

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea Magistrale in Matematica

**Definizione e validazione di uno strumento per la
misurazione dell'impatto delle variabili
redazionali di un item matematico sulla
performance degli studenti**

Tesi di Laurea in Didattica della Matematica

Relatore:
Chiar.mo Prof.
GIORGIO BOLONDI

Presentata da:
REBECCA BONINSEGNA

Sessione I
Anno Accademico 2015/2016

Alla mia famiglia.

Indice

Introduzione	1
1 Presentazione del problema	3
1.1 Revisione degli aspetti linguistici e numerici che incrementano le difficoltà nei Word Problems	10
1.1.1 Complessità linguistica	10
1.1.2 Complessità numerica e studi numerici	12
1.1.3 Connessioni fra fattori linguistici e matematici	17
1.1.4 Differenze individuali e fattori sociali	19
2 Lo strumento statistico	22
2.1 IRT (Item Response Theory)	22
2.2 Il modello di Rasch	23
2.2.1 Stima dei parametri	25
2.2.2 Proprietà di sufficienza del modello di Rasch	26
2.3 Strumenti statistici e indici utilizzati per la valutazione degli ap- prendimenti	26
2.3.1 L' α di Cronbach	27
2.3.2 Il modello di Rasch e il placement degli item	28
3 Metodologia e piano di validazione dello strumento	31
3.1 La prova INVALSI di partenza	31
3.2 Il test variato	32
3.3 Selezione del campione e raccolta dei dati	44

4	La validazione dello strumento	46
5	Analisi dei risultati	52
6	Conclusioni	77
	Bibliografia	79

Introduzione

Questa tesi descrive la ricerca condotta tra l'autunno e l'inverno di quest'anno da un gruppo di ricercatori in didattica della matematica relativamente all'influenza che le variazioni redazionali di un quesito matematico hanno sulle performance degli studenti.

Lo scopo della ricerca è quella di strutturare e validare una metodologia e uno strumento che permettano di individuare e quantificare l'influenza delle variazioni del testo sulle prestazioni dello studente.

Si è sentita l'esigenza di condurre uno studio di questo tipo poichè è sempre più evidente il profondo legame tra il linguaggio e l'apprendimento della matematica. La messa a punto di questo strumento aprirebbe le porte a una serie di ricerche più approfondite sulle varie tipologie di variazioni numeriche e/o linguistiche fin'ora individuate. Infatti le considerazioni che sono state fatte su di esse raramente hanno fatto uso di software statistici o di rigorose metodologie per un'analisi quantitativa del caso; esse hanno piuttosto utilizzato metodologie di tipo qualitativo.

Nel primo capitolo è presentato il quadro teorico di riferimento relativo agli studi condotti fino ad ora nell'ambito della didattica della matematica, dai quali emerge la grossa influenza che la componente linguistica ha sulla comprensione e la trasmissione della matematica [4], [1]. Si farà quindi riferimento alle ricerche passate volte all'individuazione e alla schematizzazione delle variazioni redazionali dei Word Problems [5]. Infine verrà brevemente descritta una recente ricerca nell'ambito delle variazioni linguistiche condotta da un punto di vista qualitativo [3]. Nel secondo capitolo invece, si passerà alla descrizione teorica relativa allo strumen-

to statistico utilizzato. Si tratta del modello di Rasch appartenente alla famiglia dei modelli statistici dell'Item Response Theory. Esso è particolarmente utilizzato nella ricerca in didattica per la valutazione degli apprendimenti e nelle valutazioni nazionali ed internazionali¹. Inoltre verranno presentati i parametri e i grafici che saranno poi utilizzati per l'analisi dei risultati.

Il terzo capitolo sarà dedicato alla descrizione dettagliata della sperimentazione svolta: la scelta del test originale, la stesura del test variato e il disegno di somministrazione alle classi.

Nel quarto capitolo verrà descritta e validata la nuova metodologia utilizzata per l'analisi dei risultati.

Nel quinto sarà eseguita un'analisi puntuale di come lo strumento ha messo in evidenza le differenze per ogni item variato.

Infine verranno tratte le conclusioni complessive dello studio condotto.

¹INVALSI, OCSE-PISA

Capitolo 1

Presentazione del problema

Negli ultimi decenni la ricerca ha mostrato che è sempre più difficile parlare di apprendimento e insegnamento della matematica senza prendere in considerazione gli aspetti comunicativi ed in particolare quelli linguistici (anche relativamente alla lingua comune). È infatti ormai vastissima la letteratura nella quale si analizza e sottolinea lo stretto e complesso legame tra il linguaggio nella sua forma comune e in quella specifica e la matematica. Ad esempio, D'Amore [4] descrive alcune situazioni che spesso si verificano nel contesto scolastico riguardo l'utilizzo del linguaggio matematico:

Molti ricercatori ma anche molti insegnanti confermano di aver trovato presso gli allievi “definizioni” di parallelogrammo del tipo: «Un quadrilatero è detto parallelogrammo quando ha i lati a due a due»; altre volte si trovano invece affermazioni del tipo: «La retta a è parallela» senza dire a che cosa, come se la proprietà di parallelismo fosse unaria; ancora mi è stato segnalato da una collega: «Un insieme si dice equipotente quando...».

Egli ritiene che questi eventi non siano altro che la manifestazione fenomenologica di una (non così nascosta) mancanza di comprensione e di senso. D'Amore conclude affermando che:

La lingua comune ed il linguaggio della Matematica entrano duramente in opposizione tra loro, in Didattica, costituendo un vero e proprio

ostacolo sia alla comprensione sia ad un uso il più possibile “naturale” e spontaneo del linguaggio matematico atteso dall’insegnante.

L’interazione tra il linguaggio e le sue forme e la matematica entrano in gioco anche nella semplice risoluzione di un problema, infatti, quando gli studenti devono affrontare un compito di matematica, le loro prestazioni sono indubbiamente influenzate dalla formulazione della domanda. In particolare, se pensiamo a uno studente posto di fronte ad un item di un test standardizzato, il fatto che questo non sia stato scritto dall’insegnante al quale sono abituati, può giocare un ruolo tutt’altro che irrilevante sulla sua performance. Risulta quindi interessante, se non necessario, comprendere quali siano gli aspetti della formulazione del testo che interferiscono con la risoluzione del problema, in che modo agiscano e con quale impatto.

Per stabilire se un testo matematico, ad esempio l’enunciato di un problema, sia davvero comprensibile, si dovrà condurre un attento esame critico dei suoi contenuti. A tale proposito, Laborde [1] formulò le seguenti *variabili redazionali* come principali aspetti da valutare all’interno dell’enunciato di un problema:

- Grado di esplicitazione ottenuto dall’impaginazione, dalla punteggiatura e dalle strutture sintattiche impiegate.
- Complessità sintattica.
- Densità dell’enunciato.
- Ordine delle informazioni fornite.
- Differenza tra la forma in cui le informazioni sono date e quella in cui le si deve trattare nella risoluzione.
- Grado di esplicitazione degli oggetti intermedi utili alla risoluzione del problema.

In questa ricerca si fa riferimento alle variabili redazionali elencate da Laborde in quanto, a differenza del precedente lavoro di Duval, includono anche modifiche

non verbali, come per esempio l'introduzione di immagini.

Bagni ([1]) propone l'analisi degli enunciati di due problemi, e lo fa seguendo punto per punto l'impostazione di Labord. Il primo problema è il seguente:

Calcolare il rapporto delle altezze di due rettangoli equivalenti con le basi rispettivamente di 12 cm e di 15 cm.

Per quanto riguarda la punteggiatura essa non viene utilizzata. Il citato "rapporto delle altezze" potrebbe essere riferito ad entrambe le situazioni: altezza maggiore : altezza minore oppure altezza minore : altezza maggiore e ciò può essere causa di perplessità. L'enunciato appare complesso, esso è infatti costituito da un solo periodo, con una proposizione subordinata; inoltre il termine "rispettivamente" richiede un certo livello di attenzione da parte dell'allievo, che, soprattutto se all'inizio della scolarizzazione, potrebbe non conoscerne il significato. La densità dell'enunciato è molto alta, tutte le parole utilizzate giocano un ruolo fondamentale per la risoluzione del problema. Per quanto riguarda la presentazione dei dati e del quesito, l'enunciato del problema pone immediatamente la domanda e anche i dati vengono forniti nell'ordine inverso rispetto a quello in cui essi possono essere utilizzati. Infine uno dei dati più importanti, ossia l'equivalenza dei due triangoli, è particolarmente celato, essendo posto come aggettivo dei triangoli e potrebbe passare quasi inosservato.

Viene quindi considerato il secondo enunciato, il quale vuole essere una riformulazione semplificata di quello appena analizzato. Il suo enunciato è il seguente:

Il rettangolo A ha la base di 12 cm; il rettangolo B ha la base di 15 cm.
L'area del rettangolo A è uguale all'area del rettangolo B. Calcolare il rapporto tra l'altezza del rettangolo A e l'altezza del rettangolo B.

Dall'analisi emerge innanzitutto che l'enunciato è ora costituito da quattro proposizioni distinte, nelle quali la punteggiatura è ampiamente usata. Il termine "rispettivamente" viene sostituito da due proposizioni, inoltre il "rapporto delle altezze" richiesto viene indicato esplicitamente. L'enunciato è più lungo del precedente, ma decisamente meno denso. L'ordine di presentazione dei dati segue la sequenza nel quale essi possono essere utilizzati e anche la domanda è posta alla

fine dell'enunciato ed è inoltre chiaramente evidenziata da un punto a capo. Anche l'informazione relativa all'uguaglianza delle aree è ora ben evidenziata, essa costituisce infatti una delle 4 proposizioni. Verrebbe quindi da affermare che la seconda formulazione risulta decisamente più semplice della prima, e probabilmente è così, ma ciò che è più interessante è una considerazione che Bagni espone al termine dell'analisi:

Dobbiamo dunque concludere che il testo del problema, così riformulato, è "migliore" del testo precedente? No, non possiamo affermarlo, in quanto non esiste, in assoluto, un criterio per stabilire quando un testo sia da considerare "migliore" di un altro. [...] È tuttavia importante sottolineare che gli ostacoli connessi alla lettura e all'interpretazione del testo costituiscono un aspetto fondamentale di un problema, un elemento spesso decisivo a proposito della sua difficoltà.

Dunque, i due problemi hanno forse la stessa risoluzione, ma non la stessa difficoltà. Non sono lo stesso problema.

Questa osservazione può essere considerata rivoluzionaria: la formulazione del problema è parte integrante di esso e gioca un ruolo fondamentale tanto che due problemi con lo stesso procedimento risolutivo possono considerarsi di due difficoltà diverse. Allora per un insegnante che desidera testare certe abilità e conoscenze dello studente, è fondamentale riflettere, non solo sul procedimento matematico, ma anche su quali capacità sono richieste per la risoluzione di un problema con una certa formulazione piuttosto che un'altra.

L'interesse per i fattori linguistici e matematici che contribuiscono alla complessità di un problema, quando esso è presentato sotto forma di un breve racconto piuttosto che in annotazioni matematiche, hanno suscitato l'interesse dei ricercatori fino dagli anni '70 del secolo scorso.

Nel 2015 Daroczy, Wolska, Maurers e Nuerk hanno pubblicato una interessante revisione della letteratura nel settore [5]. La revisione ha visto la collaborazione di ricercatori nel settore linguistico e numerico al fine di fornire una panoramica sistematica degli aspetti linguistici e numerici rilevanti per la risoluzione di un Word Problem (WP), come pure delle loro interazioni.

Fattori linguistici	Fattori matematici
Fattori strutturali <ul style="list-style-type: none"> • Proprietà quantitative <i>e.g., numero di lettere, lunghezza delle parole e delle frasi, presenza e numerosità di vocaboli complessi</i> • Complessità strutturale e del vocabolario <i>e.g., parole polisemiche, struttura della frase, voci passive, posizione e formulazione della domanda</i> 	Proprietà numeriche <ul style="list-style-type: none"> • Numeri a una o più cifre • Tipo di numeri <i>e.g., frazione, numeri decimali, ruolo dei numeri (addendo, minuendo, sottraendo, ecc...)</i> • Ordine di grandezza dei numeri
Fattori semantici <ul style="list-style-type: none"> • Parole o frasi suggeritrici di un certo procedimento <i>e.g., meno di, in più, supera, ecc...</i> • Relazioni tra gli oggetti <i>e.g., relazione funzionale (fiori/vasi), relazione categoriale (margherite/tulipani)</i> • Presenza di distrattori 	Operazioni richieste <i>e.g., addizioni, sottrazioni, moltiplicazioni, divisioni, operazioni con il riporto, ecc...</i>
	Strategie matematiche risolutive <ul style="list-style-type: none"> • Calcolo con numeri elevati • Posizionamento dell'incognita • Calcolo a mente • Sequenza con cui vengono presentati i numeri
	Individuazione e reperimento delle informazioni Altri processi e rappresentazioni numeriche
Fattori di interazione	
Coerenza lessicale Work memory	
Fattori generali Abilità ed aspetti sociali <ul style="list-style-type: none"> • Disabilità o ritardi dell'apprendimento • Risolutori più o meno abili • Bambini o adulti 	

Nella sezione 1.1 verrà riportato il lavoro di revisione svolto dagli autori. Nella tabella sono schematizzati i principali fattori matematici, linguistici e generali che entrano in gioco nella risoluzione dei Word Problems, individuati e studiati negli ultimi 50 anni.

L'effetto delle variazioni sulle prestazioni degli studenti non è facile da individuare ed analizzare in quanto risulta impossibile ricreare l'ottimale situazione da studiare. Uno studente, infatti, dovrebbe rispondere a due domande molto simili, senza che la strategia adottata e la risposta data alla prima incida sulla seconda. Ciò risulta impossibile a meno che la modifica non sia sostanziale, rischiando però

di alterare la natura della domanda. Un metodo di analisi potrebbe essere una discussione in classe o con un gruppo di studenti successivamente allo svolgimento di una prova. Questo tipo di sperimentazione restituirebbe però un suggerimento indiretto di interpretazioni a posteriori delle difficoltà di alcuni studenti in un compito particolare. Quello che si vuole direttamente analizzare, invece, è il modo in cui le variazioni influenzano le risposte degli studenti durante il test, ossia cosa farebbe lo stesso studente posto di fronte alla stessa domanda diversamente formulata? E ancora, una variazione può davvero causare un comportamento significativamente diverso nel momento in cui lo studente risponde al test? In questo caso, un risultato qualitativo basata su un approccio interattivo non è adeguato per rispondere alla domanda. Branchetti e Viale [3] hanno proposto una metodologia basata sulla IRT (Item Response Theory) e sul modello di Rasch (Rasch, 1960) per studiare tale problema. Hanno condotto una sperimentazione su un campione di circa 200 studenti di scuola secondaria di primo grado, ai quali sono state sottoposte versioni sintatticamente complicate di esercizi della prova somministrata dall'Istituto Nazionale per la Valutazione del Sistema di Istruzione (INVALSI) al termine della scuola secondaria di primo grado dell'anno scolastico 2009/2010. Gli item selezionati sono stati quindi modificati e le variazioni hanno riguardato fattori linguistici ed in particolare sintattici: è stata aumentata la lunghezza delle frasi, sono stati introdotti incisi, forme impersonali, subordinate al gerundio, participi presenti al fine di aumentare la densità sintattica. Gli autori hanno confrontato le risposte ottenute con quelle della prova del 2009. Le analisi eseguite sono state di natura qualitativa, hanno riguardato solo alcuni casi particolari di variazione, e sono state svolte senza il supporto di alcun software.

Lo studio che verrà descritto nei prossimi capitoli, ha avuto lo scopo di migliorare la metodologia e lo strumento di analisi, mediante tecniche statistiche e un più accurato piano di validazione. Le nostre domande di ricerca sono:

- i. Come possiamo misurare l'impatto di una variazione?
- ii. Una variazione può causare cambiamenti significativi nelle risposte di una popolazione o nella prestazione di un particolare gruppo di studenti?

- iii. Come possiamo far emergere dall'analisi dei dati la rilevanza di una variante sulla popolazione o su alcuni sottogruppi?

1.1 Revisione degli aspetti linguistici e numerici che incrementano le difficoltà nei Word Problems

1.1.1 Complessità linguistica

Il concetto di complessità linguistica è stato affrontato sia dal punto di vista dell'evoluzione del linguaggio (Nichols, 1990), sia dal punto di vista dell'acquisizione del linguaggio (Bulté and Housenm 2012). Secondo questi studi è utile analizzare la complessità linguistica rispetto alla complessità delle proposizioni (la quantità di significato che deve essere espresso) e a quella del periodo (l'interazione degli elementi che partecipano al discorso).

La complessità linguistica può essere esaminata rispetto a tutti gli aspetti del sistema linguistico, dagli aspetti morfologici e lessicali delle parole, a come queste parole possono essere combinate fra loro sintatticamente per formare frasi, fino alla struttura complessiva del testo.

Lingue diverse hanno poi strutture diverse: l'inglese per esempio affida all'ordine delle parole la codifica di funzioni grammaticali mentre altre lingue si affidano a un ricco repertorio morfologico. Ne consegue che: 1) la complessità dei WPs è diversa a seconda della lingua: ciò fa sì che una modifica porta ad aumentare la complessità in una lingua può non avere effetti in un'altra lingua; 2) la performance di uno studente in un WP presentato in una lingua straniera può essere condizionato dalle differenze fra la lingua madre e quella in cui il problema è presentato.

Di seguito saranno analizzati gli aspetti linguistici che maggiormente influenzano la comprensione dei WPs.

Fattori strutturali

I primi studi a riguardo di fattori strutturali affrontavano la complessità strutturale basandosi sulle proprietà quantitative del testo, come ad esempio il numero delle lettere delle parole, la lunghezza delle frasi o la proporzione di parole complesse (lunghe) (Searle et al, 1974 – Nesher, 1976 – Lepik, 1990). Ulteriori elementi

linguistici sono poi stati studiati nel contesto delle difficoltà di comprensione per gli studenti di lingue straniere (in particolare studenti di inglese):

- A livello di vocaboli le difficoltà di comprensione che determinano una difficoltà a risolvere WPs derivano dalla presenza di parole poco famigliari (non appartenenti al linguaggio comune), parole polisemiche, idiomatiche o riferimenti lessicali culturalmente specifici.
- A livello di struttura della frase i fattori che si sono dimostrati avere un ruolo nella complessità linguistica sono l'eccessiva lunghezza del discorso, il numero di frasi introdotte da preposizioni, frasi con participi, voci passive o proposizioni complesse come relative, subordinate, avverbiali, o condizionali.
- A livello della struttura del discorso, in particolare dell'organizzazione del periodo hanno particolare importanza la corrispondenza fra l'ordine in cui i dati numerici sono presentati e l'ordine in cui devono essere utilizzati per la risoluzione del problema: problemi ordinati migliorano le prestazioni (Searle et al., 1974). Inoltre, sono stati evidenziati risultati migliori anche nel caso di formulazioni più semplici della domanda o ponendo la domanda all'inizio (Cummins et al., 1988).

Fattori semantici

Un fattore ampiamente studiato e che si è dimostrato strettamente correlato con la difficoltà di risoluzione di un WP, è la presenza o l'assenza di segnali verbali espliciti, ovvero di termini la cui semantica suggerisca un'operazione matematica utilizzabile per la risoluzione del problema. Questi segnali comprendono congiunzioni ("e" per l'addizione), avverbi ("più di", "meno di" per la sottrazione), distributivi ("ciascuno") per la moltiplicazione o la divisione. Alcuni studi che utilizzano strumenti per la rilevazione dei movimenti oculari hanno evidenziato che i soggetti ai quali vengono sottoposti testi contenenti tali segnali verbali, tendono a focalizzarsi su di essi traducendoli direttamente nella operazione matematica che richiamano ((Hegarty et al., 1992; van der Shoot et al., 2009).

Dal momento che i segnali verbali possono condurre a interpretazioni errate, anche

piccole differenze nella formulazione del discorso possono determinare importanti variazioni delle prestazioni dei soggetti (LeBlanc and Weber-Russel, 1996). Questo è particolarmente importante per i bambini più piccoli che collegano parole come “unire”, “aggiungere”, “ottenere”, “trovare” e “togliere” con concetti di somma sottrazione ecc... Un problema può così essere facilitato aggiungendo segnali verbali che rendano più evidenti le relazioni semantiche così che risulti più esplicita anche la sottostante relazione matematica.

Anche le relazioni fra gli oggetti descritti nei problemi possono costituire una difficoltà. Solitamente i problemi con la divisione trattano di oggetti correlati funzionalmente (ad es. vasi e tulipani), mentre l’addizione tratta più spesso di oggetti correlati per categoria (tulipani e margherite). Quindi anche le relazioni che intercorrono fra gli oggetti sembrano riflettere relazioni ed operazioni matematiche. Per questo, spesso, le proprietà semantiche vengono considerate più insidiose di quelle sintattiche (Yeap and Kaur 2001).

1.1.2 Complessità numerica e studi numerici

La complessità numerica può influenzare la prestazione nei WPs almeno in tre modi:

- Agendo direttamente sulla struttura aritmetica: WPs con strutture aritmetiche più complesse sono più difficili indipendentemente dalla difficoltà linguistica.
- Aumentando il carico cognitivo: problemi aritmetici più complessi comportano un maggior carico cognitivo. Per esempio si ritiene che i problemi col riporto richiedano maggiori risorse di work memory. Se a ciò aggiungiamo strutture linguistiche più complesse il carico cognitivo aumenterà ulteriormente. Per questo l’alta complessità aritmetica unita alla complessità linguistica possono portare all’insorgere di nuove difficoltà non spiegabili da un semplice effetto dell’una o dell’altra.
- Agendo sulla strategia risolutiva. Thevenot e Oakhill (2005) hanno analizzato l’influenza della gestione di numeri a più cifre sulle strategie di soluzione

dei WPs. Essi hanno dimostrato che se un soggetto è posto di fronte a numeri con tre cifre, adatterà strategie risolutive differenti da quelle che avrebbe adottato in caso di numeri con un numero di cifre inferiore. Gli autori motivano tale comportamento con la ricerca di una strategia che richieda meno sforzo da un punto di vista cognitivo. Risultati simili sono stati ottenuti da Brissiaud e Sander (2010) che manipolando le dimensioni e l'ordine di grandezza dei numeri coinvolti nelle operazioni, sono arrivati a definire tipi diversi di strategie risolutive.

Nell'articolo le variabili numeriche studiate dai ricercatori vengono classificate in 5 categorie: Proprietà numeriche, operazioni richieste, strategie matematiche risolutive, individuazione e reperimento delle informazioni, altri processi e rappresentazioni numeriche.

Proprietà numeriche

La complessità aritmetica è stata raramente studiata in modo sistematico come fattore isolato. Proprietà come il numero di cifre, la parità del numero e il ruolo di esso all'interno dell'operazione sono raramente considerate nei WPs anche se si è notato che, soprattutto nei bambini, influenzino ampiamente la prestazione e la scelta strategica per la risoluzione del problema aritmetico. Per esempio in un problema come: "il numero intermedio Y è esattamente a metà fra X e Z ?" la parità di X , Y e Z condiziona la prestazione. Se X e Z hanno parità diversa, il quesito risulta più facile da risolvere rispetto a quello in cui la parità è concorde (Nuerk et al., 2002). La spiegazione di questo fatto si ritiene sia legata ad un cambiamento di strategia risolutiva, infatti se la parità non è concorde ($X=21$, $Y=25$, $Z=28$) è impossibile che il numero intermedio sia a metà fra gli altri due perché il punto medio fra due numeri con parità non concorde non è un numero intero (e nell'esperimento venivano utilizzati solo numeri interi). Perciò, dopo aver individuato la discordanza di parti, i partecipanti all'esperimento modificano la loro strategia e non procedono nel tentativo di calcolare se il numero intermedio dato (Y) è esattamente il punto medio. Uno studio successivo (Wood et al., 2008) ha confermato tale riscontro: nei casi di discordanza è stato osservato una maggiore attivazio-

ne della corteccia prefrontale ventrolaterale destra, che si attiva nelle situazioni di cambiamenti cognitivi o quando i partecipanti producono soluzioni alternative. Quindi è certo che la parità del numero influisca sulle strategie risolutive.

Ciò non avviene solo nel caso di problemi di biscazione ma anche nello svolgimento di operazioni standard come l'addizione e la sottrazione. Una revisione di Hines del 2013, suggerisce che la parità influisca sulla difficoltà dell'addizione e della sottrazione ma non della moltiplicazione, e che i quesiti che contengono numeri dispari risultino più difficili di quelli con numeri pari.

Una seconda proprietà che influisce sulle prestazioni riguarda il numero di cifre che compongono il numero. Nella maggior parte degli studi, in particolare sui bambini, vengono utilizzati numeri a una cifra e raramente viene chiarito il perché sia stato scelto un numero specifico. In alcuni casi sono stati presentati studi contenenti frazioni, numeri interi e decimali ma il loro effetto è stato comunque poco indagato.

Per quanto riguarda i numeri a una o più cifre, la ricerca degli ultimi 15 anni ha dimostrato che le strategie che richiedono la gestione di numeri a una cifra non possono essere facilmente trasferiti alla gestione di numeri a più cifre (Nuerk et al.). Secondo gli esperti non è quindi opportuno, all'interno di una ricerca, mischiare numeri a una o più cifre vista la differenza nella gestione.

Anche il ruolo del numero all'interno dell'operazione riveste un ruolo importante (De Corte et al 1988). Nel caso della moltiplicazione, per esempio, si è riscontrato che la prestazione del bambino è molto influenzata dalle proprietà del moltiplicatore (intero, decimale, maggiore di 1 o minore di 1) mentre quelle del moltiplicando hanno scarse ripercussioni. De Corte et al. hanno evidenziato che due problemi di moltiplicazione con la stessa struttura matematica e semantica ma che utilizzano numeri diversi possono dar luogo a difficoltà molto differenti.

La rilevanza dei dati nella risoluzione di un problema è stata osservata da De Corte e Verschaffel all'interno del loro studio sui movimenti oculari del 1988. Essi hanno osservato una notevole focalizzazione sui dati del problema: il 25% del tempo complessivo impiegato per la risoluzione del quesito era speso nelle due piccole aree contenenti i numeri.

Operazioni richieste

Le operazioni aritmetiche sono state ampiamente studiate. La maggior parte degli errori sembrano essere originati dalla scarsa comprensione del testo del problema (Schumacher and Fuchs, 2012). Si hanno quindi calcoli svolti in maniera corretta ma applicati ad una rappresentazione errata del problema (Lewis and Mayer, 1987). Si tratta in questo caso di una difficoltà di tipo linguistico, mentre un'ulteriore causa d'errore viene da problemi nel calcolo (Raduan, 2010; Kingsdorf and Krawec, 2014). Le operazioni più spesso utilizzate negli esperimenti sui WPs sono addizione e sottrazione, Riley et al. (1983) hanno classificato i problemi nel caso di addizioni e sottrazioni elementari. Dopo gli anni '80 e '90 ci si è principalmente concentrati su problemi con la moltiplicazione o con operazioni miste ed anche in questo caso è stata presentata una classificazione dei quesiti con moltiplicazione e divisione sulla base del tipo di numeri utilizzati (interi positivi, frazioni e decimali).

Come già sottolineato la ricerca nell'ambito dei problemi con addizione e sottrazione è molto vasta, in particolare su ciò che rende difficile questa tipologia di quesiti. E' noto che le operazioni col riporto complicano l'operazione a più cifre, sia per i bambini che per gli adulti (Nuerk et al., 2015). Sempre gli studi sui movimenti oculari suggeriscono che in un compito con risposte a scelta multipla, i bambini delle elementari eseguono sempre il calcolo cercando un riscontro tra le alternative fornite, mentre gli adulti utilizzano anche la strategia del rifiuto del risultato scorretto. All'interno delle operazioni con riporto si possono identificare, per gli adulti, almeno tre diversi problemi cognitivi: il calcolo della somma delle unità, l'individuazione della cifra da riportare e l'esecuzione del riporto. Le difficoltà che creano i problemi con il riporto è imputata da alcuni autori, come Goebel, anche a fattori linguistici. Inoltre queste operazioni richiedono risorse maggiori di work memory (Ashcraft, 1995; Furst and Hitch, 2000). Se il carico cognitivo e la richiesta di work memory sono elevati, vi possono essere ulteriori problematiche nei processi di elaborazione per la risoluzione del problema.

Strategie matematiche risolutive

Anche i cambiamenti nella scelta delle strategie risolutive sono stati ampiamente studiati e sono originati sia da fattori linguistici come per esempio le parole utilizzate, le categorie semantiche e la formulazione delle frasi, sia da fattori numerici come l'ordine di grandezza (Thevenot and Oakhill, 2005), il ruolo del numero all'interno dell'operazione, la collocazione della variabile (Garcia and al., 2006). Quest'ultimo particolare è stato studiato anche da un punto di vista della work memory da Swanson nel 2004.

Sono emersi numerosi aspetti interessanti, per esempio, persino per gli adulti, $4 + 2 = 6$ e $2 + 4 = 6$, che sono rappresentazioni matematicamente equivalenti, possono psicologicamente avere significati diversi (Kaput, 1079), ed anche per i bambini l'ordine in cui vengono presentati i numeri (dal più piccolo al più grande o viceversa), influenzano la risoluzione di semplici problemi con addizioni e sottrazioni.

Individuazione e reperimento delle informazioni

Una delle difficoltà nell'ambito dei WPs è l'individuazione, all'interno del testo, delle informazioni significative necessarie per giungere alla soluzione corretta. Informazioni di secondaria importanza distraggono e rendono più difficile il riconoscimento delle relazioni matematiche che il testo contiene. Queste informazioni aggiuntive possono essere presentate sotto forma di un numero o di una relazione espressa a parole, e possono portare a compiere una serie di passaggi inutili, soprattutto per il fatto che gli studenti ritengono che in un problema debbano essere utilizzati tutti i dati.

In generale i problemi a due passaggi sono molto più soggetti ad errori di quelli ad un solo passaggio (Muth, 1992), ma ciò non è da attribuire unicamente ad una maggior complessità aritmetica: di fatto i problemi a più passaggi implicano spesso una maggior complessità sotto il profilo linguistico.

Altri aspetti importanti rispetto alla cognizione numerica come ad esempio il senso dei numeri e dei simboli influenzano la risoluzione dei WPs e restano problemi aperti (MacGregor and Price, 1999).

1.1.3 Connessioni fra fattori linguistici e matematici

Le influenze linguistiche sulla competenza numerica e aritmetica sono numerose. Si è visto che i bambini cresciuti con una corretta struttura numero-parola hanno migliori prestazioni in numerosi tipi di prove, dal contare verbalmente fino ai problemi aritmetici. O'Neil (2013) ritiene che lingua e matematica abbiano probabilmente la stessa radice e che “le abilità richieste non sono così separate come pensiamo” e anche MacGregor e Price (1999) sostengono che fra lingua e matematica vi sia un profondo legame: “l'abilità cognitiva che guida l'elaborazione dei simboli è il collegamento fra matematica e lingua”. Bassok et al. (1998) avevano scoperto che le relazioni semantiche fra oggetti nei WPs matematici, era correlata positivamente con le operazioni aritmetiche che avevano questi oggetti come argomento. I correlati neurali (ovvero l'attività cerebrale che corrisponde ed è necessaria per una particolare esperienza) della visualizzazione e della verbalizzazione durante studi su WPs aritmetici suggerisce che il calcolo mentale è influenzato dai processi linguistici.

Per la loro stessa natura i WPs necessitano di una solida connessione fra le conoscenze linguistiche e matematiche. Per questa ragione non dovrebbe sorprenderci che nella loro risoluzione si compiano molti più errori rispetto a quelli compiuti risolvendo semplici problemi numerici (Koedinger and Nathan, 2004). I bambini sono in grado di risolvere parecchi tipi di problemi con somme e differenze e sono in grado di comprendere alcuni concetti numerici prima di cominciare la scuola. Questo porta però molto spesso a dare implicitamente per scontato che lo studente che si accinge a risolvere un problema abbia sempre le abilità aritmetiche necessarie, e ciò può portare all'idea sbagliata che vi siano fattori diversi dalla semplice abilità di calcolo a causare difficoltà nella risoluzione dei WPs (Reusser, 1993). Infine è stato anche riportato che le difficoltà nel risolvere i WPs potrebbe non essere dovuto né alla mancanza di comprensione nella lettura né ad abilità matematiche in generale (Hegarty et al., 1995).

La coerenza lessicale

Uno dei pochi fattori spesso studiati è l'incoerenza lessicale. Alcuni WPs contengono espressioni linguistiche come “meno” o “più” ai quali lo studente associa rispettivamente la sottrazione e l'addizione (Hegarty et al., 1995), gli studenti quindi tendono a cercare indicatori e parole chiave dai quali dedurre l'operazione aritmetica da compiere.

Lo studente posto di fronte a un WP costruisce un modello mentale del problema per pianificarne la risoluzione. Solitamente i risolutori di successo sono più propensi ad utilizzare una strategia di questo tipo: si focalizzano sui nomi delle variabili e sui termini relazionali, rileggono meno frequentemente il testo (risultato emerso dagli studi sui movimenti oculari). I risolutori meno abili invece fanno più spesso affidamento su una traduzione diretta delle parole chiave nell'operazione aritmetica; gli studi sui movimenti oculari mostrano come essi si focalizzino infatti su di essi rileggendo più volte il testo. Questo metodo conduce a molti errori nel momento in cui si incorre in testi lessicalmente incoerenti, ossia in cui, per esempio, la parola chiave “più” è associata alla sottrazione e “meno” all'addizione. Come esempio di incoerenza lessicale si consideri il seguente problema: “In drogheria una bottiglia di olio costa 7 euro, cioè 2 euro in più che al supermercato. Quanto costa una bottiglia di olio al supermercato?”. Le difficoltà a comprendere il testo e a trovare la soluzione corretta sono legate al fatto che l'avverbio “più” evoca il concetto di addizione, ma la soluzione giusta non è $7 + 2$ bensì $7 - 2$. Verschaffel et al. (1992) riscontrarono analoghi effetti dell'incoerenza sui tempi di reazione nei bambini ma non negli adulti. Nesher nel 1976 e Lean et al. nel 1990 in esperimenti con bambini e studenti senza problemi di apprendimento, dimostrarono che l'incoerenza semantica nei confronti dell'operazione matematica richiesta è un fattore importante nel determinare la difficoltà di un compito; inoltre essa aumenta i tempi di risposta (Hegarty et al. 1992). Nella rivisitazione di problemi a un solo passaggio effettuata da Verschaffel nel 1994 le ipotesi fatte sulla coerenza lessicale hanno una chiara conferma: gli studenti facevano circa il 13% in più di errori nell'individuare l'operazione corretta nei problemi con linguaggio incoerente e la difficoltà a comprendere tale linguaggio era maggiore quando l'operazione

matematica corretta sarebbe stata una somma. Non vi è però una letteratura che chiarisca se l'effetto dell'incoerenza riguarda sia gli studenti delle scuole di secondo grado che della scuola primaria.

Quando né la capacità linguistica né l'abilità aritmetica riescono da sole a spiegare il fallimento nei WPs, una possibile spiegazione è che complessità linguistica e numerica facciano affidamento sulle stesse risorse, per esempio entrambi gli aspetti richiedano una maggiore work memory. Infatti, solitamente, la prestazione nella risoluzione di un problema è correlata alla capacità di ridurre il carico mnemonico di informazioni inutili e irrilevanti (Passolunghi and Sieghel 2001).

Vi sono inoltre probabili influenze determinate da istruzioni fornite, per esempio dall'insegnante, su come i ragazzi devono risolvere i problemi, oppure dal metodo di valutazione adottato.

1.1.4 Differenze individuali e fattori sociali

Negli studi internazionali dell'OCSE-PISA viene normalmente utilizzata la nozione di mathematics literacy, definita come “la capacità di un individuo di utilizzare e interpretare la matematica e 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.

Molti studi si sono focalizzati sulle differenze individuali e di gruppo fra studenti con e senza problemi di apprendimento (Kingsdorf and Krawec, 2014) e tra bambini con e senza disabilità di sviluppo (Neef et al., 2003). Hegarty et al. (1995) distinguono strategie ambito-specifiche che i risolutori abili e non abili sviluppano con la pratica e come queste strategie spieghino le differenze individuali nella prestazione, sottolineando che non è possibile una completa generalizzazione nello studio delle difficoltà nell'ambito dei WPs, poichè studenti diversi hanno difficoltà con WPs diversi (Powell and Fuchs, 2014). Inoltre sono state studiate competenze

generali, come il QI, e conoscenze e procedimenti ambito-specifici al fine di fornire un resoconto completo sul problem solving e le abilità cognitive di base. Nel processo di risoluzione sono implicate abilità visive, matematiche, di lettura e abilità metacognitive. Boonen nel 2014 ha indagato il ruolo dell'abilità spaziale e della comprensione della lettura in quanto i bravi risolutori di WPs non selezionano numeri e parole chiave relazionali ma creano una rappresentazione visiva (Boonen et al., 2013).

Anche fattori sociali quali l'istruzione scolastica, gli insegnanti, i compagni meritano attenzione in quanto la modalità di risposta, i criteri di giudizio, la presenza di illustrazione nel testo o i modelli risolutivi utilizzati dagli insegnanti influenzano in gran parte le prestazioni degli alunni.

Per quanto sia noto che complessità linguistica e numerica, come pure le loro interazioni, contribuiscono a rendere più o meno difficile un problema, spesso i ricercatori si sono focalizzati su uno o l'altro di questi aspetti a seconda del campo di ricerca da cui provenivano.

Per meglio comprendere le difficoltà nell'ambito dei WPs sarebbe auspicabile manipolare isolatamente le diverse variabili ma ciò è stato raramente fatto. Per esempio in alcuni studi gli spunti linguistici sono di categorie diverse (avverbi, verbi, sostantivi, ecc...) e l'effetto della complessità delle strutture sintattiche non è stata tenuta in considerazione. Spesso sono trascurati aspetti importanti come le proprietà di base dei numeri (ad esempio la grandezza, il valore posizionale nei numeri a più cifre, ecc...) o la complessità del calcolo (effetto riporto, ecc...).

Durante un'analisi congiunta di processi linguistici e numerici si deve tener conto che le difficoltà possono essere non solo una somma delle difficoltà numeriche e di quelle linguistiche ma, come nel caso della work memory, si possono originare difficoltà date dall'interazione di questi due ambiti.

Infine si suggerisce di affrontare lo studio attraverso un approccio psicologico di differenziazione. Studenti diversi possono avere difficoltà diverse: studenti deboli sotto il profilo linguistico possono avere problemi con WPs linguisticamente complessi e studenti deboli in aritmetica avranno problemi con la complessità arit-

metica. Solo differenziazioni a livello dell'item (per quanto riguarda la complessità linguistica e numerica e le loro interazioni) e a livello individuale (per quanto riguarda abilità numeriche, linguistiche, e cognitive generali) sarà possibile comprendere perché un particolare bambino ha difficoltà con un particolare problema.

Capitolo 2

Lo strumento statistico

2.1 IRT (Item Response Theory)

La teoria della risposta all'item si sviluppò all'inizio degli anni '60, ed è composta da un insieme di modelli e teorie per l'analisi e la misura di tratti latenti¹. Si definisce *abilità*, e viene indicata con Θ quella variabile latente che si assume influenzi e determini la riuscita di un soggetto nel rispondere a uno o più item. Ogni soggetto possiede un certo livello di abilità perciò ognuno, sottoposto alla stessa prova, potrà fornire una prestazione diversa. L'abilità non è l'unico elemento che viene tenuto in considerazione dall'IRT, infatti, altri elementi influenzano la performance dell'individuo, e possono essere classificati come qualità caratterizzanti per l'item. Tali parametri sono:

- il *livello di difficoltà* dell'item, definito come il livello di abilità richiesto affinché un soggetto abbia lo 0,5 di probabilità di rispondere correttamente all'item;
- il *livello di discriminazione* dell'item, ovvero la capacità di un item di discriminare tra soggetti che hanno differenti livelli di abilità;

¹Si dice tratto latente ogni caratteristica di un individuo o cose per le quali non sia disponibile un adeguato strumento di osservazione e di misura

- *l'incidenza del caso*. La presenza di questo parametro fa sì che anche un individuo con livelli bassissimi di abilità abbia una probabilità maggiore di zero di risponder correttamente all'item, per esempio provando a indovinare.

È proprio il numero di parametri presi in considerazione che distingue tra loro i metodi dell'IRT; nella prossima sezione verrà analizzato il metodo di Rasch, per il quale, oltre al tratto latente Θ viene considerato un solo parametro, ossia la difficoltà dell'item.

I modelli dell'IRT, applicati a dati dicotomici² hanno come obiettivo quello di calcolare la probabilità che un soggetto ha di rispondere correttamente a ciascun item del test somministrato, in funzione dell'abilità posseduta dal soggetto stesso e dei parametri dell'item utilizzato [2].

2.2 Il modello di Rasch

Il modello logistico ad un parametro o modello di Rasch è il modello più semplice dell'IRT, e prende il nome dal matematico danese che negli anni '60 lo teorizzò. Egli, approcciandosi all'analisi dei test da un punto di vista probabilistico, introdusse un modello che spiega l'interazione tra l'abilità di uno studente e la difficoltà dell'item nel processo di risposta.

Si supponga di somministrare un test composto da k items ad un numero N di soggetti. La variabile aleatoria X_{ij} con $i = 1, 2, \dots, N$ e $j = 1, 2, \dots, k$ indica la risposta data dall' i -esima persona alla j -esima domanda. X_{ij} potrà assumere i valori nell'insieme $\{0, 1\}$ ed in particolare sarà $X_{ij} = 1$ se la risposta dell' i -esimo studente alla j -esima domanda è corretta altrimenti si avrà $X_{ij} = 0$. Si consideri la matrice $N \times k$ che rappresenta i risultati ottenuti dalla somministrazione

$$\Lambda = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{pmatrix} \quad (2.1)$$

²Per item dicotomici si intendono quegli item che, pur avendo più di un'alternativa, prevedono una sola risposta corretta e la cui codifica può essere ricondotta ad un'alternativa 0, 1.

dove $x_{ij} \in \{0, 1\}$ in quanto rappresenta il valore assunto dalla variabile aleatoria X_{ij} . Lo scopo è quello di estrarre da questi dati una misura dell'abilità A_i per ogni studente, e, per ogni item, una misura del livello di difficoltà D_j . Rasch ipotizza che la probabilità che lo studente i -esimo risponda correttamente all'item j -esimo sia funzione del rapporto $\frac{A_i}{D_j}$. Le ipotesi su $f\left(\frac{A_i}{D_j}\right)$ sono le seguenti:

1. $f : \mathbb{R} \longrightarrow [0, 1]$;
2. $\lim_{A_i \rightarrow 0} f\left(\frac{A_i}{D_j}\right) = 0, \quad \lim_{A_i \rightarrow \infty} f\left(\frac{A_i}{D_j}\right) = 1,$
 $\lim_{D_j \rightarrow 0} f\left(\frac{A_i}{D_j}\right) = 1, \quad \lim_{D_j \rightarrow \infty} f\left(\frac{A_i}{D_j}\right) = 0;$
3. fissato A_j , f è funzione monotona crescente di D_j , e, fissato D_j , f è funzione monotona crescente di A_i .

Nel modello di Rasch:

$$f\left(\frac{A_i}{D_j}\right) = P(X_{ij} = 1 | A_i, D_j) = \frac{\frac{A_i}{D_j}}{1 + \frac{A_i}{D_j}}, \quad (2.2)$$

Questa scelta per f è la “più semplice”, nel senso che non esistono funzioni polinomiali che verificano le tre proprietà e, tra le funzioni razionali, quella proposta ha grado minimo per numeratore e denominatore.

Infine, per uniformare il modello di Rasch con gli altri modelli dell'IRT, si pone $\Theta_i = \ln A_i$ e $b_j = \ln D_j$, ottenendo:

$$P(X_{ij} = 1 | \Theta_i, b_j) = \frac{e^{(\Theta_i - b_j)}}{1 + e^{(\Theta_i - b_j)}}. \quad (2.3)$$

Se 2.3 è la probabilità per lo studente i -esimo di rispondere correttamente alla j -esima domanda, $P(X_{ij} = 0 | \Theta_i, b_j) = 1 - \frac{e^{(\Theta_i - b_j)}}{1 + e^{(\Theta_i - b_j)}} = \frac{1}{1 + e^{(\Theta_i - b_j)}}$, sarà la probabilità, per lo stesso studente, di rispondere in modo errato alla stessa domanda. A questo punto il modello di Rasch assume l'indipendenza dell'insieme delle variabili aleatorie X_{ij} ³ ossia, si assume che siano indipendenti fra loro le risposte date da uno stesso soggetto a diversi item, nonchè le risposte date da soggetti diversi allo

³Due variabili aleatorie X e Y si dicono indipendenti se e solo se $P(X \cap Y) = P(X)P(Y)$

stesso item. Pertanto la probabilità di ottenere come una matrice delle risposte come in Tebella 2.1 è

$$P(X = \Lambda | \Theta, b) = \prod_{i=1}^N \prod_{j=i}^k \frac{e^{x_{ij}(\Theta_i - b_j)}}{1 + e^{(\Theta_i - b_j)}}.$$

Risulta evidente che la probabilità che un individuo risponda correttamente all'item dipende unicamente dall'abilità del soggetto e dalla difficoltà della domanda.

2.2.1 Stima dei parametri

Il primo passo da compiere è quello di stimare la variabile latente Θ e il parametro di difficoltà b . Il metodo di stima più utilizzato in letteratura è il principio di massima verosimiglianza che a sua volta, a seconda della funzione massimizzata, si suddivide in metodo di massima verosimiglianza congiunta, condizionata o marginale. Questi processi consistono nell'individuare i valori dei parametri incogniti $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_N)$ e $b = (b_1, b_2, \dots, b_k)$ che massimizzano la funzione scelta. Per l'analisi dei risultati della sperimentazione è stato utilizzato il software *ACER ConQuest* il quale utilizza il metodo di Massima Verosimiglianza Marginale (MML), che risulta migliore rispetto agli altri metodi (Massima Verosimiglianza Congiunta JML, Massima Verosimiglianza Condizionata CML) [10], [9].

Il metodo di massima verosimiglianza marginale (MML)

Nel Metodo di massima verosimiglianza marginale si assume che i soggetti rappresentino un campione casuale di una popolazione la cui abilità è distribuita secondo una certa funzione di densità $G(\Theta)$.

Sia Θ_i l'abilità di un fissato soggetto s_i , x_{i*} sarà dunque il vettore delle risposte date da s_i . Si avrà

$$P(X_{i*} = x_{i*} | G, b) = \int_{-\infty}^{+\infty} \prod_{j=1}^k \frac{e^{x_{ij}(\Theta_i - b_j)}}{1 + e^{(\Theta_i - b_j)}} dG(\Theta)$$

La funzione di verosimiglianza marginale da massimizzare sarà dunque:

$$L(\Theta, b) = \prod_{j=1}^N \int_{-\infty}^{+\infty} \prod_{i=1}^k \frac{e^{x_{ij}(\Theta_i - b_j)}}{1 + e^{(\Theta_i - b_j)}} dG(\Theta)$$

che rappresenta la probabilità di osservare una matrice delle risposte come in Tabella 2.1.

2.2.2 Proprietà di sufficienza del modello di Rasch

Consideriamo la Matrice 2.1, chiamiamo punteggio grezzo dell' i -esimo soggetto il valore $r_i = \sum_{j=1}^k x_{ij}$ il quale rappresenta il punteggio totale raggiunto, chiamiamo invece punteggio grezzo dell'item il valore $s_j = \sum_{i=1}^N x_{ij}$ che conta il numero di soggetti che hanno risposto correttamente al j -esimo item.

La proprietà di sufficienza del modello di Rasch afferma che:

L'insieme dei punteggi grezzi di soggetti ed item è sufficiente per la stima di tutti i parametri del modello, ossia:

- i. la probabilità di ottenere un punteggio grezzo per l' i -esimo soggetto $r = r_i$, non dipende dai parametri Θ_i dei soggetti;*
- ii. la probabilità di ottenere un punteggio grezzo per il j -esimo item $s = s_j$, non dipende dalla difficoltà b_j .*

Ciò significa che secondo il modello di Rasch, il vettore delle risposte di ogni soggetto non contiene maggiori informazioni, al fine di stimare la sua capacità, di quanto non ne contenga già il valore sintetico r_i . Da ciò segue che non è necessario ricordare i valori specifici delle risposte di tutti i soggetti agli item, ma solamente i loro punteggi grezzi.

2.3 Strumenti statistici e indici utilizzati per la valutazione degli apprendimenti

L'Item Response Theory insieme alla precedente Teoria Classica dei Test (TCT), vengono ampiamente utilizzate nella ricerca in didattica, e nella valutazione degli

apprendimenti nazionali ed internazionali.

Nel rapporto tecnico delle rilevazioni nazionali degli apprendimenti del 2013/2014 [6] vengono illustrati gli strumenti tecnico-statistici principali e gli indici, utilizzati dall'INVALSI, per valutare le caratteristiche misuratorie di una prova standardizzata. Tra questi verranno trattati quelli che interessano questa ricerca:

- l' α di Cronbach per la valutazione dell'attendibilità del test;
- il modello di Rasch per la stima della difficoltà degli item e dell'abilità degli studenti;
- placement degli item rispetto al punteggio di Rasch.

2.3.1 L' α di Cronbach

L'attendibilità è la proprietà di un test che riguarda l'accuratezza con cui un test misura una certa caratteristica, ossia esprime mediante un numero la precisione con cui lo strumento statistico utilizzato misura tale caratteristica, ed è inversamente proporzionale all'errore presente nella misura stessa [2]. Il coefficiente di attendibilità varia nell'intervallo $[0, 1]$ più l'indice è vicino a 1, tanto più è alta la coerenza complessiva della prova. Nel caso di questa ricerca la stima dell'attendibilità farà riferimento al concetto di coerenza interna del test, ovvero al grado di accordo tra gli item e il test nel suo complesso. Perché la coerenza interna sia buona è necessario che gli item siano ben correlati tra loro. Nel caso di un test per misurare l'abilità matematica di uno studente ciò significa che gli item dovrebbero misurare la stessa variabile latente, ossia proprio l'abilità dell'alunno. Esistono vari coefficienti di attendibilità e quindi vari modi per calcolarli a seconda dello studio che si sta conducendo; in questo caso verrà utilizzato l' α di Cronbach che viene calcolato in funzione delle varianze degli item che compongono il test e della varianza del test stesso. La formula per calcolare questo coefficiente è:

$$\alpha = \frac{N}{N-1} \left(1 - \frac{\sum_{i=1}^N \sigma_i^2}{\sigma^2} \right)$$

dove N è il numero di item che compongono il test, σ_i^2 la varianza dell' i -esimo item e σ^2 la varianza del test.

Il calcolo dell' α di Cronbach permette quindi di avere una valutazione sintetica e generale sulla coerenza complessiva della prova e quindi della sua attendibilità in funzione delle informazioni che da essa provengono. Solitamente è considerato adeguato un valore superiore a 0,70. L' α di Cronbach costituisce però solo un'analisi preliminare dell'affidabilità e della coerenza di una prova standardizzata. Risulta quindi necessario approfondire ulteriormente tali aspetti con l'utilizzo di altri strumenti, come ad esempio il Modello di Rasch.

2.3.2 Il modello di Rasch e il placement degli item

Il modello di Rasch è uno strumento grazie al quale è possibile verificare se la prova proposta agli studenti sia caratterizzata da un grado di affidabilità sufficiente per ottenere misurazioni precise dei livelli di apprendimento degli studenti stessi. Un altro aspetto interessante di questo metodo statistico è la possibilità di ottenere una scala a intervalli, sulla quale vengono collocati gli item in base alla loro difficoltà; sulla medesima scala viene posizionato il livello della performance di ciascun alunno. La possibilità di ordinare sullo stesso continuum sia la difficoltà degli item sia l'abilità degli studenti è molto importante dal punto di vista interpretativo, poiché consente di comprendere quali e quanti sono gli allievi che mostrano livelli di apprendimento superiori o inferiori rispetto alla difficoltà di una determinata domanda o di un insieme di quesiti, e di conseguenza comprendere ciò che gli allievi conoscono e sono in grado di fare. Un esempio di ciò lo troviamo nella Figura 2.1. Il grafico presenta nella parte sinistra la scala di valori utilizzati per il parametro di difficoltà dell'item; essa varia convenzionalmente nell'intervallo $[-4, 4]$ oppure $[-3, 3]$. Le domande che presentano un parametro di difficoltà negativo sono considerate facili, mentre quelle con parametro superiore a 0 risultano difficili. La performance di ogni studente è collocata su questa scala di valori (ogni crocetta rappresenta un numero n di studenti) e segue una distribuzione a campana. Nella parte destra, invece, sono collocate le domande a seconda del loro grado di difficoltà. La lettura di questo grafico ci permette di individuare, scelta

3	XXXXXXX	
	XXXXXXXXXX	
	XXXXXXXXXXXXXXXX	
2	XXXXXXXXXXXXXXXXXX	
	XXXXXXXXXXXXXXXXXXXX	
	XXXXXXXXXXXXXXXXXXXXXXXX	
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	
1	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	8 19
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	13
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	28
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	16 20
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	32
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	30
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	25
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	10 29 35
0	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	12 14 21
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	3 5
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	31 36
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	22
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	15
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	7 9 37 39
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	4
-1	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	24 38
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	11 18 23
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	33
	XXXXXXXXXXXXXXXXXXXXXXXXXXXX	17 34
	XXXXXXXXXXXXXXXXXXXX	
	XXXXXXXXXXXXXXXX	6
	XXXXXXX	
-2		1 2
	XXXXXX	
	XXXX	
	XX	26
	X	
-3		27
	X	

Figura 2.1: Esempio di placemente di una prova INVALSI.

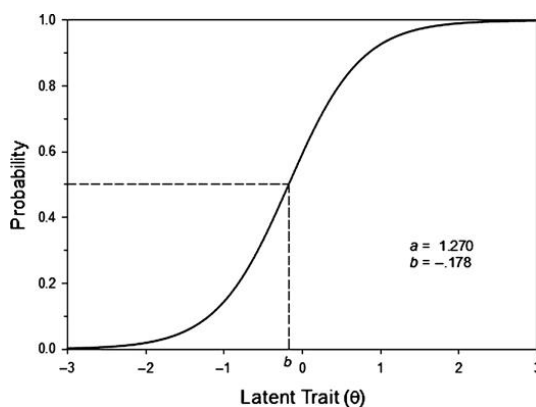


Figura 2.2: Esempio di Item Characteristic Curve

una domanda, quanti studenti hanno una scarsa probabilità di rispondere correttamente alla domanda stessa e quanti invece hanno una probabilità elevata di dare una risposta corretta: gli studenti che si trovano in una posizione inferiore a quella occupata dall'item scelto, hanno una probabilità inferiore al 50% di rispondere correttamente, mentre, coloro che sono posizionati al di sopra della stessa, hanno una probabilità superiore al 50% di rispondere correttamente.

Infine con i risultati del modello di Rasch è possibile rappresentare ciascun item tramite una funzione non lineare monotona crescente, che descriva il variare della probabilità di risposta corretta all'item al variare dell'abilità del rispondente. Queste curve sono chiamate ICC (Item Characteristic Curve) ed un esempio è riportato nella Figura 2.3.2. La lettura di questo grafico permette di individuare la principale caratteristica dell'item, ossia la sua difficoltà b . Tale caratteristica è individuata dal punto sulla scala delle abilità al quale corrisponde una probabilità di risposta del 50%. Si nota che una traslazione della curva verso destra corrisponde ad un aumento della difficoltà; viceversa, una diminuzione della difficoltà corrisponderà a una traslazione verso sinistra.

Capitolo 3

Metodologia e piano di validazione dello strumento

In questo capitolo verranno presentate le modalità con cui si è svolta la sperimentazione: le caratteristiche della popolazione campionaria, la stesura del test variato, le modalità di raccolta dei dati ed infine il piano per la validazione dello strumento.

3.1 La prova INVALSI di partenza

La prima fase della sperimentazione ha riguardato la selezione di una delle prove standardizzate dell'INVALSI. Il fatto di partire da un test già somministrato in passato per ricerche di natura analoga a quella qui descritta, è stata fatta per due principali motivi: per garantire l'attendibilità e l'efficacia della prova utilizzata per la misurazione del tratto latente, e per avere un termine di paragone con cui confrontare i risultati ottenuto dall'analisi del nuovo campione. Nella sezione 4 si vedrà infatti che il primo passo per la validazione dello strumento implementato sarà il paragone tra i risultati ottenuti dal nostro campione e quelli del CN (sigla con la quale indicheremo il Campione Nazionale) e del CE (campione rappresentativo della regione Emilia-Romagna) del 2013.

Il test selezionato fu somministrato a maggio del 2013 a 590.728 studenti italiani

frequentanti la Classe I della Scuola Secondaria di primo grado; si tratta quindi di un test di livello L06. Le prove INVALSI di livello L06, che si svolgevano al termine dell'anno scolastico, furono abolite nel 2014. Questa è la principale ragione per cui è stato selezionato un test di tale livello: si è in tal modo scongiurata la possibilità che gli studenti avessero già svolto la prova, magari in preparazione all'esame che avrebbero dovuto condurre a fine anno.

Nel 2013, il test INVALSI, che indicheremo con TO (Test Originale), è stato svolto da 590.728 studenti italiani. È stato poi selezionato un campione rappresentativo (CN) di 27.000 studenti, e su di esso sono state svolte le analisi statistiche; 1.528 individui del CN componevano il campione dell'Emili-Romagna.

Il test originale TO è costituito di 48 item ed è in tutto identico alla prova INVALSI del 2013.

3.2 Il test variato

A partire dal test originale, sono stati selezionati 7 dei 48 item. Gli item scelti sono stati modificati in modo tale che ciascuno di essi presentasse una e una sola variazione ben identificata.

Sostituendo quindi i 7 item modificati ai 7 originali si è ottenuta una seconda prova di 48 item - 41 dei quali risultano identici a quelli della prova precedente - che chiameremo TV (Test Variato).

L'insieme formato dai 41 item in comune prenderà il nome di CoreTest e verrà indicato con CT. Il CT sarà fondamentale per permettere il confronto tra i risultati del test originale e del test variato.

Verrà ora mostrata, per ciascuno dei 7 item, la modifica apportata, identificandone la natura.

Variazione Item D3

D3. Quante cifre ha il risultato della seguente moltiplicazione?

$1001 \cdot 20002$

Risposta: cifre

Figura 3.1: Item D3 originale

D3. Quante cifre ha il risultato della seguente moltiplicazione?

$1001 \cdot 20002$

A. 8

B. 9

C. 5

D. 2

Figura 3.2: Item D3 variato

L'item D3, nella sua versione originale (Figura 3.1), chiede allo studente di contare il numero delle cifre del prodotto tra i numeri che si leggono nel riquadro. In questo caso si è modificata la tipologia di domanda, ossia si è passati da una domanda aperta, che dà spazio a qualsiasi tipo di risposta da parte dello studente, a una domanda chiusa nella quale egli può scegliere tra quattro alternative proposte (Figura 3.2).

Variazione Item D8.b

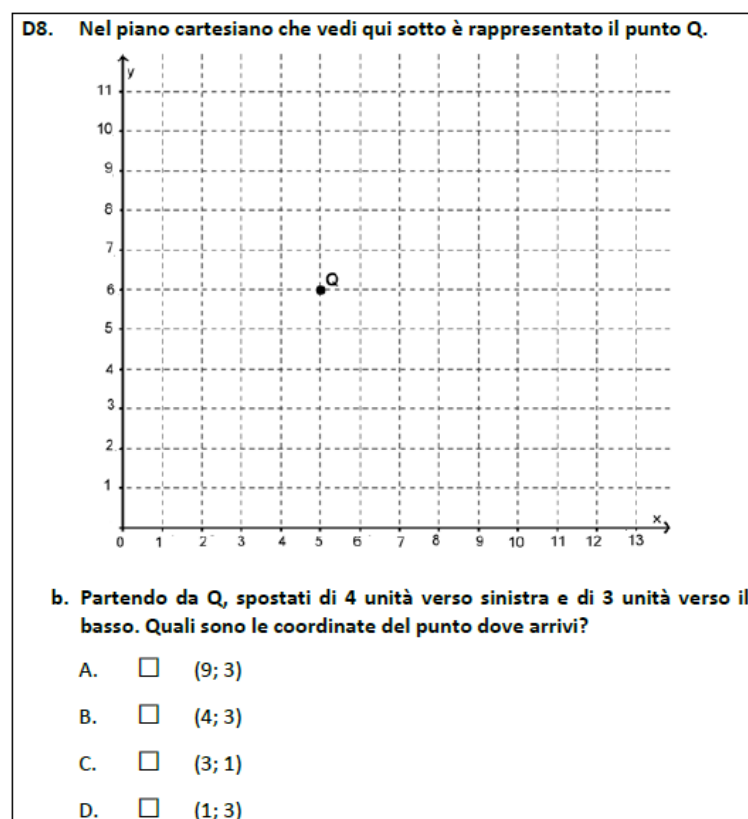


Figura 3.3: Item D8.b originale

La domanda D18 presenta al suo interno due quesiti; la modifica ha riguardato solamente il punto b mostrato in Figura 3.3. Al punto b si chiedeva, partendo da un punto assegnato Q, di eseguire una serie di movimenti - che venivano indicati - per raggiungere un punto P del quale si sarebbero poi dovute individuare le coordinate (Figura 3.3).

Sostanzialmente il testo della domanda è stato riorganizzato facendo in modo di invertire il percorso che lo studente deve seguire per giungere al punto di cui determinare le coordinate.

D8. Nel piano cartesiano che vedi qui sotto è rappresentato il punto Q.

[...]

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)

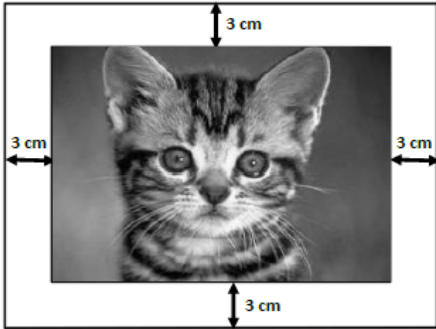
D. (1; 3)

Figura 3.4: Item D8.b variato

Vediamo infatti (Figura 3.4) che viene indicato come punto di partenza il punto P di cui non si conoscono ancora le coordinate e i movimenti da compiere sono presentati in maniera inversa rispetto ai precedenti.

Variazione Item D14

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

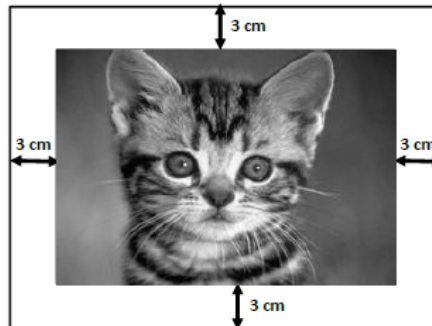
C. 28 cm x 18 cm

D. 25 cm x 18 cm

Figura 3.5: Item D14 originale

L'Item D14 presentava un problema di geometria piana. Nella prima parte il quesito contestualizzava il problema e forniva i dati: su un cartoncino veniva incollata un'immagine lasciando un bordo di 3 cm su ogni lato. Il quesito presentava inoltre un'immagine allo scopo di facilitare la comprensione del testo (Figura 3.5). Infine veniva posta la domanda che chiedeva di calcolare l'area del cartoncino in questione. La variazione apportata in questo caso va semplicemente a modificare la posizione dell'immagine rispetto alla formulazione della domanda. Come si può osservare in Figura 3.6, la foto non è più posizionata tra i dati e la domanda ma si colloca all'interno della presentazione dei dati.

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

Figura 3.6: Item D14 variato

Variazione Item D16

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

Figura 3.7: Item D16 originale

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

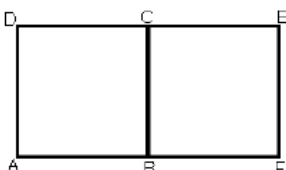
D. 8

Figura 3.8: Item D16 variato

La domanda D16 presentava un problema di aritmetica. La variazione apportata ha interessato la sintassi del testo, nel tentativo di complicare la struttura del periodo. Conducendo un'analisi del testo originale, riportato in Figura 3.7, si osserva che esso è formato da due periodi. Mentre il primo è formato da un'unica frase reggente, il secondo presenta la principale, nella quale è posta la domanda, e una subordinata relativa.

Variazione Item D18

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

Figura 3.9: Item D18 originale

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

Figura 3.10: Item D18 variato

L'item D18 presenta un problema di geometria in cui è chiesto di calcolare il perimetro di un rettangolo formato dall'affiancamento di due quadrati identici dei quali è noto il perimetro. La formulazione originale comprendeva anche un'immagine illustrativa della situazione descritta come si nota in Figura 3.9.

La modifica apportata in questo caso è stata quella di eliminare la figura illustrativa, lasciando invariato il testo della domanda.

Variazione Item D22

L'Item modificato, nella sua versione originale, chiedeva di stimare il prodotto tra due numeri decimali come in Figura 3.11. La modifica in questo caso è di tipo numerico ed interessa la tipologia dei numeri coinvolti. I numeri decimali del cui prodotto era chiesta una stima, sono stati sostituiti da numeri interi di 2 ordini di grandezza più grandi. Ovviamente la stessa modifica è stata apportata anche alle alternative di risposta (Figura 3.12).

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

Figura 3.11: Item D22 originale

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

Figura 3.12: Item D22 variato

Variazione Item D27

<p>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?</p> <p>A. <input type="checkbox"/> 150 g</p> <p>B. <input type="checkbox"/> 260 g</p> <p>C. <input type="checkbox"/> 510 g</p> <p>D. <input type="checkbox"/> 520 g</p>
--

Figura 3.13: Item D27 originale

<p>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?</p> <p>A. <input type="checkbox"/> 150 g</p> <p>B. <input type="checkbox"/> 260 g</p> <p>C. <input type="checkbox"/> 510 g</p> <p>D. <input type="checkbox"/> 520 g</p>
--

Figura 3.14: Item D27 variato

Anche all'item D27 è stata apportata una variazione a livello della sintassi. Il problema aritmetico chiedeva di calcolare il peso di uno di due quaderni uguali conoscendo il peso totale dello zaino in cui esso era contenuto, insieme ad altri libri il cui peso è noto. Il testo è formato da tre periodi. Il primo contiene una principale e tre subordinate relative di pari livello. Gli altri due periodi sono formati da una sola frase principale ed il secondo di essi esprime la domanda.

1° periodo reggente sub. relativa
Nello zaino di Chiara ci sono il libro di scienze, / (che) pesa mezzo chilo, / il libro di
 sub. Relativa coordinata 2° periodo
matematica, / (che) pesa 980 g, / (e) due quaderni uguali. // Libri e quaderni pesano in
 3° periodo
tutto due chilogrammi. // Quanto pesa ciascun quaderno?

1° periodo
Nello zaino di Chiara ci sono il libro di scienze, il libro di matematica e due
 2° periodo reggente sub. condizionale sub. dichiarativa
quaderni uguali. // Quanto pesa ciascun quaderno, / sapendo / (che) il libro di scienze
 sub. dichiarativa sub. dichiarativa sub. dichiarativa
pesa mezzo chilo, / il libro di matematica pesa 980 g, / (e che) libri e quaderni pesano in
tutto due chilogrammi?

L'item variato (Figura 3.14) contiene due periodi. Il primo è costituito da una sola frase principale mentre il secondo è un periodo complesso costituito da una principale, nella quale viene espressa la domanda, una subordinata condizionale da cui dipendono tre dichiarative di pari livello, contenenti i dati.

3.3 Selezione del campione e raccolta dei dati

La ricerca è stata svolta su un campione di 40 Classi seconde della scuola secondaria di primo grado, durante il primo quadrimestre, per un totale di 777 studenti. Le classi sono state selezionate casualmente all'interno della Regione Emilia-Romagna e ad ognuna di esse è stato assegnato un codice che è stato poi riportato su ciascun fascicolo. Il numero era composto da quattro cifre: le prime due indicanti la scuola di provenienza e le seconde due la sezione. Le scuole che hanno collaborato con la ricerca sono elencate in tabella 3.3. Al momento

Scuola	Città
Istituto Comprensivo 5	Bologna
Istituto Comprensivo di Rubiera	Rubiera (RE)
Scuola San Giovanni Bosco	Imola (BO)
Istituto Comprensivo Filippo Zappi	Mercato Saraceno (FC)
Istituto Comprensivo di Santa Sofia	Santa Sofia (FC)
Istituto Comprensivo di Argelato	Argelato (BO)
Istituto Comprensivo di Crevalcore	Crevalcore (BO)
Scuola Visitandine Malpighi	Castel San Pietro Terme (BO)
Scuola Malpighi	Bologna
Istituto Comprensivo Italo Calvino	Fabbrico (RE)
Istituto Comprensivo Italo Calvino	Rolo (RE)
Istituto Comprensivo 7 - L.Orsini	Imola (BO)

della somministrazione ogni classe è stata suddivisa in due gruppi seguendo la numerazione del registro di classe; in particolare i gruppi sono stati formati dagli studenti con numeri pari e da quelli con numeri dispari. Questo tipo di suddivisione è del tutto casuale e permette di avere, al termine della raccolta dei dati, due sottogruppi equivalenti. A ciascun membro di uno dei due sottoinsiemi - non necessariamente uno in particolare dei due - è stato consegnato il test TO, mentre a ciascuno degli studenti dell'altro gruppo è stato consegnato il test TV.

La somministrazione è sempre stata presieduta da una delle ricercatrici e si è svolta per ciascuna classe nelle seguenti modalità:

1. Numerazione e distribuzione dei fascicoli come sopra indicato.
2. Lettura ad alta voce del regolamento della prova¹.
3. Svolgimento della prova. Il tempo a disposizione era di 75 min. Durante lo svolgimento non è stata data la possibilità agli studenti di fare domande.
4. Ritiro dei fascicoli².

I fascicoli sono stati corretti seguendo la "Grigli di correzione"³, ed in fine è stato creato il dataset inserendo i risultati in un unico file excel, come mostrato in Figura 3.15. Al termine della sperimentazione la numerosità delle due sottopopolazioni PO e PV è risultata rispettivamente di 380 e 397 studenti.

ID STUDENTE (sc.rcd.-stud.)	SOMM.	FASCICOLO (O/V)	SESSO (M/F)	CITTA (IT/ST)	D1_a	D1_b	D1_c1	D1_c2	D2_a	D2_b	D2_c	D3	D4	D5	D6_a	D6_b	D6_c	D7_a	D7_b	D8_a	D8_b
133601	Rebecca	V	M	IT	A	A	B	A	D	C	B	A	B	B	A	D	B	A	C	A	C
133602	Rebecca	O	M	IT	A	A	A	B	C	D	B	A	B	D	A	D	B	C	B	B	C
133603	Rebecca	V	M	IT	A	A	A	B	D	A	B	A	B	D	A	B	B	A	C	B	C
133604	Rebecca	O	F	IT	A	A	B	A	C	D	A	A	B	D	B	B	B	D	B	C	C
133605	Rebecca	V	F	IT	A	B	A	A	B	D	M	C	B	D	A	B	B	A	A	B	C
133606	Rebecca	O	M	IT	A	B	A	A	C	B	B	B	A	A	B	B	B	B	C	B	C
133607	Rebecca	V	M	IT	A	A	A	B	A	C	M	A	B	D	B	A	A	B	C	A	D
133608	Rebecca	O	F	IT	A	A	B	A	D	A	B	B	B	D	A	B	B	A	B	B	A
133609	Rebecca	V	F	IT	A	A	A	B	C	B	B	A	B	D	M	B	B	B	D	A	D
133610	Rebecca	O	F	IT	B	A	B	A	C	D	B	A	B	D	B	B	A	B	A	B	D
133611	Rebecca	V	F	IT	A	A	A	A	D	B	B	A	B	D	B	B	M	C	A	D	D
133612	Rebecca	O	F	IT	A	A	A	B	B	A	A	B	A	D	B	B	B	D	C	A	D
133613	Rebecca	V	M	IT	A	A	A	B	C	B	A	A	B	D	A	B	D	C	B	C	C
133614	Rebecca	O	F	IT	A	A	A	B	C	B	B	A	B	B	A	B	B	A	A	A	D
133615	Rebecca	V	F	IT	A	A	A	B	C	B	B	A	B	D	A	B	A	B	C	B	C
133616	Rebecca	O	M	IT	A	A	A	A	D	D	B	A	B	D	M	B	B	B	B	B	C
133617	Rebecca	V	F	IT	A	B	B	A	C	A	M	A	B	B	A	B	D	C	A	D	D
133618	Rebecca	O	M	IT	M	B	A	B	A	A	B	B	B	B	A	B	C	B	B	C	C
133619	Rebecca	V	M	IT	A	A	A	A	A	B	M	C	A	B	B	B	B	C	B	B	B
133620	Rebecca	O	F	IT	A	B	A	A	D	B	M	B	A	B	A	A	B	D	M	A	A
133621	Rebecca	V	M	IT	A	A	A	B	C	A	B	A	A	B	B	B	D	B	A	B	B
133622	Rebecca	O	F	IT	A	A	A	B	C	B	A	B	B	D	A	B	A	B	B	A	D
133623	Rebecca	V	F	IT	A	A	A	A	D	C	B	A	B	D	A	B	B	B	B	A	A
133624	Rebecca	O	M	IT	A	A	A	B	D	B	A	A	B	D	A	B	B	B	C	B	D
133701	Rebecca	O	M	IT	A	A	A	B	D	B	A	B	B	B	B	B	C	C	B	C	C
133702	Rebecca	V	M	IT	B	A	A	B	D	B	B	B	B	A	B	D	B	A	C	B	B
133703	Rebecca	O	F	IT	A	A	A	B	D	B	A	A	B	D	A	B	B	B	C	B	C
133704	Rebecca	V	M	IT	A	A	A	A	D	A	B	A	B	A	A	B	B	B	C	A	A
133706	Rebecca	V	M	IT	A	A	A	B	D	B	M	A	B	D	A	D	B	A	B	A	A
133707	Rebecca	O	F	IT	A	A	A	B	B	B	A	B	B	D	A	B	B	B	B	A	C
133708	Rebecca	V	F	IT	A	A	B	A	C	B	B	A	B	D	A	B	M	B	B	B	C

Figura 3.15: Dataset.

¹Il regolamento era riportato all'inizio di ciascun fascicolo ed era identico a quello utilizzato per le prove INVALSI. Anche in questo caso la scelta è stata fatta per salvaguardare il più possibile l'analogia tra le due ricerche

²Le correzioni ed i risultati delle prova sono stati riconsegnati ai docenti di ciascuna classe, una volta terminata la somministrazione in tutte le sezioni appartenenti allo stesso istituto.

³GRIGLIA DI CORREZIONE 2013, Matematica Classe I Scuola secondaria di primo grado, fascicolo 1, INVALSI.

Capitolo 4

La validazione dello strumento

Verrà ora presentata e validata la metodologia attraverso la quale si vuole rispondere alla domanda: come si può valutare l'impatto di una variazione nella formulazione del testo, sulle prestazioni degli studenti, in un test standardizzato? Come già detto si è partiti da un test TO composto da 48 item somministrato ad un un campione di 2700 studenti, rappresentativo degli studenti italiani con una scolarizzazione di livello 6. 7 di questi item sono stati modificati secondo i criteri precedentemente descritti, mentre i restanti 41 sono andati a formare quello che è stato definito CoreTest (CT).

Il software utilizzato per eseguire l'analisi mediante il metodo di Rasch, è ConQuest. I primi risultati ottenuti sono stati calcolati avendo come input il dataset dell'intera popolazione P (777 studenti) sui 41 item del CT. Lo stesso metodo di analisi è stato poi eseguito, sempre sul CT, relativamente alla popolazione PO composta dai 380 studenti ai quali era stato assegnato il test originale, ed infine, è stato preso in esame il CT sulla popolazione PV composta dai 397 studenti ai quali era sato assegnato il test variato.

A questo punto, poichè si disponeva già dei risultati del CoreTest sul campione PE della Regione Emilia-Romagna e su quello nazionale, è stato possibile condurre una preliminare verifica dell'attendibilità dei risultati raccolti.

Innanzitutto si sono confrontati i valori ottenuti per l' α di Cronbach riportati nella tabella 4.1 Il risultato ottenuti per questo parametro indicano un livello di cor-

relazione interna adeguato (si ricorda che valori superiori a 0,70 sono considerati buoni per questo indice), ed inoltre esso si discosta di poco dal valore ottenuto nell'analisi del test del 2013.

Per una maggior sicurezza rispetto all'attendibilità e alla consistenza dei risultati

	α di Cronbach
CE	0,85
P	0,86

Tabella 4.1: Risultati dell' α di Cronbach

del metodo adottato, si sono poi confrontati i placement ottenuti dall'output Rasch e quelli derivanti dall'analisi del test originale.

Andando a confrontare i grafici a doppia campana (Figura 4.1), si riscontra una buona somiglianza tra l'andamento dei due studi. Infatti, considerando la zona relativa ai valori più alti di difficoltà (tra 0 e 4), si nota in tutti e tre i grafici la presenza degli item D10-a, D2-c, D11 (nel grafico compaiono con i numeri 19, 7, 22, come indica la legenda). Spostandoci a un livello intermedio (intorno allo 0) notiamo il ripetersi di numerosi Item, per esempio il D7-b, D26-a, D29 (numeri 14, 37, 40 del grafico); ed infine per i livelli di difficoltà più bassi (tra 0 e -4) si osserva il persistere dei quesiti D1-a, D1-b, D10-b (numeri 1, 2, 20).

Questa somiglianza nella distribuzione delle domande rispetto alla loro difficoltà permette di concludere che il campione sul quale è stata svolta la sperimentazione può essere considerato rappresentativo della popolazione degli studenti di sesto livello, al pari del campione nazionale selezionato dall'INVALSI, in quanto non si discosta significativamente da esso. Inoltre questo risultato fa sì che sia possibile e lecito un confronto tra questi due insiemi di dati.

Infine la provata rappresentatività del campione rispetto alla popolazione studentesca italiana permette di generalizzare le considerazioni fatte sulla popolazione campionaria.

Terminata questa parte di analisi e considerazioni, si è passati all'elaborazione dei dati riguardanti l'intero test TO, ossia si sono prese in esame le risposte date a tutti i 48 item nella loro versione originale. Poiché si era già in possesso dei risul-

tati relativi alle popolazioni CN e CE, l'analisi è stata svolta solo in riferimento alla popolazione P.

Osservando la Figura¹4.2 notiamo che le osservazioni precedenti, relativamente al posizionamento degli item all'interno dei tre grafici, acquistano di precisione. Questo risultato è considerato come una ulteriore e più precisa conferma della bontà dello strumento e del metodo adottato.

In seguito è stato analizzato, utilizzando lo stesso software, il dataset relativo al test variato sulla popolazione PV.

Oltre ai grafici del placement di cui si è parlato fino ad ora, Rasch ha restituito una stima dell'abilità di ciascuno studente e il parametro di difficoltà di ogni item. È quindi possibile ricondurre i risultati ottenuti su un'unica scala - utilizzando i 41 item comuni alle due versioni del test - e costruire così i distractor plot relativi a ciascuno dei 14 item, i 7 originali e i 7 variati, al fine di metterli a confronto.

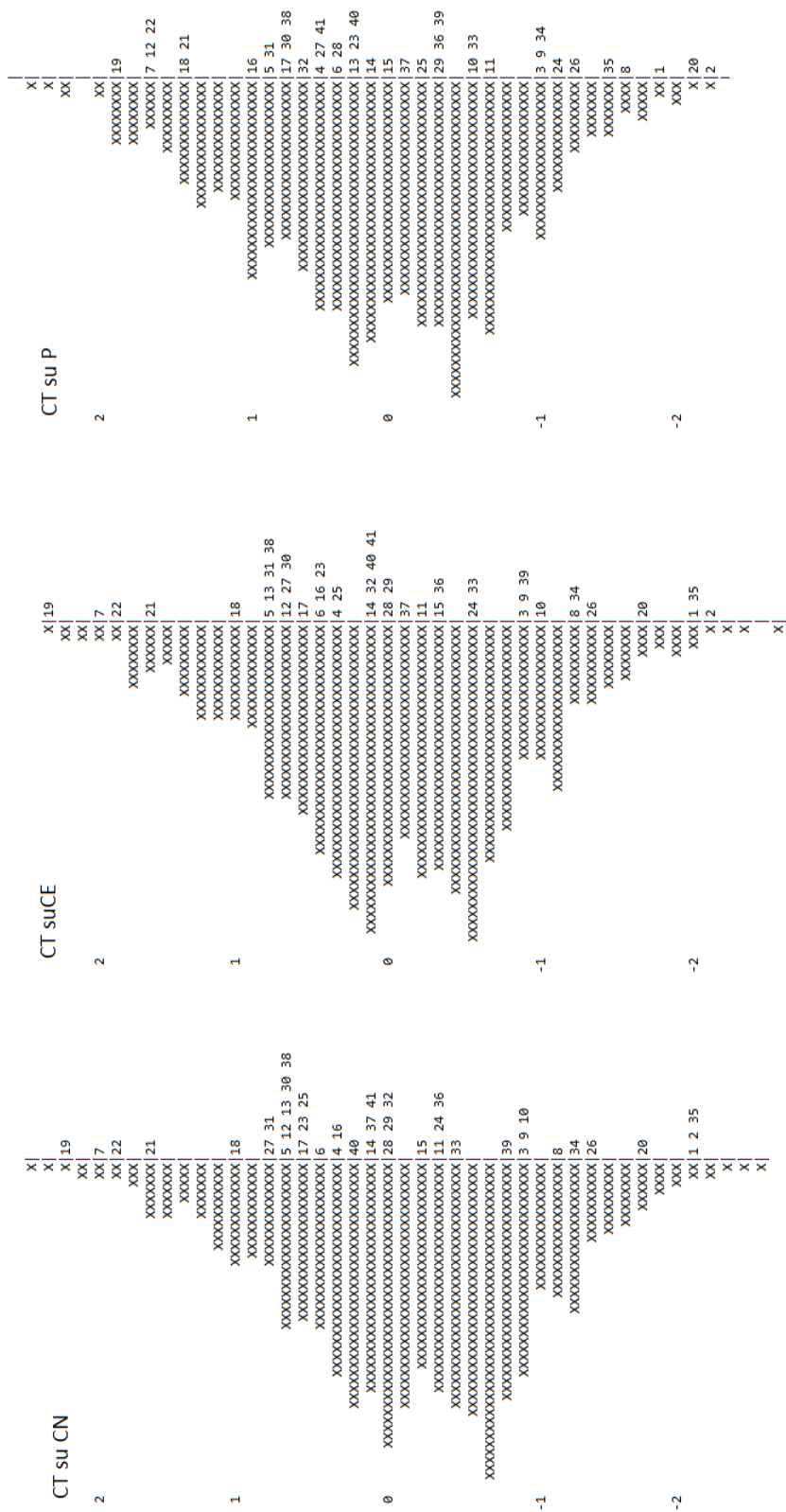
L'utilizzo di questo metodo permette di superare la problematica principale di uno studio di questo tipo. Come già detto, è infatti impossibile ricreare le condizioni ideali nelle quali studiare l'influenza delle variazioni testuali sulla risposta data, in quanto, sottoponendo allo stesso studente due test simili, egli, nel rispondere al secondo, sarebbe inevitabilmente influenzato dalla risposta data al primo.

Questo modo di condurre l'analisi equivale a somministrare un unico test composto da 55 item (41 del CT, 7 originali e 7 variati) all'intera popolazione P, a una parte della quale viene chiesto di rispondere ai 48 item costituenti il test originale, mentre, alla parte rimanente è chiesto di rispondere ai 48 item costituenti il test variato.

I risultati ottenuti da questa procedura, i parametri di difficoltà degli item e i distractor plot costruiti, hanno registrato l'effettiva presenza di un'influenza della performance dello studente al variare della formulazione della domanda. L'analisi delle variazioni sarà trattata nel prossimo capitolo.

Durante lo studio si è sperimentata un'ulteriore procedura con la quale si potrebbe ottenere lo stesso risultato. Si tratta di una particolare procedura di test

¹È importante notare che, nelle legende delle Figure 4.2, 4.1 agli stessi item non sono assegnati gli stessi numeri sul grafico.



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	

Figura 4.1: grafico a doppia campana di CT.



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
D1_a	D1_b	D1_c1	D1_c2	D2_a	D2_b	D2_c	D3	D4	D5	D6_a	D6_b	D6_c	D7_a	D7_b	D8_a	D8_b	D9_a	D9_b	D9_c	D10_a	D10_b	D10_c	D11
25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
D12	D13	D14	D15	D16	D17_a	D17_b	D18	D19	D20_a	D20_b	D21_a	D21_b	D22	D23	D24	D25_a	D25_b	D26_a	D26_b	D27	D28	D29	D30

Figura 4.2: grafico a doppia campana di TO.

Item	Difficoltà	Errore	Item	Difficoltà	Errore	Item	Difficoltà	Errore	Item	Difficoltà	Errore
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			

Figura 4.3: Output del metodo di ancoraggio statistico.

equating, chiamata ancoraggio, che, mediante la calibrazione simultanea, permette di stimare il parametro di difficoltà di ciascun item e l'abilità di ogni studente, considerando i risultati di entrambi i test contemporaneamente. La Figura 4.3 mostra i risultati del test equating, ed in particolare i parametri di difficoltà delle domande e il relativo errore statistico. Tali valori hanno mostrato nuovamente l'effettivo cambiamento dei risultati ottenuti a seguito delle modifiche apportate, proprio come la procedura precedente. Si può osservare, per esempio, che la variazione sull'item D3 lo ha reso sensibilmente più facile - il risultato era prevedibile in quanto si è passati da una risposta aperta a una chiusa-. Andando invece a complicare la sintassi della domanda si nota un aumento della difficoltà come per esempio nell'item D16.

Nell'ambito di questa analisi il confronto tra i valori del parametro di difficoltà dei 7 item originali con i rispettivi item variati, sono stati utilizzati solo come conferma dell'effettiva influenza sulla performance degli studenti, mentre le ulteriori considerazioni sono state fatte in relazione ai risultati ottenuti con le analisi precedentemente descritte.

Capitolo 5

Analisi dei risultati

In questo capitolo verranno presentati i risultati ed i distractor plot relativi ai 7 item modificati; in particolare sarà presentata un'analisi approfondita dell'item D22.

Analisi risultati item D22

La variazione apportata all'item D22 è una variazione di tipo numerico, in particolare si è operato sull'ordine di grandezza e sulla tipologia dei numeri utilizzati. Come mostrato in Figura 3.11, 3.12, la domanda chiede di stimare il risultato del prodotto di due numeri. Nel caso dell'item originale si tratta di due numeri decimali, mentre nell'item variato i numeri coinvolti sono interi e il loro ordine di grandezza è stato aumentato. Inoltre è importante notare che le alternative di risposta nell'item variato, corrispondono alle alternative dell'item originale, anch'esse mutate nell'ordine di grandezza.

Lo strumento di ricerca precedentemente descritto, consente di indagare quanto la variazione apportata abbia influenzato la performance di ciascuno studente, mediante il confronto su un'unica scala dei parametri stimati da Rasch. La ricerca può essere approfondita andando ad osservare i risultati ottenuti differenziandoli a seconda del genere, della nazionalità o della lingua madre; In questo modo si è potuto analizzare quali siano state le categorie di studenti maggiormente influenzate dalle variazioni redazionali e, in un secondo momento, di ipotizzarne il motivo.

Dal momento che l'item è stato formulato come una domanda a scelta multipla, l'analisi a priori sulle possibili strategie adottate dagli studenti nel rispondere alla domanda, può riguardare solo la scelta degli studenti in una rosa di quattro possibilità. Per quanto riguarda l'item originale, osservando le possibili risposte, si possono ipotizzare le seguenti strategie:

- Considerare solo la parte intera del numero (distrattore A);
- Approssimare entrambi i numeri per eccesso (distrattore D);
- Approssimare per eccesso un termine e per difetto l'altro (distrattori B e C).

Per quanto riguarda invece la versione modificata dell'item, sempre da un'analisi a priori, si potrebbe pensare che uno studente posto di fronte al prodotto tra due numeri interi, seppur grandi, si senta "più a suo agio" ed autorizzato ad adottare come strategia la risoluzione in colonna del prodotto. In questo caso è più raro che lo studente tenti l'utilizzo di approssimazione, mentre, nel momento in cui i numeri in gioco sono decimali, è plausibile che la dinamica venga capovolta.

item:D22 Originale

Cases for this item 380 Discrimination 0.15
 Item Threshold(s): 0.16 Weighted MNSQ 1.18
 Item Delta(s): 0.16

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	58	15.26	-0.05	-0.90(.370)	-0.09	0.75
B	0.00	81	21.32	0.07	1.41(.158)	0.04	0.90
C	1.00	175	46.05	0.15	2.96(.003)	0.10	0.81
D	0.00	42	11.05	-0.09	-1.76(.078)	-0.20	0.81
M	0.00	24	6.32	-0.25	-4.93(.000)	-0.76	0.72

Passiamo ora all'analisi dei risultati. Osservando la tabella sopra riportata, si può notare che l'item così riformulato risulta significativamente più semplice. La percentuale di risposte corrette¹, infatti, passa dal 46% al 59%.

¹L'alternativa corretta è contrassegnata da un 1 nella colonna dello *Score*. in questo caso la C

item:D22 Variato

Cases for this item 397 Discrimination 0.22
 Item Threshold(s): -0.43 Weighted MNSQ 1.15
 Item Delta(s): -0.43

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	44	11.08	-0.11	-2.28(.023)	-0.29	0.69
B	0.00	51	12.85	-0.02	-0.40(.686)	-0.04	0.87
C	1.00	234	58.94	0.22	4.43(.000)	0.14	0.87
D	0.00	46	11.59	-0.06	-1.19(.233)	-0.16	0.72
M	0.00	22	5.54	-0.20	-4.01(.000)	-0.64	0.67

Nella stessa direzione si modifica il parametro di difficoltà dell'item - indicato come *Item Delta(s)* - che passa da 0,16 a -0,43. Anche nel caso in cui il parametro venga calcolato con il metodo di ancoraggio statistico il risultato che si ottiene è il medesimo, ossia una diminuzione della difficoltà del quesito, che passa da 0,10 a -0,51, come mostra la Figura 4.3. Di questi ultimi valori sappiamo anche che si presentano con un errore standard di 0,11, e ciò rassicura rispetto alla significatività statistica del risultato.

Al fine di approfondire le osservazioni fatte fino ad ora, può essere interessante analizzare i distractor plot costruiti in funzione della capacità degli studenti, valutati sul CT con il metodo di Rasch. Un'analisi di questo tipo è utile per indagare il modo in cui sono distribuite le differenze di performance ossia identificare sugli studenti di quale livello hanno agito le variazioni apportate.

La prima evidente differenza tra le due curve che indicano la risposta corretta (in verde), è che quella relativa all'item variato, è traslata verso l'alto indicando un innalzamento della probabilità di rispondere correttamente a sottolineare la diminuzione della difficoltà del quesito. La curva non risulta semplicemente traslata, l'andamento, infatti, è piuttosto differente tra una e l'altra, in particolare si osserva un aumento della regolarità nella versione modificata rispetto a quella originale. Tornando ai risultati sopra riportati si nota che se i distrattori A e D diminuiscono o aumentano solo di alcuni punti percentuali, la risposta B (in blu) perde circa

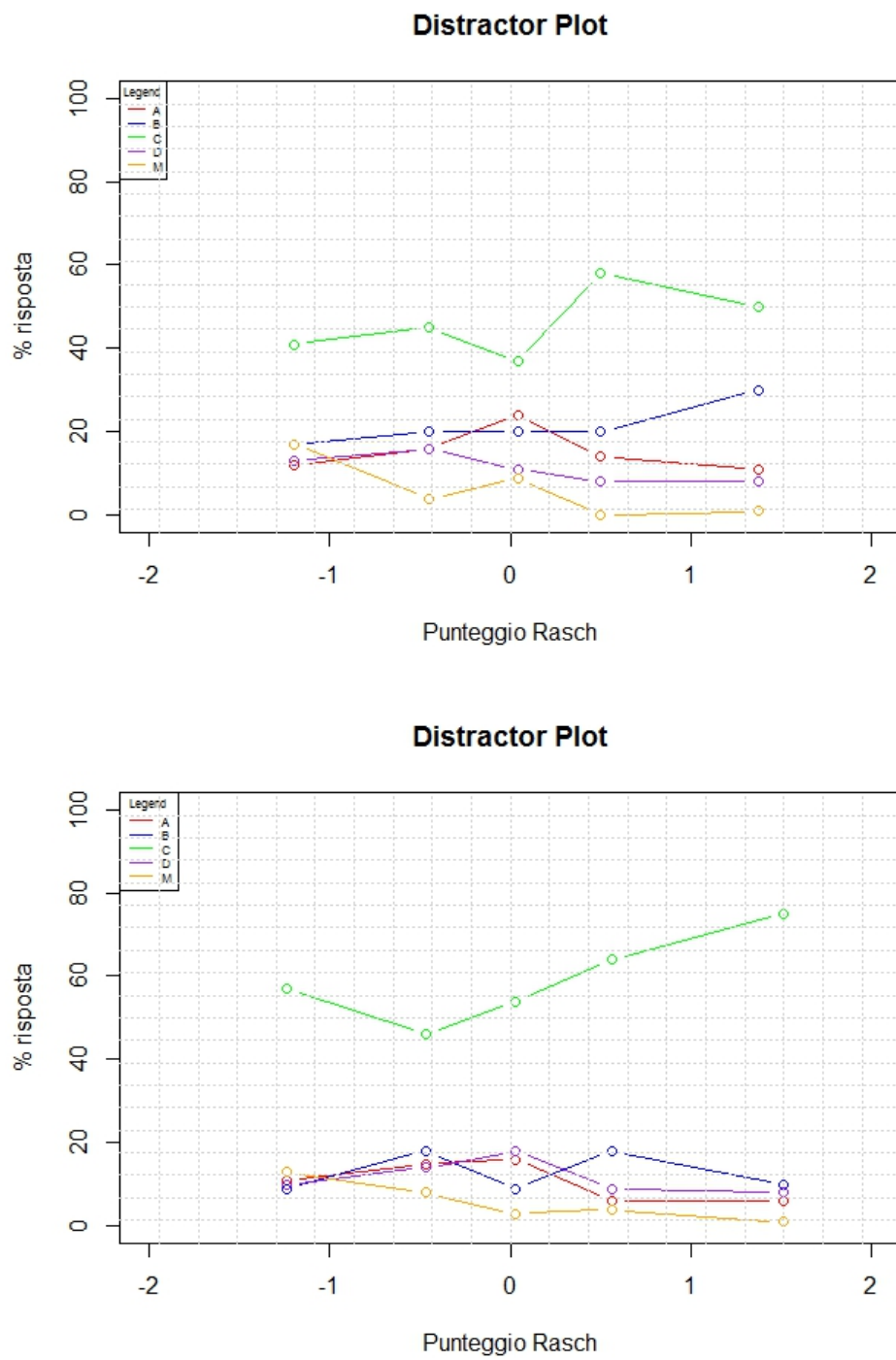


Figura 5.1: Distractor plot item D22 originale e variato

l'8% dopo la modifica dell'item. Osservando il distractor plot dell'item originale si nota che questo distrattore era particolarmente appetibile per gli studenti con un alto livello d'abilità; con l'introduzione della variazione esso, invece, perde d'interesse. Anche la percentuale di studenti che non rispondono all'item non si è significativamente modificata. Ciò può significare che, nonostante la diversa difficoltà delle due versioni, quasi tutti gli studenti si ritengono abbastanza abili per poter tentare una risposta.

Da entrambi i grafici emerge che il quesito non ha un alto livello di discriminazione, ossia l'item non segna una netta distinzione tra studenti con alti livelli di abilità e studenti con livelli più bassi. Questa osservazione è confermata dai risultati dell'analisi condotta sull'intero test: il parametro di discriminazione dell'item D22 è di 0,15 per quanto riguarda la prova originale e 0,22 nella forma variata. Il fatto che nella sua forma variata il parametro discriminante dell'item risulta più alto, fa emergere un ulteriore ed interessante risultato dello studio delle variazioni redazionali: il miglioramento delle proprietà statistiche dell'item.

La variazione ha un impatto molto diverso a seconda del sesso. Nella tabella sottostante sono riportate le percentuali di risposta dei maschi e delle femmine per ciascuna delle versioni dell'item. Nel quesito originale 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, infatti, il 18% in più degli studenti sceglie la risposta corretta. Per quanto riguarda le femmine si osserva che solo il 6% in più risponde correttamente. Questo può significare che maschi e femmine applicano diverse strategie per risolvere l'item e che la strategia utilizzata dalle studentesse rimane probabilmente invariata al variare della tipologia del numero.

Label	% di M Orig.	% di M Var.	% di F Orig.	% di F Var.
A	16	10	12	11
B	23	12	20	13
C	44	62	50	56
D	11	11	10	13
M	5	4	8	7

Infatti essendo meno abituato a calcoli con numeri decimali il ragazzo può sentirsi più insicuro nell'intraprendere questa strada e quindi optare per una delle strategie di approssimazione, forse quella di cui trova un riscontro più immediato tra le possibili risposte. Perciò, benchè l'ordine di grandezza del numero non modifichi la natura del problema, ci si può aspettare che gli studenti abbiano la tendenza ad applicare strategie differenti a seconda della tipologia del numero considerato.

Analisi risultati item D3

item:D3 Originale

Cases for this item 380 Discrimination 0.33
 Item Threshold(s): 0.09 Weighted MNSQ 1.04
 Item Delta(s): 0.08

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	1.00	181	47.63	0.33	6.83(.000)	0.26	0.79
B	0.00	166	43.68	-0.18	-3.53(.000)	-0.21	0.76
M	0.00	32	8.42	-0.26	-5.15(.000)	-0.72	0.81

item:D3 Variato

Cases for this item 397 Discrimination 0.35
 Item Threshold(s): -0.50 Weighted MNSQ 1.01
 Item Delta(s): -0.50

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	1.00	240	60.45	0.35	7.53(.000)	0.23	0.81
B	0.00	82	20.65	-0.18	-3.72(.000)	-0.29	0.68
C	0.00	52	13.10	-0.11	-2.27(.024)	-0.36	0.87
D	0.00	14	3.53	-0.20	-3.98(.000)	-0.83	0.90
M	0.00	9	2.27	-0.16	-3.27(.001)	-0.72	0.50

La variazione apportata all'item D3 è una variazione della tipologia della domanda che passa da una risposta aperta a una chiusa, in cui lo studente può scegliere tra 4 alternative di risposta delle quali una sola è esatta. L'item (Figura 3.1 3.2) chiedeva di contare il numero delle cifre del risultato ottenuto moltiplicando due numeri.

I risultati riportati nella tabella mostrano una diminuzione della difficoltà dell'item; il livello di difficoltà passa infatti da uno 0,08 nella versione originale a un -0,50 nella versione modificata. Per quanto riguarda i risultati ottenuti con il metodo di ancoraggio (tabella 4.3) anch'essi mostrano una diminuzione di tale parametro,

entrambi con un'accettabile errore statistico di 0,11. Vediamo inoltre che anche le percentuali di risposta mutano significativamente: dal 47,63% si supera la metà con un 60,45%.

Dal distractor plot dell'item variato emerge che il distrattore più frequentemente selezionato, con una percentuale di risposta del 20,65%, è l'alternativa B: "9 cifre" che corrispondono alla somma delle cifre che compongono i due fattori da moltiplicare e non al numero di cifre del risultato del prodotto. Tale distrattore comunque non sembra essere particolarmente insidioso per gli studenti, ed agisce in maniera pressochè incondizionata rispetto ai livelli di abilità del rispondente.

Andando a indagare tra le risposte date dagli studenti all'item originale - nel quale non comparivano alternative di risposta -, si nota che la maggior parte delle risposte errate riportano il risultato dell'operazione (giusto o errato che sia) e non il numero corrispondente alle sue cifre. Tale comportamento potrebbe essere una delle conseguenze del contratto didattico, per cui uno studente posto di fronte a un'operazione pensa prevalentemente a calcolarne il risultato, incurante della consegna dell'esercizio. Questa osservazione è molto interessante: il crollo della difficoltà dell'item potrebbe essere spiegato dal fatto che tra le alternative non era contemplata la risposta che la maggior parte dei ragazzi avrebbe voluto fornire, inducendoli quindi a riprendere in mano il testo della domanda e identificare l'errore commesso.

Per quanto riguarda il grado di discriminazione dell'item esso rimane pressochè invariato tra le due versioni.

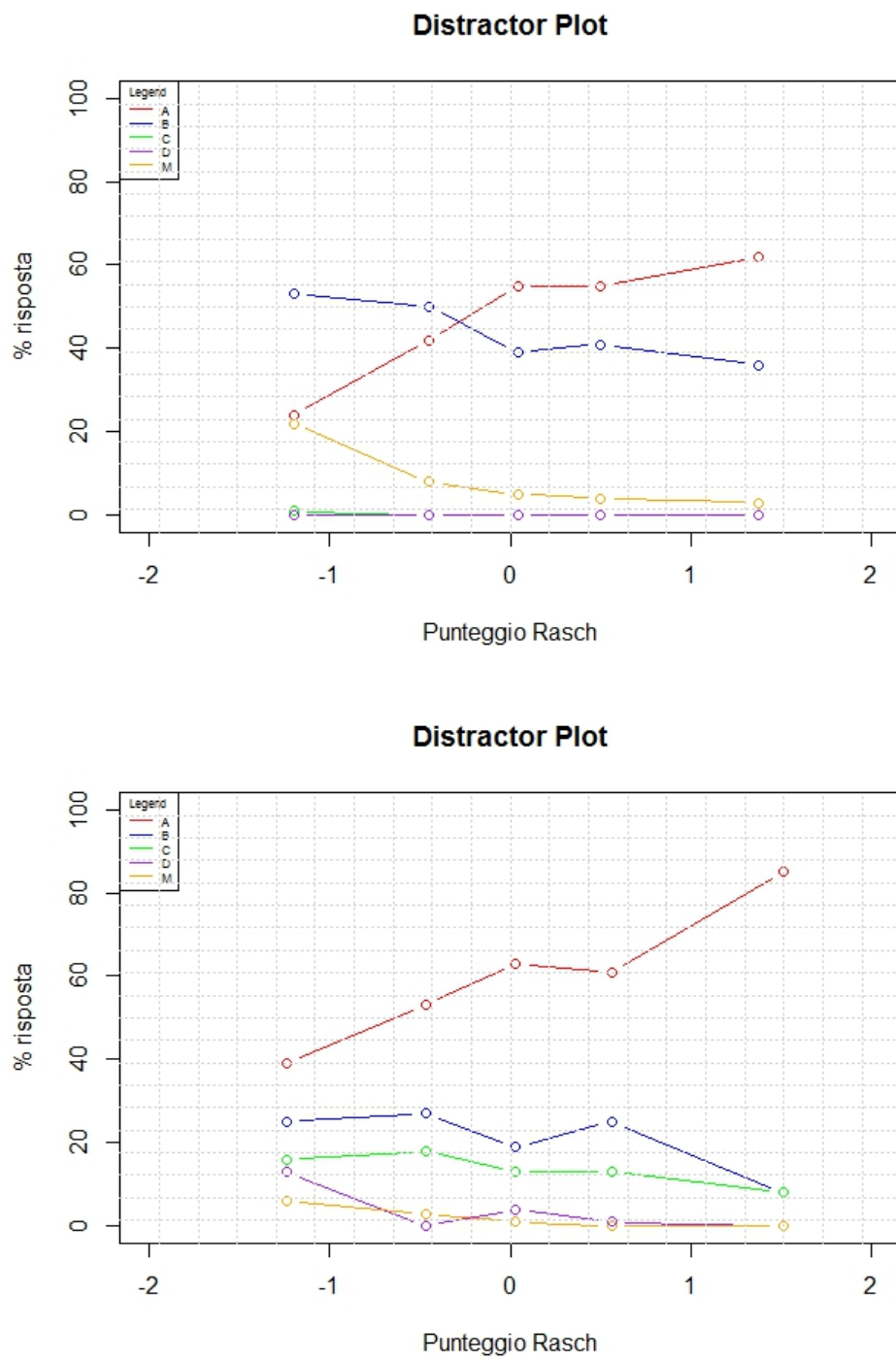


Figura 5.2: Distractor plot item D3 originale e variato

Analisi risultati item D8.b

item:D8_b Originale

Cases for this item 380 Discrimination 0.30
 Item Threshold(s): 0.52 Weighted MNSQ 1.04
 Item Delta(s): 0.52

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	43	11.32	-0.12	-2.35(.019)	-0.29	0.84
B	0.00	14	3.68	-0.23	-4.53(.000)	-1.06	0.60
C	0.00	166	43.68	-0.10	-1.89(.060)	-0.09	0.76
D	1.00	146	38.42	0.30	6.18(.000)	0.25	0.83
M	0.00	11	2.89	-0.11	-2.16(.031)	-0.58	0.73

item:D8_b Variato

Cases for this item 397 Discrimination 0.39
 Item Threshold(s): 0.82 Weighted MNSQ 1.01
 Item Delta(s): 0.82

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	52	13.10	-0.19	-3.75(.000)	-0.40	0.66
B	0.00	92	23.17	-0.29	-6.11(.000)	-0.42	0.76
C	0.00	106	26.70	0.09	1.71(.087)	0.07	0.86
D	1.00	131	33.00	0.39	8.51(.000)	0.44	0.74
M	0.00	16	4.03	-0.19	-3.77(.000)	-0.68	0.49

Il punto b dell'item D18 originale chiedeva di identificare le coordinate di un punto P raggiunto a partire da un punto Q dato, seguendo una serie di indicazioni riportate nel testo dell'esercizio. Come si può leggere nella Figura 3.3 le indicazioni venivano fornite nell'ordine in cui dovevano essere compiute. Nonostante possa apparire un esercizio piuttosto semplice, il parametro di difficoltà è di 0,52 secondo il metodo di Rasch e di 0,47 secondo l'ancoraggio (tabella 4.3). La percentuale di risposte esatte è solo del 38,42%, e anche il distractor plot mostra che

solo intorno a un'abilità dello 0,5 la risposta corretta è preferita rispetto alle altre. Dall'analisi dei risultati e del grafico appare chiara quella che dev'essere la difficoltà principale per gli studenti, ossia l'ordine con cui devono essere riportate le coordinate nella scrittura analitica. Infatti è il distrattore C, che presentava gli stessi valori delle coordinate dell'alternativa esatta ma in posizioni invertite - (3,1) anziché (1,3) -, a raggiungere il 43,68% di risposte, mentre le altre alternative rimangono a percentuali inferiori al 12%. Anche osservando il distractor plot appare evidente l'importanza del distrattore C, che viene scelto in maniera pressoché uniforme da tutti i livelli di abilità con una probabilità superiore al 35%. Tale difficoltà era riscontrabile anche osservando le risposte date al punto a dello stesso item. In questo caso era chiesto semplicemente di scrivere le coordinate del punto Q posizionato nel grafico. Pochissimi sono stati gli alunni che hanno riportato la scrittura completa e corretta: (5,6); più spesso si trovavano le cifre invertite se non addirittura elencate senza alcun tipo di parentesi, ponendo sempre prima la coordinata y seguita dalla coordinata x.

Era quindi facile immaginare che, complicando l'enunciato dell'esercizio, la situazione non poteva che peggiorare, poichè la principale problematica riscontrata dagli studenti nella versione originale non era tanto la comprensione del testo, quanto l'ignoranza rispetto al linguaggio specifico e convenzionale sull'argomento. La percentuale di risposte esatte si è abbassata al 33%, mentre il parametro di difficoltà, per entrambi i metodi di stima, è intorno a 0,80.

L'aspetto più interessante riguarda ancora l'analisi dei distrattori: la percentuale del distrattore C passa al 26,7%, significativamente inferiore alla precedente, mentre il distrattore B, che nella versione originale raggiungeva il 3,68%, passa al 23,17%. Il distrattore B riportava le coordinate (4,3), i cui valori non sono altro che l'entità degli spostamenti da compiere verso destra e verso l'alto (Figura 3.4). Passando al distractor plot notiamo che l'andamento di questo distrattore è decrescente, ossia la probabilità di scelta diminuisce all'aumentare dell'abilità dello studente. Si può infatti supporre che questa alternativa non fosse quasi considerata nella versione originale dell'item, in quanto gli spostamenti da compiere erano elencati in modo chiaro e lineare. La modifica apportata ha complicato l'item nella

direzione della strategia risolutiva, facendo perdere quasi completamente l'orientamento ai meno abili, che a questo punto non hanno potuto invertire le coordinate del punto d'arrivo, in quanto non erano riusciti nemmeno ad individuarlo. È d'altra parte interessante osservare che la variazione non ha avuto quasi alcun effetto sugli studenti con alti livelli di abilità; essi quindi sono stati in grado di superare la difficoltà del testo ed allo stesso tempo, in accordo con quanto osservato sull'item originale, sono rimasti vittima di un errore di tutt'altra natura riguardante l'ordine delle coordinate.

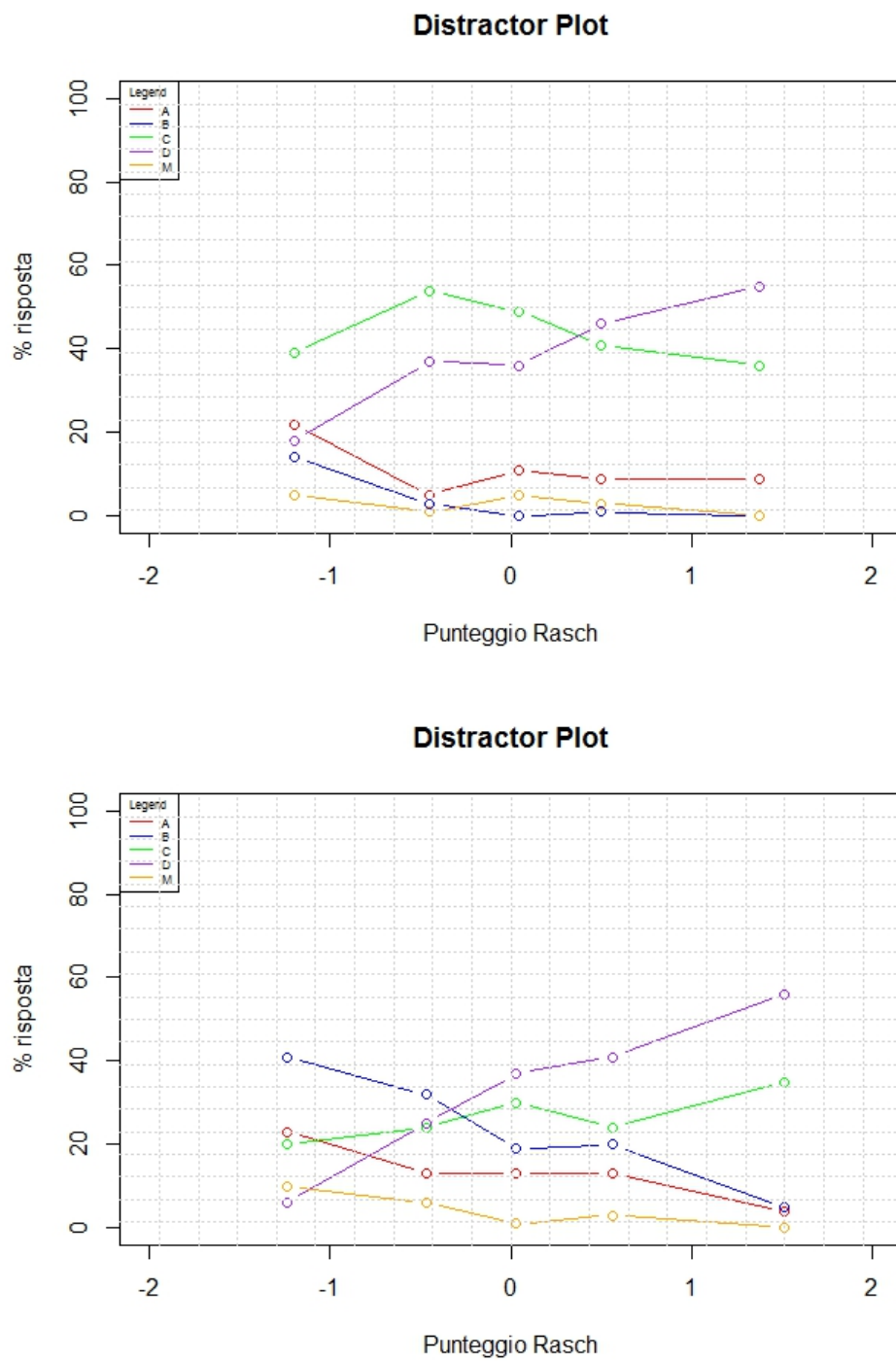


Figura 5.3: Distractor plot item D8.b originale e variato

Analisi risultati item D14

item:D14 Originale

Cases for this item 380 Discrimination 0.35
 Item Threshold(s): 0.57 Weighted MNSQ 1.02
 Item Delta(s): 0.57

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	1.00	142	37.37	0.35	7.23(.000)	0.32	0.86
B	0.00	18	4.74	-0.18	-3.56(.000)	-0.69	0.70
C	0.00	28	7.37	-0.19	-3.74(.000)	-0.51	0.61
D	0.00	182	47.89	-0.12	-2.27(.024)	-0.12	0.72
M	0.00	10	2.63	-0.15	-2.86(.005)	-0.82	1.08

item:D14 Variato

Cases for this item 397 Discrimination 0.35
 Item Threshold(s): 0.91 Weighted MNSQ 1.04
 Item Delta(s): 0.91

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	1.00	124	31.23	0.35	7.54(.000)	0.42	0.84
B	0.00	23	5.79	-0.22	-4.54(.000)	-0.71	0.58
C	0.00	41	10.33	-0.14	-2.91(.004)	-0.30	0.78
D	0.00	204	51.39	-0.12	-2.46(.014)	-0.12	0.78
M	0.00	5	1.26	-0.06	-1.22(.222)	-0.48	0.88

I risultati ottenuti dalla variazione dell'item D14 sono particolari. La modifica apportata è una semplice variazione redazionale, infatti, come già detto nella sezione 3.2, è stato cambiato di posto all'immagine illustrativa che invece di essere posta tra i dati e la domanda viene interposta alla presentazione dei primi.

Il risultato ha mostrato un aumento del parametro di difficoltà che passa da 0,57 a 0,91 secondo la stima con il metodo di Rasch, e da 0,52 a 0,88 secondo l'ancoraggio (tabella 4.3). Il dato rispecchia le percentuali di risposte corrette che nel caso del

test originale sono il 37,37% mentre per l'item variato calano al 31,23%.

Per quanto riguarda i distrattori, hanno tutti percentuali di risposta basse ad eccezione dell'alternativa D, la cui percentuale è più alta persino di quella della risposta corretta, e che passa da un 47,89% a un 51,39% nell'item modificato. Quest'alternativa riportava il risultato che si otteneva sommando ad entrambe le dimensioni della foto una sola volta lo spessore della cornice. Dai distractor plot notiamo che la scelta di questo distrattore è stata fatta in buona percentuale da tutti i livelli di abilità in entrambe le formulazioni del quesito, ma in particolar modo dagli studenti con un livello intorno allo zero.

L'andamento delle due curve rimane praticamente invariato, si nota solo un abbassamento della curva relativa alla risposta corretta, la A, che sta a significare appunto l'aumento della difficoltà dell'item.

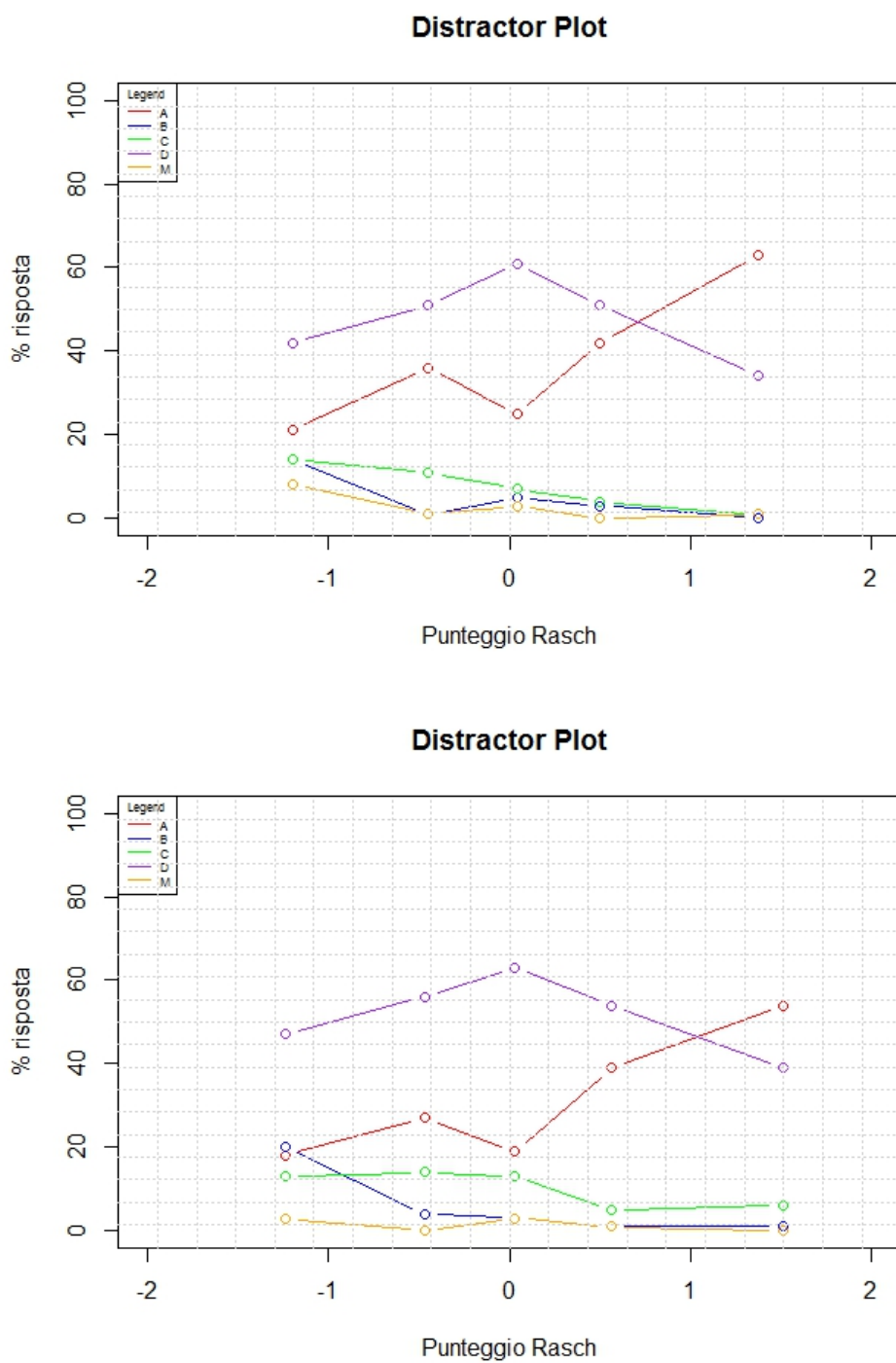


Figura 5.4: Distractor plot item D14 originale e variato

Analisi risultati item D16

item:D16 Originale

Cases for this item 380 Discrimination 0.31
 Item Threshold(s): 0.37 Weighted MNSQ 1.05
 Item Delta(s): 0.37

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	49	12.89	-0.15	-2.96(.003)	-0.38	0.67
B	1.00	158	41.58	0.31	6.34(.000)	0.26	0.89
C	0.00	108	28.42	-0.06	-1.17(.244)	-0.08	0.71
D	0.00	54	14.21	-0.18	-3.50(.001)	-0.41	0.81
M	0.00	11	2.89	-0.08	-1.58(.115)	-0.18	0.84

item:D16 Variato

Cases for this item 397 Discrimination 0.25
 Item Threshold(s): 0.82 Weighted MNSQ 1.12
 Item Delta(s): 0.82

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	43	10.83	-0.18	-3.62(.000)	-0.41	0.63
B	1.00	131	33.00	0.25	5.09(.000)	0.23	0.84
C	0.00	132	33.25	0.10	1.96(.051)	0.12	0.81
D	0.00	62	15.62	-0.22	-4.43(.000)	-0.40	0.90
M	0.00	29	7.30	-0.11	-2.17(.031)	-0.28	0.68

Nell'item D16 la variazione ha riguardato l'aspetto linguistico, in particolare è stata complicata la struttura dei periodi.

La difficoltà dell'item originale è pari a 0.37 e la percentuale di risposte corrette si aggira intorno al 41%. Dal distractor plot emerge che i distrattori hanno particolare effetto sugli studenti con un basso livello di abilità, ma che, con l'aumentare delle capacità, la probabilità di rispondere correttamente all'item si fa sempre più alta fino a un 65%.

L'andamento cala drasticamente una volta che viene inserita la variazione: il parametro di difficoltà arriva a 0,80 e le percentuali di risposta corretta cala al 33%. Ad aumentare è la percentuale degli studenti che hanno scelto il distrattore C, la cui percentuale raggiunge il 33,25%. Il distractor plot mostra infatti quasi un intreccio tra le due curve rappresentative del distrattore B e C. L'aspetto più particolare è che la variazione va ad influire maggiormente sugli studenti più abili, la cui probabilità di scegliere la risposta esatta cala di più del 20%, mentre per gli altri la riduzione non supera il 10%.

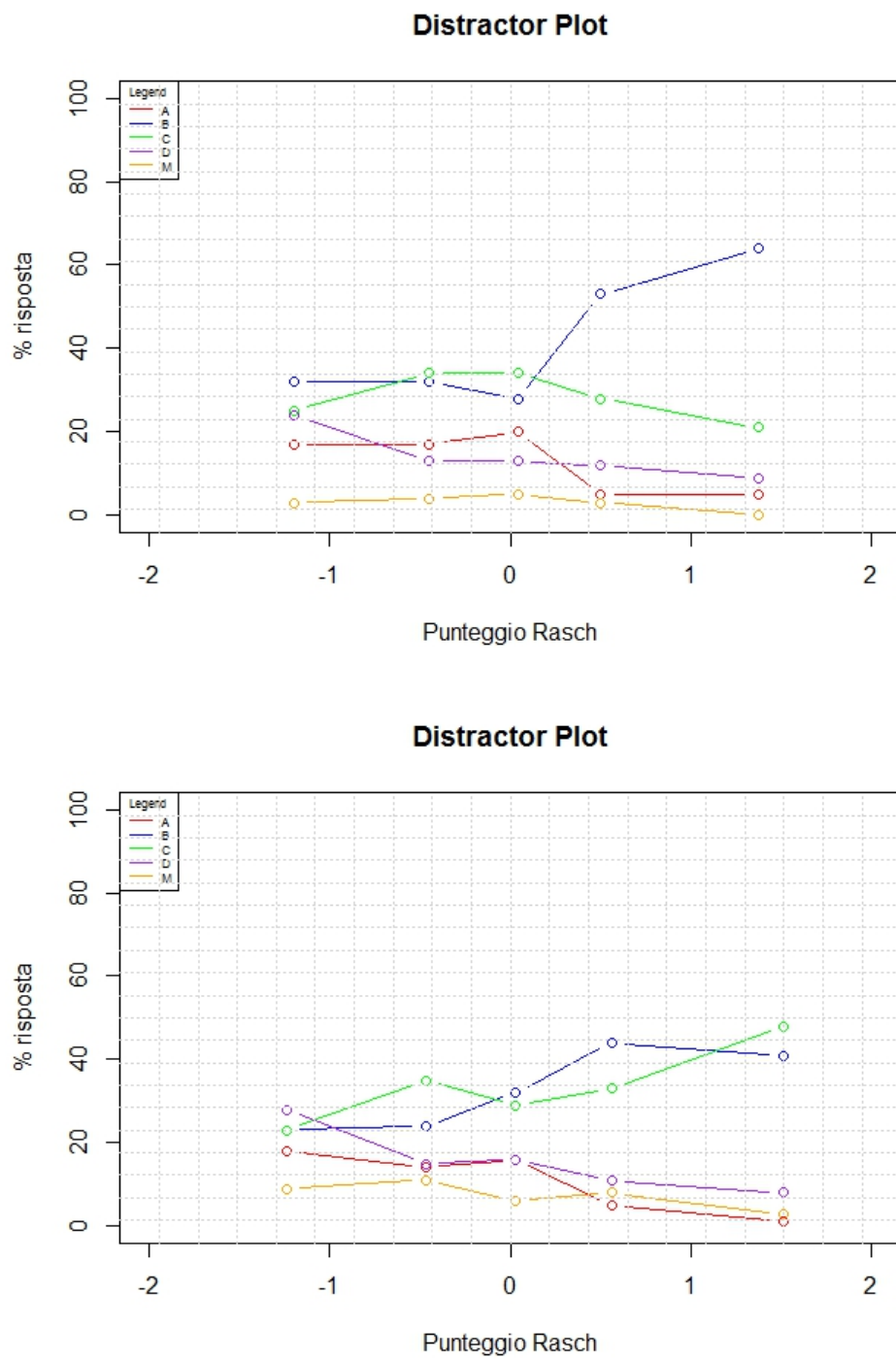


Figura 5.5: Distractor plot item D16 originale e variato

Analisi risultati item D18

item:D18 Originale

Cases for this item 380 Discrimination 0.56
 Item Threshold(s): 0.73 Weighted MNSQ 0.87
 Item Delta(s): 0.73

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	1.00	130	34.21	0.56	13.12(.000)	0.60	0.71
B	0.00	201	52.89	-0.28	-5.59(.000)	-0.25	0.67
M	0.00	49	12.89	-0.38	-7.98(.000)	-0.80	0.69

item:D18 Variato

Cases for this item 397 Discrimination 0.60
 Item Threshold(s): 1.13 Weighted MNSQ 0.80
 Item Delta(s): 1.12

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	1.00	109	27.46	0.60	15.05(.000)	0.75	0.69
B	0.00	169	42.57	-0.09	-1.89(.059)	-0.11	0.68
M	0.00	119	29.97	-0.49	-11.04(.000)	-0.57	0.68

L'item D18 chiedeva di calcolare il perimetro di un rettangolo (Figura 3.9, 3.10) e la variazione apportata è stata quella di rimuovere la rappresentazione geometrica.

Il quesito è risultato piuttosto difficile già nella sua versione originale. Il parametro di difficoltà dell'item è di 0,73 e le percentuali di risposta corretta sono il 34,21% contro la percentuale delle risposte errate di 52,89%. Con l'eliminazione dell'immagine si verifica un calo delle risposte corrette ed allo stesso tempo di quelle sbagliate, in quanto aumenta significativamente la percentuale delle risposte mancate che raggiunge il 29,97%. Di conseguenza anche il parametro di difficoltà aumenta fino a 1,12, il che significa che per studenti di livello inferiore a tale valore è più alta la probabilità di rispondere in modo errato (o decidere di non risponde-

re).

I risultati dei distractor plot sono molto chiari, infatti il grado di discriminazione di questo item è molto buono, tra lo 0,56 e lo 0,6. La curva relativa alla risposta corretta nella versione originale cresce molto rapidamente all'aumentare dell'abilità. Per quanto riguarda la curva della risposta errata, essa ha il suo massimo per il livello di abilità -1 e poi decresce rapidamente. Gli studenti meno abili, invece, tendono a omettere la risposta, fenomeno sempre meno probabile con l'aumento delle capacità.

Altrettanto chiaro è il distractor plot dell'item variato: la curva relativa alla risposta corretta cresce ancora rapidamente e quella delle risposte omesse cala in maniera pressoché speculare alla precedente. L'intersezione di queste due curve si ha all'altezza del livello di abilità 0 per il quale le due alternative si presentano con una probabilità intorno al 20%. In corrispondenza di questo livello di abilità si ha invece il picco di risposte errate che raggiunge il 60%.

Semplificando la situazione si potrebbe dire che: gli studenti più bravi non hanno avuto difficoltà a risolvere il problema nemmeno dopo l'introduzione della variazione, forse perché in grado di ricostruirsi da sé la figura omessa; gli studenti con abilità intermedie tentano una risposta al quesito che nella maggior parte dei casi risulta errata; infine, gli studenti con un livello più basso di competenze scelgono di non rispondere affatto alla domanda posta. Si può quindi concludere che in questo caso la variazione ha influito sulle performance degli studenti, andando ad accentuare gli atteggiamenti già riscontrati nel distractor plot dell'item originale.

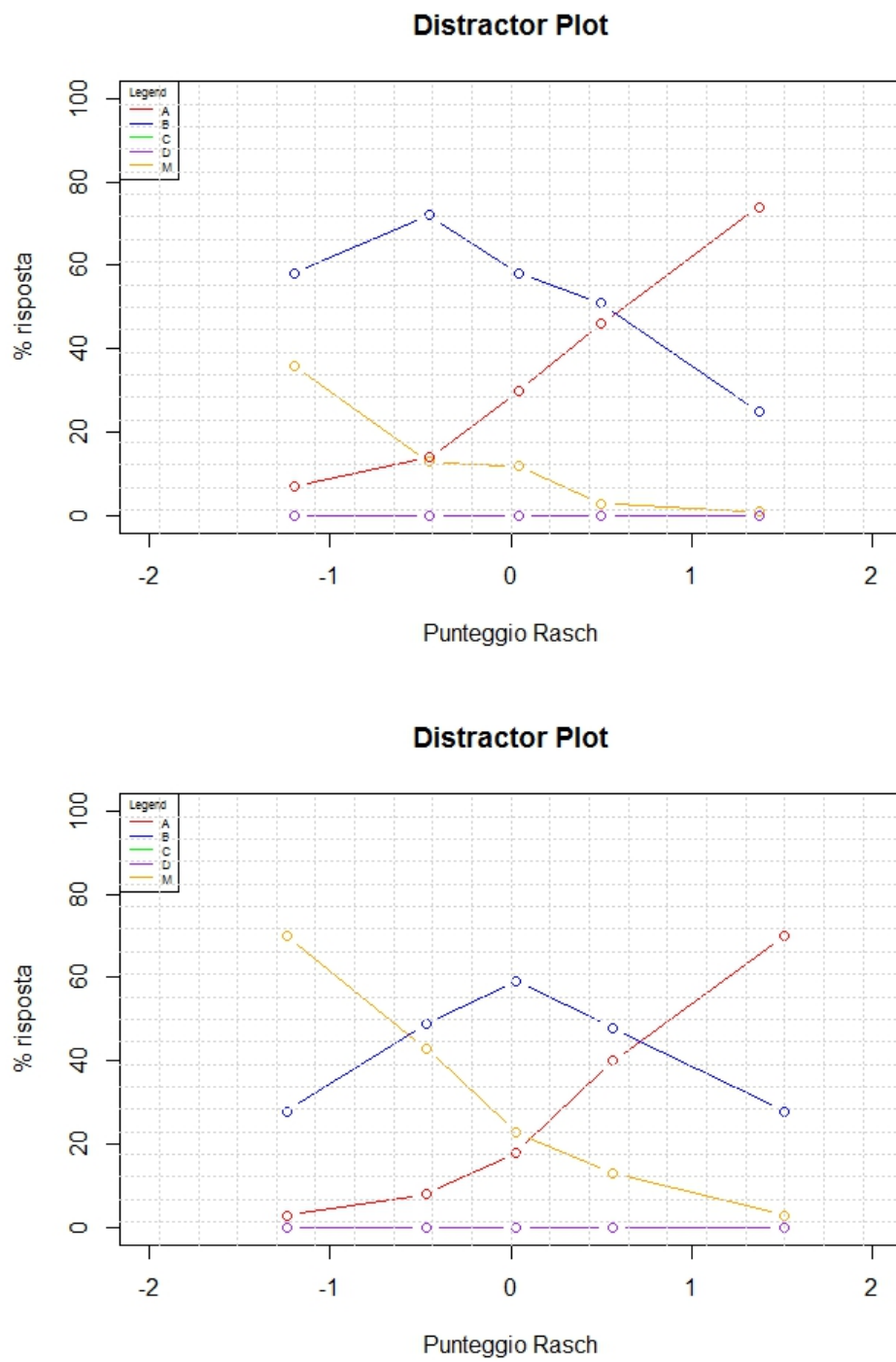


Figura 5.6: Distractor plot item D18 originale e variato

Analisi risultati item D27

item:D27 Originale

Cases for this item 380 Discrimination 0.40
 Item Threshold(s): 0.59 Weighted MNSQ 0.95
 Item Delta(s): 0.59

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	87	22.89	-0.18	-3.61(.000)	-0.29	0.70
B	1.00	140	36.84	0.40	8.56(.000)	0.41	0.83
C	0.00	48	12.63	-0.10	-1.93(.055)	-0.21	0.66
D	0.00	49	12.89	-0.01	-0.14(.890)	-0.11	0.78
M	0.00	56	14.74	-0.23	-4.66(.000)	-0.49	0.79

item:D27 Variato

Cases for this item 397 Discrimination 0.28
 Item Threshold(s): 0.44 Weighted MNSQ 1.10
 Item Delta(s): 0.44

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
A	0.00	45	11.34	-0.13	-2.56(.011)	-0.24	0.54
B	1.00	161	40.55	0.28	5.89(.000)	0.27	0.88
C	0.00	56	14.11	-0.11	-2.20(.028)	-0.24	0.71
D	0.00	73	18.39	0.09	1.78(.076)	0.10	0.84
M	0.00	62	15.62	-0.26	-5.40(.000)	-0.50	0.79

La variazione apportata all'item D27 è di natura linguistica. Anche in questo caso si è agito nell'intento di complicare la sintassi del periodo, unendo periodi separati e utilizzando forme gerundive, tipiche del linguaggio matematico (Figura 3.13, 3.14).

il risultato ottenuto è stato però inaspettato: l'item variato è risultato più semplice dell'originale. Si passa infatti da un parametro di difficoltà di 0,59 a uno di 0,44. Andando a osservare le percentuali di risposte si ha nella prima formulazione un

36,84% che diventa un 40,55% nella versione modificata.

Escludendo l'alternativa corretta, l'item originale presenta percentuali di risposta distribuite piuttosto equamente tra ciascuno dei distrattori, con eccezione dell'alternativa A che spicca sulle altre con una percentuale del 22,89%. Osservando il distractor plot di questo item si deduce che quest'ultimo distrattore agisce soprattutto sugli studenti con un livello d'abilità basso, mentre le alternative C e D sono frequenti quasi allo stesso modo per tutti i livelli di abilità. Infine la risposta corretta cresce abbastanza rapidamente con l'aumentare delle competenze. La mancanza di risposta è frequente soprattutto per livelli d'abilità inferiore a -1. Passando all'analisi dei distrattori dell'item variato, si nota che il distrattore più frequente nella precedente analisi perde in percentuale, divenendo l'alternativa meno scelta di tutte. Le percentuali di A, C, D, M si aggirano tutte tra il 10% e il 20%; la più gettonata è la D con un 18,39%. Quest'ultimo distrattore indica come risposta il valore di 520 g, che rappresenta il peso complessivo dei due quaderni, ed è quindi il risultato che si ottiene dalla sottrazione tra il peso dello zaino e la somma del peso dei libri di scienze e matematica, prima di effettuare la divisione per 2. Effettivamente nella versione variata del testo, l'informazione della presenza di due quaderni identici rimane piuttosto celata. Il dato viene esplicitato nel primo periodo che però, nel contesto dell'intero testo, risulta come un'introduzione di minima rilevanza. L'informazione viene ripetuta con l'uso del "ciascun quaderno" all'interno della domanda, che viene però posta all'inizio del secondo periodo ed è seguita dalla parte più sostanziosa del testo, parte in cui vengono presentati i dati numerici. Lo studente probabilmente tende a focalizzare la sua attenzione soprattutto su questa tipologia di dati, tralasciando o dimenticando le altre informazioni; in questo modo dimentica di effettuare l'operazione di divisione.

Andando ad analizzare i distractor plot è subito evidente che l'item deve risultare di bassa difficoltà, infatti l'alternativa corretta, per tutti i livelli di abilità, risulta la più probabile. Per quanto riguarda la scelta del distrattore D, l'errore è stato commesso più frequentemente da studenti con un livello di abilità medio e alto, mentre risulta quasi nulla la percentuale per i livelli più bassi per i quali si ripresenta la tendenza a non rispondere.

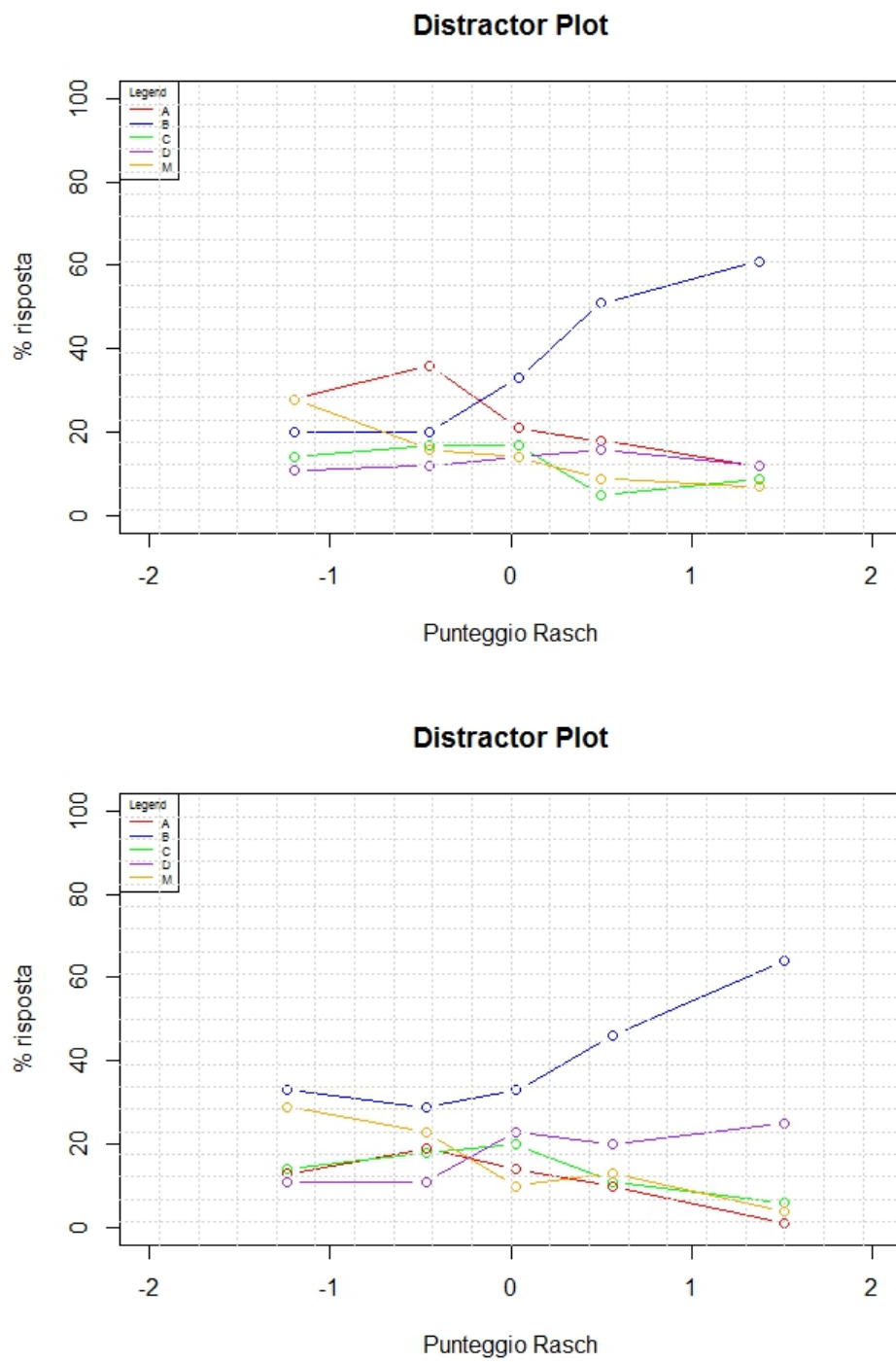


Figura 5.7: Distractor plot item D27 originale e variato

Capitolo 6

Conclusioni

Lo strumento e la metodologia presentati in questa tesi si sono rivelati effettivamente utili ed efficaci per lo studio dell'influenza delle variazioni redazionali sulla performance degli studenti. La procedura, infatti, ha efficacemente fornito informazioni sugli effetti delle variazioni su studenti di livelli di abilità diversi.

Come mostrato nel primo capitolo, sono moltissimi i tipi di variazioni già individuate che influenzano la comprensione e la risoluzione dei quesiti di un test standardizzato, e più in generale legate alla didattica della matematica. La maggior parte di questi fattori sono già stati ampiamente studiati da un punto di vista qualitativo. Lo strumento implementato permette un approfondimento da un punto di vista quantitativo, basato su risultati affidabili ottenuti tramite l'uso di software statistici.

Allo stesso tempo, questa nuova metodologia che dà la possibilità di confrontare simultaneamente le percentuali di risposta, i parametri di difficoltà degli item e i distractor plot relativi ai distrattori, potrebbero condurre all'individuazione di nuovi tipi di variazioni e di effetti non ancora studiati dai ricercatori.

Uno studio di questo tipo può essere naturalmente arricchito da studi di tipo qualitativo, come per esempio l'intervista a studenti presi singolarmente o a gruppi. In tali occasioni si potrebbe chiedere loro di fare un confronto tra le due versioni del test, invitandoli ad indicare quali sono state le difficoltà nell'affrontare le due formulazioni, se e in che modo la risposta alla seconda versione ha migliorato o

eventualmente peggiorato la loro performance.

In sintesi l'implementazione di questa metodologia è risultata efficace allo scopo prefissato, offrendo così uno strumento affidabile per futuri studi finalizzati a particolari tipologie di variazione redazionali.

Bibliografia

- [1] Bagni G.T.: Appunti di didattica della matematica, cap.5 ostacoli ed apprendimenti, Udine, 2005.
- [2] Barbaranelli C., Natali E.: I test psicologici: teorie e modelli psicometrici, Carrocci Editore, Roma, 2005.
- [3] Branchetti L., Viale M.: Tra italiano e matematica: il ruolo della formulazione sintattica nella comprensione del testo matematico. In Ostinelli M.: La didattica dell'italiano. Problemi e prospettive. Proceedings of the conference Quale didattica dell'italiano? Problemi e prospettive, Locarno, 2014.
- [4] D'Amore B.: Lingua, Matematica e Didattica. La matematica e la sua didattica. 1, 28-47, 2000.
- [5] Daroczy G., Wolska M., Meurers W.D., Nuerk H.C.: Word problems: a review of linguistic and numerical factors contributing to their difficulty, *Front Psychol*, Published online, 2015.
- [6] INVALSI: Rilevazione nazionale degli apprendimenti 2013-2014, Rapporto tecnico.
- [7] Mâsse L., Allen D., Wilson M., Williams G.: Health education research: Introducing equating methodologies to compare test scores from two different self-regulation scales Vol.21 (Supplement 1) Pages i110–i120, Advance Access publication, 22 August 2006.

-
- [8] Patrick Meyer J.: Applied measurement with jMetrik, Rutledge, New York, 2014.
- [9] Plazzi G.: Tesi di Dottorato: La costruzione di una scala di misura per l'apprendimento della matematica nella scuola primaria, Università degli Studi di Milano-Bicocca, Corso di dottorato in Statistica, 2009.
- [10] Wu M., Adams R., Wilson M., Haldane S.: Acer ConQuest version 2.0, ACER Press, Camberwell, 2007.