D8 nella versione originale (D7b livello 6 del 2013) e nella versione variata.

Infine alcuni quesiti sono stati anche creati ex novo per la sperimentazione: i questiti D10 e D11 ad esempio avranno anche l'obiettivo di verificare la validità di un nuovo formato di risposta, mentre i quesiti D9 e D15 provengono da un articolo di didattica (Sbaragli, 2012) al fine di indagare una particolare misconcezione.

Studiando i quesiti che mostrano maggiore gender gap è emersa una difficoltà maggiore delle ragazze nell'operare con i numeri decimali e con le percentuali che è già stata trattata anche nel paragrafo 4.4.3. In particolare è emersa una maggiore incidenza sulle studentesse di alcune misconcezioni legate all'ordinamento dei decimali e al calcolo con le percentuali che scaturiscono dall'ampliamento dell'insieme dei numeri naturali all'insieme dei numeri razionali.

Si è deciso quindi di proporre un quesito finalizzato allo studio di un'altra misconcezione, sempre causata dall'ampliamento dell'insieme dei numeri naturali ai numeri razionali. Gli studenti, operando esclusivamente con i numeri naturali, tendono a crearsi un modello di moltiplicazione come operazione "che accresce" (Sbaragli, 2012) ovvero in cui il risultato è maggiore dei fattori; diventa difficile per gli studenti superare questa immagine nel momento in cui si trovano svolgere una moltiplicazione in cui uno dei due fattori (o entrambi) sono un numero compreso tra 0 e 1.

L'item D9, inserito proprio a partire dall'articolo sopra citato, ha lo scopo di approfondire l'incidenza di tale misconcezione anche negli ultimi anni della scuola secondaria di primo grado e capire se, anche in questo caso, le ragazze mostrano maggiori difficoltà.

FASCICOLO 1 e FASCICOLO 2	*1CG_NEW_v1* **D9. Qual è il risultato di $4 \times 0{,}5$?** A. □ **8** B. □ **4** C. □ **2** D. □ **20**
FASCICOLO 3 e FASCICOLO 4	*1CG_NEW_v2* **D9. Qual è il risultato di $0{,}5 \times 4$?** A. □ **8** B. □ **4** C. □ **2** D. □ **20**

Figura 5.4: Item D9 nelle due versioni proposte.

Le due diverse versioni vogliono indagare se la misconcezione agisce in maniera diversa a seconda dell'ordine dei fattori: è possibile infatti che la misconcezione sia legata anche all'ordine in cui sono presentati i fattori e che il primo fattore sia quello che viene confrontato con il risultato. In questo caso si potrebbe riscontrare una maggiore incidenza della misconcezione nella prima versione dell'item.

5.3.1 Sperimentazione

Una volta messi a punto i fascicoli è stato svolto un primo *pilot study* in 5 classi terze. Questo ha permesso di evidenziare refusi e criticità delle nuove domande e delle domande già somministrate che però avevano subito variazioni. Pur non avendo una significatività in termini statistici, questi primi risultati hanno permesso di avere una prima informazione sulla difficoltà delle diverse forme degli item e capire se la distribuzione dei quesiti nei diversi fascicoli fosse omogenea.

In base a questo studio preliminare sono stati modificati i fascicoli fino ad arrivare alla versione definitiva (riportata in appendice), utilizzata nel *main study*.

La scelta del campione per il *main study* è stata svolta internamente all'Istituto INVALSI seguendo le procedure utilizzate normalmente nell'individuazione del campione per le prove nazionali (Falorsi & Righi, 2008) e tenendo conto della necessità di applicare ai fascicoli tecniche di *test equating* e *linking* basate sulla Rasch Analysis.

Per effettuare la procedura di ancoraggio tra i fascicoli e permettere quindi il confronto tra le diverse forme degli item, la letteratura indica come numerosità ottimale 400 studenti per ogni forma. Si è quindi deciso di somministrare i test in 113 classi terze, per un totale di più di 2000 alunni. Il campione è stato studiato in modo da essere il più possibile rappresentativo anche a livello nazionale e, per facilitare la somministrazione dei fascicoli, le scuole sono state individuate in 4 regioni considerate più rappresentative: Campania, Lazio, Emilia Romagna e Lombardia. Inoltre nel campionamento si è tenuto conto dello status socio-economico-culturale degli studenti, estrapolando i dati dal *main study* INVALSI del 2015 e basandosi sull'SC-index (Cascella) che risulta alternativo all'ESCS.

Una volta effettuato l'inserimento dei dati di tutti i fascicoli verranno considerati nelle analisi solamente gli studenti non indicati come BES in quanto, durante lo svolgimento della prova, hanno avuto la possibilità di adoperare strumenti compensativi e dispensativi che potrebbero aver modificato i processi attivati nel rispondere agli item (ad esempio, perde completamente di significato il quesito D9 nel caso in cui lo studente abbia a disposizione la calcolatrice).

Le analisi iniziali saranno mirate allo studio delle proprietà misuratorie dei fascicoli e degli item all'interno dei diversi fascicoli, al fine di verificarne la validità e operare i primi confronti tra i risultati dei diversi fascicoli.

La procedura di *test equating* sarà applicata attraverso l'uso del modello di Rasch e in particolare servendosi del software RUMM2020. Attraverso questa procedura si avrà la possibilità di porre tutti gli item dei 4 fascicoli sulla medesima scala di difficoltà e comprendere, quindi, se le variazioni applicate hanno causato una differenza statisticamente significativa in termini di difficoltà.

Gli studenti saranno tutti considerati sulla medesima scala di abilità, indipendentemente dal fascicolo a cui hanno risposto. Sarà quindi possibile confrontare i distractor plot delle diverse forme in cui una stessa domanda è stata inserita nei fascicoli e studiare eventuali differenze nell'andamento dei distrattori e della risposta corretta in funzione del livello di abilità degli studenti.

Si potrà analizzare l'impatto delle variazioni su sottogruppi specifici di studenti e, in particolare, in base al genere. Questo potrà essere realizzato sia attraverso il confronto delle percentuali di risposta ai due item, sia facendo uso dei risultati dell'ancoraggio. Considerando il livello di abilità degli studenti fornita dall'ancoraggio e suddividendo gli studenti in base al genere, sarà possibile costruire i distractor plot delle diverse forme, separatamente per maschi e femmine. In questo modo sarà quindi possibile evidenziare se l'impatto di una variazione è stato maggiore per uno dei due gruppi di studenti e vedere se ha influito nello stesso modo in funzione dei livelli di abilità.

Infine sarà utile correlare i dati del test di matematica con le informazioni di contesto fornite dal questionario studente. In particolare, sarà interessante utilizzare il questionario relativo alla *math anxiety* per verificare la presenza, documentata in letteratura, di una significativa differenza di genere relativa a questo costrutto. Risulta inoltre particolarmente interessante la possibilità di correlare il livello di ansia degli studenti con il livello di abilità in matematica ottenuto sull'intero test o su determinati item.

Capitolo 6

Conclusioni e Possibili sviluppi

Le prove standardizzate internazionali e nazionali hanno come obiettivo principale quello di monitorare gli apprendimenti degli studenti durante il percorso scolastico al fine di avere indicazioni per rendere più efficaci le politiche educative e intervenire nel caso in cui vengano riscontrate particolari criticità. In questa direzione l'importanza di questi test sta crescendo notevolmente negli anni, come testimoniato da Figel (2009):

> *Information on pupil performance is key to the successful implementation of targeted education policies and it is not surprising that in the past two decades national tests have emerged as an important tool for providing a measure of educational achievement.*

L'apporto che queste prove possono dare anche nel campo della ricerca in didattica della matematica viene ancora sottovalutato (come evidenziato nel capitolo 3) e, in particolare in Italia, l'uso che viene fatto delle prove INVALSI nel campo della ricerca risulta ancora limitato.

Questa tesi ha messo in luce in che modo fino ad ora sono state utilizzate le prove standardizzate per la ricerca in didattica a livello internazionale e, attraverso il confronto con le prove INVALSI, ha mostrato le potenzialità dell'uso di queste prove nella ricerca in didattica.

L'analisi dei test INVALSI attraverso l'uso dei principali indici della Teoria Classica del Test e del modello di Rasch (introdotti nel capitolo 2), ha permesso di far emergere interessanti macro fenomeni e di studiarli approfonditamente. In particolare si è visto quanto studi basati sui test INVALSI possano essere fondamentali per lo studio di differenze nell'apprendimento della matematica tra sottogruppi della popolazione (ad esempio studenti madrelingua/non madrelingua e maschi/femmine). Lo studio di queste differenze non può limitarsi però esclusivamente a una analisi del gap riscontrato sull'intera prova ma è importante che si focalizzi anche sui singoli item e i loro cluster.

Le ricerche riportate in questa tesi propongono una metodologia che va in questa direzione, basata sull'applicazione del modello di Rasch e del *test-equating*.
Una volta constatata la presenza del gap (di cittadinanza e/o di genere) sull'intera prova, si è fatto uso dell'indice presentato nel paragrafo 4.2.1 per analizzare come tale divario si distribuisca sui singoli item. In questo modo è stato possibile osservare che il gap di genere e quello di cittadinanza presentano caratteristiche molto diverse:

- Il gap in matematica tra studenti italiani e di origine immigrata è presente in tutti i quesiti anche se alcuni presentano una differenza più marcata;
- Il gap in matematica tra maschi e femmine non è uniforme ma risulta particolarmente significativo per alcuni item, mentre per altri la differenza è nulla o addirittura a favore delle femmine.

L'analisi si è quindi potuta concentrare sui quesiti ritenuti più significativi a seguito di questa prima indagine. Il modello di Rasch fornisce per ogni item di una prova un *distractor plot* (paragrafo 2.5) che permette di studiare l'andamento delle risposte degli studenti in funzione del livello di abilità ottenuto sull'intera prova e di confrontare l'andamento empirico della risposta corretta con la curva teorica fornita dal modello. L'analisi di questi grafici può portare a importanti considerazioni in termini di *fit*, di discriminazione e di scelta dei distrattori che possono essere poi interpretate attraverso le lenti della didattica (come mostrato nel capitolo 4). Infine, attraverso l'analisi del DIF e dei *distractor plot* ottenuti separando i due gruppi di studenti, è possibile analizzare, per un determinato item, su quali livelli di abilità il divario risulti più marcato e se vi siano differenze nell'andamento dei distrattori che possano essere la causa del divario sulla risposta corretta.

Nel capitolo 4 si è visto come le evidenze fornite sui singoli item da questo tipo di analisi possano essere particolarmente utili per fornire una interpretazione in chiave didattica delle differenze riscontrate. In particolare, per quanto riguarda il gender gap (paragrafo 4.4), sono stati analizzati in questo modo numerosi quesiti, tratti da prove INVALSI di diversi livelli scolastici, alla luce della letteratura presente sul tema e di alcuni costrutti studiati in didattica. Questa analisi mostrato interessanti evidenze e ha permesso di fare nuove ipotesi sulla natura del gender gap, ipotesi che dovranno essere approfondite in ricerche future.

Nonostante il tema delle differenze di genere risulti particolarmente complesso e legato a una molteplicità di fattori (riportati nel paragrafo 4.4.2), le analisi svolte supportano l'ipotesi che le principali cause di questo gap siano da ricercare nel contesto sociale e culturale, includendo anche il contesto classe in cui lo studente apprende e le dinamiche strettamente legate al *milieu*.

Le analisi svolte a livello dei singoli quesiti risultano particolarmente interessanti se precedute da analisi delle differenze nelle performance di sottogruppi di studenti anche sull'intera prova. Nei rapporti tecnici delle prove standardizzate spesso il gap tra gruppi di studenti viene indicato come differenza tra i punteggi medi ottenuti, risulta però interessante anche studiare il gap in funzione del livello di abilità degli studenti, cosa possibile solo attraverso l'analisi di prove su larga scala. In questa tesi il gap è stato analizzato longitudinalmente attraverso l'applicazione del modello di Rasch all'intera prova e osservando la distribuzione percentuale degli studenti in funzione del parametro di abilità ottenuto applicando il modello di Rasch all'intera prova. I risultati per quanto riguarda il gender gap confermano quanto osservato anche in studi precedenti (e.g. OECD, 2016a; Di Tommaso et al., 2016): non si notano differenze nella distribuzione per i livelli bassi, mentre sono molte le ragazze che ottengono livelli di abilità medi e solo poche, rispetto ai maschi, riescono a raggiungere livelli alti di abilità.

Una ulteriore potenzialità delle ricerche in didattica basate sulle prove standardizzate consiste nella possibilità di analizzare prove somministrate alla stessa popolazione in diversi momenti del percorso scolastico. Ricerche longitudinali di questo tipo, effettuate utilizzando i dati INVALSI, permettono di seguire una stessa coorte di studenti dalla primaria fino alla secondaria di secondo grado, analizzando l'evoluzione degli apprendimenti e anche dei gap riscontrati negli anni. Nel capitolo 4 sono descritte le analisi svolte sia nel caso delle differenze di cittadinanza sia per le differenze di genere. Con questo approccio è stato possibile analizzare l'andamento complessivo del gap negli anni e soprattutto trovare, all'interno delle diverse prove, cluster di quesiti che hanno evidenziato maggiori difficoltà per un sottogruppo degli studenti nei diversi livelli scolastici. Per quanto riguarda le differenze di genere, ad esempio, questo tipo di analisi ha portato a individuare particolari contenuti (confronto di numeri decimali, operazioni con le percentuali, stima…) che hanno messo in luce un forte gap a favore dei maschi anche in prove di diversi livelli scolastici.

Infine l'ultimo capitolo della tesi (capitolo 5) mostra come, attraverso l'uso di prove standardizzate e del modello di Rasch, sia stato possibile sviluppare una metodologia di ricerca per analizzare l'impatto di una variazione nella formulazione di un quesito di matematica superando i principali ostacoli incontrati dalle ricerche che in precedenza si erano occupate di questo tema (paragrafo 5.2.2). Inoltre, questo nuovo approccio ha permesso di analizzare in che modo una variazione influenza le risposte degli studenti in termini di livello di abilità e se ha un impatto maggiore su un determinato sottogruppo della popolazione.

In conclusione, questa tesi ha mostrato che un'analisi quantitativa di prove standardizzate, compiuta tenendo conto dei limiti che queste prove hanno per natura, può dare alla ricerca in didattica della matematica apporti significativi e importanti indicazioni per successive ricerche, anche di tipo qualitativo, in diverse direzioni.

Bibliografia[30]

Abbott, S., Guisbond, L., Levy, J., & Sommerfeld, M. (2014). The glossary of education reform. *Great School Partnership*.

Anderson, L. W., & Postlethwaite, T. N. (2007). Program evaluation: Large-scale and small-scale studies. *International Institute for Educational Planning Series Booklet*, *8*.

Barbaranelli, C., & Natali, E. (2005). *I test psicologici: teorie e modelli psicometrici.* Roma: Carrocci Editore.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of autism and developmental disorders*, *34*(2), 163-175.

Bell, K. N., & Norwood, K. (2007). Gender equity intersects with mathematics and technology: Problem-solving education for changing times. *Gender in the classroom*, 225-258.

Bishop, A. J., & Forgasz, H. J. (2007). Issues in access and equity in mathematics education. *Second handbook of research on mathematics teaching and learning*, *2*, 1145-1167.

Boaler, J. (1997). Reclaiming school mathematics: The girls fight back. *Gender and Education*, *9*(3), 285-305.

Bolondi, G. & Viale, M. (2014). Abilità linguistiche e discipline scientifiche: un'esperienza di formazione del corpo insegnante nel Polo dell'Emilia-Romagna del progetto «I Lincei per una nuova didattica nella scuola». *Atti del Convegno Nazionale GISCEL, Roma*.

Bolondi, G., Branchetti, L., & Ferretti, F. (2014). Correlazioni tra componenti della competenza linguistica e capacità di lavoro su un testo matematico: gli studenti del Liceo Scientifico alle prese con le prove dell 'Esame di Stato.

Branchetti, L., & Viale, M. (2014). Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico. *La didattica dell'italiano. Problemi e prospettive*, 139.

[30] La bibliografia dei singoli articoli inseriti nella tesi si trova in calce a ogni articolo. La bibliografia qui riportata si riferisce a tutta la tesi ad esclusione degli articoli inseriti.

Byrnes, J. P. (2005). Gender differences in math. *Gender differences in mathematics: An integrative psychological approach*, 73-98.

Byrnes, J. P., & Takahira, S. (1993). Explaining gender differences on SAT-math items. *Developmental Psychology*, *29*(5), 805.

Calkins, S. D. (2007). The emergence of self-regulation: Biological and behavioral control mechanisms supporting toddler competencies. *Socioemotional development in the toddler years: Transitions and transformations*, 261-284.

Cargnelutti, E., Tomasetto, C. & Passolunghi, M. C. (2016). How is anxiety related to math performance in young students? A longitudinal study of Grade 2 to Grade 3 children, *COGNITION & EMOTION*, 2017, 31, pp. 755 – 764.

Cascella, C. (2014). *Un approccio stocastico all'analisi dei gruppi per studiare la dimensionalità delle prove INVALSI*. XVII Congresso nazionale Associazione Nazionale della Valutazione (AIV). Napoli.

Cascella, C. (2017). Male and female social roles in daily file, pupils' perceptions and gender gap in Math test scores. Some empirical evidences from Italian primary school. (In press)

Croker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* Holt, Rinehart & Winston, New York.

D'Amore B. (2000). Lingua, Matematica e Didattica. La matematica e la sua didattica. 1, 28- 47.

Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and brain functions*, *8*(1), 33.

Di Tommaso, M. L., Mendolia, S., & Contini, D. (2016). The Gender Gap in Mathematics Achievement: Evidence from Italian Data. *IZA Discussion paper*, n.10053, Bonn.

Doepke, M., Tertilt, M., & Voena, A. (2011). The economics and politics of women's rights. Annual Review of Economics. Vol. 4 (339-372).

Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years?. *Frontiers in psychology*, *7*.

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journal of Social Issues*, *46*(2), 183-201.

Falorsi, P. D., & Righi, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology*, *34*(2), 223-234.

Farré, L., & Vella, F. (2013). The intergenerational transmission of gender role attitudes and its implications for female labour force participation. *Economica*, *80*(318), 219-247.

Fennema, E., Carpenter, T. P., Jacobs, V. R., Franke, M. L., & Levi, L. W. (1998). New perspectives on gender differences in mathematics: A reprise. *Educational Researcher*, *27*(5), 19-21.

Ferretti, Federica (2015) *L'effetto "età della Terra". Contratto didattico e principi regolativi dell'azione degli studenti in matematica.* [Dissertation thesis], Alma Mater Studiorum Università di Bologna. Dottorato di ricerca in Matematica, 27 Ciclo. DOI 10.6092/unibo/amsdottorato/7213.

Figel, J. (2009). Preface. In *National testing of pupils in Europe: Objectives, organisation and use of results*. Brussels: Education, Audiovisual and Culture Executive Agency Eurydice. http://www.eurydice.org. Retrieved from 27 March 2012.

Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210-240.

Forgasz, H. J. (2010). International perspectives on gender and mathematics education. IAP.

Forgasz, H. J., & Rivera, F. (2012). Towards Equity in Mathematics Education. Springer.

Fredricks, J. A., & Eccles, J. S. (2002). Children's competence and value beliefs from childhood through adolescence: growth trajectories in two male-sex-typed domains. *Developmental psychology*, *38*(4), 519.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of experimental child psychology*, *75*(3), 165-190.

Gallagher, A. M., & Kaufman, J. C. (Eds.). (2004). *Gender differences in mathematics: An integrative psychological approach*. Cambridge University Press.

Gallagher, A. M., & Kaufman, J. C. (2005). Gender differences in mathematics: Where we are and what we need to know. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics* (pp. 316-332). New York, NY: Cambridge University Press.

Giberti, C., Zivelonghi, A., & Bolondi, G. GENDER DIFFERENCES AND DIDACTIC CONTRACT: ANALYSIS OF TWO INVALSI TASKS ON POWERS PROPERTIES. *International Group for the Psychology of Mathematics Education*, 275.

González de San Román, A., & De La Rica, S. (2016). Gender gaps in PISA test scores: The impact of social norms and the mother's transmission of role attitudes. *Estudios de Economía Aplicada*, *34*(1).

Gould, S. L. (1996). Strategies used by secondary school students in learning new concepts which require spatial visualization. Unpublished doctoral dissertation, Teachers College, Columbia University, New York.

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *SCIENCE-NEW YORK THEN WASHINGTON-*, *320*(5880), 1164.

Halpern, D. F., Beninger, A. S., & Straight, C. A. (2011). Sex differences in intelligence. *The cambridge handbook of intelligence*, 253-272.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Helwig, R., Anderson, L., & Tindal, G. (2001). Influence of elementary student gender on teachers' perceptions of mathematics achievement. *The Journal of Educational Research*, *95*(2), 93-102.

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for research in mathematics education*, 33-46.

Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology*, *26*(3), 276-295.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036.

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, *321*(5888), 494-495.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis.

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences*, *106*(22), 8801-8807.

INVALSI (2013). Nota sullo svolgimento delle prove del SNV 2011-2012 per gli allievi di origine immigrata. Retrieved July 2017 from http://www.invalsi.it/snv2012/documenti/istruzioni/Nota_sugli%20alunni_di_origine_immigrata_2012.pdf

INVALSI (2015). Rilevazione nazionale degli apprendimenti 2014-2015. Rapporto risultati.

Retrieved July 2017 from http://www.invalsi.it/invalsi/doc_evidenza/2015/034_Rapporto_Prove_INVALSI_2015.pdf

INVALSI (2016a). Rilevazione nazionale degli apprendimenti 2015-2016. Rapporto risultati. Retrieved July 2017 from http://www.invalsi.it/invalsi/doc_evidenza/2016/07_Rapporto_Prove_INVALSI_2016.pdf

INVALSI (2016b). Rilevazione nazionale degli apprendimenti 2015-2016. Rapporto tecnico. Retrieved July 2017 from http://www.invalsi.it/invalsi/doc_evidenza/2016/002_Rapporto_tecnico_2016.pdf

INVALSI (2017). Rilevazione nazionale degli apprendimenti 2016-2017. Rapporto risultati. Retrieved July 2017 from https://invalsi-areaprove.cineca.it/docs/file/Rapporto_Prove_INVALSI_2017.pdf

ISTAT (2015). L'integrazione scolastica e sociale delle seconde generazioni. Statistiche Report. Retrieved July 2017 from http://www.istat.it/it/immigrati/prodotti-editoriali/istruzione

Jacobs, J. E., & Bleeker, M. M. (2004). Girls' and boys' developing interests in math and science: Do parents matter?. *New directions for child and adolescent development*, *2004*(106), 5-21.

Jacobs, J. E., & Eccles, J. S. (1992). The influence of parent stereotypes on parent and child ability beliefs in three domains. *Journal of Personality and Social Psychology*, *63*(6), 932-944.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of personality and social psychology*, *63*(6), 932.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child development*, *73*(2), 509-527.

Johnson, R. B. & Onwuegbuzie, A. J. (2004).Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33, 7: 14-26.

Klasen, S. (2002). Low schooling for girls, slower growth for all? Cross-country evidence on the effect of gender inequality in education on economic development. *The World Bank Economic Review*, *16*(3), 345-373.

Kenney, P. A., & Schloemer, C. G. (2001). Assessment of student achievement, overview. *Encyclopedia of mathematics education*, 50-56.

Kesici, S., Sahin, I., & Akturk, A. O. (2009). Analysis of cognitive learning strategies and computer attitudes, according to college students' gender and locus of control. *Computers in Human Behavior*, *25*(2), 529-534.

Kolen, M. J., & Brennan, R. L. (2013). *Test equating: Methods and Practices*. Springer Science & Business Media.

Langer, E. J. (1997). The power of mindful learning. Reading, MA: Addison Wesley

Leder, G. C. (1992). Mathematics and gender: Changing perspectives.

Leder, G., & Forgasz, H. (2008). Mathematics education: new perspectives on gender. *ZDM – The International Journal on Mathematics Education*, 40(4), 513-518.

Leder, G., & Lubienski, S. (2015). Large-Scale Test Data: Making the Invisible Visible. In *Diversity in Mathematics Education* (pp. 17-40). Springer International Publishing.

Linacre, J. M. (2004). Test validity and Rasch measurement: construct, content, etc. *Rasch measurement transactions*, *18*(1), 970-971.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis.

Lord, F. M., & Novick, M. R. (2008). *Statistical theories of mental test scores*. IAP.

Lawton, C. A., & Hatcher, D. W. (2005). Gender differences in integration of images in visuospatial memory. *Sex roles*, *53*(9-10), 717-725.

Lawton, C. A., & Hatcher, D. W. (2005). Gender differences in integration of images in visuospatial memory. *Sex roles*, *53*(9-10), 717-725.

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, *41*(1), 63-76.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for research in mathematics education*, 520-540.

Ma, X., & Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: A meta-analysis. *Journal for research in mathematics education*, 26-47.

Marrs, H., & Sigler, E. A. (2012). Male academic performance in college: The possible role of study strategies. *Psychology of Men & Masculinity*, *13*(2), 227.

McClelland, M. M., & Cameron, C. E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development*, *2011*(133), 29-44.

Maffia A., Giberti C. (2016). *Ricerca in Didattica della matematica e PISA: percorsi battuti e nuove piste da esplorare.* Pubblicazione di approfondimento sui temi relativi al PISA 2012. ISBN-13: 9788891748447

Marsh, H. W., & O'Mara, A. J. (2008). Self-concept is as multidisciplinary as it is multidimensional. *Self-processes, learning, and enabling human potential. Dynamic new approaches*, 87-115.

Mackinnon, N. (1990). Sophie Germain: or was Gauss a feminist? *Math Gaz*, 74(470), 346–351.

Matthews, J. S., Ponitz, C. C., & Morrison, F. J. (2009). Early gender differences in self-regulation and academic achievement. *Journal of educational psychology*, *101*(3), 689.

Middleton, J. A., Cai, J., & Hwang, S. (2015). Why mathematics education needs large-scale research. In *Large-scale studies in mathematics education*(pp. 1-13). Springer International Publishing.

MIUR (2012). Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. *Annali della Pubblica istruzione*.

Morley, L., Leach, F., & Lugg, R. (2009). Democratizing higher education in Ghana and Tanzania: Opportunity structures and social inequalities. *International Journal of Educational Development*, 29, 56- 64.

Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied psychological measurement*, *24*(3), 211-223.

Mullis, I.V.S. & Martin, M.O. (Eds.) (2013). TIMSS 2015 Assessment Frameworks. Retrieved from Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 International Results in Mathematics. *TIMSS & PIRLS International Study*.

OECD (2009), Technical Report for the OECD Programme for International Student Assessment 2006, OECD, Paris.

OECD (2012a). ITALY – Country Note –Results from PISA 2012. OECD Publishing.

OECD (2012b). Results in Focus: What 15-year-olds know and what they can do with what they know. PISA, OECD Publishing.

OECD (2013a). *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance to Succeed* (Volume II), PISA, OECD Publishing.

OECD (2013b), PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III), PISA, OECD Publishing.

OECD, (2015a). *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*. PISA, OECD Publishing.

OECD (2015b). *PISA 2015 Mathematics Framework.* PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy, OECD publishing, Paris.

OECD (2015c). *Can the performance gap between immigrant and non-immigrant students be closed?*. PISA in Focus, No. 53, OECD Publishing, Paris.

OECD, (2016a). PISA 2015 Results (Volume I): *Excellence and Equity in Education*. OECD Publishing, Paris.

OECD, (2016b). *ITALY – Country Note – Results from PISA 2015*. OECD Publishing.

OECD (2016c). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*, OECD Publishing, Paris.

Ongaki, N. M., & Musa, F. W. (2014). *Enhancing Socio-Economic Equity in Accessing Quality Education: A Case of Form One Selection Policy in KISII County, Kenya*. The International Journal of Business & Management, 2(11), 157.

Osterlind, S. J. (1983). *Test item bias* (Vol. 30). Sage.

Owens, T. L. (2013). Thinking beyond league tables: A review of key PISA research questions. *PISA, power, and policy: The emergence of global educational governance*, 27-49.

Pajares, F. (2005). Gender differences in mathematics self-efficacy beliefs. *Gender differences in mathematics: An integrative psychological approach*, 294-315.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of educational psychology*, *86*(2), 193.

Primi, C., Busdraghi, C., Tomasetto, C., Morsanyi, K., & Chiesi, F. (2014). Measuring math anxiety in Italian college and high school students: validity, reliability and gender invariance of the Abbreviated Math Anxiety Scale (AMAS). *Learning and Individual Differences*, *34*, 51-56.

Rash, G. (1960). Probabilistic models for some intelligence and attainment tests. *Copenhagen:*

Danish Institute for Educational Research.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. *Danish Yearbook of Philosophy,* 14, 58–94.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, *35*(1), 83-92.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, *138*(2), 353.

Riegle-Crumb, C. (2005). The cross-national context of the gender gap in math and science. *The social organization of schooling*, 227-243.

Romanes, G. J. (1887). *Mental differences between men and women*. Nineteenth Century 21:654–672.

Ruffing, S., Wach, F. S., Spinath, F. M., Brünken, R., & Karbach, J. (2015). Learning strategies and general cognitive ability as predictors of gender-specific academic achievement. *Frontiers in psychology*, *6*.

Sbaragli, S. (2012). Il ruolo delle misconcezioni nella didattica della matematica. *Bolondi B., Fandiño Pinilla MI (2012). I quaderni della didattica. Metodi e strumenti per l'insegnamento e l'apprendimento della matematica*, 121-139.

Schunk, D. H., & Zimmerman, B. J. (1997). Social origins of self-regulatory competence. *Educational psychologist*, *32*(4), 195-208.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. Cambridge University Press.

Schultz, T. P. (2002). Why governments should invest more to educate girls. *World Development*, *30*(2), 207-225.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *American Psychologist*, *60*(9), 950.

Spinath, B., & Spinath, F. M. (2005). Development of self-perceived ability in elementary school: The role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development*, *20*(2), 190-204.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational psychology*, *92*(1), 144.

Tomasetto, C., (2013). Matematica per i maschi, italiano per le femmine: Stereotipi di genere e atteggiamenti verso le materie scolastiche tra genitori e figli, *IN-MIND ITALIA*, 5, pp. 19 – 24.

Tomasetto, C., Mirisola, A., Galdi, S., & Cadinu, M. (2015). Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemporary Educational Psychology*, *42*, 186-198.

Traub, R. E. (1994). *Reliability for the social sciences*. Thousands Oaks, Sage (CA).

UNESCO (2008). *Education for all global monitoring report 2009.* Oxford: Oxford University Press.

Virtanen, P., & Nevgi, A. (2010). Disciplinary and gender differences among higher education students in self-regulated learning strategies. *Educational Psychology*, *30*(3), 323-347.

Weis, M., Heikamp, T., & Trommsdorff, G. (2013). Gender differences in school achievement: The role of self-regulation. *Frontiers in Psychology*, *4*, 442. http://doi.org/10.3389/fpsyg.2013.00442

Winkelmann, H., van den Heuvel-Panhuizen, M., & Robitzsch, A. (2008). Gender differences in the mathematics achievements of German primary school students: Results from a German large-scale study. *ZDM*, *40*(4), 601-616.

Wößmann, L., & Schütz, G. (2006). Efficiency and equity in European education and training systems. *Analytical Report for the European Commission prepared by the European Expert Network on Economics of Education,(Bruselas: Comisión Europea, 2006).*

Yee, D. K., & Eccles, J. S. (1988). Parent perceptions and attributions for children's math achievement. *Sex Roles*, *19*(5), 317-333.

Zan, R. (2007). *Difficoltà in matematica: osservare, interpretare, intervenire*. Springer Science & Business Media.

ZHU, Y. (2016). *Equity in mathematics education: what did timss and pisa tell us in the last two decades?.* 13th International Congress on Mathematical Education Hamburg, 24-31 July 2016.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-type (Ordinal) Item Scores.* Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Appendice

L'appendice contiene i questionari utilizzati per le diverse ricerche presentate in questa tesi.

In particolare, in Appendice 1 sono riportati i materiali utilizzati per la sperimentazione qualitativa legata all'articolo "Gender differences and didactic contract: analysis of two INVALSI tasks on power properties" (Giberti et al., 2016).

L'Appendice 2 è relativa alla sperimentazione Variazioni 1 e contiene i due questionari su cui la sperimentazione si è basata.

Infine, in Appendice 3 sono riportati i quattro fascicoli costruiti per il progetto Variazioni 2.

8.1 Appendice 1

Questionari somministrati prima delle interviste per indagare le risposte ai quesiti analizzati nell'articolo "Gender differences and didactic contract: analysis of two INVALSI tasks on power properties" (Giberti et al., 2016).

8.2 Appendice 2

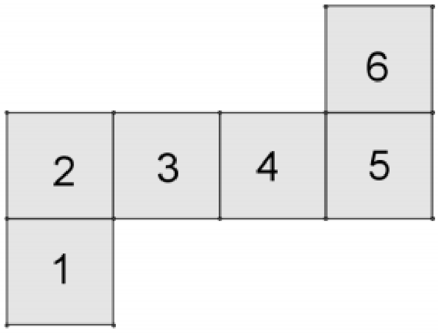
Fascicoli costruiti per il progetto Variazioni 1 (paragrafo 5.2).

8.3 Appendice 3

Fascicoli costruiti per il progetto Variazioni 1 (paragrafo 5.3).

Classe: .... Num elenco: .... Nome: ....

1 **La seguente figura rappresenta uno sviluppo piano di un cubo.**



Quale tra le seguenti coppie è formata da facce opposte del cubo?

A. ☐ 1 e 4

B. ☐ 2 e 5

C. ☐ 3 e 5

D. ☐ 4 e 6

---

2 **Nell'insieme dei numeri reali, la disequazione $x^2 > 0$ è verificata**

A. ☐ per ogni $x \neq 0$

B. ☐ per ogni x

C. ☐ solo per ogni $x < 0$

D. ☐ solo per ogni $x > 0$

---

3 **L'espressione $10^{37} + 10^{38}$ è anche uguale a**

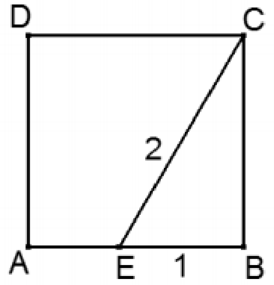
☐ A. 20^{75}

☐ B. 10^{75}

☐ C. $11 \cdot 10^{37}$

☐ D. $10^{37 \cdot 38}$

4 ABCD è un quadrato, il segmento EC è lungo 2 dm e il segmento EB è lungo 1 dm.



La superficie del quadrato ABCD misura

A. ☐ 3 dm^2

B. ☐ 4 dm^2

C. ☐ 5 dm^2

D. ☐ $4\sqrt{3}$ dm^2

---

5 Nelle ultime elezioni svoltesi in un paese europeo è andato a votare il 70% degli aventi diritto al voto. Di questi il 20% ha votato per il partito *A.* Quale percentuale di aventi diritto al voto ha votato per il partito *A*?

A. ☐ 60%

B. ☐ 50%

C. ☐ 20%

D. ☐ 14%

Prova di Matematica

Scuola___________________

Classe___________________

Studente_______________

Data___________________

ISTRUZIONI

Troverai nel fascicolo 30 domande di matematica. La maggior parte delle domande ha quattro possibili risposte, ma una sola è quella giusta. Prima di ogni risposta c'è un quadratino con una lettera dell'alfabeto: A, B, C, D.

Per rispondere, devi mettere una crocetta nel quadratino accanto alla risposta (una sola) che ritieni giusta, come nell'esempio seguente.

Esempio 1

Quanti giorni ci sono in una settimana?

A. ☒ Sette

B. ☐ Sei

C. ☐ Cinque

D. ☐ Quattro

Se ti accorgi di aver sbagliato, puoi correggere: devi scrivere **NO** accanto alla risposta sbagliata e mettere una crocetta nel quadratino accanto alla risposta che ritieni giusta, come nell'esempio seguente.

Esempio 2

Quanti minuti ci sono in un'ora?

NO A. ☒ 30 minuti

B. ☐ 50 minuti

C. ☒ 60 minuti

D. ☐ 100 minuti

In alcuni casi le domande chiedono di scrivere la risposta e/o il procedimento, oppure prevedono una diversa modalità di risposta. In questo caso il testo della domanda ti dice come rispondere. Leggilo dunque sempre con molta attenzione.

Puoi usare il righello graduato e/o la squadra, il compasso e il goniometro ma non la calcolatrice.

Non scrivere con la matita, ma usa soltanto una penna nera o blu.

Ricordati che puoi disegnare o scrivere sulle figure e puoi usare gli spazi bianchi del fascicolo per fare calcoli, se ti serve.

Per fare una prova, ora rispondi a questa domanda.

In quale delle seguenti sequenze i numeri sono scritti dal più grande al più piccolo?

A. ☐ 2; 5; 4; 8

B. ☐ 8; 5; 4; 2

C. ☐ 2; 4; 8; 5

D. ☐ 2; 4; 5; 8

Hai a disposizione un'ora e quindici minuti (in totale 75 minuti) per rispondere alle domande. L'insegnante ti dirà quando cominciare a lavorare. Quando l'insegnante ti comunicherà che il tempo è finito, posa la penna e chiudi il fascicolo.

Se finisci prima, puoi chiudere il fascicolo e aspettare la fine, oppure puoi controllare le risposte che hai dato.

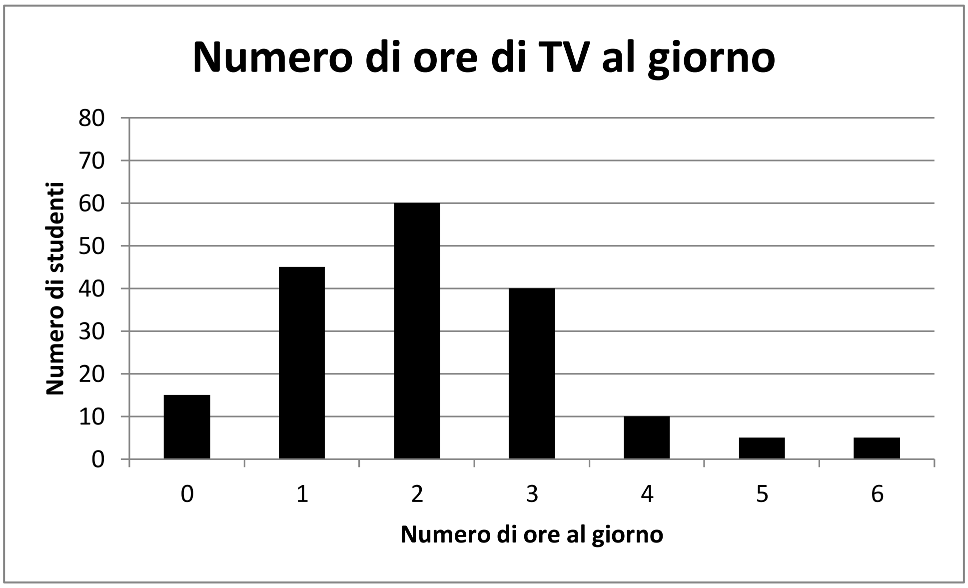
NON GIRARE LA PAGINA FINCHÉ NON TI SARÀ DETTO DI FARLO!

D1. **Eleonora ha condotto un'indagine sul numero di ore al giorno in cui gli studenti di I media della sua scuola guardano la TV.**
Ha riportato i dati nella seguente tabella:

Numero di ore al giorno	0	1	2	3	4	5	6
Numero di studenti	20	45	75	60	10	5	5

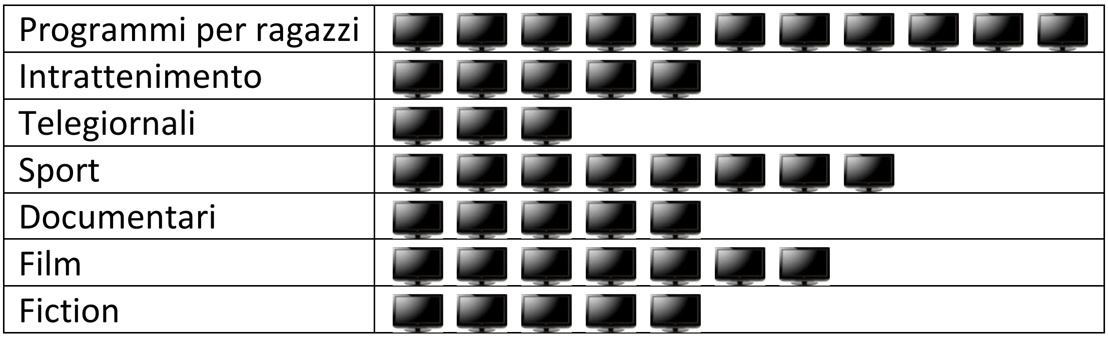
Successivamente, ha costruito con i dati della tabella il seguente grafico, ma ha commesso alcuni errori.

a. **Correggi tu il grafico, modificando le colonne che Eleonora ha sbagliato a disegnare.**



CONTINUA ALLA PAGINA A FIANCO

b. **Eleonora ha poi svolto un'altra indagine sui programmi TV preferiti dagli studenti di I media della sua scuola e ha riportato i dati nel seguente ideogramma.**



 = 5 bambini

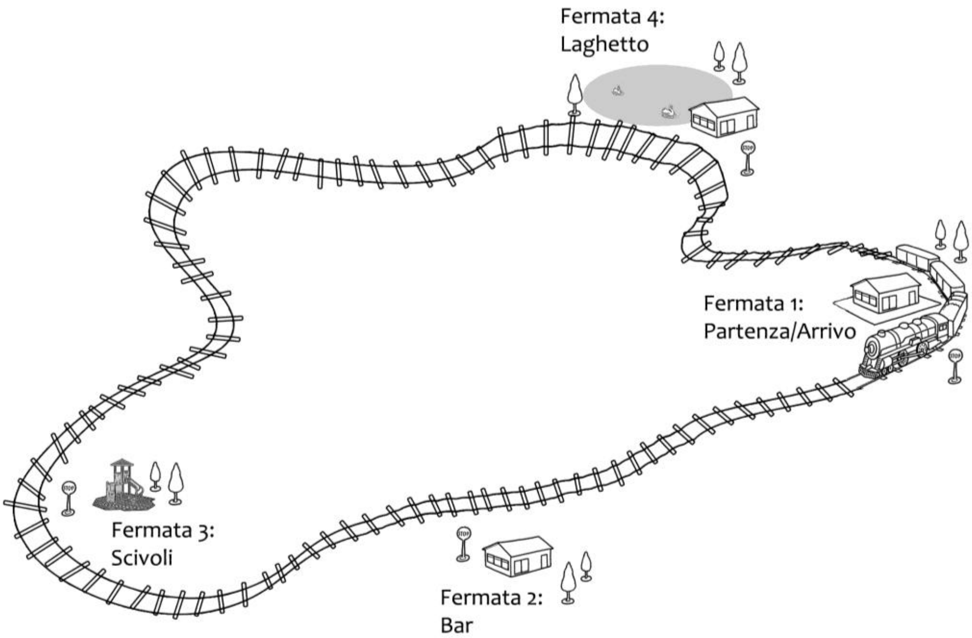
Usando i dati dell'ideogramma, compila tu la seguente tabella. Alcune caselle sono già state riempite.

Tipo di programma	Programmi per ragazzi	Intratteni-mento	……..	……..	……..	Film	……..
Numero di studenti	……..	……..	15	……..	……..	……..	……..

c. **Rispondi ora alle seguenti domande.**

		Sì	**No**
1.	Si può calcolare la media aritmetica del numero di ore al giorno in cui gli studenti guardano la TV?	☐	☐
2.	Si può calcolare la media aritmetica dei programmi preferiti dagli studenti?	☐	☐

D2. All'interno di un parco giochi ci si può spostare con un trenino che fa il seguente percorso:



Dalle 10:00 in poi, ogni mezz'ora, dalla fermata 1 parte una corsa del trenino. Il trenino impiega 5 minuti per andare da una fermata alla successiva, con l'eccezione del tratto tra la terza e la quarta, dove impiega 10 minuti.

a. Dove si trova il trenino alle 10:45?

A. ☐ Tra la seconda e la terza fermata

B. ☐ Alla terza fermata

C. ☐ Tra la terza e la quarta fermata

D. ☐ Alla quarta fermata

b. Quanti giri ha completato il trenino alle 12:00?

A. ☐ 5

B. ☐ 4

C. ☐ 3

D. ☐ 2

CONTINUA NELLA PAGINA A FIANCO

c. **Se il parco giochi chiude alle 18:00, quanti giri in totale fa il trenino in un giorno? Scrivi come fai per trovare la risposta e poi riporta sotto il risultato.**

……..

……..

………

Risultato: ……… giri

D3. Quante cifre ha il risultato della seguente moltiplicazione?

$$1\,001 \cdot 20\,002$$

Risposta: …… cifre

D4. Marta e il nonno camminano insieme lungo un sentiero. Ogni 2 passi fatti dal nonno, Marta ne fa 3 per restargli al fianco. Quando il nonno ha fatto 40 passi, quanti passi ha fatto Marta?

A. ☐ 80

B. ☐ 60

C. ☐ 40

D. ☐ 20

D5. **Osserva la figura 1.**

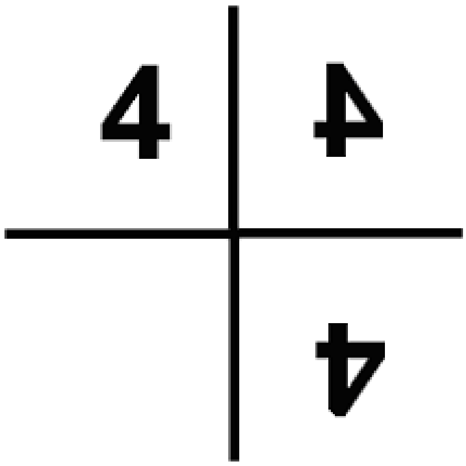


Figura 1

Osserva ora la figura 2 dove il 4 è stato sostituito con il 5.

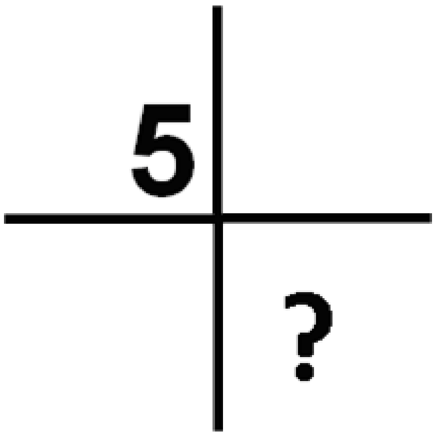


Figura 2

Che cosa ci sarà al posto del punto interrogativo?

A. ☐ 

B. ☐ 

C. ☐ **5**

D. ☐ 

D6. **Nel gioco della "morra cinese" i due giocatori devono mostrare contemporaneamente uno dei seguenti simboli con la mano:**

		
Forbice	Carta	Sasso

Le diverse combinazioni che si possono formare sono mostrate nella seguente tabella.

			
	Carta Carta	Carta Sasso	Carta Forbice
	Sasso Carta	Sasso Sasso	Sasso Forbice
	Forbice Carta	Forbice Sasso	Forbice Forbice

Le regole del gioco sono le seguenti:
Ogni segno ne batte un altro, secondo questo schema:

1. **Il sasso spezza le forbici (vince il sasso)**
2. **Le forbici tagliano la carta (vincono le forbici)**
3. **La carta avvolge il sasso (vince la carta)**

a. **Cerchia sulla tabella le combinazioni in cui vincono le forbici.**

b. **Considera l'insieme di tutte le combinazioni: le coppie formate da "carta" e "sasso" rappresentano**

A. ☐ $\frac{1}{9}$ di tutte le combinazioni

B. ☐ $\frac{2}{9}$ di tutte le combinazioni

C. ☐ $\frac{1}{3}$ di tutte le combinazioni

D. ☐ $\frac{2}{3}$ di tutte le combinazioni

CONTINUA NELLA PAGINA SUCCESSIVA

c. Cristina sostiene che la probabilità che escano due simboli uguali è minore della probabilità che escano due simboli diversi. Sei d'accordo con Cristina? Scegli una delle possibili risposte e completa la frase.

☐ **Sì, sono d'accordo con Cristina perché** ……………………………………………

……………………………………………………………………………………………………

☐ **No, non sono d'accordo con Cristina perché** ………………………………

……………………………………………………………………………………………………

D7. Nina è alla stazione ferroviaria di Napoli e deve andare a Roma. A causa del maltempo, molti treni sono in ritardo. Ecco cosa si legge sul tabellone elettronico delle partenze:

DESTINAZIONE	orario	ritardo	binario
Roma Termini	8:23	60 min	4
Bari centrale	8:32	35 min	3
Roma Termini	8:47	25 min	2
Reggio Calabria	8:49	10 min	1
Salerno	8:51		5
Roma Termini	8:53	15 min	7
Roma Termini	9:23		6

a. Nina decide di prendere il treno per Roma che partirà per primo. Da quale binario partirà Nina?

A. ☐ Dal binario 4

B. ☐ Dal binario 7

C. ☐ Dal binario 2

D. ☐ Dal binario 6

CONTINUA NELLA PAGINA A FIANCO

b. Il treno su cui Nina sale è composto dalla locomotiva e da 9 vagoni.

Quanto è lungo all'incirca il treno di Nina?

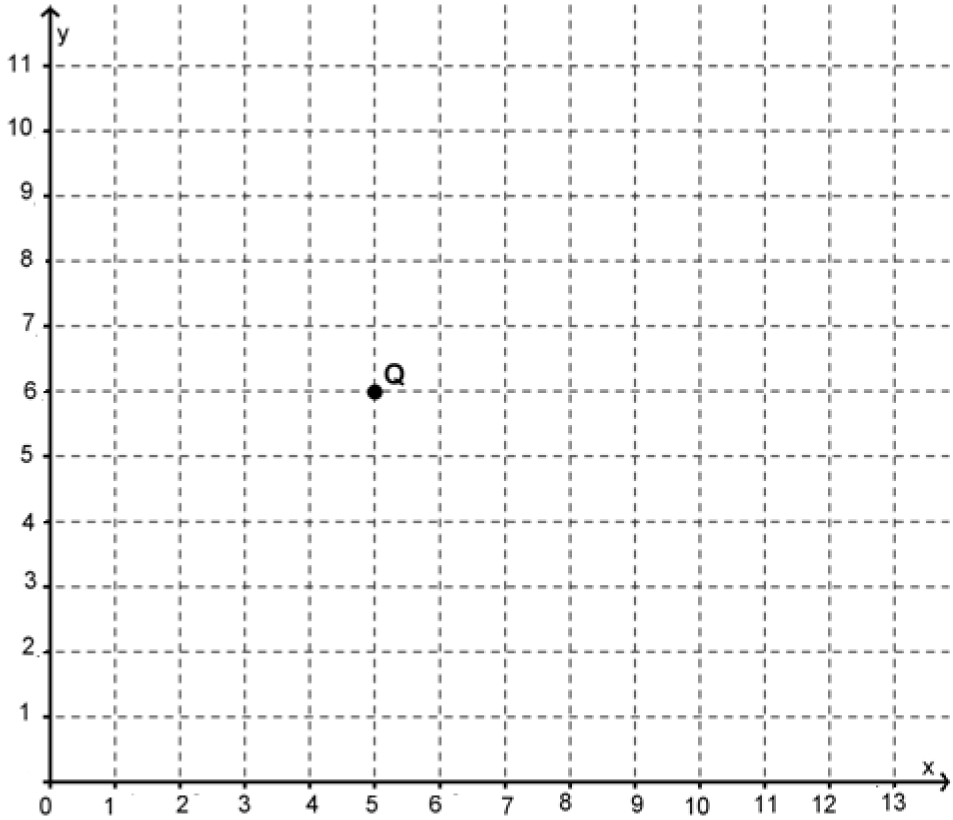
A. ☐ Circa 10 m

B. ☐ Circa 50 m

C. ☐ Circa 250 m

D. ☐ Circa 1000 m

D8. Nel piano cartesiano che vedi qui sotto è rappresentato il punto Q.



a. Scrivi le coordinate del punto Q.

Risposta:…………………………..

b. Partendo da Q, spostati di 4 unità verso sinistra e di 3 unità verso il basso. Quali sono le coordinate del punto dove arrivi?

A. ☐ (9; 3)

B. ☐ (4; 3)

C. ☐ (3; 1)

D. ☐ (1; 3)

D9. Mario va da casa a scuola con passo regolare e senza fermarsi. Fa 90 passi al minuto e conta in tutto 540 passi. La lunghezza del passo di Mario è 60 cm.

a. Quanto è lungo il percorso che Mario fa per andare da casa a scuola?

A. ☐ 324 m

B. ☐ 486 m

C. ☐ 3,24 km

D. ☐ 4,86 km

b. Quanto tempo impiega Mario per andare da casa a scuola?

Risposta: ………… minuti

c. Giulio, un compagno di classe di Mario, impiega 5 minuti per andare a piedi a scuola. Sulla base di questa informazione, si può sapere se Giulio abita più lontano o più vicino alla scuola rispetto a Mario? Scegli una delle due risposte e completa la frase.

☐ **Si può sapere perché** ……………………………………………………………….

……………………………………………………………………………………………….

……………………………………………………………………………………………….

☐ **Non si può sapere perché** ………………………………………………………

……………………………………………………………………………………………….

……………………………………………………………………………………………….

D10. All'ingresso del palazzo delle Mostre è esposto questo cartello con gli orari di apertura.

Mostra	**Lunedì**	**Martedì**	**Mercoledì**	**Giovedì**	**Venerdì**	**Sabato**	**Domenica**
Pittura			9-12		9-18	15-18	9-18
Scultura				9-12	15-18	9-18	15-18
Fotografia			9-18	9-18		9-12	9-18

a. In quali pomeriggi la mostra di Fotografia è chiusa?

A. ☐ Lunedì, Martedì e Venerdì

B. ☐ Mercoledì, Giovedì e Domenica

C. ☐ Venerdì e Sabato

D. ☐ Lunedì, Martedì, Venerdì e Sabato

b. Gianluca vuole visitare nella stessa mattina la mostra di Scultura e di Fotografia. In quali giorni della settimana potrà farlo?

A. ☐ Giovedì e Sabato

B. ☐ Mercoledì, Giovedì, Venerdì e Sabato

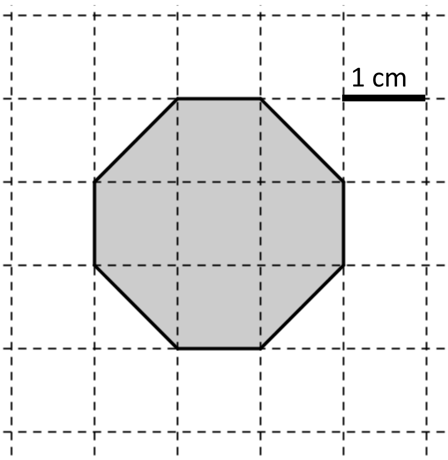
C. ☐ Mercoledì, Venerdì e Domenica

D. ☐ Mercoledì, Giovedì e Sabato

c. In quale giorno e in quale fascia oraria sono aperte contemporaneamente tutte e tre le mostre?

Giorno: ………………………… **Fascia oraria:** dalle ………. alle ……….

D11. Giulio dice che l'ottagono rappresentato in figura ha il perimetro di 8 cm.



Giulio ha ragione? Scegli una delle due risposte e completa la frase.

☐ **Giulio ha ragione perché** ....

....

....

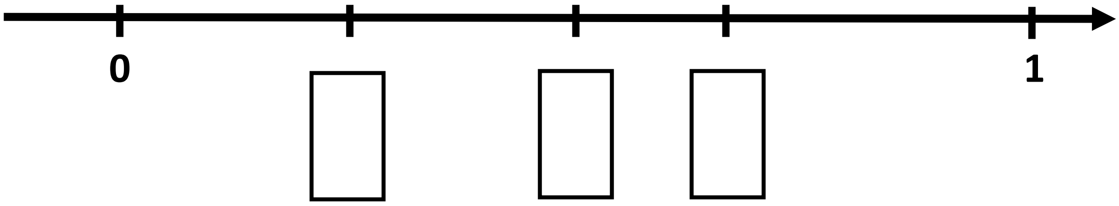
☐ **Giulio non ha ragione perché** ....

....

....

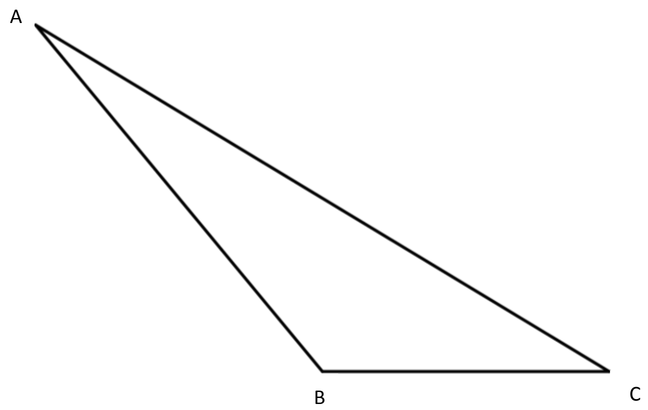
D12. Scrivi nei riquadri i seguenti numeri, posizionandoli correttamente sulla retta.

$$\frac{1}{2} \qquad \frac{1}{4} \qquad \frac{2}{3}$$

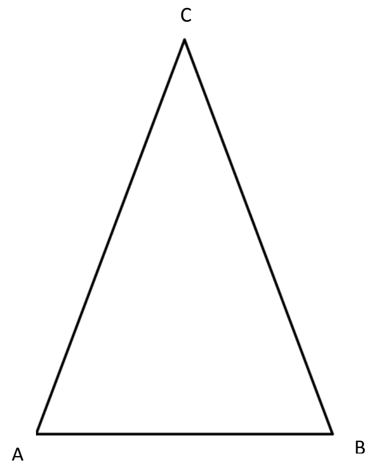


D13. Indica quale dei seguenti triangoli corrisponde a questa descrizione:

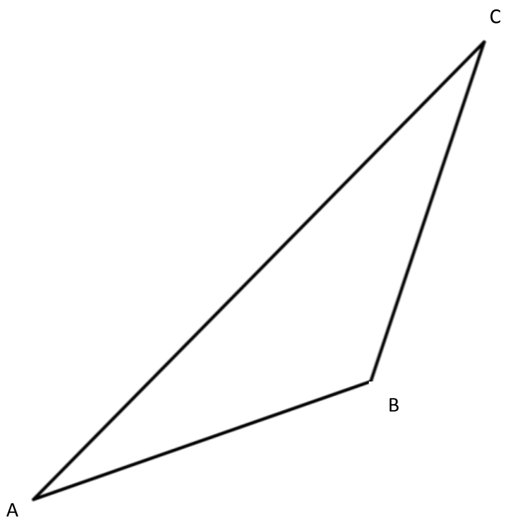
"ABC è un triangolo isoscele ottusangolo con angolo ottuso in B."



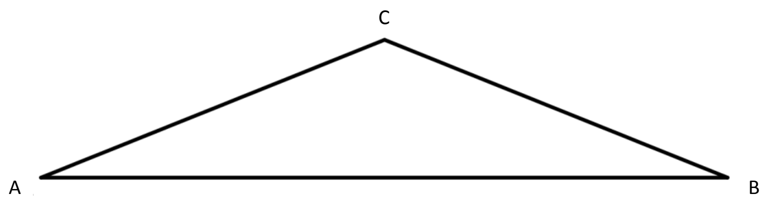
Triangolo 1



Triangolo 2



Triangolo 3



Triangolo 4

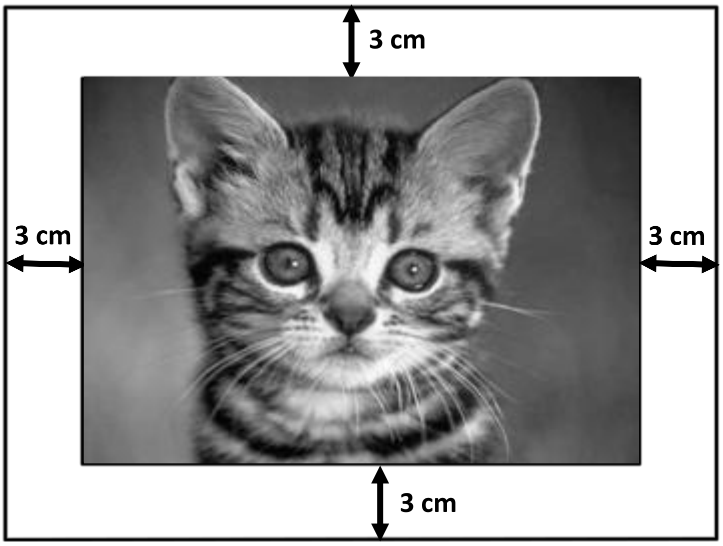
A. ☐ Triangolo 1

B. ☐ Triangolo 2

C. ☐ Triangolo 3

D. ☐ Triangolo 4

D14. **Franco incolla una fotografia rettangolare di dimensioni 22 cm x 15 cm su un cartoncino. Attorno alla fotografia resta una cornice larga 3 cm, come vedi in figura.**



Quali sono le dimensioni del cartoncino?

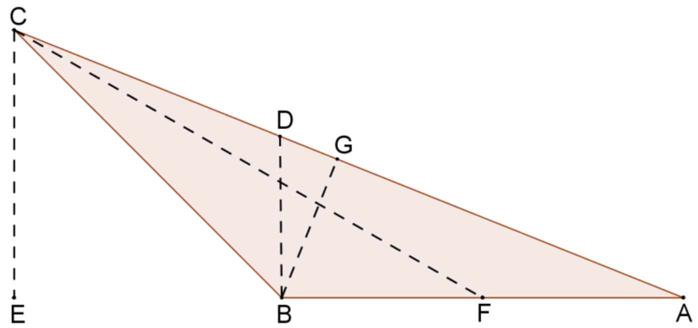
A. ☐ 28 cm x 21 cm

B. ☐ 25 cm x 21 cm

C. ☐ 28 cm x 18 cm

D. ☐ 25 cm x 18 cm

D15. Osserva la figura.



Quale, tra le seguenti coppie di segmenti, rappresenta due delle altezze del triangolo ABC?

A. ☐ CE e CF

B. ☐ BD e BG

C. ☐ CE e BG

D. ☐ CF e BD

D16. Una scatola di cioccolatini contiene 15 cioccolatini al latte e 25 cioccolatini fondenti. Con 100 cioccolatini al latte e 180 fondenti, qual è il numero massimo di scatole con la stessa composizione della precedente che si possono riempire?

A. ☐ 5

B. ☐ 6

C. ☐ 7

D. ☐ 8

D17. Gianni partecipa a un torneo.
Il regolamento del torneo stabilisce che:

- **ogni giocatore gioca 5 partite e parte con un punteggio iniziale di 100 punti;**
- **a ogni partita vinta, il punteggio raggiunto raddoppia;**
- **a ogni partita persa, il punteggio raggiunto si dimezza.**

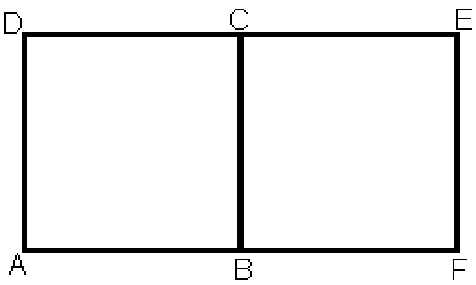
Gianni perde la seconda e la quarta partita, vince tutte le altre.

a. Completa la tabella.

	Punteggio di Gianni
Punteggio iniziale	100
Partita 1	200
Partita 2	……
Partita 3	……
Partita 4	……
Partita 5	……

b. Se Gianni avesse vinto tutte le partite, quale sarebbe stato il suo punteggio finale?
Risposta: …………………………….

D18. Il rettangolo AFED è formato da due quadrati congruenti ABCD e BFEC con un lato in comune.



Il perimetro di ciascuno dei quadrati misura 24 cm. Quanto misura il perimetro del rettangolo AFED?
Scrivi i calcoli che fai per trovare la risposta e poi riporta sotto il risultato.

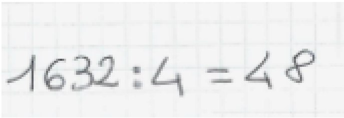
……

……

……

Risultato: ………… cm

D19. **Andrea svolge sul quaderno questa divisione:**



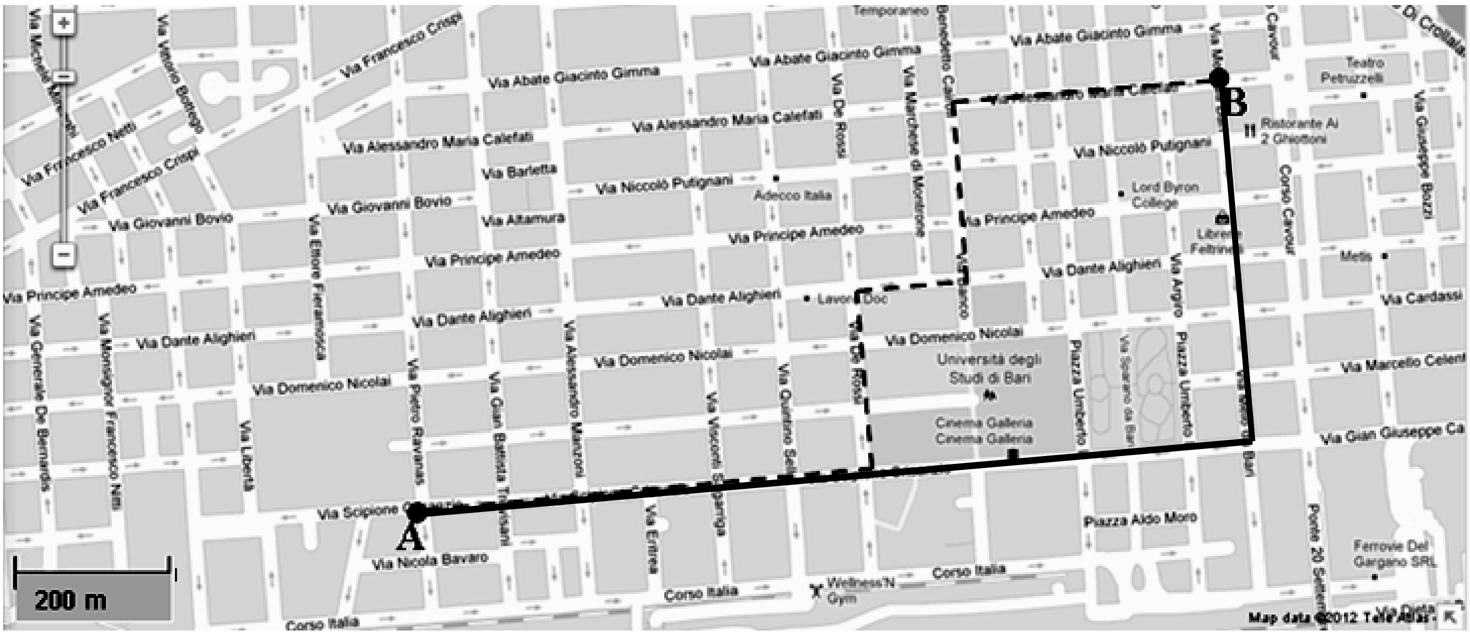
Il risultato ottenuto da Andrea è sbagliato. Quale errore può aver fatto?

Risposta: ……………………………………………………………………………………………………

……….

D20. **Osserva la mappa e utilizza la scala riportata in basso a sinistra per rispondere alle domande.**

a. **Quanto è lungo il percorso indicato dalla linea tratteggiata per andare da A a B?**



A. ☐ Circa 1,5 km

B. ☐ Circa 3 km

C. ☐ Circa 4,5 km

D. ☐ Circa 6 km

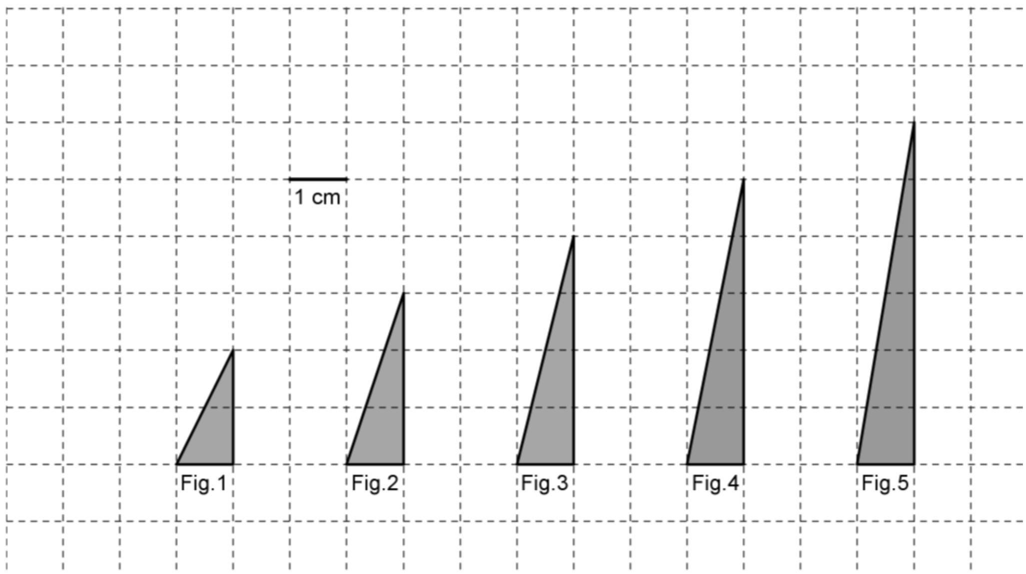
CONTINUA NELLA PAGINA A FIANCO

b. Completa la frase che segue, inserendo una delle seguenti parole:

maggiore / minore / uguale

Per andare da A a B, la lunghezza del percorso indicato dalla linea nera continua è ………………………… rispetto alla lunghezza del percorso indicato dalla linea tratteggiata.

D21. Osserva i seguenti triangoli.



a. Da un triangolo al successivo l'area del triangolo:

A. ☐ Raddoppia

B. ☐ Triplica

C. ☐ Aumenta di 1 cm^2

D. ☐ Aumenta di 0,50 cm^2

b. Se l'altezza dei triangoli continua ad aumentare di 1 cm da una figura alla successiva, quanti centimetri misurerà l'altezza del triangolo della figura 100?

A. ☐ 102

B. ☐ 101

C. ☐ 100

D. ☐ 99

D22. Quale dei seguenti numeri interi è più vicino al risultato di questa moltiplicazione?

4,82 x 9,95

A. ☐ 36

B. ☐ 42

C. ☐ 48

D. ☐ 50

D23. In quale dei seguenti gruppi i numeri sono disposti in ordine crescente?

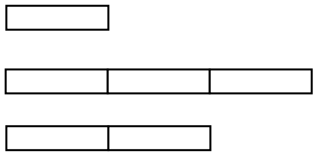
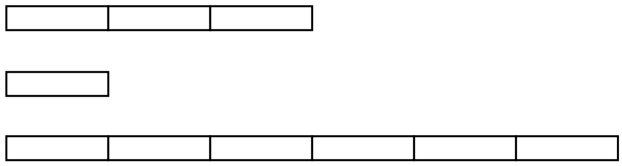
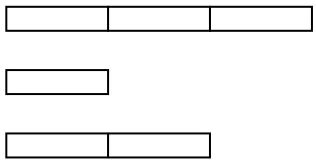
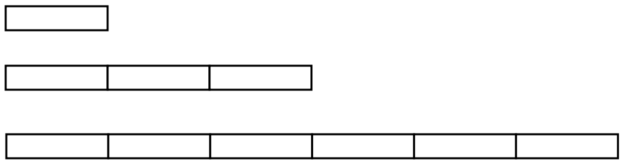
A. ☐ 3,5 ; 3,043 ; 3,28 ; 3,124

B. ☐ 3,5 ; 3,28 ; 3,124 ; 3,043

C. ☐ 3,043 ; 3,5 ; 3,124 ; 3,28

D. ☐ 3,043 ; 3,124 ; 3,28 ; 3,5

D24. Piero, Luigi e Giovanni sono fratelli. Piero ha il triplo degli anni di Luigi. Giovanni ha il doppio dell'età di Piero.
Indica qual è la rappresentazione grafica corretta della relazione tra gli anni di Piero, Luigi e Giovanni.

A. ☐	Età di Piero Età di Luigi Età di Giovanni	
B. ☐	Età di Piero Età di Luigi Età di Giovanni	
C. ☐	Età di Piero Età di Luigi Età di Giovanni	
D. ☐	Età di Piero Età di Luigi Età di Giovanni	

D25. Nella borraccia di Michele, piena per metà, ci sono 0,6 litri di acqua.

a. Michele beve la metà dell'acqua contenuta nella borraccia. Quanta acqua rimane?

A. ☐ 0,03 litri

B. ☐ 0,3 litri

C. ☐ $\frac{1}{2}$ litro

D. ☐ 1,2 litri

b. Michele riempie completamente la borraccia. Quanta acqua contiene ora?

Risposta: ……………………………… litri

D26. Alla fine di ogni mese, il numero degli iscritti al sito Internet <u>www.miseisimpatico.org</u> raddoppia rispetto al numero degli iscritti alla fine del mese precedente. Al termine del primo mese di attività gli iscritti sono 5.

a. Quale delle seguenti espressioni permette di calcolare il numero degli iscritti al termine del terzo mese?

A. ☐ $5 \cdot 5 \cdot 5$

B. ☐ $5 \cdot 2 \cdot 2$

C. ☐ $5 + 5 + 5$

D. ☐ $5 \cdot 2 \cdot 3$

b. Quando vengono superati i 100 iscritti?

A. ☐ Alla fine del terzo mese

B. ☐ Alla fine del quinto mese

C. ☐ Alla fine del sesto mese

D. ☐ Alla fine dell'ottavo mese

D27. Nello zaino di Chiara ci sono il libro di scienze, che pesa mezzo chilo, il libro di matematica, che pesa 980 g, e due quaderni uguali. Libri e quaderni pesano in tutto due chilogrammi. Quanto pesa ciascun quaderno?

A. ☐ 150 g

B. ☐ 260 g

C. ☐ 510 g

D. ☐ 520 g

D28. La seguente tabella mostra i risultati di un'inchiesta sugli animali domestici posseduti dagli abitanti di Gerlandia.

		CANE	
		SÌ	**NO**
GATTO	**SÌ**	85	109
	NO	42	96

Quante persone hanno il gatto, ma non il cane?

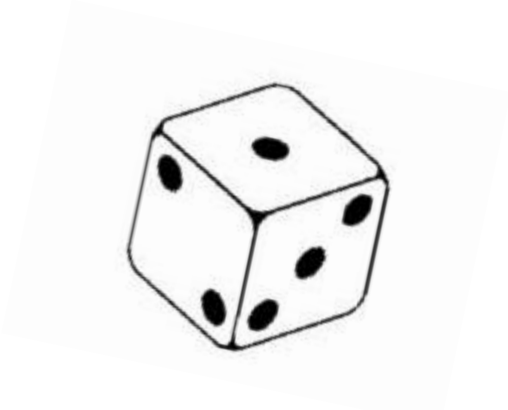
A. ☐ 42

B. ☐ 85

C. ☐ 96

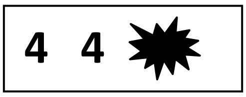
D. ☐ 109

D29. Marco lancia due volte un dado con le facce numerate da 1 a 6, come quello che vedi in figura.



La somma dei numeri usciti è 5. Quali numeri <u>non</u> possono essere usciti nel primo lancio?
Risposta: ……………………………………………………………

D30. Nel numero del riquadro la cifra finale è nascosta da una macchia.



Cerchia tutte le cifre che, messe al posto della macchia, rendono il numero divisibile per 3.

0 1 2 3 4 5 6 7 8 9

Prova di Matematica

Scuola__________________

Classe__________________

Studente______________

Data__________________

ISTRUZIONI

Troverai nel fascicolo 30 domande di matematica. La maggior parte delle domande ha quattro possibili risposte, ma una sola è quella giusta. Prima di ogni risposta c'è un quadratino con una lettera dell'alfabeto: A, B, C, D.

Per rispondere, devi mettere una crocetta nel quadratino accanto alla risposta (una sola) che ritieni giusta, come nell'esempio seguente.

Esempio 1

Quanti giorni ci sono in una settimana?

A. ☒ Sette

B. ☐ Sei

C. ☐ Cinque

D. ☐ Quattro

Se ti accorgi di aver sbagliato, puoi correggere: devi scrivere **NO** accanto alla risposta sbagliata e mettere una crocetta nel quadratino accanto alla risposta che ritieni giusta, come nell'esempio seguente.

Esempio 2

Quanti minuti ci sono in un'ora?

NO A. ☒ 30 minuti

B. ☐ 50 minuti

C. ☒ 60 minuti

D. ☐ 100 minuti

In alcuni casi le domande chiedono di scrivere la risposta e/o il procedimento, oppure prevedono una diversa modalità di risposta. In questo caso il testo della domanda ti dice come rispondere. Leggilo dunque sempre con molta attenzione.

Puoi usare il righello graduato e/o la squadra, il compasso e il goniometro ma non la calcolatrice.

Non scrivere con la matita, ma usa soltanto una penna nera o blu.

Ricordati che puoi disegnare o scrivere sulle figure e puoi usare gli spazi bianchi del fascicolo per fare calcoli, se ti serve.

Per fare una prova, ora rispondi a questa domanda.

In quale delle seguenti sequenze i numeri sono scritti dal più grande al più piccolo?

A. ☐ 2; 5; 4; 8

B. ☐ 8; 5; 4; 2

C. ☐ 2; 4; 8; 5

D. ☐ 2; 4; 5; 8

Hai a disposizione un'ora e quindici minuti (in totale 75 minuti) per rispondere alle domande. L'insegnante ti dirà quando cominciare a lavorare. Quando l'insegnante ti comunicherà che il tempo è finito, posa la penna e chiudi il fascicolo.

Se finisci prima, puoi chiudere il fascicolo e aspettare la fine, oppure puoi controllare le risposte che hai dato.

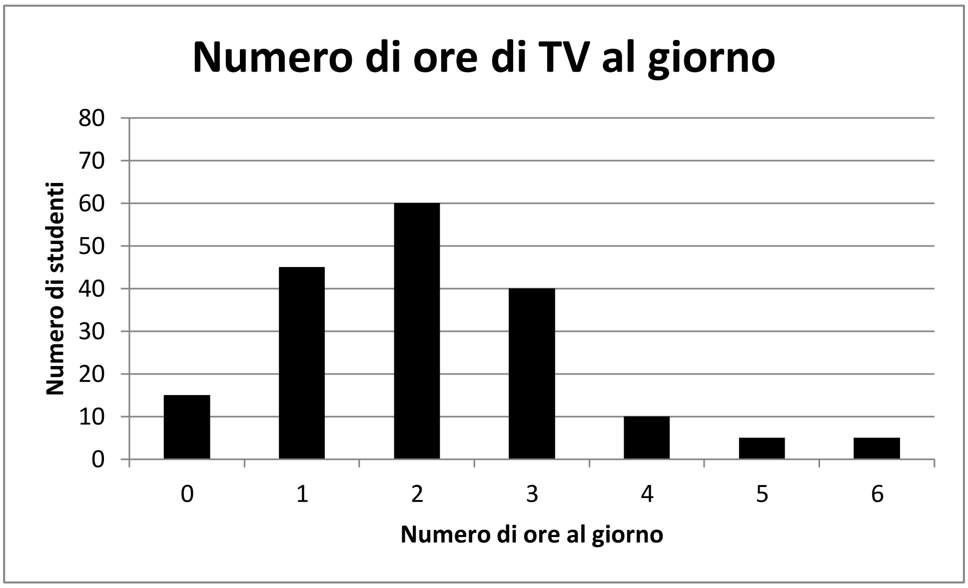
NON GIRARE LA PAGINA FINCHÉ NON TI SARÀ DETTO DI FARLO!

D1. **Eleonora ha condotto un'indagine sul numero di ore al giorno in cui gli studenti di I media della sua scuola guardano la TV.**
Ha riportato i dati nella seguente tabella:

Numero di ore al giorno	0	1	2	3	4	5	6
Numero di studenti	20	45	75	60	10	5	5

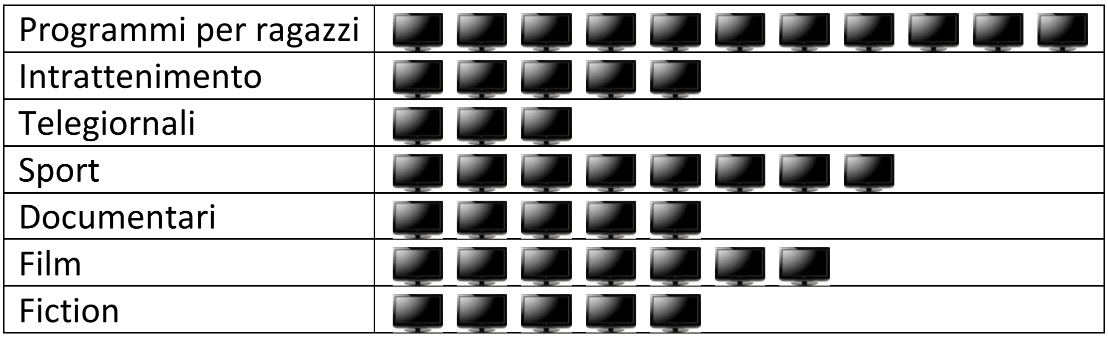
Successivamente, ha costruito con i dati della tabella il seguente grafico, ma ha commesso alcuni errori.

a. **Correggi tu il grafico, modificando le colonne che Eleonora ha sbagliato a disegnare.**



CONTINUA ALLA PAGINA A FIANCO

b. **Eleonora ha poi svolto un’altra indagine sui programmi TV preferiti dagli studenti di I media della sua scuola e ha riportato i dati nel seguente ideogramma.**



 = 5 bambini

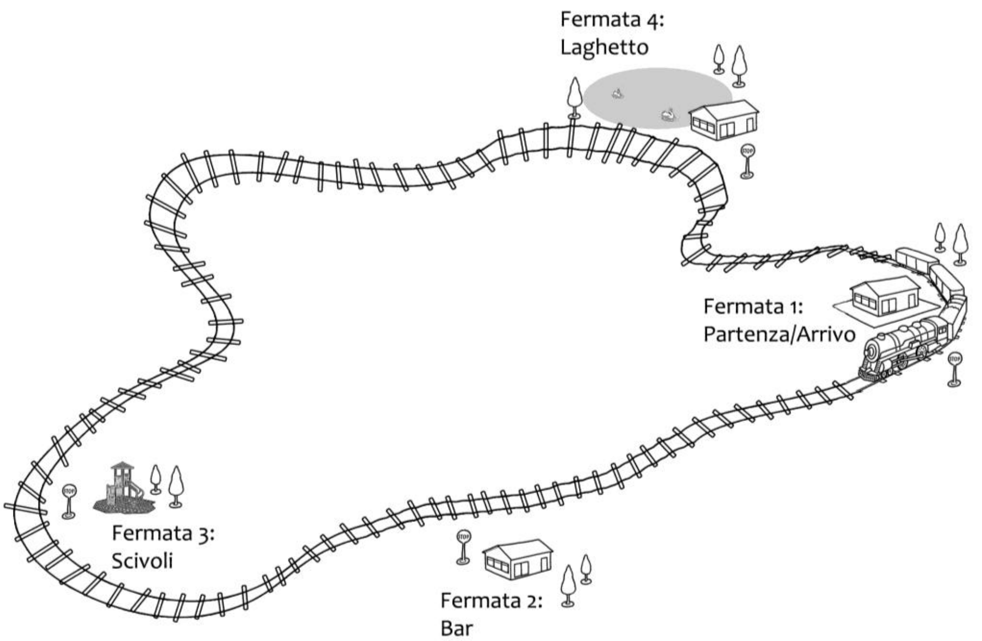
Usando i dati dell’ideogramma, compila tu la seguente tabella. Alcune caselle sono già state riempite.

Tipo di programma	Programmi per ragazzi	Intratteni-mento	……..	……..	……..	Film	……..
Numero di studenti	……..	……..	15	……..	……..	……..	……..

c. **Rispondi ora alle seguenti domande.**

		Sì	**No**
1.	Si può calcolare la media aritmetica del numero di ore al giorno in cui gli studenti guardano la TV?	☐	☐
2.	Si può calcolare la media aritmetica dei programmi preferiti dagli studenti?	☐	☐

D2. **All'interno di un parco giochi ci si può spostare con un trenino che fa il seguente percorso:**



Dalle 10:00 in poi, ogni mezz'ora, dalla fermata 1 parte una corsa del trenino. Il trenino impiega 5 minuti per andare da una fermata alla successiva, con l'eccezione del tratto tra la terza e la quarta, dove impiega 10 minuti.

a. **Dove si trova il trenino alle 10:45?**

A. ☐ Tra la seconda e la terza fermata

B. ☐ Alla terza fermata

C. ☐ Tra la terza e la quarta fermata

D. ☐ Alla quarta fermata

b. **Quanti giri ha completato il trenino alle 12:00?**

A. ☐ 5

B. ☐ 4

C. ☐ 3

D. ☐ 2

CONTINUA NELLA PAGINA A FIANCO

c. **Se il parco giochi chiude alle 18:00, quanti giri in totale fa il trenino in un giorno? Scrivi come fai per trovare la risposta e poi riporta sotto il risultato.**

………..

………..

………..

Risultato: ……… giri

D3. **Quante cifre ha il risultato della seguente moltiplicazione?**

$$1\,001 \cdot 20\,002$$

A. ☐ 8

B. ☐ 9

C. ☐ 5

D. ☐ 2

D4. **Marta e il nonno camminano insieme lungo un sentiero. Ogni 2 passi fatti dal nonno, Marta ne fa 3 per restargli al fianco. Quando il nonno ha fatto 40 passi, quanti passi ha fatto Marta?**

A. ☐ 80

B. ☐ 60

C. ☐ 40

D. ☐ 20

D5. Osserva la figura 1.

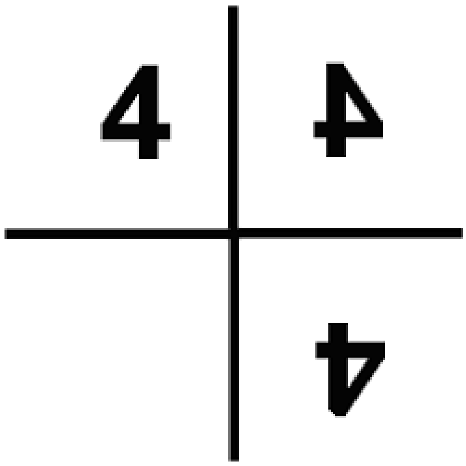


Figura 1

Osserva ora la figura 2 dove il 4 è stato sostituito con il 5.

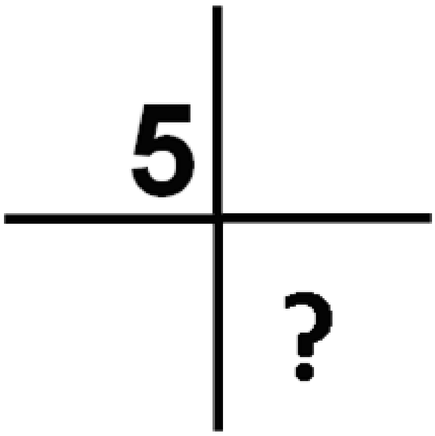


Figura 2

Che cosa ci sarà al posto del punto interrogativo?

A. ☐ 

B. ☐ 

C. ☐ **5**

D. ☐ 

D6. Nel gioco della "morra cinese" i due giocatori devono mostrare contemporaneamente uno dei seguenti simboli con la mano:

		
Forbice	Carta	Sasso

Le diverse combinazioni che si possono formare sono mostrate nella seguente tabella.

			
	Carta Carta	Carta Sasso	Carta Forbice
	Sasso Carta	Sasso Sasso	Sasso Forbice
	Forbice Carta	Forbice Sasso	Forbice Forbice

Le regole del gioco sono le seguenti:
Ogni segno ne batte un altro, secondo questo schema:

1. **Il sasso spezza le forbici (vince il sasso)**
2. **Le forbici tagliano la carta (vincono le forbici)**
3. **La carta avvolge il sasso (vince la carta)**

a. Cerchia sulla tabella le combinazioni in cui vincono le forbici.

b. Considera l'insieme di tutte le combinazioni: le coppie formate da "carta" e "sasso" rappresentano

A. ☐ $\frac{1}{9}$ di tutte le combinazioni

B. ☐ $\frac{2}{9}$ di tutte le combinazioni

C. ☐ $\frac{1}{3}$ di tutte le combinazioni

D. ☐ $\frac{2}{3}$ di tutte le combinazioni

CONTINUA NELLA PAGINA SUCCESSIVA

c. **Cristina sostiene che la probabilità che escano due simboli uguali è minore della probabilità che escano due simboli diversi. Sei d'accordo con Cristina? Scegli una delle possibili risposte e completa la frase.**

☐ **Sì, sono d'accordo con Cristina perché** ……………………………………….

……………………………………………………………………………………………….

☐ **No, non sono d'accordo con Cristina perché** …………………………….

……………………………………………………………………………………………….

D7. **Nina è alla stazione ferroviaria di Napoli e deve andare a Roma. A causa del maltempo, molti treni sono in ritardo. Ecco cosa si legge sul tabellone elettronico delle partenze:**

DESTINAZIONE	**orario**	**ritardo**	**binario**
Roma Termini	8:23	60 min	4
Bari centrale	8:32	35 min	3
Roma Termini	8:47	25 min	2
Reggio Calabria	8:49	10 min	1
Salerno	8:51		5
Roma Termini	8:53	15 min	7
Roma Termini	9:23		6

a. **Nina decide di prendere il treno per Roma che partirà per primo. Da quale binario partirà Nina?**

A. ☐ Dal binario 4

B. ☐ Dal binario 7

C. ☐ Dal binario 2

D. ☐ Dal binario 6

CONTINUA NELLA PAGINA A FIANCO

b. **Il treno su cui Nina sale è composto dalla locomotiva e da 9 vagoni.**

Quanto è lungo all’incirca il treno di Nina?

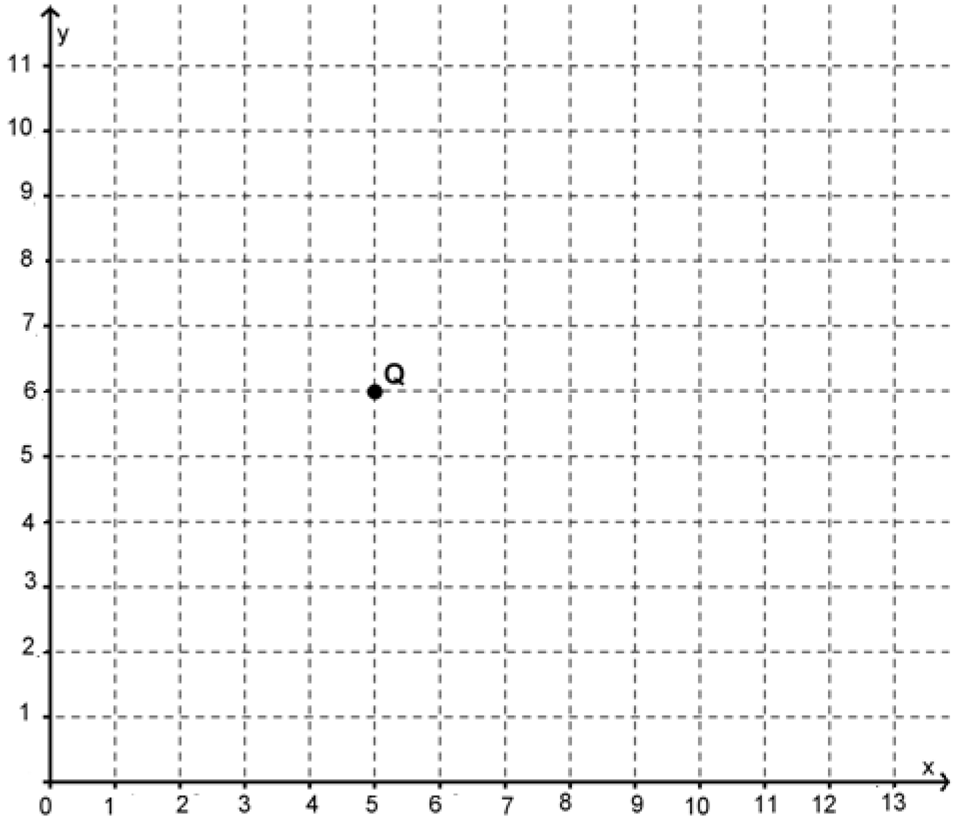
A. ☐ Circa 10 m

B. ☐ Circa 50 m

C. ☐ Circa 250 m

D. ☐ Circa 1000 m

D8. **Nel piano cartesiano che vedi qui sotto è rappresentato il punto Q.**



a. **Scrivi le coordinate del punto Q.**

Risposta:…………………………..

b. **Sei partito da un punto P, ti sei spostato di 4 unità verso destra e di 3 unità verso l'alto, e sei arrivato al punto Q. Quali sono le coordinate del punto P da cui sei partito?**

A. ☐ (9; 3)

B. ☐ (4; 3)

C. ☐ (3; 1)

D. ☐ (1; 3)

D9. Mario va da casa a scuola con passo regolare e senza fermarsi. Fa 90 passi al minuto e conta in tutto 540 passi. La lunghezza del passo di Mario è 60 cm.

a. Quanto è lungo il percorso che Mario fa per andare da casa a scuola?

A. ☐ 324 m

B. ☐ 486 m

C. ☐ 3,24 km

D. ☐ 4,86 km

b. Quanto tempo impiega Mario per andare da casa a scuola?

Risposta: ………… minuti

c. Giulio, un compagno di classe di Mario, impiega 5 minuti per andare a piedi a scuola. Sulla base di questa informazione, si può sapere se Giulio abita più lontano o più vicino alla scuola rispetto a Mario? Scegli una delle due risposte e completa la frase.

☐ **Si può sapere perché** ……………………………………………………………….

…………………………………………………………………………………………………..

…………………………………………………………………………………………………..

☐ **Non si può sapere perché** ……………………………………………………….

…………………………………………………………………………………………………..

…………………………………………………………………………………………………..

D10. All’ingresso del palazzo delle Mostre è esposto questo cartello con gli orari di apertura.

Mostra	**Lunedì**	**Martedì**	**Mercoledì**	**Giovedì**	**Venerdì**	**Sabato**	**Domenica**
Pittura			9-12		9-18	15-18	9-18
Scultura				9-12	15-18	9-18	15-18
Fotografia			9-18	9-18		9-12	9-18

a. In quali pomeriggi la mostra di Fotografia è chiusa?

A. ☐ Lunedì, Martedì e Venerdì

B. ☐ Mercoledì, Giovedì e Domenica

C. ☐ Venerdì e Sabato

D. ☐ Lunedì, Martedì, Venerdì e Sabato

b. Gianluca vuole visitare nella stessa mattina la mostra di Scultura e di Fotografia. In quali giorni della settimana potrà farlo?

A. ☐ Giovedì e Sabato

B. ☐ Mercoledì, Giovedì, Venerdì e Sabato

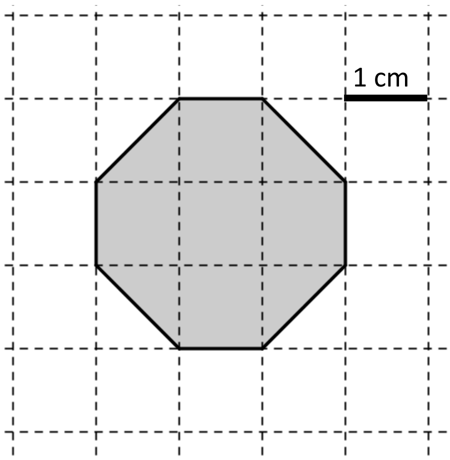
C. ☐ Mercoledì, Venerdì e Domenica

D. ☐ Mercoledì, Giovedì e Sabato

c. In quale giorno e in quale fascia oraria sono aperte contemporaneamente tutte e tre le mostre?

Giorno: ……………………… **Fascia oraria:** dalle ………. alle ……….

D11. Giulio dice che l'ottagono rappresentato in figura ha il perimetro di 8 cm.



Giulio ha ragione? Scegli una delle due risposte e completa la frase.

☐ **Giulio ha ragione perché** ………………………………………………………………….

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

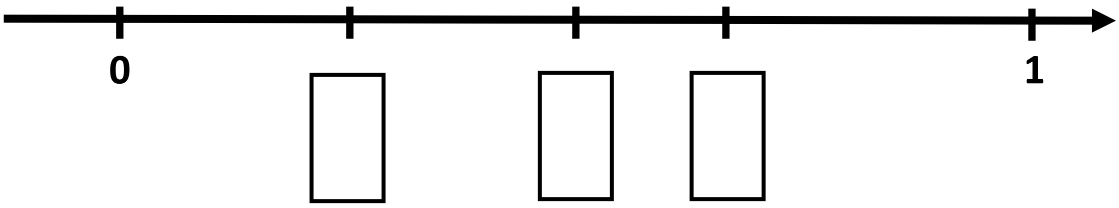
☐ **Giulio <u>non</u> ha ragione perché** …………………………………………………………..

…………………………………………………………………………………………………….

…………………………………………………………………………………………………….

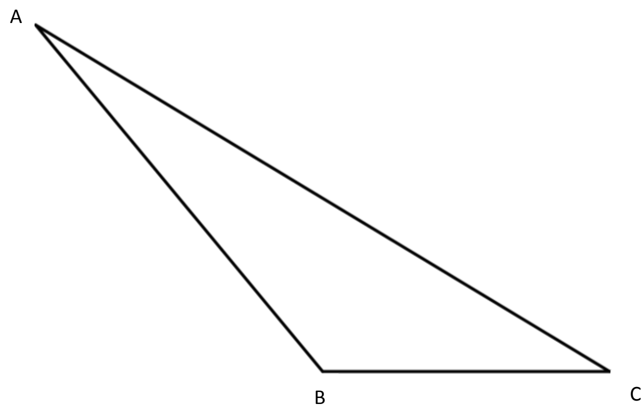
D12. Scrivi nei riquadri i seguenti numeri, posizionandoli correttamente sulla retta.

$$\frac{1}{2} \qquad \frac{1}{4} \qquad \frac{2}{3}$$

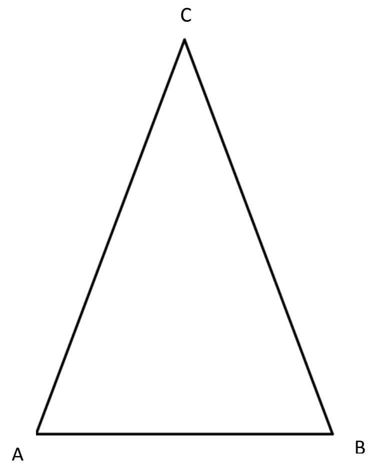


D13. Indica quale dei seguenti triangoli corrisponde a questa descrizione:

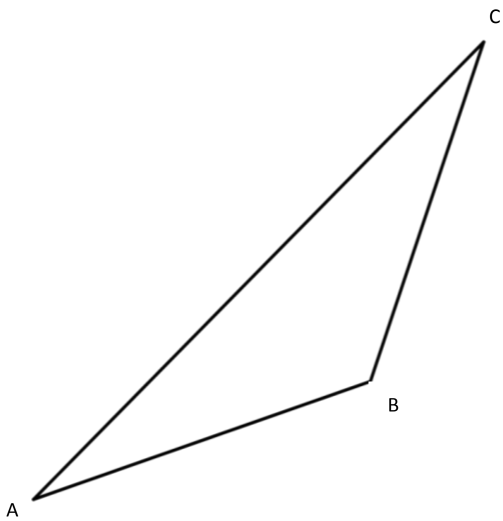
“ABC è un triangolo isoscele ottusangolo con angolo ottuso in B.”



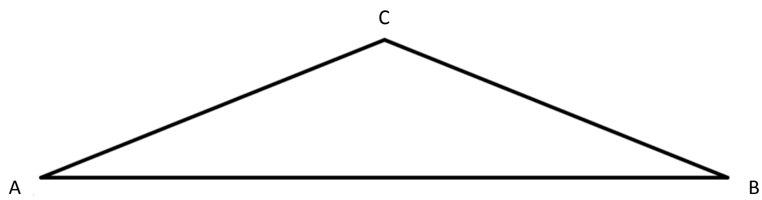
Triangolo 1



Triangolo 2



Triangolo 3



Triangolo 4

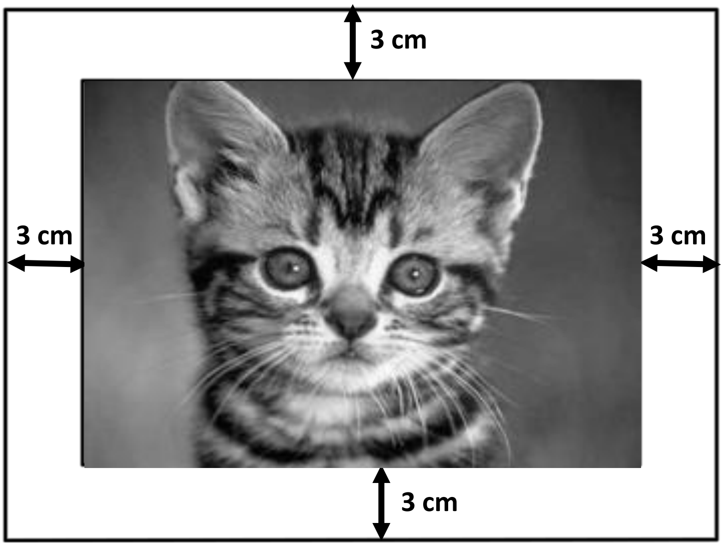
A. ☐ Triangolo 1

B. ☐ Triangolo 2

C. ☐ Triangolo 3

D. ☐ Triangolo 4

D14. **Franco incolla una fotografia rettangolare di dimensioni 22 cm x 15 cm su un cartoncino.**



Attorno alla fotografia resta una cornice larga 3 cm, come vedi in figura.
Quali sono le dimensioni del cartoncino?

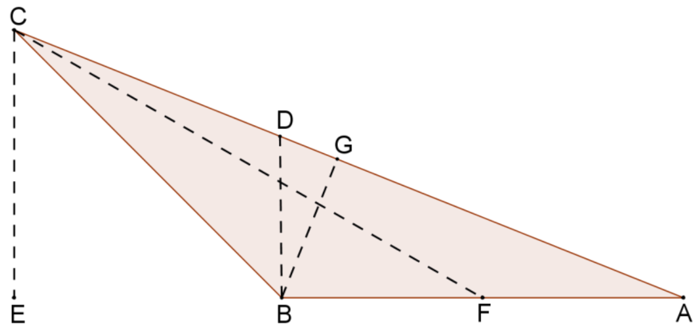
A. ☐ 28 cm x 21 cm

B. ☐ 25 cm x 21 cm

C. ☐ 28 cm x 18 cm

D. ☐ 25 cm x 18 cm

D15. Osserva la figura.



Quale, tra le seguenti coppie di segmenti, rappresenta due delle altezze del triangolo ABC?

A. ☐ CE e CF

B. ☐ BD e BG

C. ☐ CE e BG

D. ☐ CF e BD

D16. Qual è il numero massimo di scatole di cioccolatini che si possono riempire con 100 cioccolatini al latte e 180 fondenti, sapendo che ogni scatola deve contenere 15 cioccolatini al latte e 25 fondenti?

A. ☐ 5

B. ☐ 6

C. ☐ 7

D. ☐ 8

D17. Gianni partecipa a un torneo.
Il regolamento del torneo stabilisce che:

- **ogni giocatore gioca 5 partite e parte con un punteggio iniziale di 100 punti;**
- **a ogni partita vinta, il punteggio raggiunto raddoppia;**
- **a ogni partita persa, il punteggio raggiunto si dimezza.**

Gianni perde la seconda e la quarta partita, vince tutte le altre.

a. Completa la tabella.

	Punteggio di Gianni
Punteggio iniziale	100
Partita 1	200
Partita 2	……
Partita 3	……
Partita 4	……
Partita 5	……

b. Se Gianni avesse vinto tutte le partite, quale sarebbe stato il suo punteggio finale?
Risposta: ………………………….

D18. Il rettangolo AFED è formato da due quadrati congruenti ABCD e BFEC con un lato in comune.
Il perimetro di ciascuno dei quadrati misura 24 cm. Quanto misura il perimetro del rettangolo AFED?
Scrivi i calcoli che fai per trovare la risposta e poi riporta sotto il risultato.

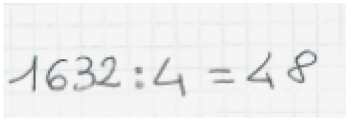
………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

Risultato: ………… cm

D19. Andrea svolge sul quaderno questa divisione:



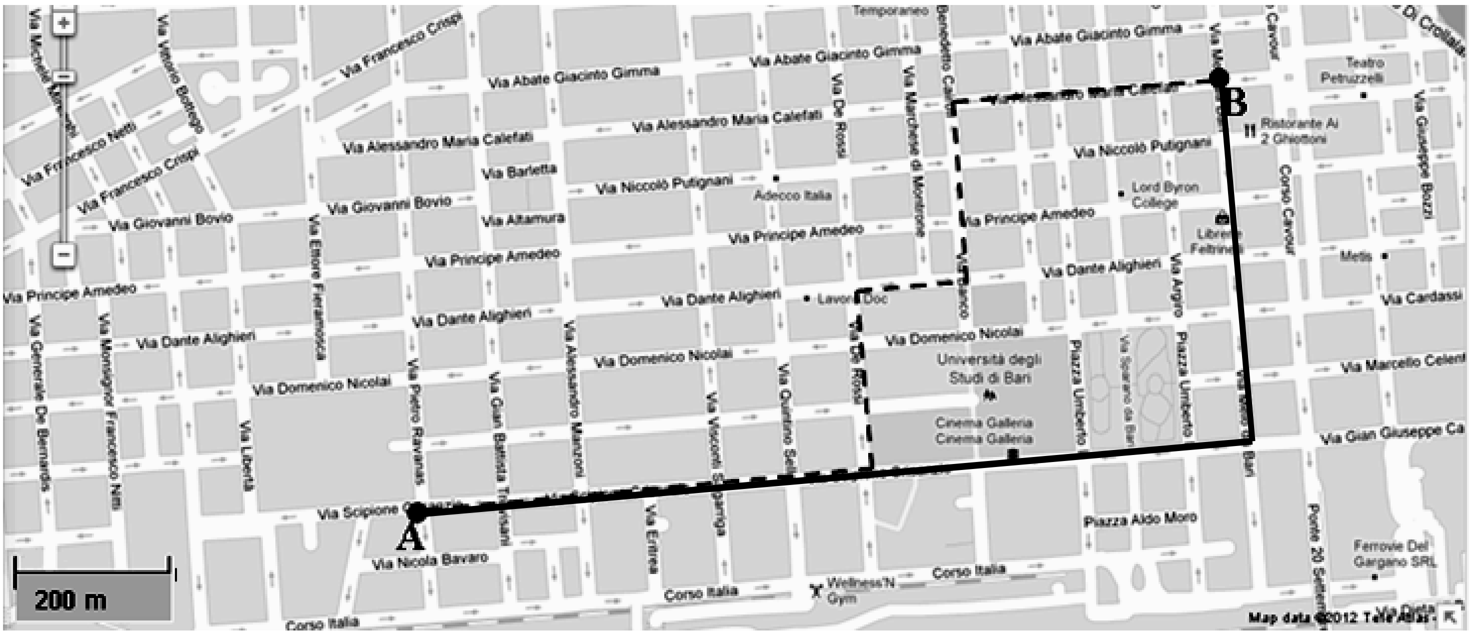
Il risultato ottenuto da Andrea è sbagliato. Quale errore può aver fatto?

Risposta: ……………………………………………………………………………………………………

……….

D20. Osserva la mappa e utilizza la scala riportata in basso a sinistra per rispondere alle domande.

a. Quanto è lungo il percorso indicato dalla linea tratteggiata per andare da A a B?



A. ☐ Circa 1,5 km

B. ☐ Circa 3 km

C. ☐ Circa 4,5 km

D. ☐ Circa 6 km

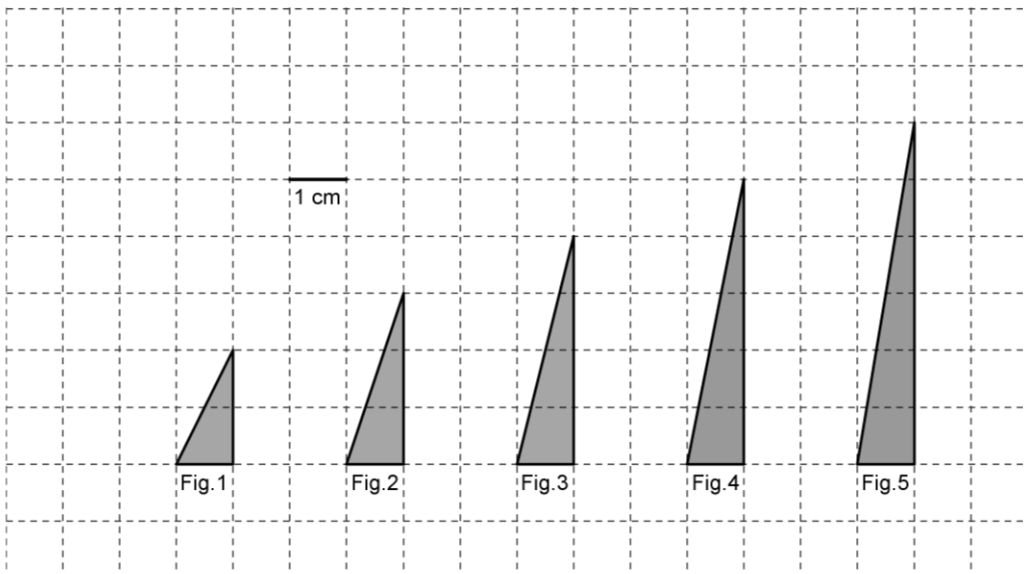
CONTINUA NELLA PAGINA A FIANCO

b. Completa la frase che segue, inserendo una delle seguenti parole:

maggiore / minore / uguale

Per andare da A a B, la lunghezza del percorso indicato dalla linea nera continua è ………………………….. rispetto alla lunghezza del percorso indicato dalla linea tratteggiata.

D21. Osserva i seguenti triangoli.



a. Da un triangolo al successivo l'area del triangolo:

A. ☐ Raddoppia

B. ☐ Triplica

C. ☐ Aumenta di 1 cm^2

D. ☐ Aumenta di 0,50 cm^2

b. Se l'altezza dei triangoli continua ad aumentare di 1 cm da una figura alla successiva, quanti centimetri misurerà l'altezza del triangolo della figura 100?

A. ☐ 102

B. ☐ 101

C. ☐ 100

D. ☐ 99

D22. Quale dei seguenti numeri è più vicino al risultato di questa moltiplicazione?

482 x 995

A. ☐ 360.000

B. ☐ 420.000

C. ☐ 480.000

D. ☐ 500.000

D23. In quale dei seguenti gruppi i numeri sono disposti in ordine crescente?

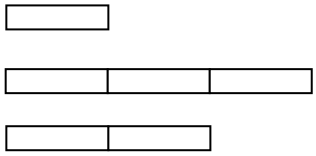
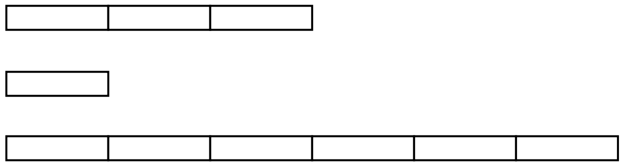
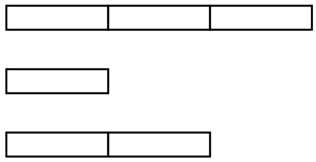
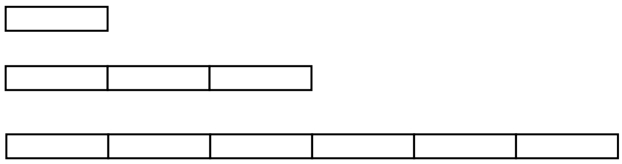
A. ☐ 3,5 ; 3,043 ; 3,28 ; 3,124

B. ☐ 3,5 ; 3,28 ; 3,124 ; 3,043

C. ☐ 3,043 ; 3,5 ; 3,124 ; 3,28

D. ☐ 3,043 ; 3,124 ; 3,28 ; 3,5

D24. Piero, Luigi e Giovanni sono fratelli. Piero ha il triplo degli anni di Luigi. Giovanni ha il doppio dell'età di Piero.
Indica qual è la rappresentazione grafica corretta della relazione tra gli anni di Piero, Luigi e Giovanni.

A. ☐	Età di Piero Età di Luigi Età di Giovanni	
B. ☐	Età di Piero Età di Luigi Età di Giovanni	
C. ☐	Età di Piero Età di Luigi Età di Giovanni	
D. ☐	Età di Piero Età di Luigi Età di Giovanni	

D25. Nella borraccia di Michele, piena per metà, ci sono 0,6 litri di acqua.

a. Michele beve la metà dell'acqua contenuta nella borraccia. Quanta acqua rimane?

A. ☐ 0,03 litri

B. ☐ 0,3 litri

C. ☐ $\frac{1}{2}$ litro

D. ☐ 1,2 litri

b. Michele riempie completamente la borraccia. Quanta acqua contiene ora?

Risposta: ……………………………… litri

D26. Alla fine di ogni mese, il numero degli iscritti al sito Internet www.miseisimpatico.org raddoppia rispetto al numero degli iscritti alla fine del mese precedente. Al termine del primo mese di attività gli iscritti sono 5.

a. Quale delle seguenti espressioni permette di calcolare il numero degli iscritti al termine del terzo mese?

A. ☐ $5 \cdot 5 \cdot 5$

B. ☐ $5 \cdot 2 \cdot 2$

C. ☐ $5 + 5 + 5$

D. ☐ $5 \cdot 2 \cdot 3$

b. Quando vengono superati i 100 iscritti?

A. ☐ Alla fine del terzo mese

B. ☐ Alla fine del quinto mese

C. ☐ Alla fine del sesto mese

D. ☐ Alla fine dell'ottavo mese

D27. Nello zaino di Chiara ci sono il libro di scienze, il libro di matematica e due quaderni uguali. Quanto pesa ciascun quaderno, sapendo che il libro di scienze pesa mezzo chilo, il libro di matematica pesa 980 g, e che libri e quaderni pesano in tutto due chilogrammi?

A. ☐ 150 g

B. ☐ 260 g

C. ☐ 510 g

D. ☐ 520 g

D28. La seguente tabella mostra i risultati di un'inchiesta sugli animali domestici posseduti dagli abitanti di Gerlandia.

		CANE	
		SÌ	**NO**
GATTO	**SÌ**	85	109
	NO	42	96

Quante persone hanno il gatto, ma non il cane?

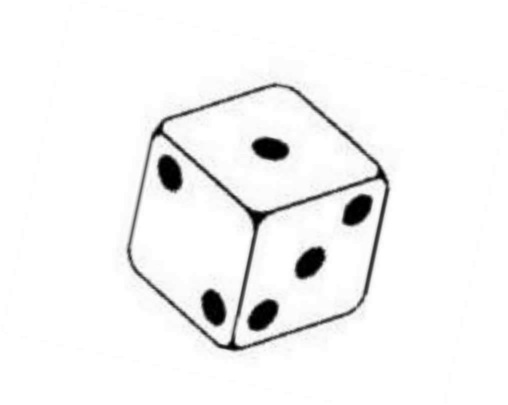
A. ☐ 42

B. ☐ 85

C. ☐ 96

D. ☐ 109

D29. Marco lancia due volte un dado con le facce numerate da 1 a 6, come quello che vedi in figura.



La somma dei numeri usciti è 5. Quali numeri <u>non</u> possono essere usciti nel primo lancio?
Risposta: ………………………………………………………………

D30. Nel numero del riquadro la cifra finale è nascosta da una macchia.

4 4

Cerchia tutte le cifre che, messe al posto della macchia, rendono il numero divisibile per 3.

0 1 2 3 4 5 6 7 8 9

Ministero dell'Istruzione dell'Università e della Ricerca

PROGETTO 'VARIAZIONI 2'

Anno Scolastico 2016 – 2017

LIVELLO 8

Prova di Matematica

Scuola Secondaria di primo grado

Classe Terza

Fascicolo 1

Classe: ....

Studente: ....

A cura di
Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione
INVALSI

Prova di Matematica - Fascicolo 1

ISTRUZIONI

Troverai nel fascicolo 30 domande di matematica. Alcune domande hanno quattro possibili risposte, ma una sola è quella giusta. Prima di ogni risposta c'è un quadratino con una lettera dell'alfabeto: A, B, C, D.

Per rispondere, devi mettere una crocetta nel quadratino accanto alla risposta (una sola) che ritieni giusta, come nell'esempio seguente.

Esempio 1

Quanti giorni ci sono in una settimana?

A. ☒ Sette

B. ☐ Sei

C. ☐ Cinque

D. ☐ Quattro

Se ti accorgi di aver sbagliato, puoi correggere: devi scrivere **NO** accanto alla risposta sbagliata e mettere una crocetta nel quadratino accanto alla risposta che ritieni giusta, come nell'esempio seguente.

Esempio 2

Quanti minuti ci sono in un'ora?

NO A. ☒ 30 minuti

B. ☐ 50 minuti

C. ☒ 60 minuti

D. ☐ 100 minuti

Altre domande chiedono di scrivere la risposta o il procedimento, oppure prevedono una diversa modalità di risposta. In questo caso il testo della domanda ti dice come rispondere. Leggilo dunque sempre con molta attenzione.

Puoi usare il righello graduato, la squadra, il compasso e il goniometro ma non la calcolatrice.

Non scrivere con la matita, ma usa soltanto una penna nera o blu.

Puoi usare le pagine bianche del fascicolo o gli spazi bianchi accanto alle domande per fare calcoli o disegni.

Per fare una prova, ora rispondi a questa domanda.

In quale delle seguenti sequenze i numeri sono scritti dal più grande al più piccolo?

A. ☐ 2; 5; 4; 8

B. ☐ 8; 5; 4; 2

C. ☐ 2; 4; 8; 5

D. ☐ 2; 4; 5; 8

Hai a disposizione 1 ora e quindici minuti (in totale 75 minuti) per rispondere alle domande. L'insegnante ti dirà quando cominciare a lavorare. Quando l'insegnante ti comunicherà che il tempo è finito, posa la penna e chiudi il fascicolo.

Se finisci prima, puoi chiudere il fascicolo e aspettare la fine, oppure puoi controllare le risposte che hai dato.

Caro studente,
ti chiediamo di rispondere alle seguenti domande.

Q1. Tu sei: maschio □ femmina □

Q2. Tu sei nato nel mese di ________________ dell'anno ________________

Q3. Hai frequentato la scuola materna?
   a. No
   b. Sì, per più di un anno
   c. Sì, per meno di un anno

Q4. Hai frequentato l'asilo nido?
   a. No
   b. Sì, per più di un anno
   c. Sì, per meno di un anno

Q5. Tu sei nato
   a. In Italia
   b. All'estero (scrivi sui puntini in quale Paese sei nato): ________________________________

Q6. Se non sei nato in Italia, quanti anni avevi quando sei arrivato in Italia? __________________ anni

Q7. In quale Paese è nata la tua mamma? __________________________

Q8. In quale Paese è nato il tuo papà? __________________________

Q9. Qual è il titolo di studio conseguito dai tuoi genitori?

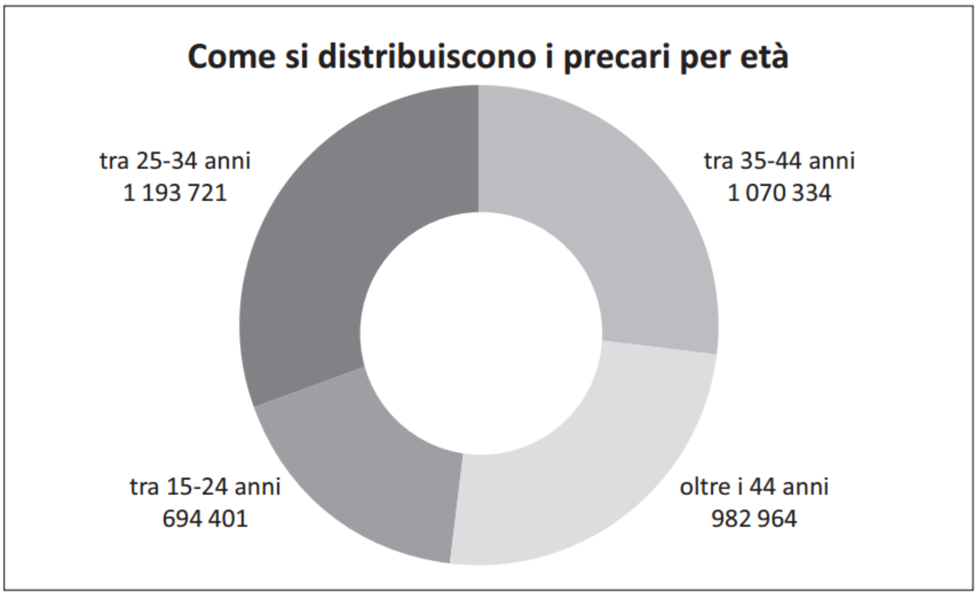
	madre	padre
A. Licenza elementare		
B. Licenza media		
C. Qualifica professionale triennale		
D. Diploma di scuola secondaria superiore (liceo, istituto tecnico o istituto professionale)		
E. Titolo di studio superiore al diploma, diverso dalla laurea (ISEF, Accademia di Belle Arti, Conservatorio)		
F. Laurea / Dottorato di ricerca / Master		
G. Non me lo ricordo		

Q10. Che lavoro fanno i tuoi genitori?

	madre	padre
A. Disoccupato/a		
B. Si occupa della casa		
C. Dirigente, docente universitario, funzionario, ufficiale militare		
D. Imprenditore, proprietario agricolo		
E. Professionista dipendente, sottufficiale militare, libero professionista (psicologo, ricercatore, medico, avvocato, commissario di polizia, ecc.)		
F. Lavoratore in proprio (commerciante, artigiano, coltivatore diretto, meccanico, sarto, ecc.)		
G. Insegnante, impiegato, militare graduato		
H. Operaio, addetto ai servizi, socio di cooperativa (tecnico, infermiere, cameriere, commessa, ecc.)		
I. Pensionato/a		
L. Non me lo ricordo		

NON GIRARE LA PAGINA
FINCHÉ NON TI SARÀ DETTO DI FARLO

A1. **Il seguente grafico rappresenta la distribuzione dei lavoratori precari in Italia suddivisi per età nell'anno 2012.**



a. Quanti sono in totale i precari?

A. ☐ **Circa due milioni**

B. ☐ **Circa tre milioni**

C. ☐ **Circa quattro milioni**

D. ☐ **Circa cinque milioni**

b. Quale percentuale rappresentano i precari che hanno tra i 25 e i 34 anni?

A. ☐ **Circa il 50%**

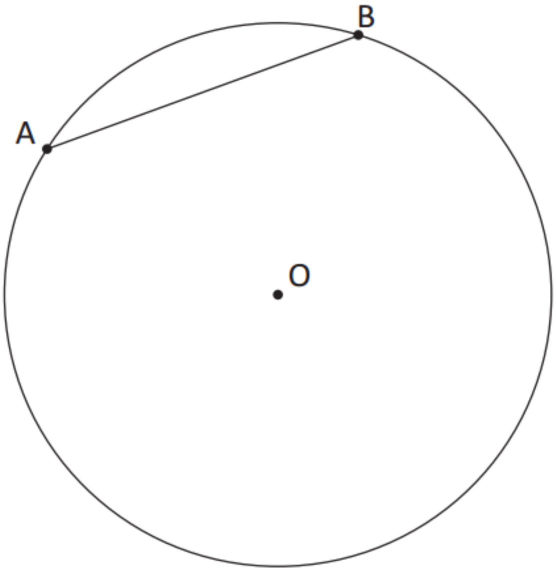
B. ☐ **Circa il 40%**

C. ☐ **Circa il 30%**

D. ☐ **Circa il 20%**

D18_PN2014

A2. Osserva la figura. AB è un cateto di un triangolo rettangolo inscritto nella circonferenza di centro O. Disegna il triangolo rettangolo.



D22_PN2013

A3. Se *n* é un numero naturale, allora il numero $n \cdot (n + 2)$

A. □ **è sempre dispari**

B. □ **è sempre pari**

C. □ **è dispari se n è pari**

D. □ **è dispari se n è dispari**

A4. **Tempo fa si è disputata la partita di pallacanestro B. Pozzo di Gotto - Brescia, finita con il punteggio di 92-94.**

La seguente tabella riassume le statistiche di tale partita per la squadra di Brescia.

		Tiri a canestro			
Nome del giocatore	**Minuti giocati**	**Tiri da 2**	**Tiri da 3**	**Tiri liberi**	**PUNTI**
Roberto	25	0	0	2	2
Clelia	23	4	0	1	9
Chiara	20	2	0	0	4
Giorgio	36	2	1	7	14
Cristina	37	3	1	1	10
Monica	30	9	1	8	29
Paolo	9	0	1	2	5
Anna	15	0	1	0	3
Maria	30	6	0	6	18
	Totale	**26**	**5**	**27**	**94**

Quanti sono i giocatori che hanno realizzato un numero di punti superiore alla media?

Risposta: ………………………………………………………………………………………

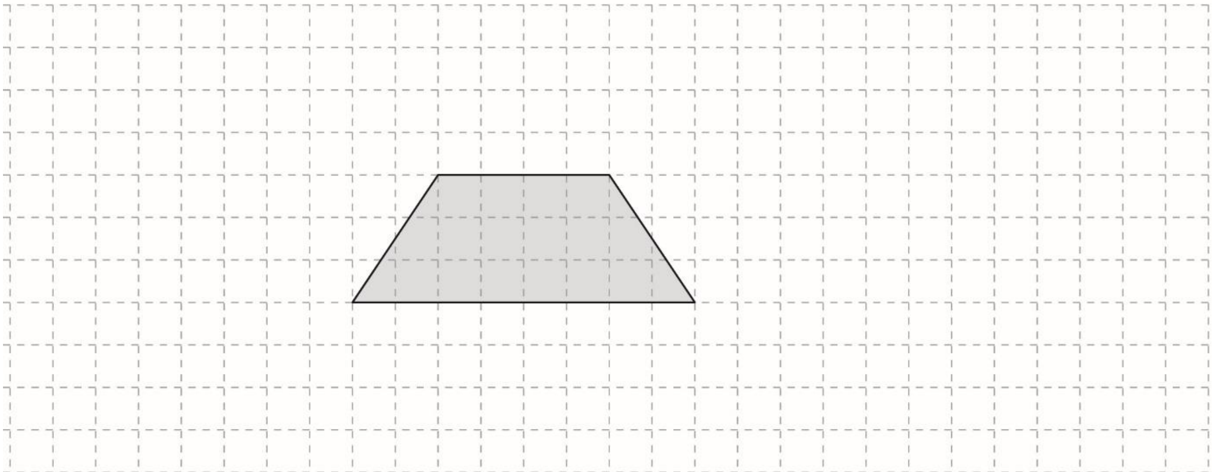
A5. **Il Signor Carlo scende dal tram all'incrocio di *via Pietro Micca* con *via Antonio Giuseppe Bertola* (nella mappa che vedi qui sotto il punto è contrassegnato da un asterisco).**

Percorre 200 metri di *via Bertola* e all'incrocio con *via 20 Settembre* svolta a sinistra; dopo aver camminato per 150 metri, raggiunge l'incrocio con *via Pietro Micca*. Da lì decide di tornare al punto di partenza per *via Pietro Micca*.
Quanti metri all'incirca percorre al ritorno?

A. □ **200 m**
B. □ **250 m**
C. □ **350 m**
D. □ **600 m**

A6. Il trapezio che vedi sotto è stato ritagliato da una figura F più grande. Il trapezio è i $\frac{3}{4}$ della figura F.

Disegna una delle possibili figure F da cui il trapezio è stato ritagliato.



D7_PN2011

Anch_1. Antonio e Giada partecipano a una gara a quiz. Per ogni risposta esatta si assegnano due punti mentre per ogni risposta sbagliata si toglie un punto. L'esito della gara è il seguente:

- **Antonio ha dato 11 risposte esatte e 9 sbagliate;**
- **Giada ha dato 6 risposte esatte e 14 sbagliate.**

Quali sono i punteggi finali dei due ragazzi?

A. □ **+ 13; +2**

B. □ **+ 13; −2**

C. □ **+ 2; + 8**

D. □ **+ 2; − 8**

D13_PN2011_v4

D1. Il numero RADICE QUADRATA DI 10 è:

A. □ **compreso tra 9 e 11**

B. □ **uguale a 5**

C. □ **compreso tra 3 e 4**

D. □ **uguale a 100**

D2. Un bicchiere contiene ¼ di litro d'acqua.

Se si vuole riempire una bottiglia da 1,5 litri, quanti bicchieri d'acqua bisogna versare nella bottiglia?

Risposta: ………………………………

D15_PN2012_originale

D3. Una grande azienda nel 2009 aveva 100 impiegati. Nell'anno 2010 il numero degli impiegati è diminuito del 20% rispetto al 2009 mentre nel 2011 è aumentato del 20% rispetto al 2010.
Al termine dei due anni gli impiegati dell'azienda sono

A. □ **diminuiti del 4%**

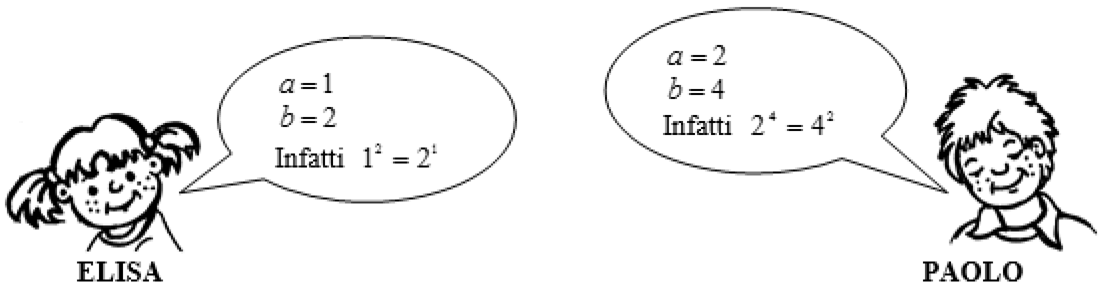
B. □ **diminuiti del 10%**

C. □ **aumentati del 4%**

D. □ **aumentati del 10%**

D18_PN2011

Anch_3. Elisa e Paolo stanno cercando di rispondere a questa domanda: "***Qual è la coppia di numeri interi a, b (diversi fra loro) tali che $a^b = b^a$?***"
Ecco le loro soluzioni.



Chi ha ragione?

A. □ **Solo Elisa**

B. □ **Solo Paolo**

C. □ **Entrambi**

D. □ **Nessuno dei due**

D5. Il rettangolo rappresenta, in scala 1:5, il piano rettangolare di un banco.

Quanti rettangoli uguali a quello disegnato servono come minimo per coprire interamente la superficie reale del piano del banco?

A. □ **25**

B. □ **20**

C. □ **10**

D. □ **5**

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?**

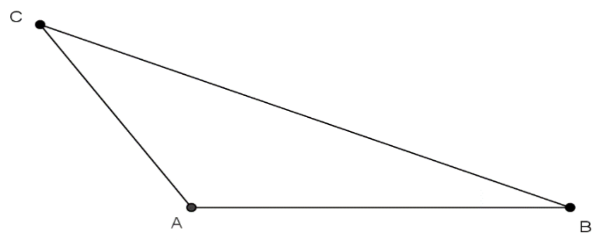
A. □ **30 cm**

B. □ **50 cm**

C. □ **60 cm**

D. □ **80 cm**

D7. Osserva il disegno.



Calcola l'area del triangolo prendendo con un righello le misure necessarie.

a. Risposta: ………….. **cm²**

b. Scrivi i calcoli che hai fatto per arrivare alla risposta.

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

D8. Nino sale su un treno composto dalla locomotiva e da 9 vagoni:

Quanto è lungo all'incirca il treno di Nino?

A. □ **Circa 10 m**

B. □ **Circa 50 m**

C. □ **Circa 250 m**

D. □ **Circa 1000 m**

D27_PN2013

Anch_7. Nella scuola Nino Bixio ci sono 600 studenti e un insegnante ogni 15 studenti.

Quale proporzione permette di trovare il numero x degli insegnanti?

A. □ **$x : 15 = 1 : 600$**

B. □ **$15 : 1 = x : 600$**

C. □ **$1 : 15 = x : 600$**

D. □ **$x : 1 = 15 : 600$**

1CG_NEW_v1

D9. Qual è il risultato di $4 \times 0,5$?

A. □ **8**

B. □ **4**

C. □ **2**

D. □ **20**

Anch_4. La formula $L = L_0 + K \times P$ esprime la lunghezza L di una molla al variare del peso P applicato. L_0 rappresenta la lunghezza in centimetri "a riposo" della molla; K indica di quanto si allunga in centimetri la molla quando le si applica una unità di peso.

Quale delle formule elencate si adatta meglio alla seguente descrizione:

"*È una molla molto corta e molto dura (cioè molto resistente alla trazione)*"?

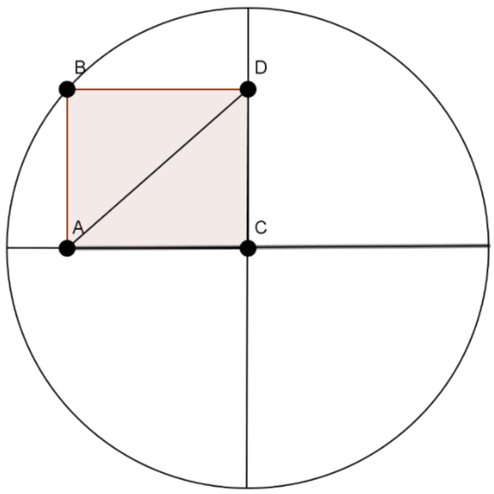
A. □ $L = 10 + 0{,}5 \times P$

B. □ $L = 10 + 7 \times P$

C. □ $L = 80 + 0{,}5 \times P$

D. □ $L = 80 + 7 \times P$

D10. Osserva la figura.



La circonferenza rappresentata in figura ha centro C e raggio 4 cm.

CABD è un rettangolo.

Mario afferma che il segmento AD misura 4 cm.

Mario ha ragione? Scegli una risposta con il suo perché.

Mario NON HA ragione perché		Mario HA ragione perché	
La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è diverso da 4	Non ci sono abbastanza dati per calcolare la lunghezza di AD	La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è 4	Le diagonali di un rettangolo sono congruenti tra di loro
□ A	□ B	□ C	□ D

D26_PN2015

Anch_8. Un gruppo di 20 amici va in pizzeria. Ciascuno di essi ordina una pizza che costa 8 euro.
Ogni 5 pizze ordinate, il proprietario non ne fa pagare una.
Quanto spendono in tutto gli amici per le pizze?

A. □ **160 euro**

B. □ **128 euro**

C. □ **120 euro**

D. □ **112 euro**

D25_PN2011

Anch_5. Quale fra le seguenti disuguaglianze è quella corretta?

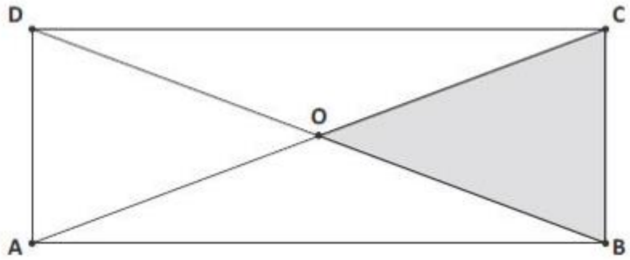
A. □ $\frac{3}{10} < \frac{3}{5} < \frac{3}{20}$

B. □ $\frac{4}{10} < \frac{3}{5} < \frac{11}{20}$

C. □ $\frac{5}{10} < \frac{3}{5} < \frac{13}{20}$

D. □ $\frac{7}{10} < \frac{3}{5} < \frac{13}{20}$

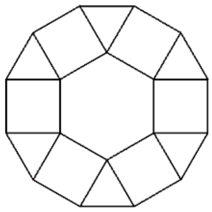
D11. In figura è rappresentato il rettangolo ABCD con le sue diagonali. Se conosci l'area del rettangolo, puoi calcolare l'area del triangolo in grigio?



Scegli una risposta con il suo perché.

NO		**SI**	
perché i quattro triangoli di vertice O non sono tutti uguali fra loro	**perché non conosco le dimensioni del rettangolo**	**perché i quattro triangoli AOB, BOC, COD, DOA sono equivalenti**	**perché i quattro triangoli AOB, BOC, COD, DOA sono isosceli**
□ A	□ B	□ C	□ D

Anch_2. Le immagini che seguono rappresentano un motivo del pavimento di una antica casa romana e la sua schematizzazione geometrica:



Il motivo, corrispondente a un dodecagono, è composto da un esagono regolare interno, sei quadrati uguali e sei triangoli equilateri uguali.

L'area dell'esagono è metà dell'area del dodecagono?
VERO □ FALSO □

L'area di ciascun triangolo è un sesto dell'area dell'esagono?
VERO □ FALSO □

D5_PN2011_originale

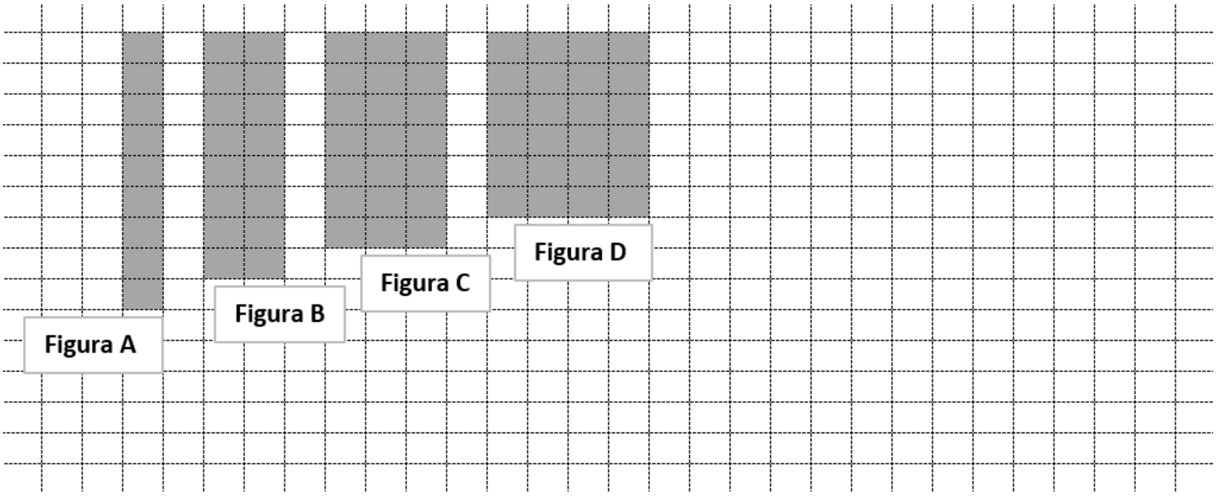
D12. Giovanni e Caterina si stanno allenando in piscina. Nuotano entrambi alla stessa velocità ma Giovanni ha cominciato più tardi ad allenarsi. Quando Giovanni ha fatto 10 vasche, Caterina ne ha fatte 30. Al termine dell'allenamento Giovanni ha fatto 50 vasche; quante ne ha fatte Caterina?

Risposta: .... vasche

D13. Quale numero puoi inserire nel quadratino per rendere vera la seguente uguaglianza?

$$\frac{7}{11} < \boxed{} < \frac{9}{11}$$

D14. Osserva la seguente sequenza di figure:

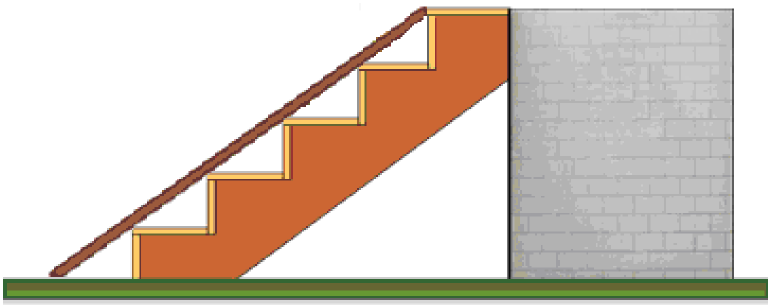


Completa le seguenti frasi scegliendo, per ciascuna, l'espressione che la rende corretta.

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

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

Anch_6. Una scala, costituita da 5 gradini profondi 24 cm e alti 18 cm l'uno, deve essere coperta da una tavola di legno utilizzata come scivolo per il trasporto di alcune merci. Qual è il procedimento corretto per trovare la lunghezza dello scivolo?



A. □ $\left(\sqrt{18^2}+\sqrt{24^2}\right)\times 5$

B. □ $\sqrt{(24+18)^2}\times 5$

C. □ $\sqrt{24^2+18^2}\times 5$

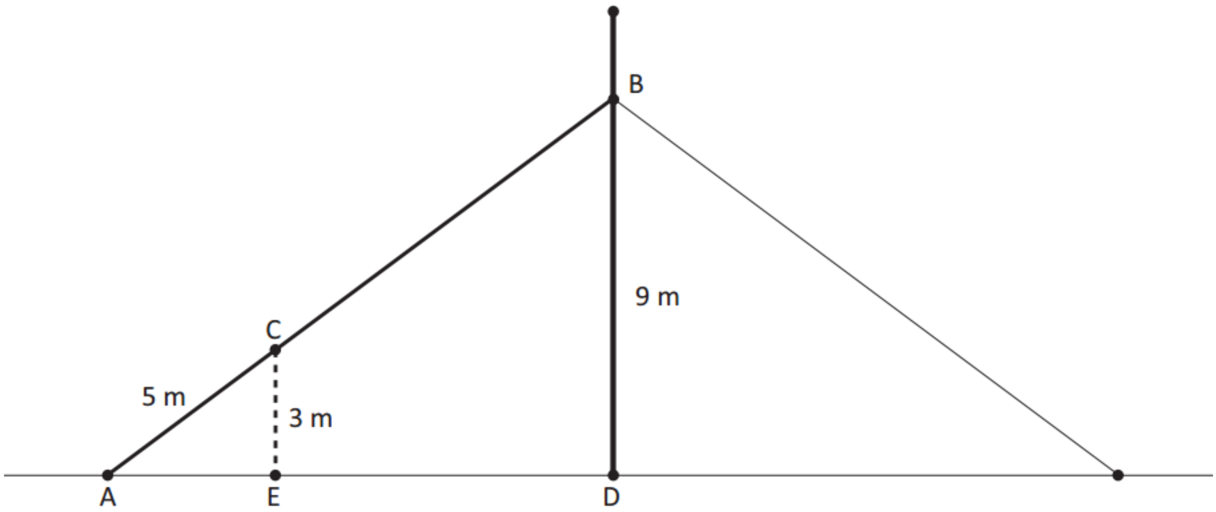
D. □ $\sqrt{\left(24^2+18^2\right)\times 5}$

D15. In ogni coppia di operazioni cerchia quella che dà il risultato maggiore:

$\boxed{8 \times 4}$	oppure	$\boxed{8 : 4}$
$\boxed{8 \times 0{,}4}$	oppure	$\boxed{8 : 0{,}4}$
$\boxed{0{,}8 \times 0{,}4}$	oppure	$\boxed{0{,}8 : 0{,}4}$

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**

D17. Giulio sa che nel negozio A e nel negozio B le bottiglie di olio della marca che preferisce hanno lo stesso prezzo.
Sua moglie gli dice che oggi, su quell'olio, nel negozio A fanno l'offerta "compri 3 e paghi 2" e nel negozio B fanno lo sconto del 40%.
Giulio deve comprare 3 bottiglie d'olio.

a. In quale negozio gli conviene comprarle?

Risposta: ……………………………….

b. Scrivi come hai fatto per trovare la risposta.

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

Grazie per aver risposto alle domande precedenti. Ti chiediamo ora, di completare la seguente sezione ricordandoti che, in questa sezione, non esistono risposte giuste o sbagliate. Leggi, quindi, tutto con attenzione e rispondi con sincerità!

Parte A – Il tuo rapporto con la scuola

A1) Qual è la materia che a scuola studi con più piacere?
Risposta: ____________________

A2) Qual è invece la materia che studi con meno piacere?
Risposta: ___________________

A3) Quanto pensi di essere bravo nelle seguenti materie? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= non sono bravo)	2 (= sono nella media)	3 (= sono bravo)	4 (= sono bravissimo)
Italiano	1	2	3	4
Matematica	1	2	3	4
Scienze	1	2	3	4

A4) Rispetto allo scorso anno, pensi che la tua abilità in Italiano, Matematica e Scienze sia cambiata? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= sono peggiorato)	2 (= non è cambiato nulla)	3 (= sono migliorato)
Italiano	1	2	3
Matematica	1	2	3
Scienze	1	2	3

A5) Quanto spesso i tuoi genitori o le persone che si prendono cura di te fanno le seguenti cose? Metti una crocetta in corrispondenza della risposta esatta, per ciascuna riga della seguente tabella:

	mai	raramente	qualche volta	spesso	sempre
Controllare se hai fatto i compiti					
Aiutarti a fare i compiti					
Elogiarti o ricompensarti per aver avuto buoni voti					
Ridurre le tue ricompense quando prendi un brutto voto					
Trovare qualcuno che ti aiuti a fare i compiti					

Parte B – Cosa pensi della Matematica?

B1) Quanto ti senti in accordo o in disaccordo con le seguenti affermazioni? Per ciascuna riga della seguente tabella, cerchia il numero che rispecchia meglio quello che pensi sapendo che:

1 = totalmente in disaccordo;
2 = parzialmente in disaccordo;
3 = non so;
4 = parzialmente d'accordo;
5 = totalmente d'accordo.

1. La Matematica per me è importante	**1**	**2**	**3**	**4**	**5**
2. La Matematica la possono capire ed imparare in tanti	**1**	**2**	**3**	**4**	**5**
3. Ai miei genitori e/o a chi si prende cura di me piace la Matematica	**1**	**2**	**3**	**4**	**5**
4. La Matematica è una della materie più interessanti che studio a scuola	**1**	**2**	**3**	**4**	**5**
5. Studiare la Matematica è divertente	**1**	**2**	**3**	**4**	**5**
6. Io ho una mente matematica	**1**	**2**	**3**	**4**	**5**
7. Sono in grado di prendere buoni voti in Matematica	**1**	**2**	**3**	**4**	**5**
8. Prendere buoni voti in Matematica mi interessa	**1**	**2**	**3**	**4**	**5**
9. In Matematica, il mio impegno viene ricompensato	**1**	**2**	**3**	**4**	**5**
10. Essere bravo in Matematica è un talento di cui bisogna essere dotati alla nascita	**1**	**2**	**3**	**4**	**5**
11. Posso imparare la Matematica anche se è una materia difficile	**1**	**2**	**3**	**4**	**5**
12. Mi piace utilizzare la Matematica che già conosco bene piuttosto che usare concetti matematici che nuovi o che ho studiato da poco	**1**	**2**	**3**	**4**	**5**
13. Sono più preoccupato/a per lo studio della Matematica che per altre materie	**1**	**2**	**3**	**4**	**5**
14. Ho spesso bisogno di aiuto con la Matematica	**1**	**2**	**3**	**4**	**5**
15. Rispetto ai miei compagni di classe, sono bravo/a in Matematica	**1**	**2**	**3**	**4**	**5**
16. Ai miei parenti e/o a chi si prende cura di me piace risolvere problemi matematici	**1**	**2**	**3**	**4**	**5**
17. Spero di non dover studiare Matematica mai più in vita mia	**1**	**2**	**3**	**4**	**5**
18. Mi piacerebbe che i miei studi futuri includessero molta Matematica	**1**	**2**	**3**	**4**	**5**
19. Mi piacerebbe tanto studiare più Matematica a scuola	**1**	**2**	**3**	**4**	**5**
20. Da grande, mi piacerebbe fare della Matematica il mio mestiere	**1**	**2**	**3**	**4**	**5**
21. La Matematica è importante per il mio futuro (dopo la scuola)	**1**	**2**	**3**	**4**	**5**

Parte C – Per favore, dicci quanto spesso fate in classe le seguenti attività. Cerchia il numero più appropriato, riga per riga, sapendo che 1 = mai; 2 = raramente; 3 = qualche volta; 4 = sempre.

	1 = mai	**2 = raramente**	**3 = qualche volta**	**4 = sempre**
C1. L'insegnante ci fa delle domande	**1**	**2**	**3**	**4**
C2. L'insegnante ci chiede di spiegare il ragionamento che abbiamo fatto per arrivare al risultato	**1**	**2**	**3**	**4**
C3. L'insegnante inizia a spiegare gli argomenti nuovi partendo da esempi tratti dal mondo reale	**1**	**2**	**3**	**4**
C4. L'insegnante ci dice di lavorare più velocemente	**1**	**2**	**3**	**4**
C5. L'insegnante utilizza il computer per spiegarci alcuni argomenti	**1**	**2**	**3**	**4**
C6. L'insegnante ci dà spesso problemi da esaminare	**1**	**2**	**3**	**4**
C7. L'insegnante si aspetta che ci ricordiamo cose imparate in passato	**1**	**2**	**3**	**4**
C8. L'insegnante ci dice dettagliatamente quali attività fare	**1**	**2**	**3**	**4**
C9. L'insegnante ci chiede se conosciamo già qualcosa dell'argomento che sta per spiegarci	**1**	**2**	**3**	**4**
C10. L'insegnante ci spiega qual è il valore futuro delle cose che ci spiega	**1**	**2**	**3**	**4**
C11. Lavoriamo spesso in gruppo	**1**	**2**	**3**	**4**
C12. Ascoltiamo l'insegnante spiegare la lezione	**1**	**2**	**3**	**4**
C13. Copiamo ciò che l'insegnante scrive sulla lavagna	**1**	**2**	**3**	**4**
C14. Ci confrontiamo spesso l'un l'altro sulle strategie di soluzione dei problemi	**1**	**2**	**3**	**4**
C15. Chiediamo agli altri studenti di spiegare le loro idee e il loro punto di vista	**1**	**2**	**3**	**4**
C17. Facciamo gli esercizi riportati nel libro	**1**	**2**	**3**	**4**
C18. Studiamo come alcuni concetti matematici si sono evoluti nel tempo	**1**	**2**	**3**	**4**
C19. Quello che impariamo a scuola è collegato con la nostra vita al di fuori della scuola	**1**	**2**	**3**	**4**
C20. Abbiamo imparato che la Matematica di basa su regole inventate	**1**	**2**	**3**	**4**
C21. Facciamo ricerche su argomenti che scegliamo noi	**1**	**2**	**3**	**4**
C22. Usiamo la calcolatrice	**1**	**2**	**3**	**4**
C23. Usiamo i computer	**1**	**2**	**3**	**4**
C24. In classe, usiamo riviste, giornali o video	**1**	**2**	**3**	**4**
C25. Discutiamo le nostre idee con l'intera classe	**1**	**2**	**3**	**4**
C26. Esponiamo il nostro lavoro all'intera classe	**1**	**2**	**3**	**4**

C27. Se utilizzate il computer, che cosa lo utilizzate prevalentemente?

Risposta: ____

____

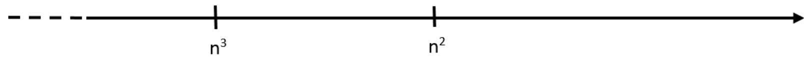
____

Parte D – Quanto ti senti sicuro/a di saper risolvere alcuni specifici problemi matematici

In questa sezione, ti chiediamo di fare una valutazione di quanto ti senti sicuro/a di saper risolvere alcuni problemi matematici. **Non ti chiediamo di fornire una risposta per ciascun quesito ma solo di dirci se e quanto pensi che saresti in grado di rispondere correttamente alle singole domande**. **Leggi quindi con grande attenzione e rispondi con la massima sincerità!**

D1. Quanto ti senti sicuro/a di saper utilizzare i numeri razionali, le loro diverse rappresentazioni, stimare la loro grandezza e/o il risultato di operazioni con i numeri razionali), come nella domanda riportata qui sotto?

Sulla seguente retta dei numeri sono ordinate due potenze di un numero razionale n.



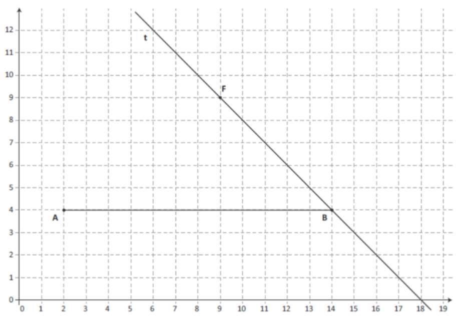
Indica con una crocetta se ciascuna delle seguenti affermazioni è vera (V) o falsa (F).

	SI	NO
Il valore di n può essere +1/2		
Il valore di n può essere -1/2		
Il valore di n può essere +3/2		
Il valore di n può essere -3/2		

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D2. Quanto ti senti sicuro/a di saper riconoscere le forme del piano e dello spazio, le loro rappresentazioni e di saper cogliere relazioni tra gli elementi, come nella domanda riportata qui sotto?

Osserva la figura

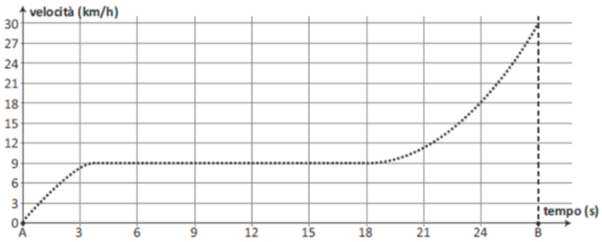


Sei in grado di disegnare la retta s perpendicolare a t passante per f e di individuare le coordinate del punto F di intersezione tra la retta *s* ed il segmento AB?

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D3. Quanto ti senti sicuro/a di saper calcolare misure statistiche di sintesi (ad esempio, la media, la moda, la mediana, ecc.), come nella domanda riportata qui sotto?

Luca percorre una strada in bicicletta e, con l'aiuto del computer, registra la propria velocità ogni decimo di secondo. Il grafico in figura rappresenta le diverse velocità raggiunte da Luca al passare del tempo.



Sei in grado di calcolare il valore modale delle velocità raggiunte da Luca dall'istante A all'istante B?

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D4. Quanto ti senti sicuro/a di saper risolvere problemi di misura e stima, come nella domanda riportata qui sotto?

Osserva l'edificio

Quanto può essere alto l'edificio?

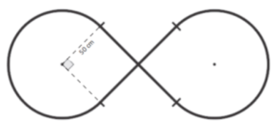
A. ☐ Meno di 10 metri
B. ☐ Tra 15 e 20 metri
C. ☐ Tra 25 e 30 metri
D. ☐ Più di 35 metri

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D5. Quanto ti senti sicuro/a di risolvere un problema matematico e di esplicitare in forma scritta il ragionamento che hai fatto, come nella domanda riportata di seguito?

La figura rappresenta lo schema di una pista formata da:

- Due archi di circonferenza di raggio 50 cm;
- Due tratti rettilinei di 100 cm ciascuno, perpendicolari tra loro nel punto medio.



Qual è la lunghezza della pista?

Scrivi i calcoli che fai per trovare la risposta e, infine, riporta il risultato.

____

____

____

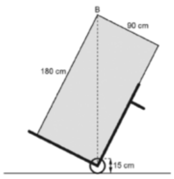
____

Risultato: circa ____________ cm

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D6. Quanto ti senti sicuro/a nel risolvere problemi come quello presentato nella seguente domanda?

Gabriele ha comprato un nuovo frigorifero. Per portarlo in cucina usa un carrello, come rappresentato in figura.



Quale espressione di permette di calcolare la massima distanza dal suolo del punto B quando il frigorifero è trasportato sul carrello?

A. ☐ $\sqrt{180^2+90^2}+7{,}5$

B. ☐ $\sqrt{180^2-90^2}+7{,}5$

C. ☐ $\sqrt{180+90}+7{,}5$

D. ☐ $\sqrt{180^2+90^2}+7{,}5$

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D7. Quanto ti senti sicuro/a di saper risolvere problemi di probabilità, come quello presentato nella seguente domanda?

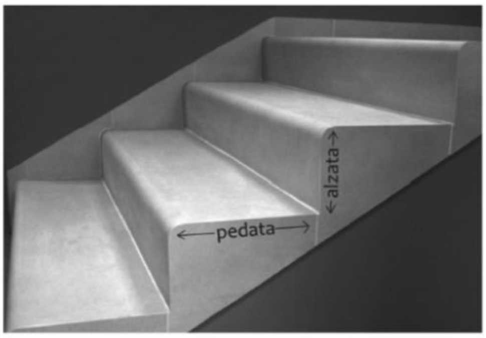
Giuseppe mette 2 palline bianche e 1 pallina nera in una busta.

Senza guardare, estrae una pallina. Qual è la probabilità che la pallina sia nera?

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D8. Quanto ti senti sicuro/a di saper utilizzare la matematica nella vita di tutti i giorni per rispondere a domande concrete, come quello presentato nella seguente domanda?

Nel disegno sottostante è rappresentata una scala.



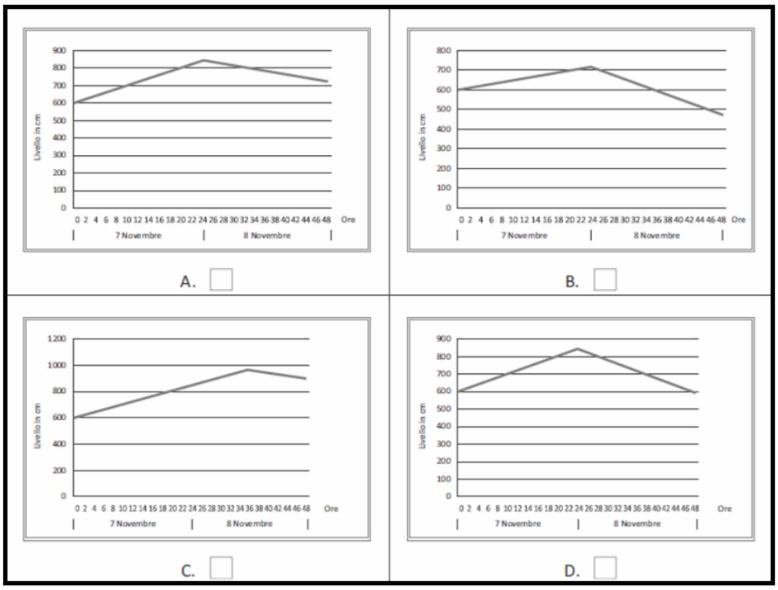
Per legge, la pedata deve essere lunga almeno 30 cm e la somma tra il doppio dell'alzata e la pedata deve essere compresa tra 62 e 64 cm (estremi inclusi).

Se la pedata di una scala misura 34 cm, il doppio dell'alzata dovrà essere compreso tra 28 e ____________ cm e, quindi, l'alzata dovrà essere compresa tra 14 cm e ____________ cm.

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D9. Quanto ti senti sicuro/a di saper risolvere problemi come quello presentato nella seguente domanda?

Oggi, il livello dell'acqua di un fiume è aumentato di circa 10 cm all'ora, per tutte le 24 ore. Il giorno successivo, il livello dell'acqua è diminuito di circa 5 cm all'ora per tutte le 24 ore. Quale tra i seguenti grafici può rappresentare meglio la situazione descritta?



Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

Parte E – Qual è il tuo rapporto con lo studio della Matematica?

Leggi attentamente le seguenti affermazioni e ripensa all'esperienza che hai avuto nel tuo percorso scolastico. Per ciascuna situazione, indica il livello di ansia che hai provato utilizzando una scala con 5 possibili risposte: da 1 (=poca ansia) a 5 (=molta ansia).

	Poca ansia				*Molta ansia*
1. Dover consultare le tavole in un manuale di matematica.	1	2	3	4	5
2. Pensare al compito di matematica del giorno dopo.	1	2	3	4	5
3. Guardare la risoluzione di un'equazione svolta dall'insegnante alla lavagna.	1	2	3	4	5
4. Fare il compito di matematica.	1	2	3	4	5
5. Avere da fare molti problemi difficili di matematica per la lezione successiva.	1	2	3	4	5
6. Ascoltare una spiegazione durante l'ora di matematica.	1	2	3	4	5
7. Ascoltare uno altro studente che spiega una formula di matematica.	1	2	3	4	5
8. Dover fare una verifica a sorpresa durante l'ora di matematica.	1	2	3	4	5
9. Iniziare un nuovo capitolo del manuale di matematica.	1	2	3	4	5

Parte F – Come sono i tuoi rapporti con tuoi compagni di scuola e di classe?

Questa è l'ultima sezione a cui ti chiediamo di rispondere. Le domande che ti facciamo, ci servono per capire quali sono i rapporti tra te e i tuoi compagni di classe. Prova quindi a pensare alla tua giornata a scuola e al rapporto che hai con i tuoi compagni, al tempo che trascorri con loro e alle attività che fate insieme. Nella maggior parte dei casi, non ti chiederemo il nome dei tuoi compagni ma solo il loro numero. Nelle ultime sette domande,

invece, ti chiediamo il tuo nome e quello di alcuni dei tuoi compagni. **In nessun caso, ti chiediamo il tuo cognome o quello dei tuoi compagni. Ad ogni modo, scegli tu a quali domande vuoi rispondere e quali invece preferisci lasciare in bianco!**

F1) Quanti studenti ci sono nella tua classe?

Risposta: _________

F2) Con quanti dei tuoi compagni di classe ti fa piacere chiacchierare durante la ricreazione, l'ora di ginnastica o quando possibile durante l'orario scolastico?

Risposta: _________

F3) Con quanti dei tuoi compagni di classe ti vedi fuori dalla scuola?

Risposta: _________

F4) Quanti dei tuoi compagni sono tuoi amici su Facebook e/o altri social networks?

Risposta: _________

F5) Con quanti compagni di classe hai l'abitudine di tenerti in contatto tramite telefonate, SMS, messaggi whatsapp, chat, ecc.?

Risposta: _________

F6) Quanti dei tuoi compagni vedi abitualmente al di fuori della scuola per

i. fare i compiti. Risposta: ___________

ii. fare attività sportive. Risposta: ___________

iii. uscire insieme (per andare al cinema, a mangiare una pizza, ecc.). Risposta: ___________

F7) Se tu avessi qualche problema, a scuola o nella tua vita privata,

i. con quanti dei tuoi compagni ti confideresti? Risposta: _________

ii. a quanti dei tuoi compagni chiederesti aiuto? Risposta: _________

iii. quanti dei tuoi compagni pensi che ti aiuterebbero? Risposta: _________

F8) Quanti dei tuoi compagni ti chiederebbero aiuto o si confiderebbero con te se avessero dei problemi?

Risposta: _________

F9) Quanti dei tuoi compagni aiuteresti se ti chiedessero aiuto?

Risposta: _________

F1. Qual è il tuo nome di battesimo? Risposta: ___________________

F2. Se si dovesse organizzare una gita, chi sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi, in ordine di preferenza:

1. _________________________
2. _________________________
3. _________________________

F3. Se si dovesse organizzare una gita, chi **non** sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

F4. Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, chi sceglieresti? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

F5. Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, con quale ti piacerebbe non stare? Indica 3 nomi

1. ___________________________
2. ___________________________
3. ___________________________

F6. Abbiamo fatto le stesse domande a tutti i tuoi compagni. Secondo te, tra i tuoi compagni, chi pensi ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

F7. Chi pensi invece che **non** ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

Hai finito!
GRAZIE!!!

Prova di Matematica - Fascicolo 2

Ministero dell'Istruzione dell'Università e della Ricerca

PROGETTO 'VARIAZIONI 2'

Anno Scolastico 2016 – 2017

LIVELLO 8

Prova di Matematica

Scuola Secondaria di primo grado

Classe Terza

Fascicolo 2

Classe: ………………….………………………………

Studente: ………………………..…………………

A cura di
Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione
INVALSI

ISTRUZIONI

Troverai nel fascicolo 30 domande di matematica. Alcune domande hanno quattro possibili risposte, ma una sola è quella giusta. Prima di ogni risposta c'è un quadratino con una lettera dell'alfabeto: A, B, C, D.

Per rispondere, devi mettere una crocetta nel quadratino accanto alla risposta (una sola) che ritieni giusta, come nell'esempio seguente.

Esempio 1

Quanti giorni ci sono in una settimana?

A. ☒ Sette

B. ☐ Sei

C. ☐ Cinque

D. ☐ Quattro

Se ti accorgi di aver sbagliato, puoi correggere: devi scrivere **NO** accanto alla risposta sbagliata e mettere una crocetta nel quadratino accanto alla risposta che ritieni giusta, come nell'esempio seguente.

Esempio 2

Quanti minuti ci sono in un'ora?

NO A. ☒ 30 minuti

B. ☐ 50 minuti

C. ☒ 60 minuti

D. ☐ 100 minuti

Altre domande chiedono di scrivere la risposta o il procedimento, oppure prevedono una diversa modalità di risposta. In questo caso il testo della domanda ti dice come rispondere. Leggilo dunque sempre con molta attenzione.

Puoi usare il righello graduato, la squadra, il compasso e il goniometro ma non la calcolatrice.

Non scrivere con la matita, ma usa soltanto una penna nera o blu.

Puoi usare le pagine bianche del fascicolo o gli spazi bianchi accanto alle domande per fare calcoli o disegni.

Per fare una prova, ora rispondi a questa domanda.

In quale delle seguenti sequenze i numeri sono scritti dal più grande al più piccolo?

A. ☐ 2; 5; 4; 8

B. ☐ 8; 5; 4; 2

C. ☐ 2; 4; 8; 5

D. ☐ 2; 4; 5; 8

Hai a disposizione 1 ora e quindici minuti (in totale 75 minuti) per rispondere alle domande. L'insegnante ti dirà quando cominciare a lavorare. Quando l'insegnante ti comunicherà che il tempo è finito, posa la penna e chiudi il fascicolo.

Se finisci prima, puoi chiudere il fascicolo e aspettare la fine, oppure puoi controllare le risposte che hai dato.

Caro studente,
ti chiediamo di rispondere alle seguenti domande.

Q1. Tu sei: maschio □ femmina □

Q2. Tu sei nato nel mese di ________________ dell'anno ________________

Q3. Hai frequentato la scuola materna?
  a. No
  b. Sì, per più di un anno
  c. Sì, per meno di un anno

Q4. Hai frequentato l'asilo nido?
  a. No
  b. Sì, per più di un anno
  c. Sì, per meno di un anno

Q5. Tu sei nato
  a. In Italia
  b. All'estero (scrivi sui puntini in quale Paese sei nato): ______________________________

Q6. Se non sei nato in Italia, quanti anni avevi quando sei arrivato in Italia? ________________ anni

Q7. In quale Paese è nata la tua mamma? __________________________

Q8. In quale Paese è nato il tuo papà? __________________________

Q9. Qual è il titolo di studio conseguito dai tuoi genitori?

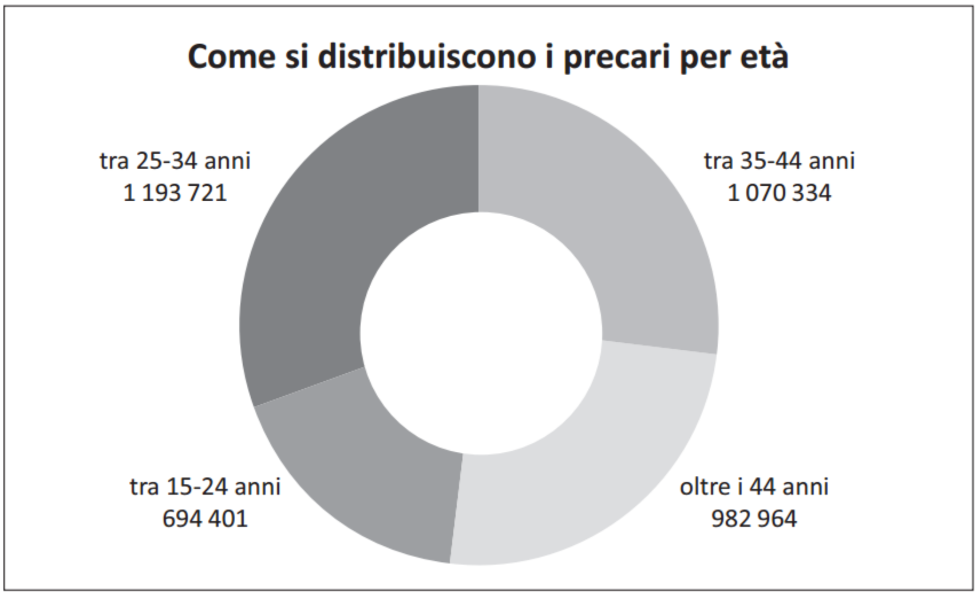
	madre	padre
A. Licenza elementare		
B. Licenza media		
C. Qualifica professionale triennale		
D. Diploma di scuola secondaria superiore (liceo, istituto tecnico o istituto professionale)		
E. Titolo di studio superiore al diploma, diverso dalla laurea (ISEF, Accademia di Belle Arti, Conservatorio)		
F. Laurea / Dottorato di ricerca / Master		
G. Non me lo ricordo		

Q10. Che lavoro fanno i tuoi genitori?

	madre	padre
A. Disoccupato/a		
B. Si occupa della casa		
C. Dirigente, docente universitario, funzionario, ufficiale militare		
D. Imprenditore, proprietario agricolo		
E. Professionista dipendente, sottufficiale militare, libero professionista (psicologo, ricercatore, medico, avvocato, commissario di polizia, ecc.)		
F. Lavoratore in proprio (commerciante, artigiano, coltivatore diretto, meccanico, sarto, ecc.)		
G. Insegnante, impiegato, militare graduato		
H. Operaio, addetto ai servizi, socio di cooperativa (tecnico, infermiere, cameriere, commessa, ecc.)		
I. Pensionato/a		
L. Non me lo ricordo		

NON GIRARE LA PAGINA
FINCHÉ NON TI SARÀ DETTO DI FARLO

A1. Il seguente grafico rappresenta la distribuzione dei lavoratori precari in Italia suddivisi per età nell'anno 2012.



a. Quanti sono in totale i precari?

A. ☐ **Circa due milioni**

B. ☐ **Circa tre milioni**

C. ☐ **Circa quattro milioni**

D. ☐ **Circa cinque milioni**

b. Quale percentuale rappresentano i precari che hanno tra i 25 e i 34 anni?

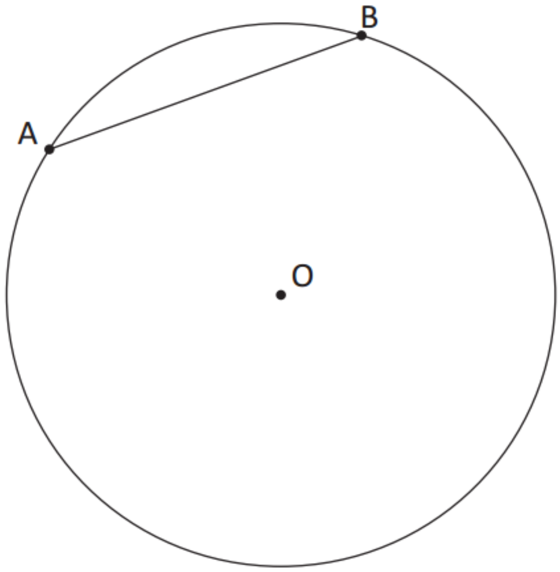
A. ☐ **Circa il 50%**

B. ☐ **Circa il 40%**

C. ☐ **Circa il 30%**

D. ☐ **Circa il 20%**

A2. Osserva la figura. AB è un cateto di un triangolo rettangolo inscritto nella circonferenza di centro O. Disegna il triangolo rettangolo.



A3. Se *n* é un numero naturale, allora il numero *n · (n + 2)*

A. □ **è sempre dispari**

B. □ **è sempre pari**

C. □ **è dispari se n è pari**

D. □ **è dispari se n è dispari**

A4. Tempo fa si è disputata la partita di pallacanestro B. Pozzo di Gotto - Brescia, finita con il punteggio di 92-94.

La seguente tabella riassume le statistiche di tale partita per la squadra di Brescia.

		Tiri a canestro			
Nome del giocatore	**Minuti giocati**	**Tiri da 2**	**Tiri da 3**	**Tiri liberi**	**PUNTI**
Roberto	25	0	0	2	2
Clelia	23	4	0	1	9
Chiara	20	2	0	0	4
Giorgio	36	2	1	7	14
Cristina	37	3	1	1	10
Monica	30	9	1	8	29
Paolo	9	0	1	2	5
Anna	15	0	1	0	3
Maria	30	6	0	6	18
	Totale	**26**	**5**	**27**	**94**

Quanti sono i giocatori che hanno realizzato un numero di punti superiore alla media?

Risposta: ……………………………………………………………………………………………

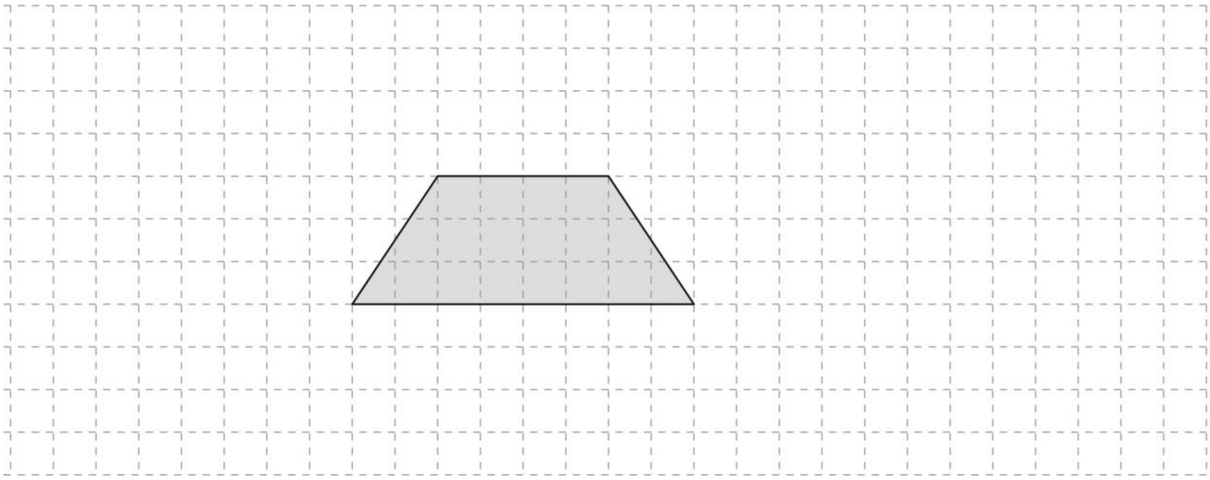
A5. **Il Signor Carlo scende dal tram all'incrocio di *via Pietro Micca* con *via Antonio Giuseppe Bertola* (nella mappa che vedi qui sotto il punto è contrassegnato da un asterisco).**

Percorre 200 metri di *via Bertola* e all'incrocio con *via 20 Settembre* svolta a sinistra; dopo aver camminato per 150 metri, raggiunge l'incrocio con *via Pietro Micca*. Da lì decide di tornare al punto di partenza per *via Pietro Micca*.
Quanti metri all'incirca percorre al ritorno?

A. □ **200 m**
B. □ **250 m**
C. □ **350 m**
D. □ **600 m**

A6. **Il trapezio che vedi sotto è stato ritagliato da una figura F più grande. Il trapezio è i $\frac{3}{4}$ della figura F.**

Disegna una delle possibili figure F da cui il trapezio è stato ritagliato.



Anch_1. **Antonio e Giada partecipano a una gara a quiz. Per ogni risposta esatta si assegnano due punti mentre per ogni risposta sbagliata si toglie un punto. L'esito della gara è il seguente:**

- **Antonio ha dato 11 risposte esatte e 9 sbagliate;**
- **Giada ha dato 6 risposte esatte e 14 sbagliate.**

Quali sono i punteggi finali dei due ragazzi?

A. □ **+ 13; +2**

B. □ **+ 13; −2**

C. □ **+ 2; + 8**

D. □ **+ 2; − 8**

D13_PN2011_v4

D1. **Il numero** $\sqrt{10}$ **è:**

A. □ **compreso tra 9 e 11**

B. □ **uguale a 5**

C. □ **compreso tra 3 e 4**

D. □ **uguale a 100**

D2. Un bicchiere contiene 1/4 di litro di acqua.

Se si vuole riempire una bottiglia da 2 litri, quanti bicchieri di acqua bisogna versare nella bottiglia?

Risposta: ……………………………………

D15_PN2012_v4

D3. **Un maglione nel mese di dicembre costa 100 euro. A gennaio, nel periodo dei saldi il suo prezzo viene diminuito del 20% rispetto al prezzo di dicembre, mentre a febbraio il suo prezzo viene aumentato del 20% rispetto al prezzo di gennaio.**

A febbraio quindi il maglione costa:

A. □ **96 euro**

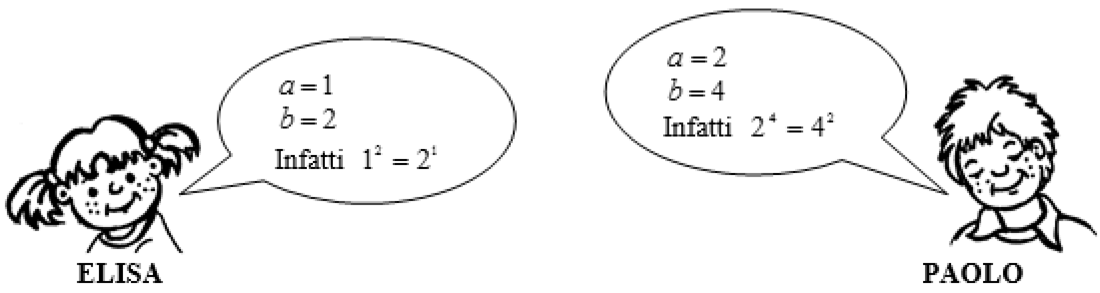
B. □ **90 euro**

C. □ **104 euro**

D. □ **110 euro**

D18_PN2011

Anch_3. Elisa e Paolo stanno cercando di rispondere a questa domanda: *"**Qual è la coppia di numeri interi a, b (diversi fra loro) tali che** $a^b = b^a$ **?**"*
Ecco le loro soluzioni.



Chi ha ragione?

A. □ **Solo Elisa**

B. □ **Solo Paolo**

C. □ **Entrambi**

D. □ **Nessuno dei due**

D5. Chiara ha un banco rettangolare e un pacco di fogli rettangolari. Il lato lungo dei fogli è un quinto del lato lungo del banco, e il lato corto dei fogli è un quinto del lato corto del banco. Quanti fogli come minimo le occorrono per ricoprire interamente la superficie del banco?

A. □ **25**

B. □ **20**

C. □ **10**

D. □ **5**

D6. **Un rettangolo ha il perimetro di 1 metro. Un lato del rettangolo misura 20 cm. Quanto misura l’altro lato?**

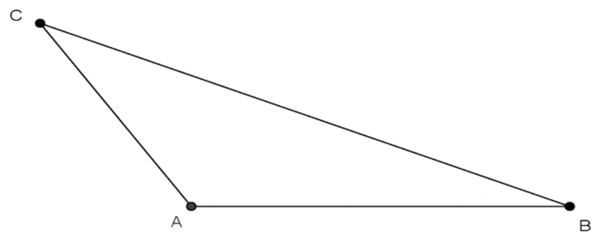
A. □ **30 cm**

B. □ **50 cm**

C. □ **60 cm**

D. □ **80 cm**

D7. Osserva il disegno.



Disegna l'altezza CH del triangolo, relativa alla base AB, e misura col righello AB e CH.

Calcola l’area del triangolo ABC.

Risposta: ………….. cm^2

D8. Nino sale su un treno composto dalla locomotiva e da 9 vagoni:

Quanto è lungo all’incirca il treno di Nino?

A. □ **Circa 10 m**

B. □ **Circa 50 m**

C. □ **Circa 250 m**

D. □ **Circa 1000 m**

D27_PN2013

Anch_7. Nella scuola Nino Bixio ci sono 600 studenti e un insegnante ogni 15 studenti.

Quale proporzione permette di trovare il numero x degli insegnanti?

A. □ **$x : 15 = 1 : 600$**

B. □ **$15 : 1 = x : 600$**

C. □ **$1 : 15 = x : 600$**

D. □ **$x : 1 = 15 : 600$**

1CG_NEW_v1

D9. Qual è il risultato di $4 \times 0{,}5$?

A. □ **8**

B. □ **4**

C. □ **2**

D. □ **20**

Anch_4. La formula $L = L_0 + K \times P$ esprime la lunghezza L di una molla al variare del peso P applicato. L_0 rappresenta la lunghezza in centimetri "a riposo" della molla; K indica di quanto si allunga in centimetri la molla quando le si applica una unità di peso.

Quale delle formule elencate si adatta meglio alla seguente descrizione: "*È una molla molto corta e molto dura (cioè molto resistente alla trazione)*" ?

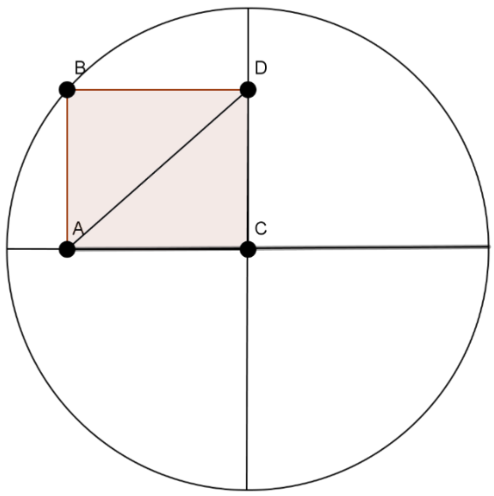
A. □ $L = 10 + 0{,}5 \times P$

B. □ $L = 10 + 7 \times P$

C. □ $L = 80 + 0{,}5 \times P$

D. □ $L = 80 + 7 \times P$

D10. Osserva la figura.



Nella circonferenza rappresentata in figura di centro C e raggio 4 cm, CABD è un rettangolo.

Mario afferma che il segmento AD misura 4 cm.

Mario ha ragione? Scegli una risposta con il suo perché.

Mario NON HA ragione perché		Mario HA ragione perché	
La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è diverso da 4	Non ci sono abbastanza dati per calcolare la lunghezza di AD	La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è 4	Le diagonali di un rettangolo sono congruenti tra di loro
□ A	□ B	□ C	□ D

D26_PN2015

Anch_8. Un gruppo di 20 amici va in pizzeria. Ciascuno di essi ordina una pizza che costa 8 euro.
Ogni 5 pizze ordinate, il proprietario non ne fa pagare una.
Quanto spendono in tutto gli amici per le pizze?

A. □ **160 euro**

B. □ **128 euro**

C. □ **120 euro**

D. □ **112 euro**

D25_PN2011

Anch_5. Quale fra le seguenti disuguaglianze è quella corretta?

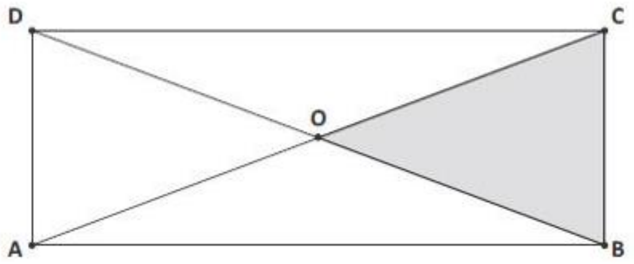
A. □ $\frac{3}{10} < \frac{3}{5} < \frac{3}{20}$

B. □ $\frac{4}{10} < \frac{3}{5} < \frac{11}{20}$

C. □ $\frac{5}{10} < \frac{3}{5} < \frac{13}{20}$

D. □ $\frac{7}{10} < \frac{3}{5} < \frac{13}{20}$

D11. In figura è rappresentato il rettangolo ABCD con le sue diagonali. Se conosci l'area del rettangolo, puoi calcolare l'area del triangolo in grigio?



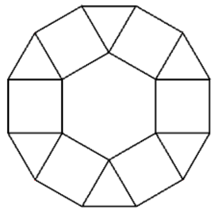
A. □ **No, perché i quattro triangoli di vertice O non sono tutti uguali fra loro**

B. □ **No, perché non conosco le dimensioni del rettangolo**

C. □ **Sì, perché i quattro triangoli di vertice O sono equivalenti**

D. □ **Sì, perché i quattro triangoli di vertice O sono isosceli**

Anch_2. Le immagini che seguono rappresentano un motivo del pavimento di una antica casa romana e la sua schematizzazione geometrica:



Il motivo, corrispondente a un dodecagono, è composto da un esagono regolare interno, sei quadrati uguali e sei triangoli equilateri uguali.

L'area dell'esagono è metà dell'area del dodecagono?
VERO □ FALSO □

L'area di ciascun triangolo è un sesto dell'area dell'esagono?
VERO □ FALSO □

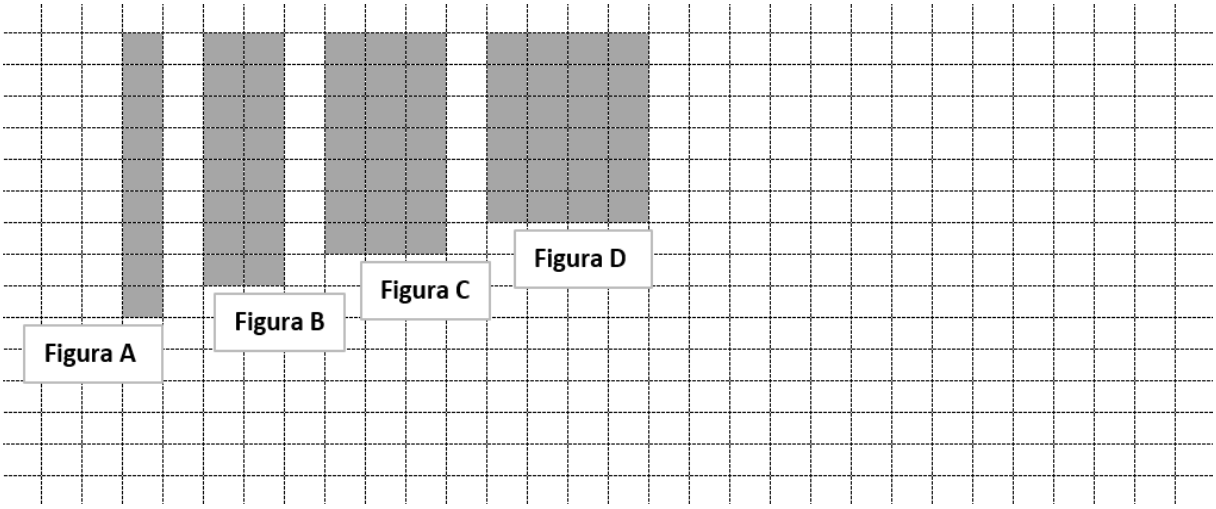
D12. Giovanni e Caterina mettono la paghetta settimanale ognuno nel proprio salvadanaio. Ricevono la stessa paghetta, ma Giovanni ha iniziato più tardi a metterla nel suo salvadanaio. Quando Giovanni ha 10 euro, Caterina ne ha 30. A Natale, quando rompono i salvadanai, Giovanni ha 50 euro; quanti euro ha Caterina?

Risposta: .... euro

D13. Quale numero puoi inserire nel quadratino per rendere vera la seguente uguaglianza?

$$\frac{2}{5} < \frac{\square}{10} < \frac{3}{5}$$

D14. Osserva la seguente sequenza di figure:



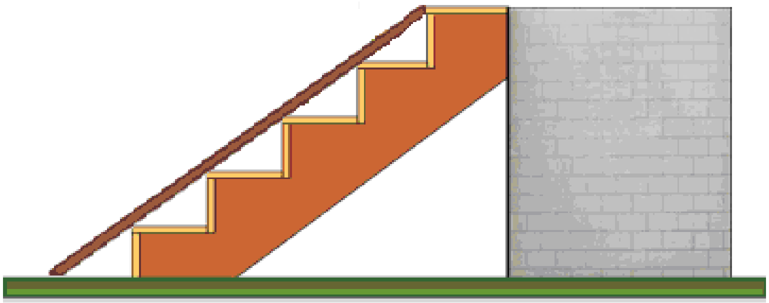
Completa le seguenti frasi scegliendo, per ciascuna, l'espressione che la rende corretta.

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);*

D22_PN2011

Anch_6. Una scala, costituita da 5 gradini profondi 24 cm e alti 18 cm l'uno, deve essere coperta da una tavola di legno utilizzata come scivolo per il trasporto di alcune merci. Qual è il procedimento corretto per trovare la lunghezza dello scivolo?



A. ☐ $\left(\sqrt{18^2}+\sqrt{24^2}\right)\times 5$

B. ☐ $\sqrt{(24+18)^2}\times 5$

C. ☐ $\sqrt{24^2+18^2}\times 5$

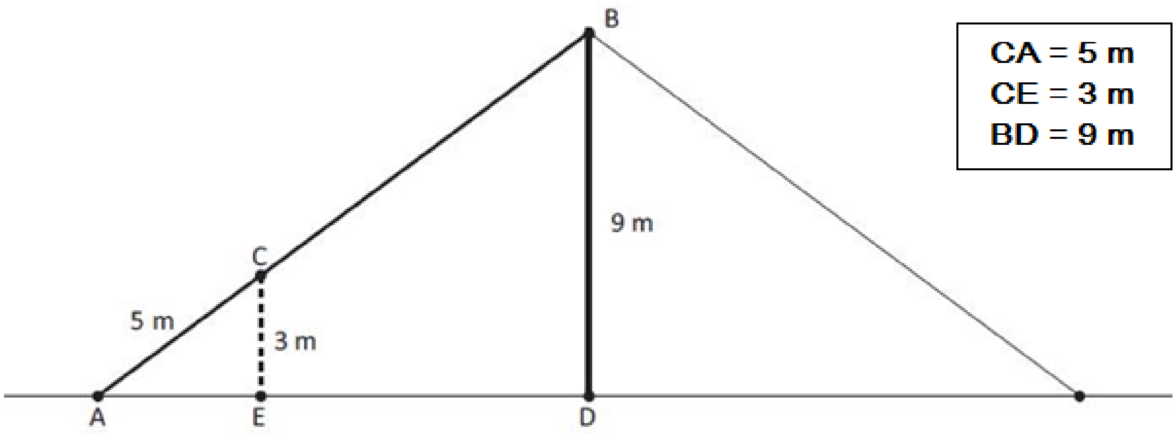
D. ☐ $\sqrt{\left(24^2+18^2\right)\times 5}$

3CG_NEW_inv

D15. In ogni coppia di operazioni cerchia quella che dà il risultato maggiore:

$\boxed{8\times 4}$	oppure	$\boxed{8:4}$
$\boxed{8\times 0{,}4}$	oppure	$\boxed{8:0{,}4}$
$\boxed{0{,}8\times 0{,}4}$	oppure	$\boxed{0{,}8:0{,}4}$

D16. Osserva il disegno.



Qual è la lunghezza del segmento AB?

Risposta: …………………… m

D17. **Giulio sa che nel negozio A e nel negozio B le bottiglie di olio della marca che preferisce hanno lo stesso prezzo.**
Sua moglie gli dice che oggi, su quell'olio, nel negozio A fanno l'offerta "compri 3 e paghi 2" e nel negozio B fanno lo sconto del 40%.
Giulio deve comprare 3 bottiglie d'olio.

a. **In quale negozio gli conviene comprarle?**

Risposta: ………………………………

b. **Scrivi come hai fatto per trovare la risposta.**

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Grazie per aver risposto alle domande precedenti. Ti chiediamo ora, di completare la seguente sezione ricordandoti che, in questa sezione, non esistono risposte giuste o sbagliate. Leggi, quindi, tutto con attenzione e rispondi con sincerità!

Parte A – Il tuo rapporto con la scuola

A1) Qual è la materia che a scuola studi con più piacere?
Risposta: ___________________

A2) Qual è invece la materia che studi con meno piacere?
Risposta: __________________

A3) Quanto pensi di essere bravo nelle seguenti materie? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= non sono bravo)	2 (= sono nella media)	3 (= sono bravo)	4 (= sono bravissimo)
Italiano	1	2	3	4
Matematica	1	2	3	4
Scienze	1	2	3	4

A4) Rispetto allo scorso anno, pensi che la tua abilità in Italiano, Matematica e Scienze sia cambiata? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= sono peggiorato)	2 (= non è cambiato nulla)	3 (= sono migliorato)
Italiano	1	2	3
Matematica	1	2	3
Scienze	1	2	3

A5) Quanto spesso i tuoi genitori o le persone che si prendono cura di te fanno le seguenti cose? Metti una crocetta in corrispondenza della risposta esatta, per ciascuna riga della seguente tabella:

	mai	raramente	qualche volta	spesso	sempre
Controllare se hai fatto i compiti					
Aiutarti a fare i compiti					
Elogiarti o ricompensarti per aver avuto buoni voti					
Ridurre le tue ricompense quando prendi un brutto voto					
Trovare qualcuno che ti aiuti a fare i compiti					

Parte B – Cosa pensi della Matematica?

E1) Quanto ti senti in accordo o in disaccordo con le seguenti affermazioni? Per ciascuna riga della seguente tabella, cerchia il numero che rispecchia meglio quello che pensi sapendo che:

1 = totalmente in disaccordo;
2 = parzialmente in disaccordo;
3 = non so;
4 = parzialmente d'accordo;
5 = totalmente d'accordo.

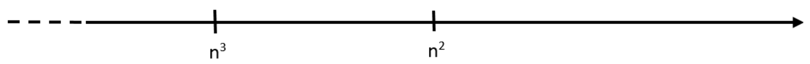
1. La Matematica per me è importante	**1**	**2**	**3**	**4**	**5**
2. La Matematica la possono capire ed imparare in tanti	**1**	**2**	**3**	**4**	**5**
3. Ai miei genitori e/o a chi si prende cura di me piace la Matematica	**1**	**2**	**3**	**4**	**5**
4. La Matematica è una della materie più interessanti che studio a scuola	**1**	**2**	**3**	**4**	**5**
5. Studiare la Matematica è divertente	**1**	**2**	**3**	**4**	**5**
6. Io ho una mente matematica	**1**	**2**	**3**	**4**	**5**
7. Sono in grado di prendere buoni voti in Matematica	**1**	**2**	**3**	**4**	**5**
8. Prendere buoni voti in Matematica mi interessa	**1**	**2**	**3**	**4**	**5**
9. In Matematica, il mio impegno viene ricompensato	**1**	**2**	**3**	**4**	**5**
10. Essere bravo in Matematica è un talento di cui bisogna essere dotati alla nascita	**1**	**2**	**3**	**4**	**5**
11. Posso imparare la Matematica anche se è una materia difficile	**1**	**2**	**3**	**4**	**5**
12. Mi piace utilizzare la Matematica che già conosco bene piuttosto che usare concetti matematici che nuovi o che ho studiato da poco	**1**	**2**	**3**	**4**	**5**
13. Sono più preoccupato/a per lo studio della Matematica che per altre materie	**1**	**2**	**3**	**4**	**5**
14. Ho spesso bisogno di aiuto con la Matematica	**1**	**2**	**3**	**4**	**5**
15. Rispetto ai miei compagni di classe, sono bravo/a in Matematica	**1**	**2**	**3**	**4**	**5**
16. Ai miei parenti e/o a chi si prende cura di me piace risolvere problemi matematici	**1**	**2**	**3**	**4**	**5**
17. Spero di non dover studiare Matematica mai più in vita mia	**1**	**2**	**3**	**4**	**5**
18. Mi piacerebbe che i miei studi futuri includessero molta Matematica	**1**	**2**	**3**	**4**	**5**
19. Mi piacerebbe tanto studiare più Matematica a scuola	**1**	**2**	**3**	**4**	**5**
20. Da grande, mi piacerebbe fare della Matematica il mio mestiere	**1**	**2**	**3**	**4**	**5**
21. La Matematica è importante per il mio futuro (dopo la scuola)	**1**	**2**	**3**	**4**	**5**

Parte C – Quanto ti senti sicuro/a di saper risolvere alcuni specifici problemi matematici

In questa sezione, ti chiediamo di fare una valutazione di quanto ti senti sicuro/a di saper risolvere alcuni problemi matematici. **Non ti chiediamo di fornire una risposta per ciascun quesito ma solo di dirci se e quanto pensi che saresti in grado di rispondere correttamente alle singole domande**. **Leggi quindi con grande attenzione e rispondi con la massima sincerità!**

C1. Quanto ti senti sicuro/a di saper utilizzare i numeri razionali, le loro diverse rappresentazioni, stimare la loro grandezza e/o il risultato di operazioni con i numeri razionali), come nella domanda riportata qui sotto?

Sulla seguente retta dei numeri sono ordinate due potenze di un numero razionale n.



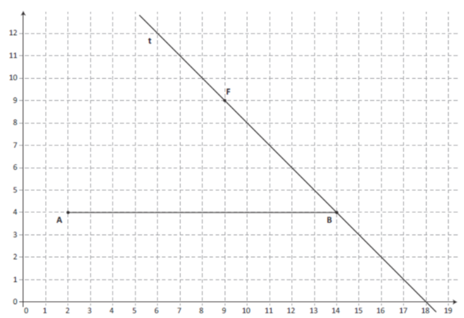
Indica con una crocetta se ciascuna delle seguenti affermazioni è vera (V) o falsa (F).

	SI	NO
Il valore di n può essere +1/2		
Il valore di n può essere -1/2		
Il valore di n può essere +3/2		
Il valore di n può essere -3/2		

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

C2. Quanto ti senti sicuro/a di saper riconoscere le forme del piano e dello spazio, le loro rappresentazioni e di saper cogliere relazioni tra gli elementi, come nella domanda riportata qui sotto?

Osserva la figura

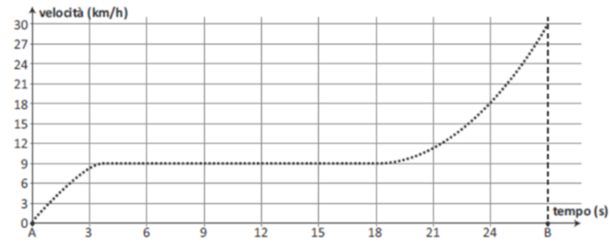


Sei in grado di disegnare la retta s perpendicolare a t passante per f e di individuare le coordinate del punto F di intersezione tra la retta *s* ed il segmento AB?

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

C3. Quanto ti senti sicuro/a di saper calcolare misure statistiche di sintesi (ad esempio, la media, la moda, la mediana, ecc.), come nella domanda riportata qui sotto?

Luca percorre una strada in bicicletta e, con l'aiuto del computer, registra la propria velocità ogni decimo di secondo. Il grafico in figura rappresenta le diverse velocità raggiunte da Luca al passare del tempo.



Sei in grado di calcolare il valore modale delle velocità raggiunte da Luca dall'istante A all'istante B?

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

C4. Quanto ti senti sicuro/a di saper risolvere problemi di misura e stima, come nella domanda riportata qui sotto?

Osserva l'edificio

Quanto può essere alto l'edificio?

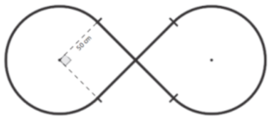
A. □ Meno di 10 metri
B. □ Tra 15 e 20 metri
C. □ Tra 25 e 30 metri
D. □ Più di 35 metri

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

C5. Quanto ti senti sicuro/a di risolvere un problema matematico e di esplicitare in forma scritta il ragionamento che hai fatto, come nella domanda riportata di seguito?

La figura rappresenta lo schema di una pista formata da:

- Due archi di circonferenza di raggio 50 cm;
- Due tratti rettilinei di 100 cm ciascuno, perpendicolari tra loro nel punto medio.



Qual è la lunghezza della pista?

Scrivi i calcoli che fai per trovare la risposta e, infine, riporta il risultato.

____

____

____

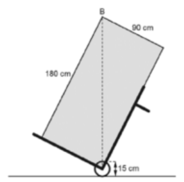
____

Risultato: circa ____________ cm

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

C6. Quanto ti senti sicuro/a nel risolvere problemi come quello presentato nella seguente domanda?

Gabriele ha comprato un nuovo frigorifero. Per portarlo in cucina usa un carrello, come rappresentato in figura.



Quale espressione di permette di calcolare la massima distanza dal suolo del punto B quando il frigorifero è trasportato sul carrello?

A. □ $\sqrt{180^2+90^2}+7{,}5$

B. □ $\sqrt{180^2-90^2}+7{,}5$

C. □ $\sqrt{180+90}+7{,}5$

D. □ $\sqrt{180^2+90^2}+7{,}5$

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

C7. Quanto ti senti sicuro/a di saper risolvere problemi di probabilità, come quello presentato nella seguente domanda?

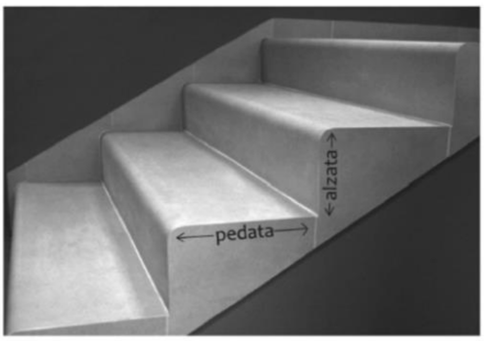
Giuseppe mette 2 palline bianche e 1 pallina nera in una busta.

Senza guardare, estrae una pallina. Qual è la probabilità che la pallina sia nera?

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

C8. Quanto ti senti sicuro/a di saper utilizzare la matematica nella vita di tutti i giorni per rispondere a domande concrete, come quello presentato nella seguente domanda?

Nel disegno sottostante è rappresentata una scala.



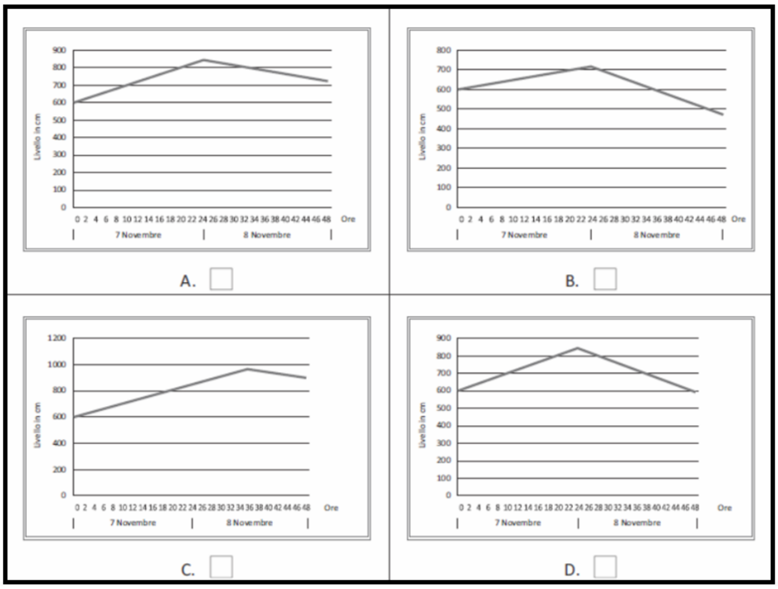
Per legge, la pedata deve essere lunga almeno 30 cm e la somma tra il doppio dell'alzata e la pedata deve essere compresa tra 62 e 64 cm (estremi inclusi).

Se la pedata di una scala misura 34 cm, il doppio dell'alzata dovrà essere compreso tra 28 e ____________ cm e, quindi, l'alzata dovrà essere compresa tra 14 cm e ____________ cm.

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

C9. Quanto ti senti sicuro/a di saper risolvere problemi come quello presentato nella seguente domanda?

Oggi, il livello dell'acqua di un fiume è aumentato di circa 10 cm all'ora, per tutte le 24 ore. Il giorno successivo, il livello dell'acqua è diminuito di circa 5 cm all'ora per tutte le 24 ore. Quale tra i seguenti grafici può rappresentare meglio la situazione descritta?



Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

Parte D – Qual è il tuo rapporto con lo studio della Matematica?

Leggi attentamente le seguenti affermazioni e ripensa all'esperienza che hai avuto nel tuo percorso scolastico. Per ciascuna situazione, indica il livello di ansia che hai provato utilizzando una scala con 5 possibili risposte: da 1 (=poca ansia) a 5 (=molta ansia).

	Poca ansia				*Molta ansia*
1. Dover consultare le tavole in un manuale di matematica.	1	2	3	4	5
2. Pensare al compito di matematica del giorno dopo.	1	2	3	4	5
3. Guardare la risoluzione di un'equazione svolta dall'insegnante alla lavagna.	1	2	3	4	5
4. Fare il compito di matematica.	1	2	3	4	5
5. Avere da fare molti problemi difficili di matematica per la lezione successiva.	1	2	3	4	5
6. Ascoltare una spiegazione durante l'ora di matematica.	1	2	3	4	5
7. Ascoltare uno altro studente che spiega una formula di matematica.	1	2	3	4	5
8. Dover fare una verifica a sorpresa durante l'ora di matematica.	1	2	3	4	5
9. Iniziare un nuovo capitolo del manuale di matematica.	1	2	3	4	5

Parte E – Come sono i tuoi rapporti con tuoi compagni di scuola e di classe?

Questa è l'ultima sezione a cui ti chiediamo di rispondere. Le domande che ti facciamo, ci servono per capire quali sono i rapporti tra te e i tuoi compagni di classe. Prova quindi a pensare alla tua giornata a scuola e al rapporto che hai con i tuoi compagni, al tempo che trascorri con loro e alle attività che fate insieme. Nella maggior parte dei casi, non ti chiederemo il nome dei tuoi compagni ma solo il loro numero. Nelle ultime domande,

invece, ti chiediamo il tuo nome e quello di alcuni dei tuoi compagni. **In nessun caso, ti chiediamo il tuo cognome o quello dei tuoi compagni. Ad ogni modo, scegli tu a quali domande vuoi rispondere e quali invece preferisci lasciare in bianco!**

E1) Quanti studenti ci sono nella tua classe?
Risposta: __________

E2) Con quanti dei tuoi compagni di classe ti fa piacere chiacchierare durante la ricreazione, l'ora di ginnastica o quando possibile durante l'orario scolastico?
Risposta: __________

E3) Con quanti dei tuoi compagni di classe ti vedi fuori dalla scuola?
Risposta: __________

E4) Quanti dei tuoi compagni sono tuoi amici su Facebook e/o altri social networks?
Risposta: __________

E5) Con quanti compagni di classe hai l'abitudine di tenerti in contatto tramite telefonate, SMS, messaggi whatsapp, chat, ecc.?
Risposta: __________

E6) Quanti dei tuoi compagni vedi abitualmente al di fuori della scuola per
  a. Fare i compiti: __________
  b. Fare attività sportive: _________
  c. Uscire insieme (per andare a mangiare una pizza, andare al cinema, ecc.): _________

E7) Se tu avessi qualche problema, a scuola o nella tua vita privata,
  a. Con quanti dei tuoi compagni ti confideresti? __________
  b. A quanti dei tuoi compagni chiederesti aiuto? _________
  c. Quanti dei tuoi compagni pensi che ti aiuterebbero? _________

E8) Quanti dei tuoi compagni pensi si confiderebbero con te o chi chiederebbero aiuto se avessero dei problemi?
Risposta: __________

E9) Quanti dei tuoi compagni aiuteresti se ti chiedessero aiuto?
Risposta: __________

E10) Qual è il tuo nome di battesimo?
Risposta: __________

E11) Se si dovesse organizzare una gita, chi sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi, in ordine di preferenza:

1. ____________________________
2. ____________________________
3. ____________________________

E12) Se si dovesse organizzare una gita, chi **non** sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

E13) Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, con quale ti piacerebbe non stare? Indica 3 nomi

1. ____________________________
2. ____________________________
3. ____________________________

E14) Abbiamo fatto le stesse domande a tutti i tuoi compagni. Secondo te, tra i tuoi compagni, chi pensi ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

E15) Chi pensi invece che **non** ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

Hai finito!
GRAZIE!!!

Prova di Matematica - Fascicolo 3

Ministero dell'Istruzione dell'Università e della Ricerca

PROGETTO 'VARIAZIONI 2'

Anno Scolastico 2016 – 2017

LIVELLO 8

Prova di Matematica

Scuola Secondaria di primo grado

Classe Terza

Fascicolo 3

Classe: ....

Studente: ....

A cura di
Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione

INVALSI

ISTRUZIONI

Troverai nel fascicolo 30 domande di matematica. Alcune domande hanno quattro possibili risposte, ma una sola è quella giusta. Prima di ogni risposta c'è un quadratino con una lettera dell'alfabeto: A, B, C, D.

Per rispondere, devi mettere una crocetta nel quadratino accanto alla risposta (una sola) che ritieni giusta, come nell'esempio seguente.

Esempio 1

Quanti giorni ci sono in una settimana?

A. ☒ Sette

B. ☐ Sei

C. ☐ Cinque

D. ☐ Quattro

Se ti accorgi di aver sbagliato, puoi correggere: devi scrivere **NO** accanto alla risposta sbagliata e mettere una crocetta nel quadratino accanto alla risposta che ritieni giusta, come nell'esempio seguente.

Esempio 2

Quanti minuti ci sono in un'ora?

NO A. ☒ 30 minuti

B. ☐ 50 minuti

C. ☒ 60 minuti

D. ☐ 100 minuti

Altre domande chiedono di scrivere la risposta o il procedimento, oppure prevedono una diversa modalità di risposta. In questo caso il testo della domanda ti dice come rispondere. Leggilo dunque sempre con molta attenzione.

Puoi usare il righello graduato, la squadra, il compasso e il goniometro ma non la calcolatrice.

Non scrivere con la matita, ma usa soltanto una penna nera o blu.

Puoi usare le pagine bianche del fascicolo o gli spazi bianchi accanto alle domande per fare calcoli o disegni.

Per fare una prova, ora rispondi a questa domanda.

In quale delle seguenti sequenze i numeri sono scritti dal più grande al più piccolo?

A. ☐ 2; 5; 4; 8

B. ☐ 8; 5; 4; 2

C. ☐ 2; 4; 8; 5

D. ☐ 2; 4; 5; 8

Hai a disposizione 1 ora e quindici minuti (in totale 75 minuti) per rispondere alle domande. L'insegnante ti dirà quando cominciare a lavorare. Quando l'insegnante ti comunicherà che il tempo è finito, posa la penna e chiudi il fascicolo.

Se finisci prima, puoi chiudere il fascicolo e aspettare la fine, oppure puoi controllare le risposte che hai dato.

Caro studente,
ti chiediamo di rispondere alle seguenti domande.

Q1. Tu sei: maschio □ femmina □

Q2. Tu sei nato nel mese di ________________ dell'anno ________________

Q3. Hai frequentato la scuola materna?
a. No
b. Sì, per più di un anno
c. Sì, per meno di un anno

Q4. Hai frequentato l'asilo nido?
a. No
b. Sì, per più di un anno
c. Sì, per meno di un anno

Q5. Tu sei nato
a. In Italia
b. All'estero (scrivi sui puntini in quale Paese sei nato): ______________________________

Q6. Se non sei nato in Italia, quanti anni avevi quando sei arrivato in Italia? _______________ anni

Q7. In quale Paese è nata la tua mamma? __________________________

Q8. In quale Paese è nato il tuo papà? __________________________

Q9. Qual è il titolo di studio conseguito dai tuoi genitori?

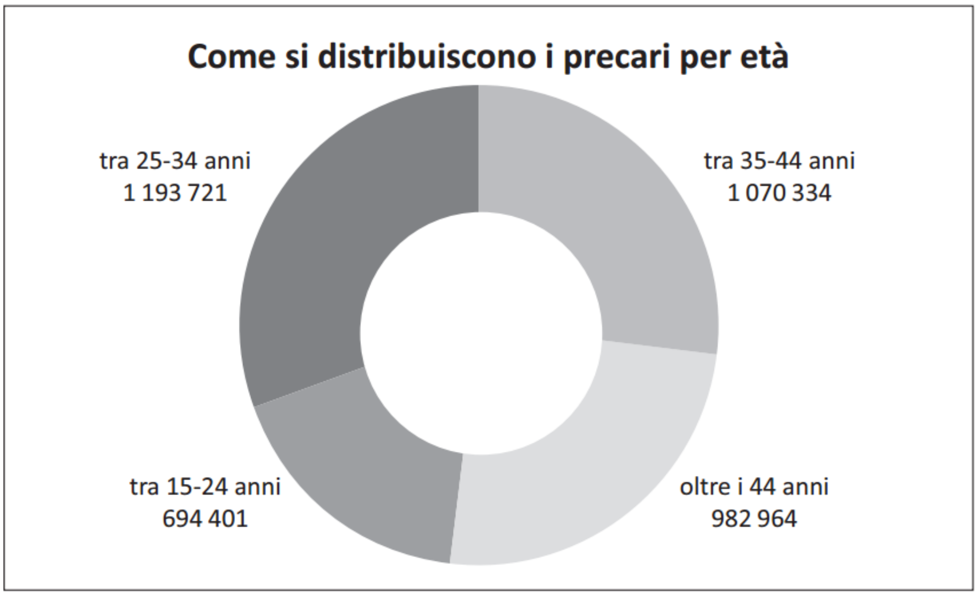
	madre	padre
A. Licenza elementare		
B. Licenza media		
C. Qualifica professionale triennale		
D. Diploma di scuola secondaria superiore (liceo, istituto tecnico o istituto professionale)		
E. Titolo di studio superiore al diploma, diverso dalla laurea (ISEF, Accademia di Belle Arti, Conservatorio)		
F. Laurea / Dottorato di ricerca / Master		
G. Non me lo ricordo		

Q10. Che lavoro fanno i tuoi genitori?

	madre	padre
A. Disoccupato/a		
B. Si occupa della casa		
C. Dirigente, docente universitario, funzionario, ufficiale militare		
D. Imprenditore, proprietario agricolo		
E. Professionista dipendente, sottufficiale militare, libero professionista (psicologo, ricercatore, medico, avvocato, commissario di polizia, ecc.)		
F. Lavoratore in proprio (commerciante, artigiano, coltivatore diretto, meccanico, sarto, ecc.)		
G. Insegnante, impiegato, militare graduato		
H. Operaio, addetto ai servizi, socio di cooperativa (tecnico, infermiere, cameriere, commessa, ecc.)		
I. Pensionato/a		
L. Non me lo ricordo		

NON GIRARE LA PAGINA
FINCHÉ NON TI SARÀ DETTO DI FARLO

A1. Il seguente grafico rappresenta la distribuzione dei lavoratori precari in Italia suddivisi per età nell'anno 2012.



a. Quanti sono in totale i precari?

A. ☐ **Circa due milioni**

B. ☐ **Circa tre milioni**

C. ☐ **Circa quattro milioni**

D. ☐ **Circa cinque milioni**

b. Quale percentuale rappresentano i precari che hanno tra i 25 e i 34 anni?

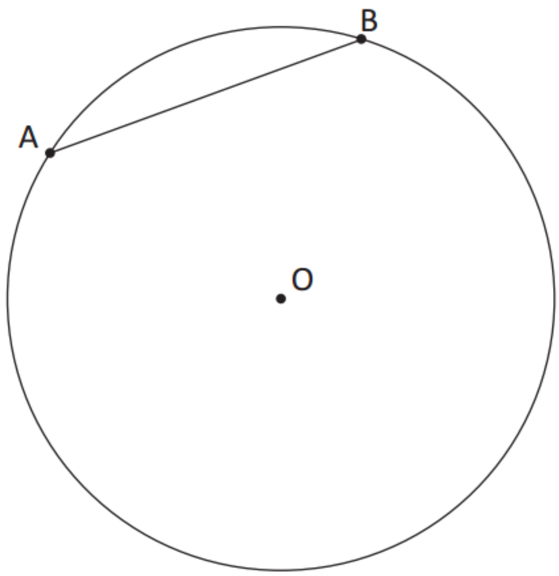
A. ☐ **Circa il 50%**

B. ☐ **Circa il 40%**

C. ☐ **Circa il 30%**

D. ☐ **Circa il 20%**

A2. **Osserva la figura. AB è un cateto di un triangolo rettangolo inscritto nella circonferenza di centro O. Disegna il triangolo rettangolo.**



A3. **Se *n* é un numero naturale, allora il numero *n · (n + 2)***

A. □ **è sempre dispari**

B. □ **è sempre pari**

C. □ **è dispari se n è pari**

D. □ **è dispari se n è dispari**

A4. **Tempo fa si è disputata la partita di pallacanestro B. Pozzo di Gotto - Brescia, finita con il punteggio di 92-94.**

La seguente tabella riassume le statistiche di tale partita per la squadra di Brescia.

		Tiri a canestro			
Nome del giocatore	**Minuti giocati**	**Tiri da 2**	**Tiri da 3**	**Tiri liberi**	**PUNTI**
Roberto	25	0	0	2	2
Clelia	23	4	0	1	9
Chiara	20	2	0	0	4
Giorgio	36	2	1	7	14
Cristina	37	3	1	1	10
Monica	30	9	1	8	29
Paolo	9	0	1	2	5
Anna	15	0	1	0	3
Maria	30	6	0	6	18
	Totale	**26**	**5**	**27**	**94**

Quanti sono i giocatori che hanno realizzato un numero di punti superiore alla media?

Risposta: ……………………………………………………………………………………………………

A5. **Il Signor Carlo scende dal tram all'incrocio di *via Pietro Micca* con *via Antonio Giuseppe Bertola* (nella mappa che vedi qui sotto il punto è contrassegnato da un asterisco).**

Percorre 200 metri di *via Bertola* e all'incrocio con *via 20 Settembre* svolta a sinistra; dopo aver camminato per 150 metri, raggiunge l'incrocio con *via Pietro Micca*. Da lì decide di tornare al punto di partenza per *via Pietro Micca*.

Quanti metri all'incirca percorre al ritorno?

A. □ **200 m**

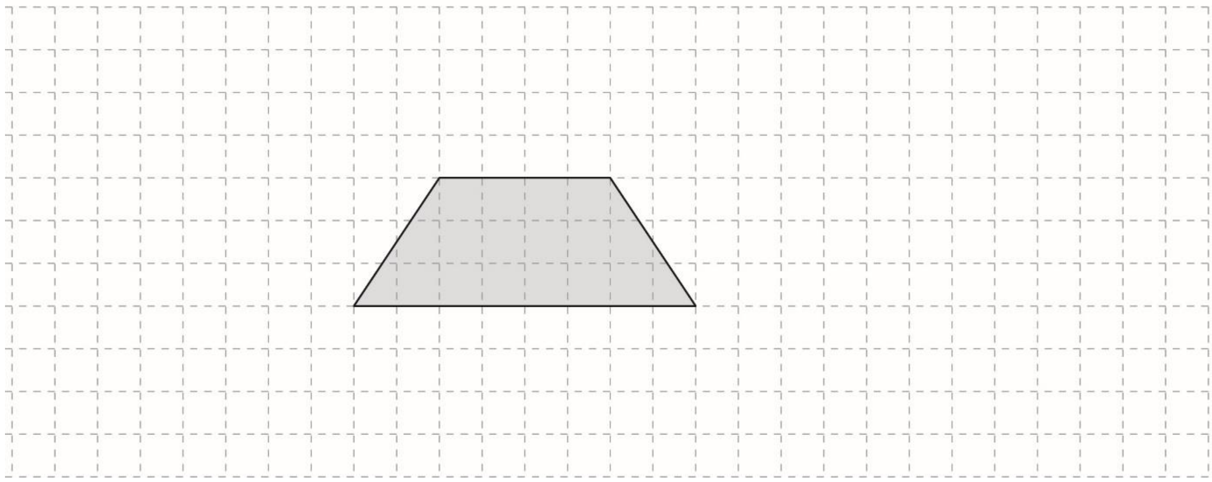
B. □ **250 m**

C. □ **350 m**

D. □ **600 m**

A6. **Il trapezio che vedi sotto è stato ritagliato da una figura F più grande. Il trapezio è i $\frac{3}{4}$ della figura F.**

Disegna una delle possibili figure F da cui il trapezio è stato ritagliato.



D7_PN2011

Anch_1. Antonio e Giada partecipano a una gara a quiz. Per ogni risposta esatta si assegnano due punti mentre per ogni risposta sbagliata si toglie un punto. L'esito della gara è il seguente:

- **Antonio ha dato 11 risposte esatte e 9 sbagliate;**
- **Giada ha dato 6 risposte esatte e 14 sbagliate.**

Quali sono i punteggi finali dei due ragazzi?

A. □ **+ 13; +2**

B. □ **+ 13; −2**

C. □ **+ 2; + 8**

D. □ **+ 2; − 8**

D13_PN2011_v2

D1. Il numero $\sqrt{30}$ è:

A. □ **compreso tra 29 e 31**

B. □ **uguale a 15**

C. □ **compreso tra 5 e 6**

D. □ **uguale a 900**

D2. Un bicchiere contiene 0,25 litri di acqua.

Se si vuole riempire una bottiglia da 1,5 litri, quanti bicchieri di acqua bisogna versare nella bottiglia?

Risposta: ………………………………….

D15_PN2012_v3

D3. **Un maglione nel mese di dicembre costa 100 euro. A gennaio, nel periodo dei saldi il suo prezzo viene diminuito del 20% rispetto al prezzo di dicembre, mentre a febbraio il suo prezzo viene aumentato del 20% rispetto al prezzo di gennaio.**

A febbraio quindi il suo prezzo, rispetto a dicembre, è

A. □ **diminuito del 4%**

B. □ **diminuito del 10%**

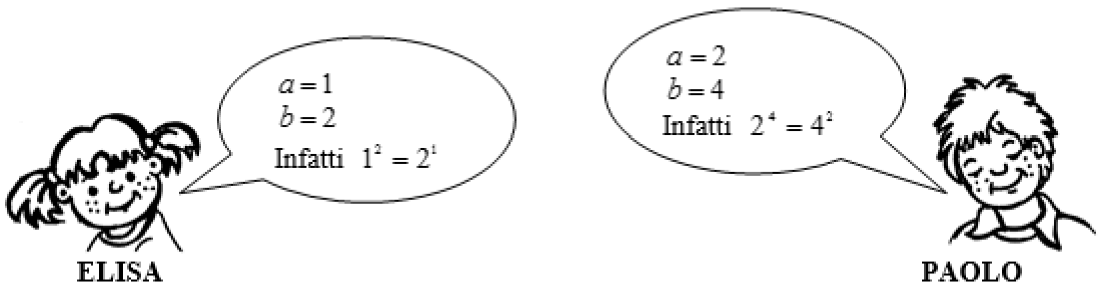
C. □ **aumentato del 4%**

D. □ **aumentato del 10%**

D18_PN2011

Anch_3. Elisa e Paolo stanno cercando di rispondere a questa domanda: "***Qual è la coppia di numeri interi a, b (diversi fra loro) tali che*** $a^b = b^a$ ***?***"

Ecco le loro soluzioni.



Chi ha ragione?

A. □ **Solo Elisa**

B. □ **Solo Paolo**

C. □ **Entrambi**

D. □ **Nessuno dei due**

D5. **Chiara ha un banco rettangolare e un pacco di fogli rettangolari. Il lato lungo del banco è 5 volte il lato lungo dei fogli, e il lato corto del banco è 5 volte il lato corto dei fogli. Quanti fogli come minimo le occorrono per ricoprire interamente la superficie del banco?**

A. □ **25**

B. □ **20**

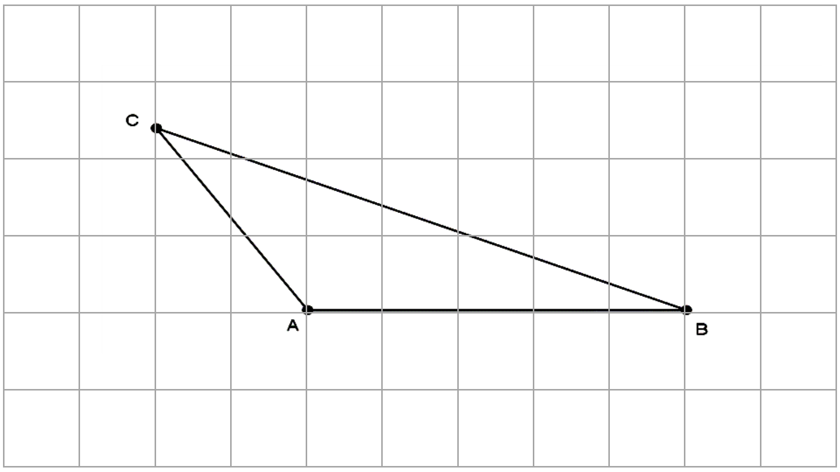
C. □ **10**

D. □ **5**

D6. Un rettangolo ha il perimetro di 1 metro. Il lato corto del rettangolo misura 20 cm. Quanto misura il lato lungo?

Risposta: …………………… cm

D7. Osserva il disegno.



a. **Calcola l'area del triangolo ABC prendendo con un righello le misure necessarie.**

Risposta: …………. cm^2

b. **Scrivi i calcoli che hai fatto per arrivare alla risposta.**

……

……

……

……

……

D8. Nina sale su un treno composto dalla locomotiva e da 9 vagoni:

Quanto è lungo all'incirca il treno di Nina?

A. □ **Circa 10 m**

B. □ **Circa 50 m**

C. □ **Circa 250 m**

D. □ **Circa 1000 m**

D27_PN2013

Anch_7. Nella scuola Nino Bixio ci sono 600 studenti e un insegnante ogni 15 studenti.

Quale proporzione permette di trovare il numero x degli insegnanti?

A. □ **x : 15 = 1 : 600**

B. □ **15 : 1 = x : 600**

C. □ **1 : 15 = x : 600**

D. □ **x : 1 = 15 : 600**

1CG_NEW_v2

D9. Qual è il risultato di 0,5 × 4 ?

A. □ **8**

B. □ **4**

C. □ **2**

D. □ **20**

Anch_4. La formula $L = L_0 + K \times P$ esprime la lunghezza L di una molla al variare del peso P applicato. L_0 rappresenta la lunghezza in centimetri "a riposo" della molla; K indica di quanto si allunga in centimetri la molla quando le si applica una unità di peso.

Quale delle formule elencate si adatta meglio alla seguente descrizione: "*È una molla molto corta e molto dura (cioè molto resistente alla trazione)*"?

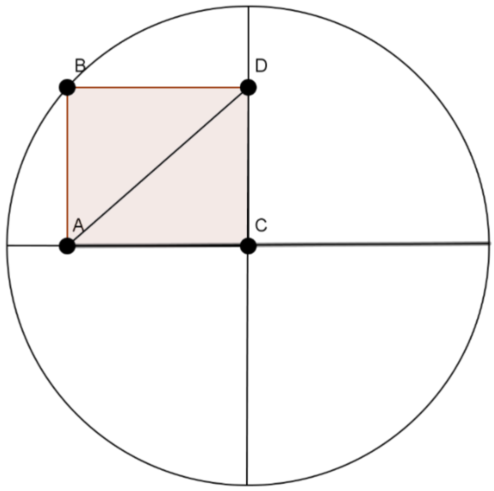
A. □ $L = 10 + 0,5 \times P$

B. □ $L = 10 + 7 \times P$

C. □ $L = 80 + 0,5 \times P$

D. □ $L = 80 + 7 \times P$

D10. Osserva la figura.



Sapendo che la circonferenza rappresentata in figura ha centro C e raggio 4 cm, Mario afferma che il segmento AD misura 4 cm, tenendo conto che CABD è un rettangolo.

Mario ha ragione? Scegli una risposta con il suo perché.

Mario NON HA ragione perché		**Mario HA ragione perché**	
La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è diverso da 4	**Non ci sono abbastanza dati per calcolare la lunghezza di AD**	**La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è 4**	**Le diagonali di un rettangolo sono congruenti tra di loro**
□ A	□ B	□ C	□ D

D26_PN2015

Anch_8. Un gruppo di 20 amici va in pizzeria. Ciascuno di essi ordina una pizza che costa 8 euro.
Ogni 5 pizze ordinate, il proprietario non ne fa pagare una.
Quanto spendono in tutto gli amici per le pizze?

A. ☐ **160 euro**

B. ☐ **128 euro**

C. ☐ **120 euro**

D. ☐ **112 euro**

D25_PN2011

Anch_5. Quale fra le seguenti disuguaglianze è quella corretta?

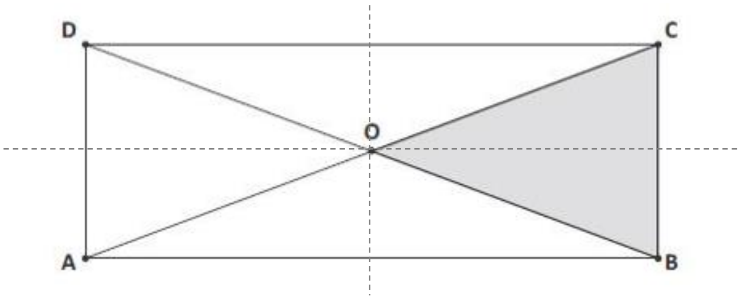
A. ☐ $\frac{3}{10} < \frac{3}{5} < \frac{3}{20}$

B. ☐ $\frac{4}{10} < \frac{3}{5} < \frac{11}{20}$

C. ☐ $\frac{5}{10} < \frac{3}{5} < \frac{13}{20}$

D. ☐ $\frac{7}{10} < \frac{3}{5} < \frac{13}{20}$

D11. In figura è rappresentato il rettangolo ABCD con le sue diagonali e i suoi assi di simmetria. Se conosci l'area del rettangolo, puoi calcolare l'area del triangolo COB in grigio?



A. □ **No, perché i quattro triangoli di vertice O non sono tutti uguali fra loro**

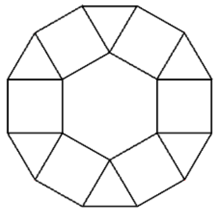
B. □ **No, perché non conosco le dimensioni del rettangolo**

C. □ **Sì, perché i quattro triangoli AOB, BOC, COD, DOA sono equivalenti**

D. □ **Sì, perché i quattro triangoli AOB, BOC, COD, DOA sono isosceli**

D9b_PN2011

Anch_2. Le immagini che seguono rappresentano un motivo del pavimento di una antica casa romana e la sua schematizzazione geometrica:



Il motivo, corrispondente a un dodecagono, è composto da un esagono regolare interno, sei quadrati uguali e sei triangoli equilateri uguali.

L'area dell'esagono è metà dell'area del dodecagono?
VERO □ FALSO □

L'area di ciascun triangolo è un sesto dell'area dell'esagono?
VERO □ FALSO □

D5_PN2011_v4

D12. Giovanni e Caterina contano le macchine che passano davanti al cancello di casa. Giovanni ha cominciato più tardi a contare. Quando Giovanni ha contato 10 macchine, Caterina ne ha contate 30. Quando rientrano in casa, Giovanni ha contato 50 macchine; quante ne ha contate Caterina?

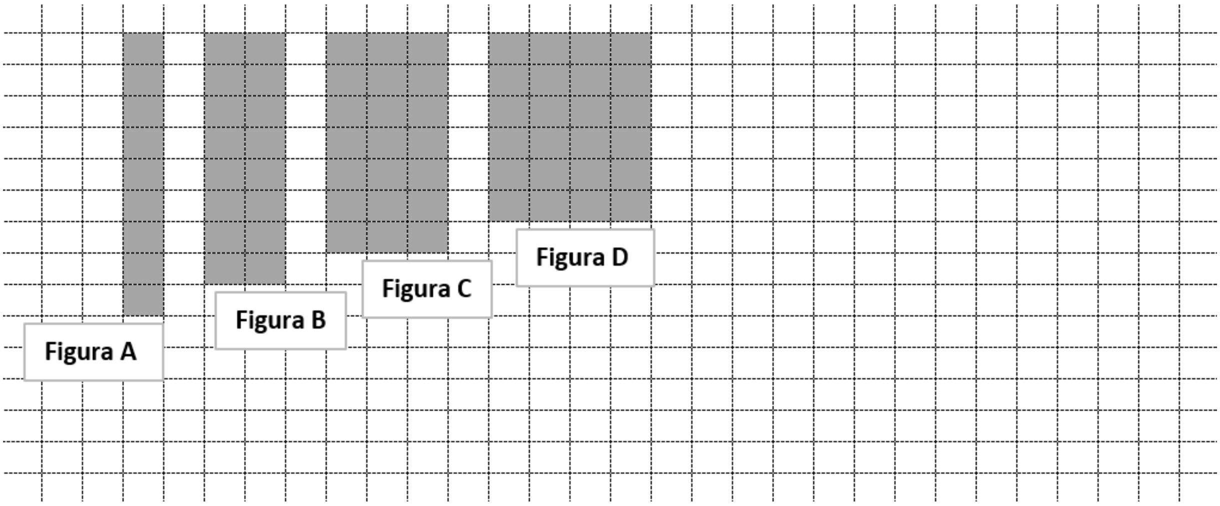
Risposta: .... macchine

D13. Quale numero intero, messo al posto di x, rende vere entrambe le seguenti disuguaglianze?

$$\frac{2}{5} < \frac{x}{10}$$

$$\frac{x}{10} < \frac{3}{5}$$

D14. Osserva la seguente sequenza di figure:



Quale delle seguenti affermazioni è vera?

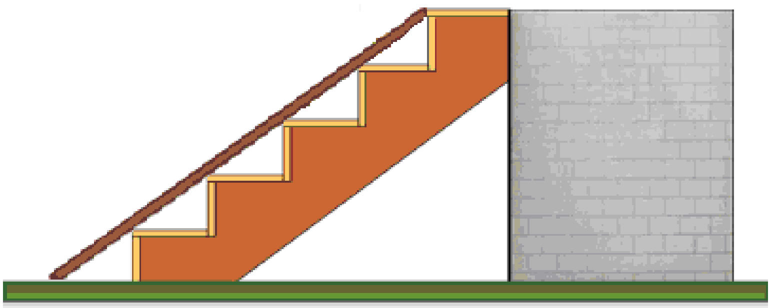
A. □ **Le aree delle figure restano sempre uguali**

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

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

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

Anch_6. Una scala, costituita da 5 gradini profondi 24 cm e alti 18 cm l'uno, deve essere coperta da una tavola di legno utilizzata come scivolo per il trasporto di alcune merci. Qual è il procedimento corretto per trovare la lunghezza dello scivolo?



A. ☐ $\left(\sqrt{18^2}+\sqrt{24^2}\right)\times 5$

B. ☐ $\sqrt{(24+18)^2}\times 5$

C. ☐ $\sqrt{24^2+18^2}\times 5$

D. ☐ $\sqrt{\left(24^2+18^2\right)\times 5}$

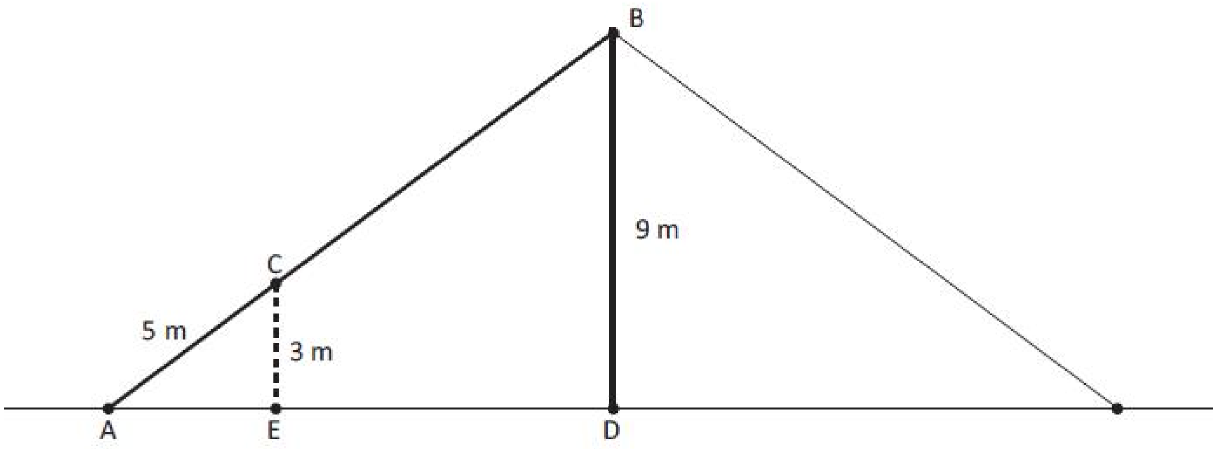
3CG_NEW_inv

D15. In ogni coppia di operazioni cerchia quella che dà il risultato minore:

8×4	oppure	$8 : 4$
$8 \times 0{,}4$	oppure	$8 : 0{,}4$
$0{,}8 \times 0{,}4$	oppure	$0{,}8 : 0{,}4$

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

D17. **Giulio sa che nel negozio A e nel negozio B le bottiglie di olio della marca che preferisce hanno lo stesso prezzo.**

Oggi, su quell'olio, i due negozi fanno due promozioni diverse.

Giulio deve comprare 3 bottiglie d'olio.

a. **In quale negozio gli conviene comprarle?**

Risposta: ……………………………….

b. **Scrivi come hai fatto per trovare la risposta.**

………

………

………

Grazie per aver risposto alle domande precedenti. Ti chiediamo ora, di completare la seguente sezione ricordandoti che, in questa sezione, non esistono risposte giuste o sbagliate. Leggi, quindi, tutto con attenzione e rispondi con sincerità!

Parte A – Il tuo rapporto con la scuola

A1) Qual è la materia che a scuola studi con più piacere?
Risposta: ____________________

A2) Qual è invece la materia che studi con meno piacere?
Risposta: ___________________

A3) Quanto pensi di essere bravo nelle seguenti materie? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= non sono bravo)	2 (= sono nella media)	3 (= sono bravo)	4 (= sono bravissimo)
Italiano	1	2	3	4
Matematica	1	2	3	4
Scienze	1	2	3	4

A4) Rispetto allo scorso anno, pensi che la tua abilità in Italiano, Matematica e Scienze sia cambiata? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= sono peggiorato)	2 (= non è cambiato nulla)	3 (= sono migliorato)
Italiano	1	2	3
Matematica	1	2	3
Scienze	1	2	3

A5) Quanto spesso i tuoi genitori o le persone che si prendono cura di te fanno le seguenti cose? Metti una crocetta in corrispondenza della risposta esatta, per ciascuna riga della seguente tabella:

	mai	raramente	qualche volta	spesso	sempre
Controllare se hai fatto i compiti					
Aiutarti a fare i compiti					
Elogiarti o ricompensarti per aver avuto buoni voti					
Ridurre le tue ricompense quando prendi un brutto voto					
Trovare qualcuno che ti aiuti a fare i compiti					

Parte B – Cosa pensi della Matematica?

B1) Quanto ti senti in accordo o in disaccordo con le seguenti affermazioni? Per ciascuna riga della seguente tabella, cerchia il numero che rispecchia meglio quello che pensi sapendo che:
1 = totalmente in disaccordo;
2 = parzialmente in disaccordo;
3 = non so;
4 = parzialmente d'accordo;
5 = totalmente d'accordo.

1. La Matematica per me è importante	**1**	**2**	**3**	**4**	**5**
2. La Matematica la possono capire ed imparare in tanti	**1**	**2**	**3**	**4**	**5**
3. Ai miei genitori e/o a chi si prende cura di me piace la Matematica	**1**	**2**	**3**	**4**	**5**
4. La Matematica è una della materie più interessanti che studio a scuola	**1**	**2**	**3**	**4**	**5**
5. Studiare la Matematica è divertente	**1**	**2**	**3**	**4**	**5**
6. Io ho una mente matematica	**1**	**2**	**3**	**4**	**5**
7. Sono in grado di prendere buoni voti in Matematica	**1**	**2**	**3**	**4**	**5**
8. Prendere buoni voti in Matematica mi interessa	**1**	**2**	**3**	**4**	**5**
9. In Matematica, il mio impegno viene ricompensato	**1**	**2**	**3**	**4**	**5**
10. Essere bravo in Matematica è un talento di cui bisogna essere dotati alla nascita	**1**	**2**	**3**	**4**	**5**
11. Posso imparare la Matematica anche se è una materia difficile	**1**	**2**	**3**	**4**	**5**
12. Mi piace utilizzare la Matematica che già conosco bene piuttosto che usare concetti matematici che nuovi o che ho studiato da poco	**1**	**2**	**3**	**4**	**5**
13. Sono più preoccupato/a per lo studio della Matematica che per altre materie	**1**	**2**	**3**	**4**	**5**
14. Ho spesso bisogno di aiuto con la Matematica	**1**	**2**	**3**	**4**	**5**
15. Rispetto ai miei compagni di classe, sono bravo/a in Matematica	**1**	**2**	**3**	**4**	**5**
16. Ai miei parenti e/o a chi si prende cura di me piace risolvere problemi matematici	**1**	**2**	**3**	**4**	**5**
17. Spero di non dover studiare Matematica mai più in vita mia	**1**	**2**	**3**	**4**	**5**
18. Mi piacerebbe che i miei studi futuri includessero molta Matematica	**1**	**2**	**3**	**4**	**5**
19. Mi piacerebbe tanto studiare più Matematica a scuola	**1**	**2**	**3**	**4**	**5**
20. Da grande, mi piacerebbe fare della Matematica il mio mestiere	**1**	**2**	**3**	**4**	**5**
21. La Matematica è importante per il mio futuro (dopo la scuola)	**1**	**2**	**3**	**4**	**5**

Parte C – Per favore, dicci quanto spesso fate in classe le seguenti attività. Cerchia il numero più appropriato, riga per riga, sapendo che 1 = mai; 2 = raramente; 3 = qualche volta; 4 = sempre.

	1 = mai	2 = raramente	3 = qualche volta	4 = sempre
C1. L'insegnante ci fa delle domande	1	2	3	4
C2. L'insegnante ci chiede di spiegare il ragionamento che abbiamo fatto per arrivare al risultato	1	2	3	4
C3. L'insegnante inizia a spiegare gli argomenti nuovi partendo da esempi tratti dal mondo reale	1	2	3	4
C4. L'insegnante ci dice di lavorare più velocemente	1	2	3	4
C5. L'insegnante utilizza il computer per spiegarci alcuni argomenti	1	2	3	4
C6. L'insegnante ci dà spesso problemi da esaminare	1	2	3	4
C7. L'insegnante si aspetta che ci ricordiamo cose imparate in passato	1	2	3	4
C8. L'insegnante ci dice dettagliatamente quali attività fare	1	2	3	4
C9. L'insegnante ci chiede se conosciamo già qualcosa dell'argomento che sta per spiegarci	1	2	3	4
C10. L'insegnante ci spiega qual è il valore futuro delle cose che ci spiega	1	2	3	4
C11. Lavoriamo spesso in gruppo	1	2	3	4
C12. Ascoltiamo l'insegnante spiegare la lezione	1	2	3	4
C13. Copiamo ciò che l'insegnante scrive sulla lavagna	1	2	3	4
C14. Ci confrontiamo spesso l'un l'altro sulle strategie di soluzione dei problemi	1	2	3	4
C15. Chiediamo agli altri studenti di spiegare le loro idee e il loro punto di vista	1	2	3	4
C17. Facciamo gli esercizi riportati nel libro	1	2	3	4
C18. Studiamo come alcuni concetti matematici si sono evoluti nel tempo	1	2	3	4
C19. Quello che impariamo a scuola è collegato con la nostra vita al di fuori della scuola	1	2	3	4
C20. Abbiamo imparato che la Matematica di basa su regole inventate	1	2	3	4
C21. Facciamo ricerche su argomenti che scegliamo noi	1	2	3	4
C22. Usiamo la calcolatrice	1	2	3	4
C23. Usiamo i computer	1	2	3	4
C24. In classe, usiamo riviste, giornali o video	1	2	3	4
C25. Discutiamo le nostre idee con l'intera classe	1	2	3	4
C26. Esponiamo il nostro lavoro all'intera classe	1	2	3	4

C27. Se utilizzate il computer, che cosa lo utilizzate prevalentemente?

Risposta: ____

____

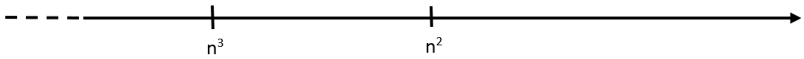
____

Parte D – Quanto ti senti sicuro/a di saper risolvere alcuni specifici problemi matematici

In questa sezione, ti chiediamo di fare una valutazione di quanto ti senti sicuro/a di saper risolvere alcuni problemi matematici. **Non ti chiediamo di fornire una risposta per ciascun quesito ma solo di dirci se e quanto pensi che saresti in grado di rispondere correttamente alle singole domande**. **Leggi quindi con grande attenzione e rispondi con la massima sincerità!**

D1. Quanto ti senti sicuro/a di saper utilizzare i numeri razionali, le loro diverse rappresentazioni, stimare la loro grandezza e/o il risultato di operazioni con i numeri razionali), come nella domanda riportata qui sotto?

Sulla seguente retta dei numeri sono ordinate due potenze di un numero razionale n.



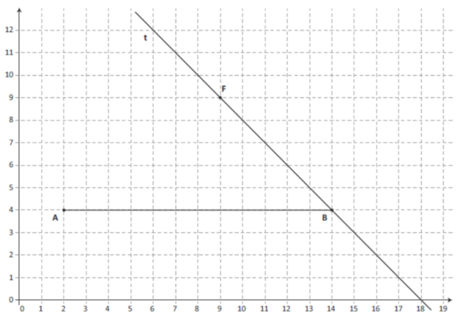
Indica con una crocetta se ciascuna delle seguenti affermazioni è vera (V) o falsa (F).

	SI	NO
Il valore di n può essere +1/2		
Il valore di n può essere -1/2		
Il valore di n può essere +3/2		
Il valore di n può essere -3/2		

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D2. Quanto ti senti sicuro/a di saper riconoscere le forme del piano e dello spazio, le loro rappresentazioni e di saper cogliere relazioni tra gli elementi, come nella domanda riportata qui sotto?

Osserva la figura

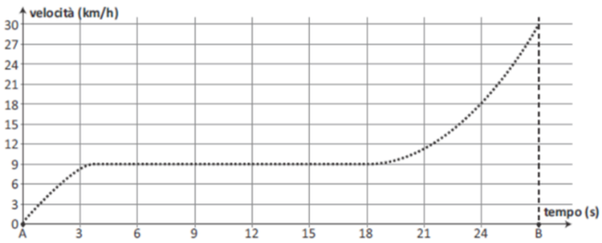


Sei in grado di disegnare la retta s perpendicolare a t passante per f e di individuare le coordinate del punto F di intersezione tra la retta *s* ed il segmento AB?

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D3. Quanto ti senti sicuro/a di saper calcolare misure statistiche di sintesi (ad esempio, la media, la moda, la mediana, ecc.), come nella domanda riportata qui sotto?

Luca percorre una strada in bicicletta e, con l'aiuto del computer, registra la propria velocità ogni decimo di secondo. Il grafico in figura rappresenta le diverse velocità raggiunte da Luca al passare del tempo.



Sei in grado di calcolare il valore modale delle velocità raggiunte da Luca dall'istante A all'istante B?

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D4. Quanto ti senti sicuro/a di saper risolvere problemi di misura e stima, come nella domanda riportata qui sotto?

Osserva l'edificio

Quanto può essere alto l'edificio?

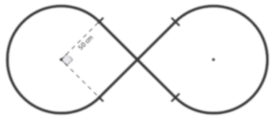
A. ☐ Meno di 10 metri
B. ☐ Tra 15 e 20 metri
C. ☐ Tra 25 e 30 metri
D. ☐ Più di 35 metri

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D5. Quanto ti senti sicuro/a di risolvere un problema matematico e di esplicitare in forma scritta il ragionamento che hai fatto, come nella domanda riportata di seguito?

La figura rappresenta lo schema di una pista formata da:

- Due archi di circonferenza di raggio 50 cm;
- Due tratti rettilinei di 100 cm ciascuno, perpendicolari tra loro nel punto medio.



Qual è la lunghezza della pista?

Scrivi i calcoli che fai per trovare la risposta e, infine, riporta il risultato.

____

____

____

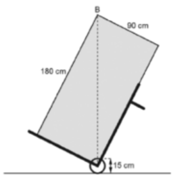
____

Risultato: circa ____________ cm

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D6. Quanto ti senti sicuro/a nel risolvere problemi come quello presentato nella seguente domanda?

Gabriele ha comprato un nuovo frigorifero. Per portarlo in cucina usa un carrello, come rappresentato in figura.



Quale espressione di permette di calcolare la massima distanza dal suolo del punto B quando il frigorifero è trasportato sul carrello?

A. □ $\sqrt{180^2 + 90^2} + 7{,}5$

B. □ $\sqrt{180^2 - 90^2} + 7{,}5$

C. □ $\sqrt{180 + 90} + 7{,}5$

D. □ $\sqrt{180^2 + 90^2} + 7{,}5$

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D7. Quanto ti senti sicuro/a di saper risolvere problemi di probabilità, come quello presentato nella seguente domanda?

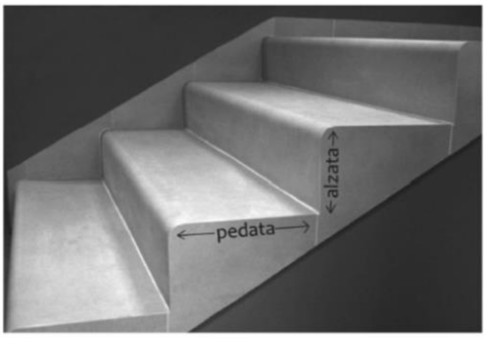
Giuseppe mette 2 palline bianche e 1 pallina nera in una busta.

Senza guardare, estrae una pallina. Qual è la probabilità che la pallina sia nera?

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D8. Quanto ti senti sicuro/a di saper utilizzare la matematica nella vita di tutti i giorni per rispondere a domande concrete, come quello presentato nella seguente domanda?

Nel disegno sottostante è rappresentata una scala.



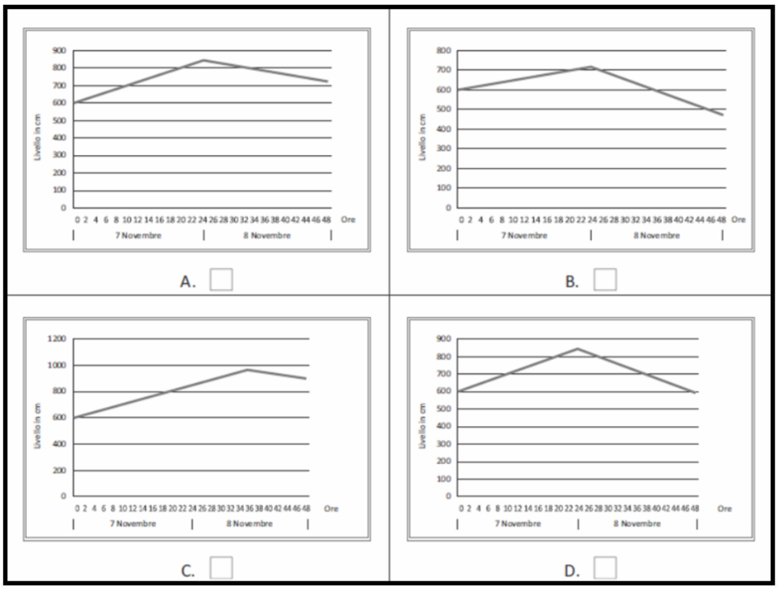
Per legge, la pedata deve essere lunga almeno 30 cm e la somma tra il doppio dell'alzata e la pedata deve essere compresa tra 62 e 64 cm (estremi inclusi).

Se la pedata di una scala misura 34 cm, il doppio dell'alzata dovrà essere compreso tra 28 e ____________ cm e, quindi, l'alzata dovrà essere compresa tra 14 cm e ____________ cm.

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D9. Quanto ti senti sicuro/a di saper risolvere problemi come quello presentato nella seguente domanda?

Oggi, il livello dell'acqua di un fiume è aumentato di circa 10 cm all'ora, per tutte le 24 ore. Il giorno successivo, il livello dell'acqua è diminuito di circa 5 cm all'ora per tutte le 24 ore. Quale tra i seguenti grafici può rappresentare meglio la situazione descritta?



Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

Parte E – Qual è il tuo rapporto con lo studio della Matematica?

Leggi attentamente le seguenti affermazioni e ripensa all'esperienza che hai avuto nel tuo percorso scolastico. Per ciascuna situazione, indica il livello di ansia che hai provato utilizzando una scala con 5 possibili risposte: da 1 (=poca ansia) a 5 (=molta ansia).

	Poca ansia				*Molta ansia*
1. Dover consultare le tavole in un manuale di matematica.	1	2	3	4	5
2. Pensare al compito di matematica del giorno dopo.	1	2	3	4	5
3. Guardare la risoluzione di un'equazione svolta dall'insegnante alla lavagna.	1	2	3	4	5
4. Fare il compito di matematica.	1	2	3	4	5
5. Avere da fare molti problemi difficili di matematica per la lezione successiva.	1	2	3	4	5
6. Ascoltare una spiegazione durante l'ora di matematica.	1	2	3	4	5
7. Ascoltare uno altro studente che spiega una formula di matematica.	1	2	3	4	5
8. Dover fare una verifica a sorpresa durante l'ora di matematica.	1	2	3	4	5
9. Iniziare un nuovo capitolo del manuale di matematica.	1	2	3	4	5

Parte F – Come sono i tuoi rapporti con tuoi compagni di scuola e di classe?

Questa è l'ultima sezione a cui ti chiediamo di rispondere. Le domande che ti facciamo, ci servono per capire quali sono i rapporti tra te e i tuoi compagni di classe. Prova quindi a pensare alla tua giornata a scuola e al rapporto che hai con i tuoi compagni, al tempo che trascorri con loro e alle attività che fate insieme. Nella maggior parte dei casi, non ti chiederemo il nome dei tuoi compagni ma solo il loro numero. Nelle ultime sette domande,

invece, ti chiediamo il tuo nome e quello di alcuni dei tuoi compagni. **In nessun caso, ti chiediamo il tuo cognome o quello dei tuoi compagni. Ad ogni modo, scegli tu a quali domande vuoi rispondere e quali invece preferisci lasciare in bianco!**

F1) Quanti studenti ci sono nella tua classe?

Risposta: _________

F2) Con quanti dei tuoi compagni di classe ti fa piacere chiacchierare durante la ricreazione, l'ora di ginnastica o quando possibile durante l'orario scolastico?

Risposta: _________

F3) Con quanti dei tuoi compagni di classe ti vedi fuori dalla scuola?

Risposta: _________

F4) Quanti dei tuoi compagni sono tuoi amici su Facebook e/o altri social networks?

Risposta: _________

F5) Con quanti compagni di classe hai l'abitudine di tenerti in contatto tramite telefonate, SMS, messaggi whatsapp, chat, ecc.?

Risposta: _________

F6) Quanti dei tuoi compagni vedi abitualmente al di fuori della scuola per

i. fare i compiti. Risposta: ___________

ii. fare attività sportive. Risposta: ___________

iii. uscire insieme (per andare al cinema, a mangiare una pizza, ecc.). Risposta: ___________

F7) Se tu avessi qualche problema, a scuola o nella tua vita privata,

i. con quanti dei tuoi compagni ti confideresti? Risposta: _________

ii. a quanti dei tuoi compagni chiederesti aiuto? Risposta: _________

iii. quanti dei tuoi compagni pensi che ti aiuterebbero? Risposta: _________

F8) Quanti dei tuoi compagni ti chiederebbero aiuto o si confiderebbero con te se avessero dei problemi?

Risposta: _________

F9) Quanti dei tuoi compagni aiuteresti se ti chiedessero aiuto?

Risposta: _________

F1. Qual è il tuo nome di battesimo? Risposta: ___________________

F2. Se si dovesse organizzare una gita, chi sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi, in ordine di preferenza:

1. _________________________
2. _________________________
3. _________________________

F3. Se si dovesse organizzare una gita, chi **non** sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

F4. Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, chi sceglieresti? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

F5. Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, con quale ti piacerebbe non stare? Indica 3 nomi

1. ____________________________
2. ____________________________
3. ____________________________

F6. Abbiamo fatto le stesse domande a tutti i tuoi compagni. Secondo te, tra i tuoi compagni, chi pensi ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

F7. Chi pensi invece che **non** ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ____________________________
2. ____________________________
3. ____________________________

Hai finito!
GRAZIE!!!

Ministero dell'Istruzione dell'Università e della Ricerca

PROGETTO 'VARIAZIONI 2'

Anno Scolastico 2016 – 2017

LIVELLO 8

Prova di Matematica

Scuola Secondaria di primo grado

Classe Terza

Fascicolo 4

Classe: ....

Studente: ....

A cura di
Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e di Formazione

INVALSI

Prova di Matematica - Fascicolo 4

ISTRUZIONI

Troverai nel fascicolo 30 domande di matematica. Alcune domande hanno quattro possibili risposte, ma una sola è quella giusta. Prima di ogni risposta c'è un quadratino con una lettera dell'alfabeto: A, B, C, D.

Per rispondere, devi mettere una crocetta nel quadratino accanto alla risposta (una sola) che ritieni giusta, come nell'esempio seguente.

Esempio 1

Quanti giorni ci sono in una settimana?

A. ☒ Sette

B. ☐ Sei

C. ☐ Cinque

D. ☐ Quattro

Se ti accorgi di aver sbagliato, puoi correggere: devi scrivere **NO** accanto alla risposta sbagliata e mettere una crocetta nel quadratino accanto alla risposta che ritieni giusta, come nell'esempio seguente.

Esempio 2

Quanti minuti ci sono in un'ora?

NO A. ☒ 30 minuti

B. ☐ 50 minuti

C. ☒ 60 minuti

D. ☐ 100 minuti

Altre domande chiedono di scrivere la risposta o il procedimento, oppure prevedono una diversa modalità di risposta. In questo caso il testo della domanda ti dice come rispondere. Leggilo dunque sempre con molta attenzione.

Puoi usare il righello graduato, la squadra, il compasso e il goniometro ma non la calcolatrice.

Non scrivere con la matita, ma usa soltanto una penna nera o blu.

Puoi usare le pagine bianche del fascicolo o gli spazi bianchi accanto alle domande per fare calcoli o disegni.

Per fare una prova, ora rispondi a questa domanda.

In quale delle seguenti sequenze i numeri sono scritti dal più grande al più piccolo?

A. ☐ 2; 5; 4; 8

B. ☐ 8; 5; 4; 2

C. ☐ 2; 4; 8; 5

D. ☐ 2; 4; 5; 8

Hai a disposizione 1 ora e quindici minuti (in totale 75 minuti) per rispondere alle domande. L'insegnante ti dirà quando cominciare a lavorare. Quando l'insegnante ti comunicherà che il tempo è finito, posa la penna e chiudi il fascicolo.

Se finisci prima, puoi chiudere il fascicolo e aspettare la fine, oppure puoi controllare le risposte che hai dato.

Caro studente,
ti chiediamo di rispondere alle seguenti domande.

Q1. Tu sei: maschio □ femmina □

Q2. Tu sei nato nel mese di _________________ dell'anno _________________

Q3. Hai frequentato la scuola materna?
   a. No
   b. Sì, per più di un anno
   c. Sì, per meno di un anno

Q4. Hai frequentato l'asilo nido?
   a. No
   b. Sì, per più di un anno
   c. Sì, per meno di un anno

Q5. Tu sei nato
   a. In Italia
   b. All'estero (scrivi sui puntini in quale Paese sei nato): _______________________________

Q6. Se non sei nato in Italia, quanti anni avevi quando sei arrivato in Italia? ___________________ anni

Q7. In quale Paese è nata la tua mamma? __________________________

Q8. In quale Paese è nato il tuo papà? __________________________

Q9. Qual è il titolo di studio conseguito dai tuoi genitori?

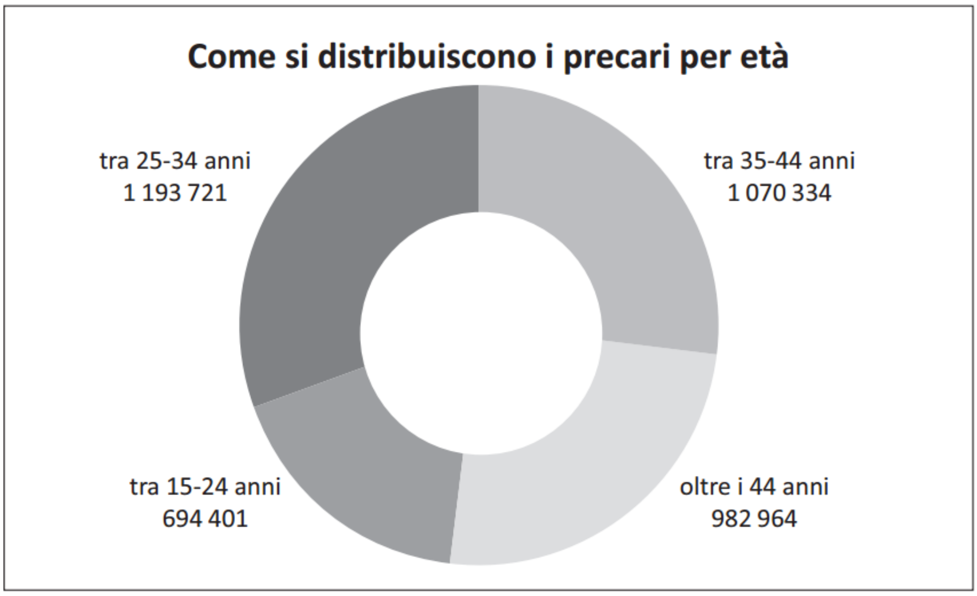
	madre	padre
A. Licenza elementare		
B. Licenza media		
C. Qualifica professionale triennale		
D. Diploma di scuola secondaria superiore (liceo, istituto tecnico o istituto professionale)		
E. Titolo di studio superiore al diploma, diverso dalla laurea (ISEF, Accademia di Belle Arti, Conservatorio)		
F. Laurea / Dottorato di ricerca / Master		
G. Non me lo ricordo		

Q10. Che lavoro fanno i tuoi genitori?

	madre	padre
A. Disoccupato/a		
B. Si occupa della casa		
C. Dirigente, docente universitario, funzionario, ufficiale militare		
D. Imprenditore, proprietario agricolo		
E. Professionista dipendente, sottufficiale militare, libero professionista (psicologo, ricercatore, medico, avvocato, commissario di polizia, ecc.)		
F. Lavoratore in proprio (commerciante, artigiano, coltivatore diretto, meccanico, sarto, ecc.)		
G. Insegnante, impiegato, militare graduato		
H. Operaio, addetto ai servizi, socio di cooperativa (tecnico, infermiere, cameriere, commessa, ecc.)		
I. Pensionato/a		
L. Non me lo ricordo		

NON GIRARE LA PAGINA
FINCHÉ NON TI SARÀ DETTO DI FARLO

A1. **Il seguente grafico rappresenta la distribuzione dei lavoratori precari in Italia suddivisi per età nell'anno 2012.**



a. Quanti sono in totale i precari?

A. ☐ **Circa due milioni**

B. ☐ **Circa tre milioni**

C. ☐ **Circa quattro milioni**

D. ☐ **Circa cinque milioni**

b. Quale percentuale rappresentano i precari che hanno tra i 25 e i 34 anni?

A. ☐ **Circa il 50%**

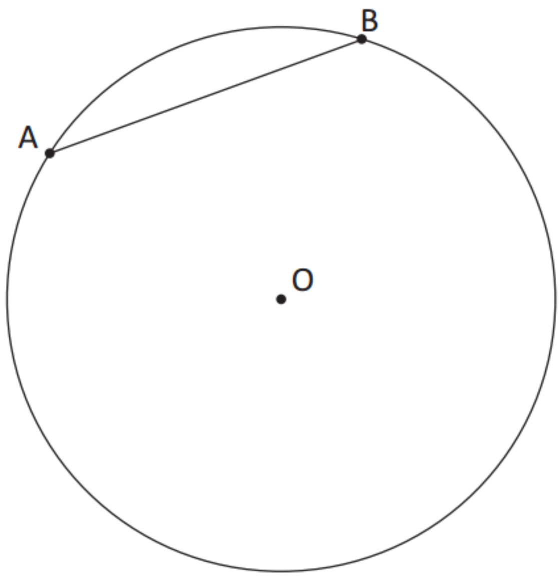
B. ☐ **Circa il 40%**

C. ☐ **Circa il 30%**

D. ☐ **Circa il 20%**

D18_PN2014

A2. **Osserva la figura. AB è un cateto di un triangolo rettangolo inscritto nella circonferenza di centro O. Disegna il triangolo rettangolo.**



D22_PN2013

A3. **Se *n* é un numero naturale, allora il numero** ***n · (n + 2)***

A. □ **è sempre dispari**

B. □ **è sempre pari**

C. □ **è dispari se n è pari**

D. □ **è dispari se n è dispari**

A4. **Tempo fa si è disputata la partita di pallacanestro B. Pozzo di Gotto - Brescia, finita con il punteggio di 92-94.**

La seguente tabella riassume le statistiche di tale partita per la squadra di Brescia.

		Tiri a canestro			
Nome del giocatore	**Minuti giocati**	**Tiri da 2**	**Tiri da 3**	**Tiri liberi**	**PUNTI**
Roberto	25	0	0	2	2
Clelia	23	4	0	1	9
Chiara	20	2	0	0	4
Giorgio	36	2	1	7	14
Cristina	37	3	1	1	10
Monica	30	9	1	8	29
Paolo	9	0	1	2	5
Anna	15	0	1	0	3
Maria	30	6	0	6	18
	Totale	**26**	**5**	**27**	**94**

Quanti sono i giocatori che hanno realizzato un numero di punti superiore alla media?

Risposta: ………………………………………………………………………………………………

A5. Il Signor Carlo scende dal tram all'incrocio di *via Pietro Micca* con *via Antonio Giuseppe Bertola* (nella mappa che vedi qui sotto il punto è contrassegnato da un asterisco).

Percorre 200 metri di *via Bertola* e all'incrocio con *via 20 Settembre* svolta a sinistra; dopo aver camminato per 150 metri, raggiunge l'incrocio con *via Pietro Micca*. Da lì decide di tornare al punto di partenza per *via Pietro Micca*.

Quanti metri all'incirca percorre al ritorno?

A. □ **200 m**

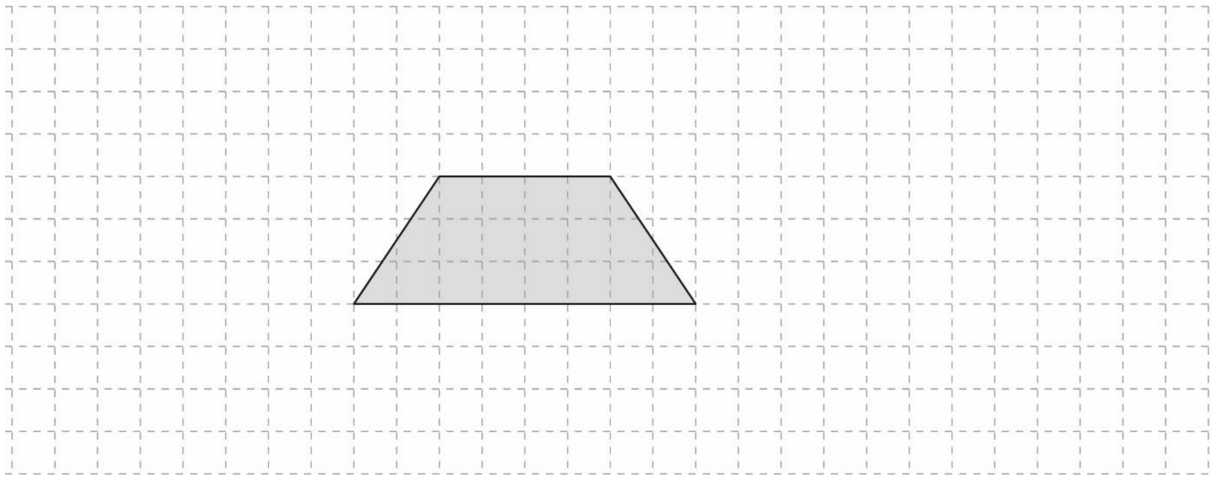
B. □ **250 m**

C. □ **350 m**

D. □ **600 m**

A6. **Il trapezio che vedi sotto è stato ritagliato da una figura F più grande. Il trapezio è i $\frac{3}{4}$ della figura F.**

Disegna una delle possibili figure F da cui il trapezio è stato ritagliato.



Anch_1. Antonio e Giada partecipano a una gara a quiz. Per ogni risposta esatta si assegnano due punti mentre per ogni risposta sbagliata si toglie un punto. L'esito della gara è il seguente:

- **Antonio ha dato 11 risposte esatte e 9 sbagliate;**
- **Giada ha dato 6 risposte esatte e 14 sbagliate.**

Quali sono i punteggi finali dei due ragazzi?

A. □ **+ 13; +2**

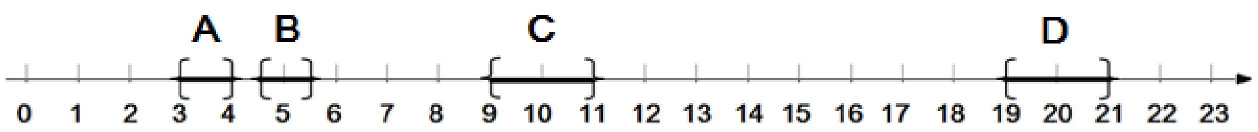
B. □ **+ 13; −2**

C. □ **+ 2; + 8**

D. □ **+ 2; − 8**

D13_PN2011_v3

D1. Osserva la retta.



In quale intervallo si trova il numero $\sqrt{10}$?

A. □ **Intervallo A**

B. □ **Intervallo B**

C. □ **Intervallo C**

D. □ **Intervallo D**

D2. Un bicchiere contiene 1/4 di litro di acqua.

Se si vuole riempire una bottiglia da 1,5 litri, quanti bicchieri di acqua bisogna versare nella bottiglia?

Risposta: …………….

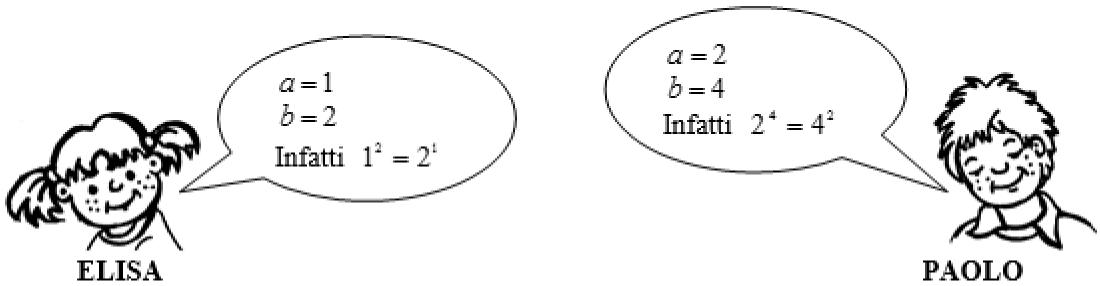
D15_PN2012_v2

D3. Una grande azienda nel 2009 aveva 100 impiegati. Nell'anno 2010 il numero degli impiegati è diminuito del 20% rispetto al 2009 mentre nel 2011 è aumentato del 20% rispetto al 2010. Al termine dei due anni gli impiegati dell'azienda sono

A. □ **96**
B. □ **90**
C. □ **104**
D. □ **110**

D18_PN2011

Anch_3. Elisa e Paolo stanno cercando di rispondere a questa domanda: “***Qual è la coppia di numeri interi a, b (diversi fra loro) tali che $a^b = b^a$?*** ”
Ecco le loro soluzioni.



Chi ha ragione?

A. □ **Solo Elisa**
B. □ **Solo Paolo**
C. □ **Entrambi**
D. □ **Nessuno dei due**

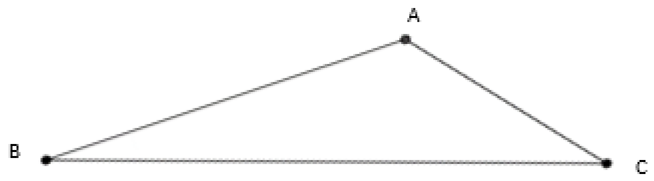
D5. **Chiara ha un banco rettangolare e un pacco di fogli rettangolari. Il lato lungo dei fogli è un quinto del lato lungo del banco, e il lato corto dei fogli è un quinto del lato corto del banco. Quanti fogli come minimo le occorrono per ricoprire interamente la superficie del banco?**

Risposta: …………….

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?

Risposta: …………………… cm

D7. Osserva il disegno.



a. Calcola l'area del triangolo ABC prendendo con un righello le misure necessarie.

Risposta: …………. cm^2

b. Scrivi i calcoli che hai fatto per arrivare alla risposta.

………

………

………

………

D8. Nino sale su un treno composto dalla locomotiva e da 9 vagoni:

Quanto è lungo all'incirca il treno di Nino?

A. □ **Circa 10 m**

B. □ **Circa 50 m**

C. □ **Circa 250 m**

D. □ **Circa 1000 m**

D27_PN2013

Anch_7. Nella scuola Nino Bixio ci sono 600 studenti e un insegnante ogni 15 studenti.

Quale proporzione permette di trovare il numero x degli insegnanti?

A. □ **x : 15 = 1 : 600**

B. □ **15 : 1 = x : 600**

C. □ **1 : 15 = x : 600**

D. □ **x : 1 = 15 : 600**

1CG_NEW_v2

D9. Qual è il risultato di 0,5 × 4?

A. □ **8**

B. □ **4**

C. □ **2**

D. □ **20**

Anch_4. La formula $L = L_0 + K \times P$ esprime la lunghezza L di una molla al variare del peso P applicato. L_0 rappresenta la lunghezza in centimetri "a riposo" della molla; K indica di quanto si allunga in centimetri la molla quando le si applica una unità di peso.

Quale delle formule elencate si adatta meglio alla seguente descrizione:

"*È una molla molto corta e molto dura (cioè molto resistente alla trazione)*"?

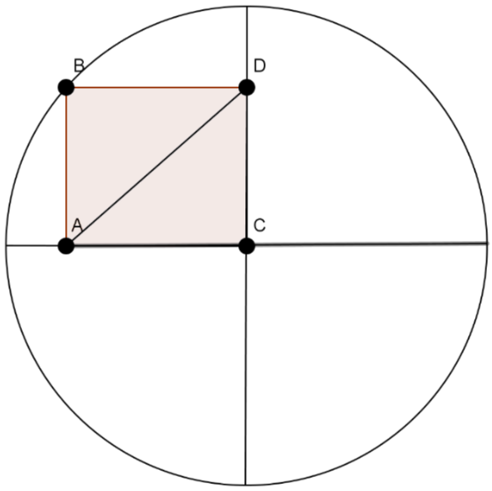
A. ☐ $L = 10 + 0{,}5 \times P$

B. ☐ $L = 10 + 7 \times P$

C. ☐ $L = 80 + 0{,}5 \times P$

D. ☐ $L = 80 + 7 \times P$

D10. Osserva la figura.



Mario afferma che il segmento AD misura 4 cm, visto che la circonferenza rappresentata in figura ha centro C e raggio 4 cm e CABD è un rettangolo.

Mario ha ragione? Scegli una risposta con il suo perché.

Mario NON HA ragione perché		Mario HA ragione perché	
La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è diverso da 4	Non ci sono abbastanza dati per calcolare la lunghezza di AD	La lunghezza di AD si può calcolare usando il teorema di Pitagora e il risultato è 4	Le diagonali di un rettangolo sono congruenti tra di loro
□ A	□ B	□ C	□ D

D26_PN2015

Anch_8. Un gruppo di 20 amici va in pizzeria. Ciascuno di essi ordina una pizza che costa 8 euro.
Ogni 5 pizze ordinate, il proprietario non ne fa pagare una.
Quanto spendono in tutto gli amici per le pizze?

A. □ **160 euro**

B. □ **128 euro**

C. □ **120 euro**

D. □ **112 euro**

D25_PN2011

Anch_5. Quale fra le seguenti disuguaglianze è quella corretta?

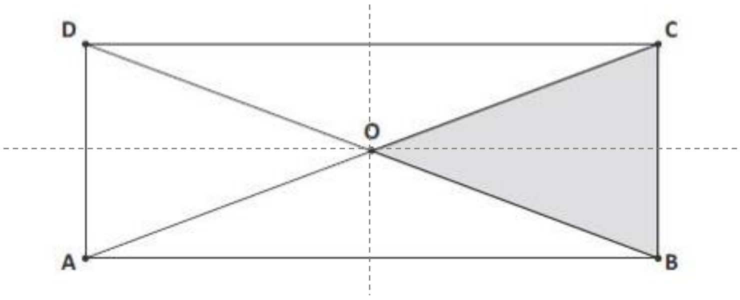
A. □ $\frac{3}{10} < \frac{3}{5} < \frac{3}{20}$

B. □ $\frac{4}{10} < \frac{3}{5} < \frac{11}{20}$

C. □ $\frac{5}{10} < \frac{3}{5} < \frac{13}{20}$

D. □ $\frac{7}{10} < \frac{3}{5} < \frac{13}{20}$

D11. In figura è rappresentato il rettangolo ABCD con le sue diagonali e i suoi assi di simmetria. Se conosci l'area del rettangolo, puoi calcolare l'area del triangolo COB in grigio?

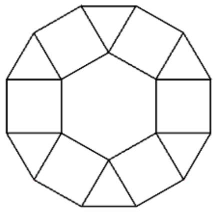


Scegli una risposta con il suo perché.

NO		**SI**	
perché i quattro triangoli di vertice O non sono tutti uguali fra loro	**perché non conosco le dimensioni del rettangolo**	**perché i quattro triangoli AOB, BOC, COD, DOA sono equivalenti**	**perché i quattro triangoli AOB, BOC, COD, DOA sono isosceli**
□ A	□ B	□ C	□ D

D9b_PN2011

Anch_2. Le immagini che seguono rappresentano un motivo del pavimento di una antica casa romana e la sua schematizzazione geometrica:



Il motivo, corrispondente a un dodecagono, è composto da un esagono regolare interno, sei quadrati uguali e sei triangoli equilateri uguali.

L'area dell'esagono è metà dell'area del dodecagono?
VERO □ FALSO □

L'area di ciascun triangolo è un sesto dell'area dell'esagono?
VERO □ FALSO □

D5_PN2011_v3

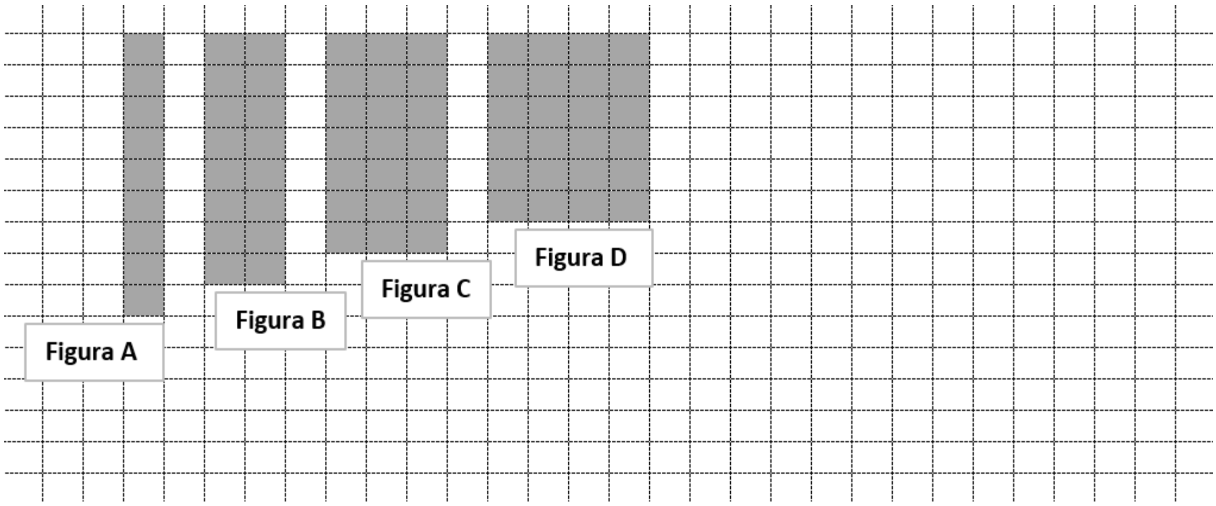
D12. Due cronometri A e B segnano il tempo. Il cronometro B è partito dopo il cronometro A. Quando B segna 10 secondi, A ne segna 30. Quando vengono fermati, B segna 50 secondi; quanti secondi segna A in quell'istante?

Risposta: .... secondi

D13. Individua un numero maggiore di 2/5 e minore di 3/5.

Risposta: ………………………………….

D14. Osserva la seguente sequenza di figure:



Quale delle seguenti affermazioni è vera?

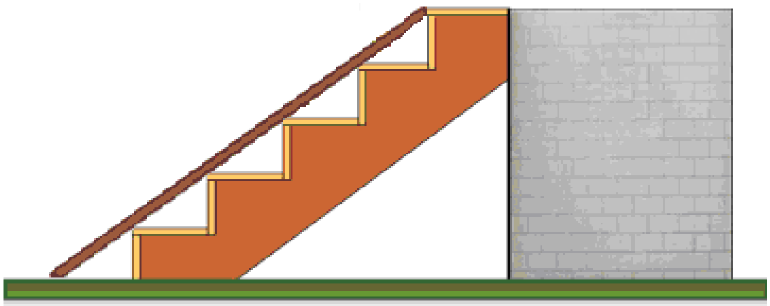
A. □ **I perimetri delle figure restano sempre uguali**

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**

Anch_6. Una scala, costituita da 5 gradini profondi 24 cm e alti 18 cm l'uno, deve essere coperta da una tavola di legno utilizzata come scivolo per il trasporto di alcune merci. Qual è il procedimento corretto per trovare la lunghezza dello scivolo?



A. □ $\left(\sqrt{18^2}+\sqrt{24^2}\right)\times 5$

B. □ $\sqrt{(24+18)^2}\times 5$

C. □ $\sqrt{24^2+18^2}\times 5$

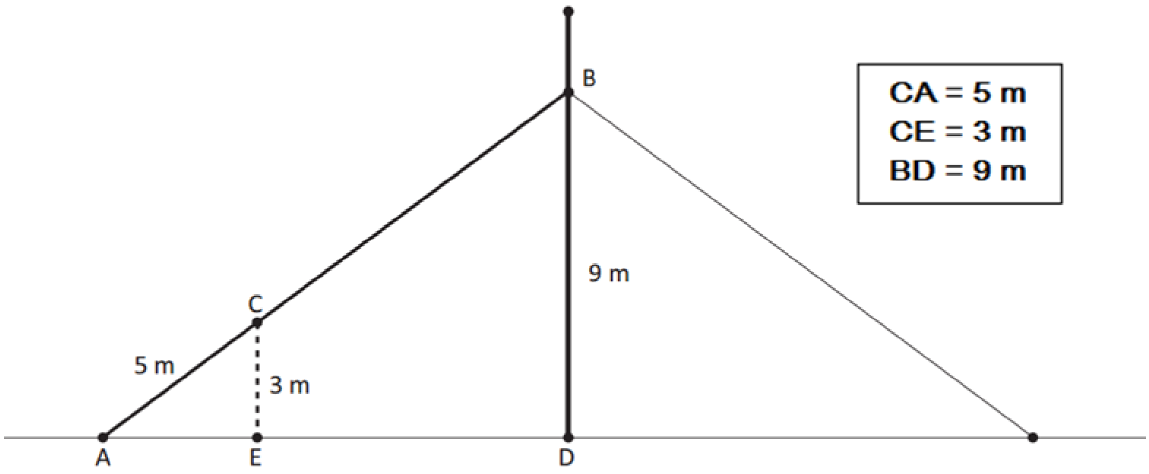
D. □ $\sqrt{\left(24^2+18^2\right)\times 5}$

3CG_NEW_inv

D15. In ogni coppia di operazioni cerchia quella che dà il risultato minore:

8×4	oppure	$8 : 4$
$8 \times 0{,}4$	oppure	$8 : 0{,}4$
$0{,}8 \times 0{,}4$	oppure	$0{,}8 : 0{,}4$

D16. Osserva il disegno.



Qual è la lunghezza del segmento AB?

Risposta: …………………… m

D17. **Giulio sa che nel negozio A e nel negozio B le bottiglie di olio della marca che preferisce hanno lo stesso prezzo.**

Oggi, su quell'olio, i due negozi fanno due promozioni diverse.

Giulio deve comprare 3 bottiglie d'olio.

a. **In quale negozio gli conviene comprarle?**

Risposta: …………………………………

b. **Scrivi come hai fatto per trovare la risposta.**

……

……

……

Grazie per aver risposto alle domande precedenti. Ti chiediamo ora, di completare la seguente sezione ricordandoti che, in questa sezione, non esistono risposte giuste o sbagliate. Leggi, quindi, tutto con attenzione e rispondi con sincerità!

Parte A – Il tuo rapporto con la scuola

A1) Qual è la materia che a scuola studi con più piacere?
Risposta: ____________________

A2) Qual è invece la materia che studi con meno piacere?
Risposta: ___________________

A3) Quanto pensi di essere bravo nelle seguenti materie? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= non sono bravo)	2 (= sono nella media)	3 (= sono bravo)	4 (= sono bravissimo)
Italiano	1	2	3	4
Matematica	1	2	3	4
Scienze	1	2	3	4

A4) Rispetto allo scorso anno, pensi che la tua abilità in Italiano, Matematica e Scienze sia cambiata? Cerchia la risposta che ritieni più appropriata in ciascuna delle righe che compongono le seguente tabella:

	1 (= sono peggiorato)	2 (= non è cambiato nulla)	3 (= sono migliorato)
Italiano	1	2	3
Matematica	1	2	3
Scienze	1	2	3

A5) Quanto spesso i tuoi genitori o le persone che si prendono cura di te fanno le seguenti cose? Metti una crocetta in corrispondenza della risposta esatta, per ciascuna riga della seguente tabella:

	mai	raramente	qualche volta	spesso	sempre
Controllare se hai fatto i compiti					
Aiutarti a fare i compiti					
Elogiarti o ricompensarti per aver avuto buoni voti					
Ridurre le tue ricompense quando prendi un brutto voto					
Trovare qualcuno che ti aiuti a fare i compiti					

Parte B – Cosa pensi della Matematica?

B1) Quanto ti senti in accordo o in disaccordo con le seguenti affermazioni? Per ciascuna riga della seguente tabella, cerchia il numero che rispecchia meglio quello che pensi sapendo che:

1 = totalmente in disaccordo;
2 = parzialmente in disaccordo;
3 = non so;
4 = parzialmente d'accordo;
5 = totalmente d'accordo.

1. La Matematica per me è importante	**1**	**2**	**3**	**4**	**5**
2. La Matematica la possono capire ed imparare in tanti	**1**	**2**	**3**	**4**	**5**
3. Ai miei genitori e/o a chi si prende cura di me piace la Matematica	**1**	**2**	**3**	**4**	**5**
4. La Matematica è una della materie più interessanti che studio a scuola	**1**	**2**	**3**	**4**	**5**
5. Studiare la Matematica è divertente	**1**	**2**	**3**	**4**	**5**
6. Io ho una mente matematica	**1**	**2**	**3**	**4**	**5**
7. Sono in grado di prendere buoni voti in Matematica	**1**	**2**	**3**	**4**	**5**
8. Prendere buoni voti in Matematica mi interessa	**1**	**2**	**3**	**4**	**5**
9. In Matematica, il mio impegno viene ricompensato	**1**	**2**	**3**	**4**	**5**
10. Essere bravo in Matematica è un talento di cui bisogna essere dotati alla nascita	**1**	**2**	**3**	**4**	**5**
11. Posso imparare la Matematica anche se è una materia difficile	**1**	**2**	**3**	**4**	**5**
12. Mi piace utilizzare la Matematica che già conosco bene piuttosto che usare concetti matematici che nuovi o che ho studiato da poco	**1**	**2**	**3**	**4**	**5**
13. Sono più preoccupato/a per lo studio della Matematica che per altre materie	**1**	**2**	**3**	**4**	**5**
14. Ho spesso bisogno di aiuto con la Matematica	**1**	**2**	**3**	**4**	**5**
15. Rispetto ai miei compagni di classe, sono bravo/a in Matematica	**1**	**2**	**3**	**4**	**5**
16. Ai miei parenti e/o a chi si prende cura di me piace risolvere problemi matematici	**1**	**2**	**3**	**4**	**5**
17. Spero di non dover studiare Matematica mai più in vita mia	**1**	**2**	**3**	**4**	**5**
18. Mi piacerebbe che i miei studi futuri includessero molta Matematica	**1**	**2**	**3**	**4**	**5**
19. Mi piacerebbe tanto studiare più Matematica a scuola	**1**	**2**	**3**	**4**	**5**
20. Da grande, mi piacerebbe fare della Matematica il mio mestiere	**1**	**2**	**3**	**4**	**5**
21. La Matematica è importante per il mio futuro (dopo la scuola)	**1**	**2**	**3**	**4**	**5**

Parte C – Per favore, dicci quanto spesso fate in classe le seguenti attività. Cerchia il numero più appropriato, riga per riga, sapendo che 1 = mai; 2 = raramente; 3 = qualche volta; 4 = sempre.

	1 = mai	2 = raramente	3 = qualche volta	4 = sempre
C1. L'insegnante ci fa delle domande	1	2	3	4
C2. L'insegnante ci chiede di spiegare il ragionamento che abbiamo fatto per arrivare al risultato	1	2	3	4
C3. L'insegnante inizia a spiegare gli argomenti nuovi partendo da esempi tratti dal mondo reale	1	2	3	4
C4. L'insegnante ci dice di lavorare più velocemente	1	2	3	4
C5. L'insegnante utilizza il computer per spiegarci alcuni argomenti	1	2	3	4
C6. L'insegnante ci dà spesso problemi da esaminare	1	2	3	4
C7. L'insegnante si aspetta che ci ricordiamo cose imparate in passato	1	2	3	4
C8. L'insegnante ci dice dettagliatamente quali attività fare	1	2	3	4
C9. L'insegnante ci chiede se conosciamo già qualcosa dell'argomento che sta per spiegarci	1	2	3	4
C10. L'insegnante ci spiega qual è il valore futuro delle cose che ci spiega	1	2	3	4
C11. Lavoriamo spesso in gruppo	1	2	3	4
C12. Ascoltiamo l'insegnante spiegare la lezione	1	2	3	4
C13. Copiamo ciò che l'insegnante scrive sulla lavagna	1	2	3	4
C14. Ci confrontiamo spesso l'un l'altro sulle strategie di soluzione dei problemi	1	2	3	4
C15. Chiediamo agli altri studenti di spiegare le loro idee e il loro punto di vista	1	2	3	4
C17. Facciamo gli esercizi riportati nel libro	1	2	3	4
C18. Studiamo come alcuni concetti matematici si sono evoluti nel tempo	1	2	3	4
C19. Quello che impariamo a scuola è collegato con la nostra vita al di fuori della scuola	1	2	3	4
C20. Abbiamo imparato che la Matematica di basa su regole inventate	1	2	3	4
C21. Facciamo ricerche su argomenti che scegliamo noi	1	2	3	4
C22. Usiamo la calcolatrice	1	2	3	4
C23. Usiamo i computer	1	2	3	4
C24. In classe, usiamo riviste, giornali o video	1	2	3	4
C25. Discutiamo le nostre idee con l'intera classe	1	2	3	4
C26. Esponiamo il nostro lavoro all'intera classe	1	2	3	4

C27. Se utilizzate il computer, che cosa lo utilizzate prevalentemente?

Risposta: ____

____

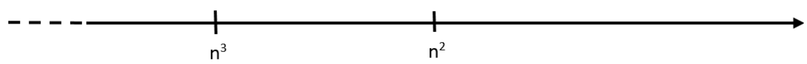
____

Parte D – Quanto ti senti sicuro/a di saper risolvere alcuni specifici problemi matematici

In questa sezione, ti chiediamo di fare una valutazione di quanto ti senti sicuro/a di saper risolvere alcuni problemi matematici. **Non ti chiediamo di fornire una risposta per ciascun quesito ma solo di dirci se e quanto pensi che saresti in grado di rispondere correttamente alle singole domande**. **Leggi quindi con grande attenzione e rispondi con la massima sincerità!**

D1. Quanto ti senti sicuro/a di saper utilizzare i numeri razionali, le loro diverse rappresentazioni, stimare la loro grandezza e/o il risultato di operazioni con i numeri razionali), come nella domanda riportata qui sotto?

Sulla seguente retta dei numeri sono ordinate due potenze di un numero razionale n.



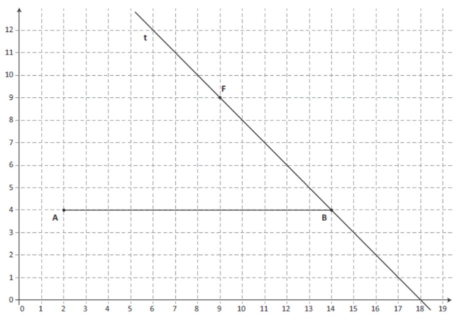
Indica con una crocetta se ciascuna delle seguenti affermazioni è vera (V) o falsa (F).

	SI	NO
Il valore di n può essere +1/2		
Il valore di n può essere -1/2		
Il valore di n può essere +3/2		
Il valore di n può essere -3/2		

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D2. Quanto ti senti sicuro/a di saper riconoscere le forme del piano e dello spazio, le loro rappresentazioni e di saper cogliere relazioni tra gli elementi, come nella domanda riportata qui sotto?

Osserva la figura

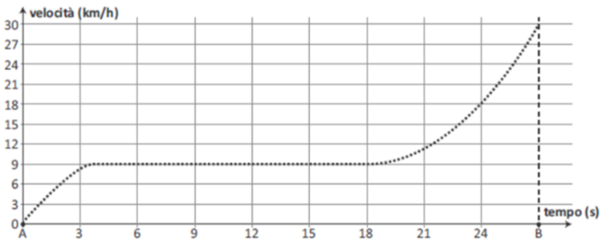


Sei in grado di disegnare la retta s perpendicolare a t passante per f e di individuare le coordinate del punto F di intersezione tra la retta *s* ed il segmento AB?

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D3. Quanto ti senti sicuro/a di saper calcolare misure statistiche di sintesi (ad esempio, la media, la moda, la mediana, ecc.), come nella domanda riportata qui sotto?

Luca percorre una strada in bicicletta e, con l'aiuto del computer, registra la propria velocità ogni decimo di secondo. Il grafico in figura rappresenta le diverse velocità raggiunte da Luca al passare del tempo.



Sei in grado di calcolare il valore modale delle velocità raggiunte da Luca dall'istante A all'istante B?

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D4. Quanto ti senti sicuro/a di saper risolvere problemi di misura e stima, come nella domanda riportata qui sotto?

Osserva l'edificio

Quanto può essere alto l'edificio?

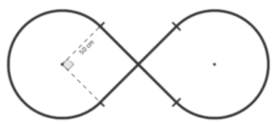
A. ◻ Meno di 10 metri
B. ◻ Tra 15 e 20 metri
C. ◻ Tra 25 e 30 metri
D. ◻ Più di 35 metri

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D5. Quanto ti senti sicuro/a di risolvere un problema matematico e di esplicitare in forma scritta il ragionamento che hai fatto, come nella domanda riportata di seguito?

La figura rappresenta lo schema di una pista formata da:

- Due archi di circonferenza di raggio 50 cm;
- Due tratti rettilinei di 100 cm ciascuno, perpendicolari tra loro nel punto medio.



Qual è la lunghezza della pista?

Scrivi i calcoli che fai per trovare la risposta e, infine, riporta il risultato.

____

____

____

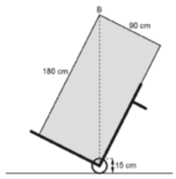
____

Risultato: circa ____________ cm

Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

D6. Quanto ti senti sicuro/a nel risolvere problemi come quello presentato nella seguente domanda?

Gabriele ha comprato un nuovo frigorifero. Per portarlo in cucina usa un carrello, come rappresentato in figura.



Quale espressione di permette di calcolare la massima distanza dal suolo del punto B quando il frigorifero è trasportato sul carrello?

A. ☐ $\sqrt{180^2+90^2}+7,5$

B. ☐ $\sqrt{180^2-90^2}+7,5$

C. ☐ $\sqrt{180+90}+7,5$

D. ☐ $\sqrt{180^2+90^2}+7,5$

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D7. Quanto ti senti sicuro/a di saper risolvere problemi di probabilità, come quello presentato nella seguente domanda?

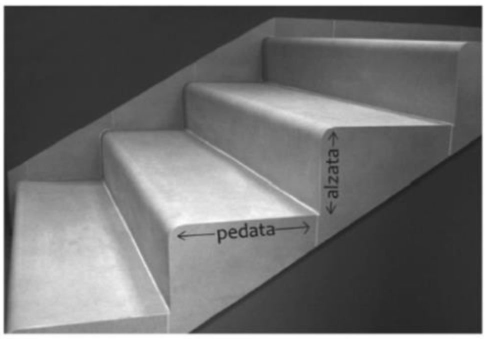
Giuseppe mette 2 palline bianche e 1 pallina nera in una busta.

Senza guardare, estrae una pallina. Qual è la probabilità che la pallina sia nera?

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D8. Quanto ti senti sicuro/a di saper utilizzare la matematica nella vita di tutti i giorni per rispondere a domande concrete, come quello presentato nella seguente domanda?

Nel disegno sottostante è rappresentata una scala.



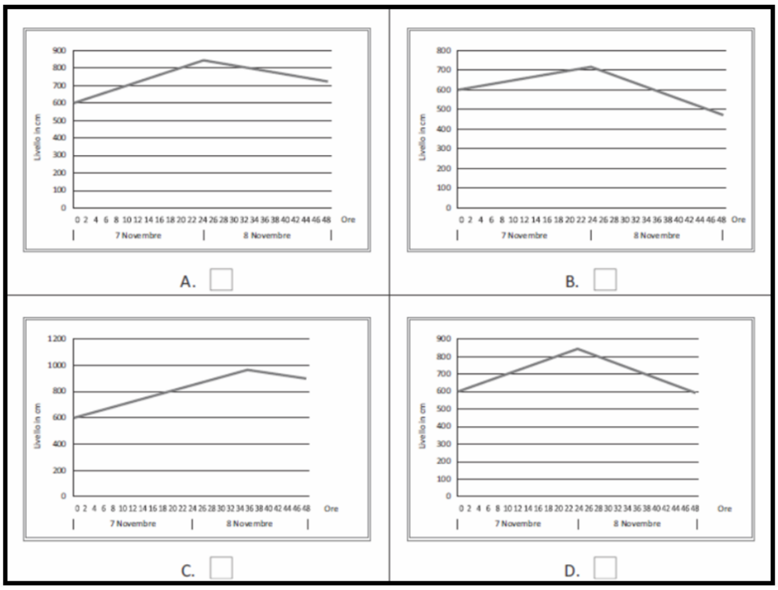
Per legge, la pedata deve essere lunga almeno 30 cm e la somma tra il doppio dell'alzata e la pedata deve essere compresa tra 62 e 64 cm (estremi inclusi).

Se la pedata di una scala misura 34 cm, il doppio dell'alzata dovrà essere compreso tra 28 e ____________ cm e, quindi, l'alzata dovrà essere compresa tra 14 cm e ____________ cm.

Per niente sicuro/a	**Non molto sicuro/a**	**Abbastanza sicuro/a**	**Molto sicuro/a**

D9. Quanto ti senti sicuro/a di saper risolvere problemi come quello presentato nella seguente domanda?

Oggi, il livello dell'acqua di un fiume è aumentato di circa 10 cm all'ora, per tutte le 24 ore. Il giorno successivo, il livello dell'acqua è diminuito di circa 5 cm all'ora per tutte le 24 ore. Quale tra i seguenti grafici può rappresentare meglio la situazione descritta?



Per niente sicuro/a	Non molto sicuro/a	Abbastanza sicuro/a	Molto sicuro/a

Parte E – Qual è il tuo rapporto con lo studio della Matematica?

Leggi attentamente le seguenti affermazioni e ripensa all'esperienza che hai avuto nel tuo percorso scolastico. Per ciascuna situazione, indica il livello di ansia che hai provato utilizzando una scala con 5 possibili risposte: da 1 (=poca ansia) a 5 (=molta ansia).

	Poca ansia				*Molta ansia*
1. Dover consultare le tavole in un manuale di matematica.	1	2	3	4	5
2. Pensare al compito di matematica del giorno dopo.	1	2	3	4	5
3. Guardare la risoluzione di un'equazione svolta dall'insegnante alla lavagna.	1	2	3	4	5
4. Fare il compito di matematica.	1	2	3	4	5
5. Avere da fare molti problemi difficili di matematica per la lezione successiva.	1	2	3	4	5
6. Ascoltare una spiegazione durante l'ora di matematica.	1	2	3	4	5
7. Ascoltare uno altro studente che spiega una formula di matematica.	1	2	3	4	5
8. Dover fare una verifica a sorpresa durante l'ora di matematica.	1	2	3	4	5
9. Iniziare un nuovo capitolo del manuale di matematica.	1	2	3	4	5

Parte F – Come sono i tuoi rapporti con tuoi compagni di scuola e di classe?

Questa è l'ultima sezione a cui ti chiediamo di rispondere. Le domande che ti facciamo, ci servono per capire quali sono i rapporti tra te e i tuoi compagni di classe. Prova quindi a pensare alla tua giornata a scuola e al rapporto che hai con i tuoi compagni, al tempo che trascorri con loro e alle attività che fate insieme. Nella maggior parte dei casi, non ti chiederemo il nome dei tuoi compagni ma solo il loro numero. Nelle ultime sette domande,

invece, ti chiediamo il tuo nome e quello di alcuni dei tuoi compagni. **In nessun caso, ti chiediamo il tuo cognome o quello dei tuoi compagni. Ad ogni modo, scegli tu a quali domande vuoi rispondere e quali invece preferisci lasciare in bianco!**

F1) Quanti studenti ci sono nella tua classe?

Risposta: _________

F2) Con quanti dei tuoi compagni di classe ti fa piacere chiacchierare durante la ricreazione, l'ora di ginnastica o quando possibile durante l'orario scolastico?

Risposta: _________

F3) Con quanti dei tuoi compagni di classe ti vedi fuori dalla scuola?

Risposta: _________

F4) Quanti dei tuoi compagni sono tuoi amici su Facebook e/o altri social networks?

Risposta: _________

F5) Con quanti compagni di classe hai l'abitudine di tenerti in contatto tramite telefonate, SMS, messaggi whatsapp, chat, ecc.?

Risposta: _________

F6) Quanti dei tuoi compagni vedi abitualmente al di fuori della scuola per

i. fare i compiti. Risposta: ___________

ii. fare attività sportive. Risposta: ___________

iii. uscire insieme (per andare al cinema, a mangiare una pizza, ecc.). Risposta: ___________

F7) Se tu avessi qualche problema, a scuola o nella tua vita privata,

i. con quanti dei tuoi compagni ti confideresti? Risposta: _________

ii. a quanti dei tuoi compagni chiederesti aiuto? Risposta: _________

iii. quanti dei tuoi compagni pensi che ti aiuterebbero? Risposta: _________

F8) Quanti dei tuoi compagni ti chiederebbero aiuto o si confiderebbero con te se avessero dei problemi?

Risposta: _________

F9) Quanti dei tuoi compagni aiuteresti se ti chiedessero aiuto?

Risposta: _________

F1. Qual è il tuo nome di battesimo? Risposta: ___________________

F2. Se si dovesse organizzare una gita, chi sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi, in ordine di preferenza:

1. _________________________
2. _________________________
3. _________________________

F3. Se si dovesse organizzare una gita, chi **non** sceglieresti tra i tuoi compagni per farlo venire insieme a te? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

F4. Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, chi sceglieresti? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

F5. Se l'insegnante ti permettesse di scegliere il tuo compagno di banco, con quale ti piacerebbe non stare? Indica 3 nomi

1. ___________________________
2. ___________________________
3. ___________________________

F6. Abbiamo fatto le stesse domande a tutti i tuoi compagni. Secondo te, tra i tuoi compagni, chi pensi ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

F7. Chi pensi invece che **non** ti abbia scelto come compagno di banco? Indica di seguito tre nomi:

1. ___________________________
2. ___________________________
3. ___________________________

Hai finito!

GRAZIE!!!