

RILEVAZIONI NAZIONALI SUGLI APPRENDIMENTI 2011-12

La rilevazione degli apprendimenti nelle classi II e V primaria, nelle
classi I e III (Prova nazionale) della scuola secondaria di primo grado e
nella II classe della scuola secondaria di secondo grado

Rapporto tecnico

Hanno collaborato alla redazione del presente rapporto: Angela Martini, Antonella Mastrogiovanni, Alessia Mattei, Laura Palmerio, Roberto Ricci (coordinatore). La redazione del rapporto è stata supervisionata da Paolo Sestito (commissario straordinario dell'INVALSI).

La costruzione e l'organizzazione delle rilevazioni nazionali sugli apprendimenti è stata realizzata con la collaborazione di Monica Amici, Andrea Biggera, Luigi Bonanni, Emanuela Cuzzucoli, Alessandra Fazio, Elisabetta Figura, Teresa Fiorino, Cristina Lasorsa, Antonella Mastrantonio, Antonella Mastrogiovanni, Alessia Mattei, Caterina Ponzio, Roberto Ricci (responsabile dell'area prove dell'INVALSI), Antonio Severoni, Maddalena Tozzi. Patrizia Falzetti (responsabile dell'area sistema informativo INVALSI), Francesca Fortini, Paola Giangiacomo hanno curato la predisposizione del dataset con i risultati delle rilevazioni e predisposto le elaborazioni proposte nel presente rapporto. Si ringraziano Paola Muzzioli, Donatella Poliandri e Sara Romiti (dell'area ricerche valutative dell'INVALSI) per la predisposizione del Questionario Studente; Alessandro Borsella, Carlo Di Giovamberardino (responsabile dei servizi tecnici dell'INVALSI), Stefano Famiglietti e Antonio Severoni per il fondamentale supporto informatico essenziale in tutte le fasi delle rilevazioni.

Le rilevazioni sono in ogni caso opera dell'intero sistema scolastico italiano: si ringraziano pertanto gli Uffici Scolastici regionali e provinciali, i Referenti regionali e provinciali, i Dirigenti scolastici, gli insegnanti e gli studenti di tutte le scuole italiane; un ringraziamento particolare va agli osservatori esterni e alle segreterie delle scuole campione i cui dati sono oggetto di questo rapporto.

INDICE

INDICE	1
Capitolo 1 – Le prove del Servizio nazionale di valutazione: dalle esperienze internazionali alle indicazioni per il curriculum	1
1.1 Somiglianze e differenze con le prove delle ricerche internazionali.....	1
1.2 Relazione con le indicazioni per il curriculum.....	5
Capitolo 2 – Il processo di costruzione delle prove	11
CAPITOLO 3 - La valutazione delle prove standardizzate del Servizio nazionale di valutazione ...	38
3.1 L’analisi degli item.....	38
3.2 L’indice α di Cronbach.....	39
3.3 La Rasch Analysis per la stima della difficoltà delle domande e delle “abilità” in italiano e in matematica degli studenti.....	41
3.4 L’indice di discriminazione statistica.....	43
Capitolo 4 - I dati sulle prove per i vari livelli coinvolti.....	45
4.1 La scuola primaria	45
4.2 La scuola secondaria di primo grado.....	53
4.2.1. La Prova nazionale	58
4.3. La scuola secondaria di secondo grado	62
Capitolo 5 – Il <i>placement</i> relativo delle domande rispetto al punteggio di Rasch dei rispondenti....	67

Capitolo 1 – Le prove del Servizio nazionale di valutazione: dalle esperienze internazionali alle indicazioni per il curriculum

1.1 Somiglianze e differenze con le prove delle ricerche internazionali

I *framework* delle prove usate nelle indagini internazionali sugli apprendimenti sono una delle fonti d'ispirazione dei Quadri di Riferimento INVALSI di Italiano e Matematica. Tuttavia, se tra le prove INVALSI e le prove delle ricerche internazionali sussistono evidenti similarità, vi sono anche differenze di non poco conto, su cui è opportuno soffermarsi. Esse riguardano sia il contenuto che la procedura di costruzione dei fascicoli di prova.

Cominciamo dal primo punto. I test delle indagini internazionali sugli apprendimenti (IEA-PIRLS, IEA-TIMSS e OCSE-PISA)¹ vertono su tre ambiti disciplinari: la comprensione della lettura di testi scritti, la matematica e le scienze. Innanzitutto, al momento, le prove INVALSI riguardano solo i primi due ambiti di competenza. Ciò detto, anche per quanto concerne le due aree di valutazione comuni, i contenuti delle prove INVALSI e delle prove internazionali divergono parzialmente. Vediamo come per ciascuno dei due ambiti distintamente considerato, cominciando dalla prova di Italiano INVALSI. La prima cosa da dire è che essa non si limita a valutare la comprensione della lettura ma valuta anche, in una apposita sezione, le competenze grammaticali degli studenti in relazione a 6 ambiti di contenuto: ortografia, morfologia, formazione delle parole, lessico e semantica, sintassi, testualità (vedi QdR di Italiano, tabella 4, pag. 24). Ulteriori differenze si riscontrano, inoltre, nella definizione degli aspetti della comprensione che la prova intende misurare. La tavola che segue pone a confronto i processi di comprensione indagati, da una parte, nelle due indagini internazionali sulla lettura, PIRLS e PISA, e dall'altra parte nelle rilevazioni dell'INVALSI.

¹ PIRLS (*Progress in International Reading Literacy Study*) e TIMSS (*Trends in International Mathematics and Science Study*) sono indagini promosse dalla IEA (*International Association for the Evaluation of Educational Achievement*): la prima si svolge ogni 5 anni ed ha per oggetto la comprensione della lettura degli alunni del quarto anno di scuola primaria, la seconda si tiene ogni 4 anni ed ha per oggetto la conoscenza matematica e scientifica degli alunni del quarto e dell'ottavo anno di scuola. PISA (*Program for International Student Assessment*), infine, è organizzata dall'OCSE e testa, ogni tre anni, le competenze di base degli studenti di 15 anni in lettura, matematica e scienze.

Tavola 1.1: Aspetti della comprensione della lettura misurati in PIRLS, PISA e nelle prove INVALSI di Italiano

	Aspetti della comprensione
IEA-PIRLS	<ul style="list-style-type: none"> - Ritrovare informazioni date esplicitamente nel testo - Fare inferenze dirette - Interpretare e integrare concetti e informazioni - Analizzare e valutare il contenuto, il linguaggio e gli elementi testuali
OCSE-PISA	<ul style="list-style-type: none"> - Accedere al testo e ritrovare informazioni - Integrare e interpretare le informazioni del testo, per ricostruirne il significato - Riflettere su e valutare il contenuto e/o la forma del testo
INVALSI	<ul style="list-style-type: none"> - Riconoscere e comprendere il significato letterale e figurato di parole ed espressioni; riconoscere le relazioni tra parole - Individuare informazioni date esplicitamente nel testo - Fare un'inferenza diretta, ricavando un'informazione implicita da una o più informazioni date nel testo e/o tratte dall'enciclopedia personale - Cogliere le relazioni di coesione e coerenza testuale - Ricostruire il significato di una parte più o meno estesa del testo, integrando più informazioni e concetti, anche formulando inferenze complesse - Ricostruire il significato globale del testo, integrando più informazioni e concetti, anche formulando inferenze complesse - Sviluppare un'interpretazione del testo, a partire dal suo contenuto e/o dalla sua forma, andando al di là di una comprensione letterale - Valutare il contenuto e/o la forma del testo alla luce delle conoscenze ed esperienze personali (riflettendo sulla plausibilità delle informazioni, sulla validità delle argomentazioni, sulla efficacia comunicativa, ecc.)

Mentre, nonostante alcune differenze nella formulazione verbale, gli aspetti che le prove di lettura PIRLS e PISA si prefiggono di valutare sono sostanzialmente gli stessi, tra gli aspetti su cui verte la prova INVALSI ne compaiono due che non trovano riscontro nei *framework* delle prove internazionali: essi sono la capacità di comprendere il significato di parole ed espressioni e la capacità di cogliere relazioni di coerenza e coesione testuale. Non a caso si tratta di dimensioni che sono strettamente legate alla semantica e alle strutture sintattiche e testuali di una particolare lingua e che, come tali, possono esser oggetto di valutazione in prove a carattere nazionale ma difficilmente potrebbero esserlo in prove che si rivolgono ai parlanti lingue diverse. I rimanenti cinque aspetti riprendono quelli valutati anche nelle prove internazionali, ma con una maggiore

articolazione e distinzione tra operazioni di comprensione legate alla lettera del testo e operazioni che vanno al di là di essa.

Per quanto riguarda la prova di matematica, le aree di contenuto – o nuclei tematici nel linguaggio usato dall’INVALSI – sono in sostanza le stesse valutate anche nelle prove TIMSS e PISA, al di là delle differenze nella loro denominazione. Si deve però sottolineare che gli argomenti specifici all’interno delle aree, nel caso delle prove INVALSI, hanno, come è ovvio che sia, una precisa relazione con quelli dei curricula previsti dalle norme vigenti in Italia (vedi QdR di Matematica, pag. 8).

La tavola che segue mostra gli ambiti di contenuto indagati dalle prove TIMSS e PISA e dalle prove INVALSI di Matematica.

Tavola 1.2: Ambiti di contenuto valutati in TIMSS, PISA e nelle prove INVALSI di Matematica

	Livello scolastico/Età	Aree di contenuto
IEA-TIMSS	4° anno	<ul style="list-style-type: none"> - Numero - Forme geometriche e misure - Rappresentazione di dati
	8° anno	<ul style="list-style-type: none"> - Numero - Algebra - Geometria - Dati e Probabilità
OCSE-PISA	15 anni	<ul style="list-style-type: none"> - Quantità - Spazio e Forma - Cambiamento e relazioni - Incertezza
INVALSI	2° anno	<ul style="list-style-type: none"> - Numeri - Spazio e figure - Dati e previsioni
	5°, 6°, 8°, 10° anno	<ul style="list-style-type: none"> - Numeri - Spazio e figure - Relazioni e funzioni - Dati e previsioni

Per ciò che riguarda la seconda dimensione delle prove di Matematica, vale a dire le operazioni cognitive implicate dai quesiti sugli argomenti in essi proposti, mentre le prove internazionali fanno riferimento a tre macro-categorie di processi (Conoscenza, Applicazione, Ragionamento), il QdR di matematica INVALSI (pag. 8) distingue otto tipi di processo che le prove mirano a valutare e che fungono da guida nella costruzione dei quesiti.

La maggiore differenza tra le prove internazionali e le prove INVALSI di Italiano e Matematica riguarda però, più che i contenuti, la procedura di costruzione dei fascicoli di prova, che a sua volta è connessa al diverso carattere, campionario in un caso e censuario nell'altro, delle rilevazioni. I fascicoli di prova delle indagini internazionali sugli apprendimenti sono costruiti a partire da insiemi di item che coprono l'intero spettro di contenuti e processi che di volta in volta si intendono misurare e che sono in un secondo momento suddivisi in blocchi o sottogruppi, equivalenti per gamma di difficoltà delle domande. Ad esempio, nel caso di TIMSS 2007 i blocchi erano 28, 14 per la Matematica e 14 per le Scienze (per un totale di 353 item per il 4° anno e di 429 per l'8° anno). I blocchi vengono quindi variamente combinati tra loro in modo da ottenere diversi fascicoli di prova (o *booklets*) con lo stesso numero totale di domande: solitamente 13 fascicoli nel caso di PIRLS e PISA, 14 nel caso di TIMSS. In questo modo i fascicoli di prova utilizzati ad ogni tornata delle indagini non sono identici per tutti gli alunni oggetto di rilevazione, anche se il livello di difficoltà complessivo di ogni fascicolo rimane all'incirca il medesimo e la distribuzione dei fascicoli agli studenti del campione di ognuno dei Paesi partecipanti avviene con rotazione sistematica, così da assicurare che ciascun blocco di domande all'interno dei fascicoli sia assegnato a uno stesso numero di studenti.

I fascicoli delle prove INVALSI - che, come già ricordato, hanno carattere censuario – sono invece unici per tutti gli studenti e costruiti in modo da rispettare una struttura organizzativa (per sezioni nel caso della prova di Italiano, per nuclei tematici nel caso della prova di matematica) che si riproduce in maniera uniforme da una rilevazione all'altra per lo stesso livello scolastico. Ciò ha un riflesso anche sul versante della successiva elaborazione statistica cui le risposte alle domande sono sottoposte, sia in sede di “prova sul campo”², sia in sede di rilevazione vera e propria. Mentre l'approccio seguito nelle indagini internazionali per l'analisi statistica delle risposte ai test è quello

² Si chiama così la fase di pre-test, o di test del test, cui tutte le prove INVALSI, prima della loro redazione definitiva, sono sottoposte. Da questo punto di vista la procedura è analoga a quella seguita nelle indagini internazionali, le cui prove sono prima testate su campioni ristretti di studenti nello studio preliminare sul campo (*field-trial*) mirato a validarle e metterle a punto e, una volta che ciò è stato fatto, sono proposte nella loro versione definitiva ai campioni nazionali dei vari Paesi partecipanti alla ricerca nella fase d'indagine vera e propria (*main study*).

dell'*Item Response Theory* (IRT)³, che si focalizza sui singoli item, l'INVALSI adotta sia questo approccio, sia quello cosiddetto della “teoria classica dei test” (CTT), che si focalizza invece sul test nel suo insieme e – sebbene parzialmente superato - è per alcuni versi più adatto ad analizzare i dati raccolti mediante uno strumento di misura costituito da una prova unitaria, come sono appunto le prove INVALSI.

1.2 Relazione con le indicazioni per il curricolo

Essendo le prove INVALSI prove nazionali che hanno lo scopo di verificare e comparare i livelli di apprendimento di tutti gli alunni italiani di determinati livelli scolari nelle due aree chiave dell'Italiano e della Matematica, esse non possono non avere un punto di riferimento anche nei piani di studio previsti per le scuole dal Ministero della PI. A normativa vigente, le indicazioni ministeriali per il curricolo della scuola primaria e secondaria di primo grado sono contenute in un documento (*Indicazioni per il curricolo per la scuola dell'infanzia e per il primo ciclo d'istruzione*, MPI, Roma 2007) pubblicato il 4 settembre 2007, originariamente allegato al DM 31/07/2007, meglio noto come “decreto Fioroni”. Per quanto riguarda la scuola secondaria superiore, al momento e in attesa che sia completata la riforma di questo grado d'istruzione, il punto di riferimento principale è costituito dall'allegato tecnico al DM 22 agosto 2007, dove sono elencate le competenze di base che devono essere raggiunte dagli studenti a conclusione dell'obbligo in relazione a quattro assi culturali, tra cui quello dei linguaggi e quello matematico. Tra le competenze dell'asse dei linguaggi figura la capacità di “leggere, comprendere e interpretare testi scritti di vario tipo”, mentre le competenze di base relative all'asse matematico sono così definite:

- utilizzare le tecniche e le procedure del calcolo aritmetico ed algebrico, rappresentandole anche sotto forma grafica;
- confrontare e analizzare figure geometriche, individuando invarianti e relazioni;
- individuare le strategie appropriate per la soluzione di problemi;
- analizzare dati e interpretarli sviluppando deduzioni e ragionamenti sugli stessi anche con l'ausilio di rappresentazioni grafiche, usando consapevolmente gli strumenti di calcolo.

A titolo di esempio, nelle tavole che seguono riprendiamo dai documenti ministeriali, distintamente per l'Italiano e la Matematica, alcuni degli obiettivi di apprendimento indicati per la classe quinta

³ Van der Linden, W.J. & Hambleton, R.K. (Eds.) (1997). *Handbook of modern item response theory*. New York: Springer

primaria e terza media, da cui risultano evidenti i collegamenti con quello che le prove INVALSI si prefiggono di valutare.

Tavola 1.3: Obiettivi di apprendimento indicati nelle indicazioni curriculari MPI per il primo ciclo d’istruzione per l’Italiano (DM 31.07.07) testati nelle prove INVALSI

V Primaria	Leggere	<ul style="list-style-type: none"> – Leggere testi narrativi e descrittivi, sia realistici sia fantastici, distinguendo l'invenzione letteraria dalla realtà. – Sfruttare le informazioni della titolazione, delle immagini e delle didascalie per farsi un’idea del testo che si intende leggere. – Ricercare informazioni in testi di diversa natura e provenienza per scopi pratici e/o conoscitivi. – Leggere semplici e brevi testi letterari sia poetici sia narrativi mostrando di riconoscere le caratteristiche essenziali che li contraddistinguono.
	Riflettere sulla lingua	<ul style="list-style-type: none"> – Riconoscere e denominare le parti principali del discorso e gli elementi basilari di una frase; individuare e usare in modo consapevole modi e tempi del verbo; riconoscere in un testo i principali connettivi (temporali, spaziali, logici); analizzare la frase nelle sue funzioni (predicato e principali complementi diretti e indiretti). – Conoscere i principali meccanismi di formazione e derivazione delle parole (parole semplici, derivate, composte, prefissi e suffissi). – Comprendere le principali relazioni tra le parole (somiglianze, differenze) sul piano dei significati. – Utilizzare il dizionario come strumento di consultazione. – Riconoscere la funzione dei principali segni interpuntivi.

Tavola 1.3: Obiettivi di apprendimento indicati nelle indicazioni curriculari MPI per il primo ciclo d’istruzione per l’Italiano (DM 31.07.07) testati nelle prove INVALSI

III Secondaria primo grado	Leggere	<ul style="list-style-type: none"> – Ricavare informazioni esplicite e implicite da testi informativi ed espositivi per documentarsi su un argomento specifico e/o per realizzare scopi pratici. – Comprendere testi letterari di vario tipo e forma (racconti, novelle, romanzi, poesie) individuando personaggi, loro caratteristiche, ruoli, relazioni e motivazione delle loro azioni; ambientazione spaziale e temporale; relazioni causali, tema principale e temi di sfondo; il genere di appartenenza e le tecniche narrative usate dall’autore. – Comprendere testi descrittivi, individuando gli elementi della descrizione, la loro collocazione nello spazio, le caratteristiche essenziali, il punto di vista dell'osservatore. – Comprendere tesi centrale, argomenti a sostegno e intenzione comunicativa di semplici testi argomentativi.
	Riflettere sulla lingua	<ul style="list-style-type: none"> – Conoscere la costruzione della frase complessa (distinguere la principale dalle subordinate) e riconoscere i principali tipi di proposizioni subordinate (relative, temporali, finali, causali, consecutive, ecc.). – Analizzare la frase complessa e visualizzare i rapporti fra le singole proposizioni rappresentandoli anche graficamente. – Stabilire relazioni tra situazione di comunicazione, interlocutori e registri linguistici. – Stabilire relazioni tra campi di discorso e forme di testo, lessico specialistico, ecc. – Riconoscere in un testo i principali connettivi e la loro funzione. – Conoscere le principali relazioni fra significati (sinonimia, contrarietà, polisemia, gradazione, inclusione). – Conoscere i principali meccanismi di derivazione. – Utilizzare strumenti di consultazione (riconoscere e capire il tipo di informazioni fornite da un dizionario per ogni voce). – Riconoscere le caratteristiche dei principali tipi testuali (narrativi, regolativi, descrittivi, argomentativi) e dei generi.

Tavola 1.4: Obiettivi di apprendimento indicati nelle indicazioni curriculari MPI per il primo ciclo d’istruzione per la Matematica (DM 31.07.07) testati nelle prove INVALSI

V Primaria	Numeri	<ul style="list-style-type: none"> – Individuare multipli e divisori di un numero. – Leggere, scrivere, confrontare numeri decimali ed eseguire le quattro operazioni con sicurezza. – Dare stime per il risultato di una operazione. – Conoscere il concetto di frazione e di frazioni equivalenti. – Utilizzare numeri decimali, frazioni e percentuali per descrivere situazioni quotidiane. – Interpretare i numeri interi negativi in contesti concreti. – Rappresentare i numeri conosciuti sulla retta e utilizzare scale graduate in contesti significativi.
	Spazio e figure	<ul style="list-style-type: none"> – Classificare figure geometriche, identificando elementi significativi e simmetrie. – Riprodurre una figura in base a una descrizione, utilizzando gli strumenti opportuni. – Utilizzare il piano cartesiano per localizzare punti. – Riconoscere figure ruotate, traslate e riflesse. – Determinare il perimetro di una figura. – Determinare l’area di rettangoli e triangoli e di altre figure per scomposizione.
	Relazioni, misure, dati e previsioni	<ul style="list-style-type: none"> – Rappresentare relazioni e dati e, in situazioni significative, utilizzare le rappresentazioni per ricavare informazioni, formulare giudizi e prendere decisioni. – Usare le nozioni di media aritmetica e di frequenza. – Rappresentare problemi con tabelle e grafici che ne esprimono la struttura. – Conoscere le principali unità di misura per lunghezze, angoli, aree, volumi/capacità, intervalli temporali, masse/pesi e usarle per effettuare misure e stime. – Passare da un’unità di misura a un’altra, limitatamente alle unità di uso più comune, anche nel contesto del sistema monetario. – In situazioni concrete, di una coppia di eventi intuire e cominciare ad argomentare qual è il più probabile, dando una prima quantificazione, oppure riconoscere se si tratta di eventi ugualmente probabili. – Riconoscere e descrivere regolarità in una sequenza di numeri o di figure.

Tavola 1.4: Obiettivi di apprendimento indicati nelle indicazioni curriculari MPI per il primo ciclo d’istruzione per la Matematica (DM 31.07.07) testati nelle prove INVALSI

III Secondaria primo grado	Numeri	<ul style="list-style-type: none"> – Eseguire addizioni, sottrazioni, moltiplicazioni, divisioni e confronti tra i numeri conosciuti (numeri naturali, numeri interi, frazioni e numeri decimali). – Dare stime approssimate per il risultato di una operazione. – Rappresentare i numeri conosciuti sulla retta. – Utilizzare scale graduate in contesti significativi. – Descrivere rapporti e quozienti mediante frazioni. – Utilizzare frazioni equivalenti e numeri decimali per denotare uno stesso numero razionale in diversi modi. – Calcolare percentuali. – Individuare multipli e divisori di un numero naturale e multipli e divisori comuni a più numeri. – Comprendere il significato e l'utilità del multiplo comune più piccolo e del divisore comune più grande, in matematica e in diverse situazioni concrete. – Usare le proprietà delle potenze anche per semplificare calcoli e notazioni. – Dare stime della radice quadrata utilizzando solo la moltiplicazione. – Eseguire mentalmente semplici calcoli, utilizzando la proprietà associativa e distributiva per raggruppare e semplificare le operazioni. – Descrivere con una espressione numerica la sequenza di operazioni che fornisce la soluzione di un problema. – Eseguire semplici espressioni di calcolo con i numeri conosciuti, essendo consapevoli del significato delle parentesi e delle convenzioni sulla precedenza delle operazioni.
	Spazio e figure	<ul style="list-style-type: none"> – Rappresentare punti, segmenti e figure sul piano cartesiano. – Conoscere definizioni e proprietà significative delle principali figure piane (triangoli, quadrilateri, poligoni regolari, cerchio). – Riprodurre figure e disegni geometrici in base a una descrizione e codificazione fatta da altri. – Riconoscere figure piane simili in vari contesti. – Conoscere il Teorema di Pitagora e le sue applicazioni in matematica e in situazioni concrete. – Calcolare l’area di semplici figure scomponendole in figure elementari, ad esempio triangoli. – Stimare per difetto e per eccesso l’area di una figura delimitata da linee curve. – Conoscere le formule per trovare l’area del cerchio e la lunghezza della circonferenza, conoscendo il raggio. – Visualizzare oggetti tridimensionali a partire da rappresentazioni bidimensionali. – Calcolare il volume delle figure tridimensionali più comuni e dare stime di quello degli oggetti della vita quotidiana. – Risolvere problemi utilizzando le proprietà geometriche delle figure.
	Relazioni e funzioni	<ul style="list-style-type: none"> – Costruire, interpretare e trasformare formule che contengono lettere per esprimere in forma generale relazioni e proprietà. – Esprimere la relazione di proporzionalità con una uguaglianza di frazioni e viceversa. – Usare il piano cartesiano per rappresentare relazioni e funzioni. – Esplorare e risolvere problemi utilizzando equazioni di primo grado.
	Misure, dati e previsioni	<ul style="list-style-type: none"> – In situazioni significative, confrontare dati al fine di prendere decisioni, utilizzando le distribuzioni delle frequenze e delle frequenze relative e le nozioni di media aritmetica e mediana. – In semplici situazioni aleatorie, individuare gli eventi elementari, discutere i modi per assegnare a essi una probabilità, calcolare la probabilità di qualche evento, scomponendolo in eventi elementari disgiunti. – Riconoscere coppie di eventi complementari, incompatibili, indipendenti.

Per concludere, è anche il caso di rilevare che, sebbene la logica di costruzione delle prove, dal punto di vista dello schema organizzativo sia la stessa per i diversi livelli scolari, fatte salve le differenze legate all'età degli alunni, le prove di seconda primaria e prima media rispetto alle altre tre, che mirano a rilevare i livelli raggiunti alla fine delle principali tappe del percorso scolastico (fine scuola primaria, fine primo ciclo, fine istruzione obbligatoria), hanno in qualche modo anche la funzione – oltre che di valutare le competenze raggiunte - di definire una linea di base a partire da cui sia possibile stabilire il progresso realizzato dagli studenti nel corso del successivo itinerario educativo.

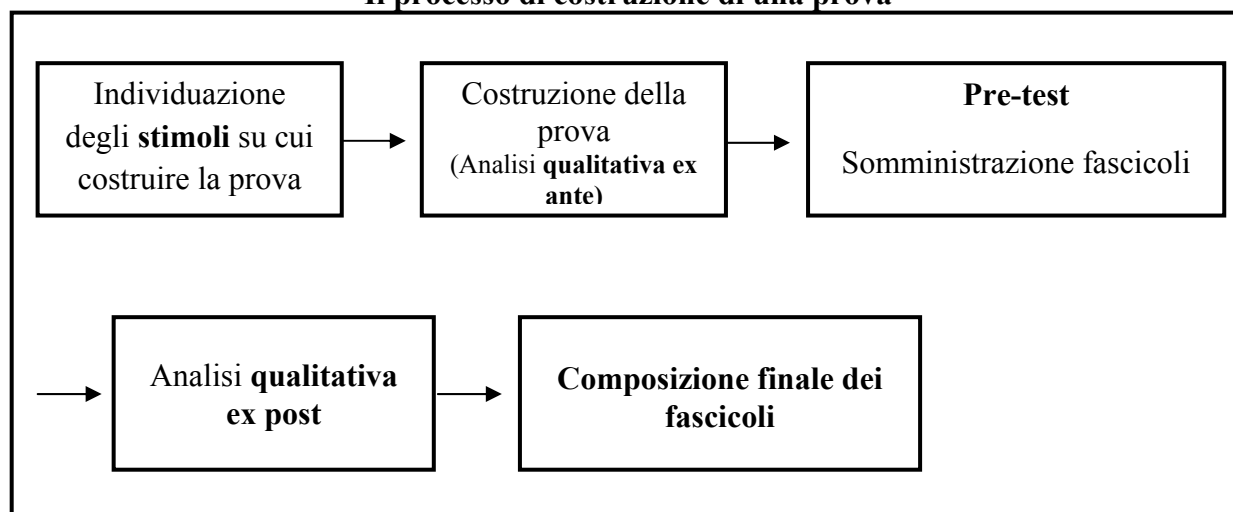
Capitolo 2 – Il processo di costruzione delle prove

Le prove standardizzate, strumenti utilizzati per la misurazione degli apprendimenti, forniscono misure sufficientemente attendibili del grado di padronanza di quegli elementi che sono l'oggetto della prova stessa.

La costruzione di prove standardizzate prevede come base di partenza la predisposizione di un quadro teorico di riferimento (QdR) in cui vengono descritti gli aspetti/ambiti oggetto della misurazione e le caratteristiche delle prove.

La costruzione di una prova standardizzata è il risultato di una attività di ideazione, reperimento di materiali, stesura, verifica, correzione ed altro ancora che ha le caratteristiche di un percorso di ricerca sperimentale che inizia con l'individuazione degli stimoli, fino alla composizione del fascicolo (Schema 1). Il tempo richiesto per la realizzazione del lungo e articolato processo tecnico scientifico di costruzione di una prova è di circa 15-18 mesi.

Il processo di costruzione di una prova



Di seguito vengono descritte nel dettaglio le fasi in cui si articola il processo di costruzione di una prova.

FASE 1

Per costruire una prova, per ogni livello scolastico interessato dal Servizio Nazionale di Valutazione (SNV), è necessario selezionare stimoli adeguati e costruire su di essi un numero consistente di quesiti.

Di norma, per la costruzione di una prova serve inizialmente un numero di quesiti superiore a quello che effettivamente comparirà nella prova che verrà poi somministrata agli studenti.

La prima fase del processo di costruzione prevede anche la formazione di un gruppo di autori (docenti di tutti i livelli scolastici) per approfondire le modalità di costruzione di una prova di tipo standardizzato (dalla selezione dello stimolo alla costruzione dei quesiti) puntando l'attenzione sulle differenze tra questa tipologia di prove e le prove che vengono utilizzate dai docenti nella pratica didattica.

FASE 2

L'analisi qualitativa ex-ante consiste nella revisione e classificazione degli stimoli (per la prova di italiano anche la tipologia di testo: narrativo, espositivo, misto ecc.) e nella verifica dei quesiti costruiti. Questo lavoro di analisi è affidato a un gruppo di esperti composto da ricercatori dell'INVALSI, esperti nazionali (docenti e disciplinari) ed internazionali nell'ambito della costruzione di prove oggettive e delle analisi statistico-psicometriche. Il gruppo di lavoro procede a una prima valutazione qualitativa delle prove, in funzione della rispondenza di queste al QdR e al livello scolastico per il quale devono essere proposte le prove.

L'obiettivo del gruppo di esperti è quello di comporre i fascicoli che dovranno poi essere pretestati. Il lavoro di analisi consiste nell'escludere stimoli e quesiti non coerenti con le finalità del Servizio Nazionale di Valutazione e nel procedere ad un primo adattamento dei quesiti stessi (modifica di alcune opzioni di risposta nel caso di domande con 4 alternative di risposta, trasformazione di item chiusi in item aperti e viceversa, modifica della domanda, ecc.) ritenuti idonei per essere inviate al pre-test.

FASE 3

La fase del pre-test riveste un'importanza notevole nell'intero processo di costruzione della prova ed è il momento in cui si hanno dei riscontri "empirici" rispetto al lavoro realizzato.

Due sono gli aspetti su cui si punta l'attenzione per la riuscita di questa fase: da una parte la composizione dei fascicoli da somministrare, dall'altra il target di popolazione a cui vengono sottoposte le prove.

Nel primo caso, è importante far ruotare i singoli quesiti e nel caso specifico della prova di italiano i diversi testi all'interno del fascicolo. Nella fase del pre-test è possibile anche sperimentare quesiti formulati diversamente ma che rilevano lo stesso aspetto/ambito.

Nel secondo caso è importante riuscire a somministrare i fascicoli a studenti con caratteristiche analoghe a quelle di coloro che dovranno svolgere le prove SNV; l'ideale è somministrare le prove del pre-test nei mesi di aprile e maggio nelle classi il cui target corrisponde a quello che realmente deve affrontare la prova (classe seconda e quinta della scuola primaria, classe prima della scuola secondaria di I grado, classe terza della scuola secondaria di I grado, classe seconda della scuola secondaria di II grado).

Il pre-test viene condotto durante l'anno scolastico precedente a quello dello studio principale. Il numero di studenti coinvolti dipende fondamentalmente da quanti fascicoli devono essere pretestati e di conseguenza è necessario un numero di allievi, per ogni livello scolastico, che consenta poi di avere una buona "tenuta" statistica dei dati raccolti. Il campione per il pre-test è rappresentativo per le diverse tipologie di scuole (soprattutto per la secondaria superiore) e per area geografica.

Le prove sono somministrate esclusivamente da personale INVALSI, l'unico che accede ai contenuti dei fascicoli per ovvi motivi di riservatezza; un procedimento ugualmente riservato viene seguito anche per la correzione delle prove.

Successivamente si procede alla costruzione del dataset (immissione dei dati) per l'analisi dei dati. Le analisi vengono realizzate attraverso l'applicazione di modelli statistico-psicometrici ascrivibili alla teoria cosiddetta "classica" dei test (*TCT*) e alla teoria della risposta (*Item Response Theory*)⁴.

La capacità misuratoria di ogni quesito viene analizzata mediante modelli statistici in grado di stabilire la coerenza di ciascuna opzione di risposta rispetto al costruito oggetto di valutazione, rispetto al livello di abilità del rispondente e rispetto alla difficoltà specifica del quesito stesso.

I richiamati modelli statistici permettono, inoltre, di valutare il cosiddetto potere discriminante di ciascun quesito, ovvero la capacità di ogni quesito di distinguere adeguatamente gli allievi in termini di abilità, in funzione della risposta fornita.

⁴ (Cfr. Boomsma, van Duijn, Snijders, 1991; Fischer, Molenaar 1995; Hambleton, Swaminathan, Rogers, 1991)

FASE 4

Sulla base delle analisi statistiche il gruppo di esperti procede alla seconda analisi qualitativa dei quesiti.

In questa fase, la più delicata, l'oggettività dei dati raccolti spesso chiarisce i dubbi e le perplessità scaturiti durante il processo di costruzione delle prove. Tuttavia, l'esperienza e la professionalità di chi legge quei dati, non solo da un punto di vista psicometrico, consentono di tenere ben presenti alcuni aspetti che i dati da soli non spiegano.

Solo quei quesiti che mostrano adeguati requisiti di chiarezza, affidabilità e validità possono essere inseriti nei fascicoli definitivi.

FASE 5

L'intero processo si conclude con la composizione dei fascicoli che saranno somministrati durante la rilevazione principale.

Inoltre, il gruppo di esperti procede alla composizione dei fascicoli definitivi tenendo in considerazione:

- i tempi di compilazione dei fascicoli, stimati sulla base dei risultati del pre-test;
- il livello complessivo di difficoltà, dati i vincoli di composizione definiti dai quadri di riferimento per la valutazione.

La struttura dei quesiti

Le prove standardizzate si caratterizzano per la chiusura degli stimoli e delle risposte⁵. L'obiettivo è quello di ridurre l'ambiguità interpretativa, che diminuisce tanto più quanto più precisi, chiari, circoscritti sono gli stimoli e i quesiti a cui si deve rispondere. Di conseguenza si facilita il lavoro di correzione che risulta tanto più univoco quanto più il numero delle risposte possibili/accettabili risulta delimitato.

Le prove standardizzate strutturalmente dovrebbero garantire a tutti i soggetti ai quali una prova viene somministrata le stesse condizioni di lavoro: stessa prova e stesso tempo a disposizione; condizioni necessarie che contribuiscono al rispetto dei requisiti della validità e della attendibilità delle rilevazioni e quindi delle misurazioni degli apprendimenti.

⁵ (Cfr. Haladyna 1994; McDonald 2002)

Non a caso le domande a risposta chiusa rappresentano la tipologia di quesiti più utilizzata nella costruzione di prove standardizzate.

I vantaggi nell'utilizzo di questa tipologia di quesiti sono molteplici:

- le modalità di correzione soddisfano il criterio della riproducibilità, l'esito della correzione è indipendente dal soggetto che la effettua riducendo quindi al minimo la percentuale di errori;
- riducono il problema delle omissioni e gli studenti le percepiscono come più agevoli;
- ogni quesito sottoposto ad analisi statistica fornisce una serie di dati (disponibili per ognuna delle alternative di risposta) che consentono di capire più facilmente il perché degli errori.

La maggior parte dei quesiti che compongono la prova sono domande a scelta multipla con 4 alternative di risposta. La difficoltà nel costruire questa tipologia di quesiti consiste proprio nel formulare le quattro alternative (una risposta corretta e tre distrattori); nello specifico i distrattori dovranno risultare comunque "plausibili" (cfr. Osterlind 1998), per evitare che lo studente arrivi alla soluzione corretta per approssimazioni successive, ossia scartando quei distrattori poco convincenti per giungere alla individuazione della risposta corretta in una condizione di maggiore o minore probabilità. Al contrario, un quesito "ben" formulato dovrebbe far attivare allo studente un procedimento logico che risulti significativo dal punto di vista delle strategie cognitive messe in atto per giungere, in una situazione di certezza, alla risposta corretta.

Il processo di costruzione delle domande a scelta multipla richiede, quindi, particolare attenzione se si vuole ottenere un quesito che abbia una "robustezza" dal punto di vista psicometrico⁶.

Nel costruire la domanda e le alternative di risposta vengono prese in considerazione alcune indicazioni generali (riportate di seguito) che, anche se non sempre vengono comunque tutte rispettate, rappresentano il riferimento certo verso cui convergere.

Indicazioni per la costruzione della domanda.

1. La domanda deve essere formulata in maniera diretta e positiva (limitare il più possibile l'uso delle negazioni).
2. Le domande devono richiedere una sola informazione.
3. La domanda deve contenere solo informazioni indispensabili.

⁶ Per una descrizione dettagliata sul formato dei quesiti si rimanda al QdR.

4. La domanda non deve lasciare dubbi sul tipo di richiesta fatta (vocabolario preciso ma al tempo stesso il più semplice possibile, evitare costruzioni complesse come ad esempio forme passive etc.).

Indicazioni per la costruzione delle alternative di risposta.

1. Le alternative di risposta devono essere legate in modo grammaticalmente corretto alla domanda.
2. Le alternative di risposta devono essere indipendenti fra loro e mutualmente esclusive.
3. Le alternative di risposta non devono contenere parti della domanda.
4. Le alternative di risposta devono avere più o meno la stessa lunghezza.
5. Le alternative di risposta devono essere formulate cercando di evitare l'uso di termini assoluti.
6. Le alternative di risposta che presentano l'opzione *nessuna delle precedenti* o simili devono essere evitate.

Indicazioni per la scelta dei distrattori.

1. I distrattori non devono essere troppo vicini alla risposta corretta.
2. I distrattori devono essere abbastanza attrattivi e plausibili (es. evitare di formulare distrattori che possono essere esclusi anche senza leggere il testo).
3. I distrattori non devono essere costruiti per trarre in inganno il rispondente.

Nella costruzione delle alternative di risposta, una volta formulata la risposta corretta, si devono costruire distrattori plausibili in modo che la risposta fornita dallo studente rappresenti, come già detto, il risultato di un articolato processo di discriminazione (tra chi padroneggia di più un certo tipo di abilità, o tratto latente, che la prova intende misurare e chi lo padroneggia meno). Se la risposta giusta venisse, a qualunque livello di abilità, sempre preferita rispetto ai distrattori, questa sarebbe una evidenza, secondo le regole che presiedono alla costruzione di domande a scelta multipla, che i distrattori non funzionano e che vanno riformulati o sostituiti. È necessario, inoltre, porre particolare attenzione nel momento della costruzione della domanda a valutare che una sola alternativa sia corretta, proprio perché le altre alternative sono state costruite sul criterio della plausibilità.

Altro elemento che deve essere considerato è la posizione della risposta corretta rispetto alle altre tre opzioni di risposta, variandola all'interno della prova (fascicolo) in modo che occupi le diverse posizioni. È importante, infatti, evitare qualsiasi forma di regolarità nella successione delle risposte corrette.

La lettura dei dati del pre-test

I risultati della prova sul campo (*pre-test*) vengono elaborati statisticamente per stabilire le proprietà misurative dell'intera prova e di ciascun singolo quesito e per poter procedere alla selezione e correzione degli stessi.

Per ognuna delle prove utilizzate nel pre-test come nelle rilevazioni nazionali vengono presi in considerazione alcuni indici statistici⁷:

- l' α di *Cronbach*, che esprime una valutazione sintetica e generale sulla coerenza complessiva di una prova e quindi sull'attendibilità delle informazioni della prova stessa, nel senso che permette di capire se le domande che compongono la prova sono tra loro coerenti e volte alla misurazione dello stesso oggetto. In generale elevati valori di α indicano che i rispondenti esprimono abilità coerenti rispetto a ciascun item appartenente a ciascuna dimensione;
- la difficoltà degli item in termini di percentuale di risposte corrette;
- il coefficiente di correlazione punto-biserial degli item che esprime la correlazione tra i punteggi ottenuti dai rispondenti su un determinato item e il punteggio totale dei rispondenti su tutti gli item;
- i parametri di difficoltà e di discriminazione degli item;
- il posizionamento relativo (*placement*) degli item rispetto al punteggio Rasch dei rispondenti che permette di osservare contemporaneamente la difficoltà delle domande e le abilità dei rispondenti. La lettura del *placement* verifica come le domande si distribuiscono lungo l'intera scala di difficoltà o se sono presenti degli intervalli troppo ampi in cui non è presente alcun item della prova.

⁷ Per una descrizione più approfondita vedi il capitolo 3 del presente rapporto.

La lettura dei dati del pre-test consente di intervenire sulle domande e sulle alternative di risposta (risposta corretta e distrattori), scartando, correggendo o modificando i quesiti dove necessario (cfr. Osterlind 1983).

Per spiegare come vengono utilizzati i parametri statistici, calcolati per ogni quesito, viene descritto il processo di trasformazione che questo può subire in seguito al pre-test.

Di seguito si presentano alcuni esempi in cui vengono riportati i quesiti nella versione utilizzata nel pre-test, con le relative elaborazioni, e gli stessi, presentati nella versione utilizzata per la rilevazione nazionale degli apprendimenti con relative nuove elaborazioni.

I primi due quesiti esemplificano casi in cui è stato necessario modificare i distrattori (domande a scelta multipla), il terzo quesito a più item esemplifica casi in cui è stato necessario modificare le domande.

Esempi di modifica dei distrattori

ESEMPIO 1: QUESITO DI II PRIMARIA - ITALIANO

Box 2.1: Quesito nella versione pre-test

Testo: La gara di barche

B11. In quale momento Piddo ha detto che la barca della squadra avversaria aveva un forellino?

- A. Quando la gara era appena iniziata
- B. Quando la gara stava per finire
- C. Quando la gara era finita
- D. Quando la gara era in pieno svolgimento

**Tabella 2.1: Analisi IRT
Tabella dati: quesito B11**

```

Item 29
-----
item:29 (B11)
Cases for this item   176   Discrimination  0.32
Item Threshold(s):   0.40   Weighted MNSQ  1.06
Item Delta(s):       0.40
-----

```

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	20	11.36	-0.15	-1.98 (.049)	-0.34	0.61
2	0.00	63	35.80	0.06	0.77 (.444)	-0.04	0.64
3	0.00	11	6.25	-0.22	-2.90 (.004)	-0.59	0.73
4	1.00	73	41.48	0.32	4.40 (.000)	0.28	0.79
9	0.00	9	5.11	-0.38	-5.48 (.000)	-1.14	1.16

```

=====

```

La prima indicazione che troviamo nella Tabella 2.1 è il numero di rispondenti ai quali è stato sottoposto l'item. Nel caso in questione, l'item 29 (quesito B11) è stato somministrato a 176 individui. L'informazione immediatamente successiva, che troviamo sulla stessa riga, è l'indice di discriminazione (*Discrimination*) che consente di valutare in termini quantitativi la capacità di una domanda di individuare gruppi di rispondenti in funzione del loro livello di preparazione. In altri termini, quanto più una domanda è *discriminativa*, tanto più essa è in grado di misurare la variazione di probabilità di fornire la risposta corretta anche per piccole variazioni di abilità del rispondente. Gli standard in letteratura indicano che l'indice di discriminazione deve raggiungere almeno il valore 0.20 e può considerarsi buono quando supera il valore di 0.25. La domanda in questione risulta avere una discriminazione decisamente buona 0.32.

Nella prima colonna (Label) sono indicate le quattro alternative di risposta (1, 2, 3, 4) e le mancate risposte (9). Nella colonna 2 (Score) viene indicata l'opzione di risposta corretta con il codice 1.00, in questo caso la risposta corretta è l'opzione 4. Le colonne 3 (Count) e 4 (% of tot) riportano rispettivamente le frequenze assolute e percentuali delle risposte registrate. Questa informazione viene letta assieme ai dati della colonna 5 (Pt Bis) che riporta i valori della correlazione punto-biseriale (la correlazione tra la probabilità di scegliere una data opzione e l'abilità complessiva del rispondente). Tale correlazione deve essere negativa per le opzioni di risposta non corrette e positiva per quella esatta. Una domanda a scelta multipla è ben formulata se, in media, coloro che rispondono correttamente a quella domanda ottengono un punteggio elevato nella prova complessivamente considerata⁸. Nell'esempio presentato, la correlazione punto-biseriale dell'opzione corretta è +0.32. Per le altre opzioni di risposta, la 1 e la 3 hanno correlazioni punto-biseriali negative, invece la 2 ha una correlazione punto-biseriale positiva (0.06). I dati della colonna 6 [$t(p)$] consentono di valutare la significatività della correlazione punto-biseriale: come si può vedere nell'esempio, i valori delle opzioni di risposta 1, e 3 (colonna 1) registrati nella colonna 5 sono significativi in senso propriamente statistico⁹, mentre il valore relativo all'opzione 2 registra una correlazione punto biseriale positiva, non significativa da un punto di vista statistico (la

⁸ Ciò vale a condizione che la prova sia coerente e quindi i risultati forniti siano attendibili. La valutazione dell'attendibilità e coerenza complessiva della prova è effettuata mediante l' α di Cronbach come precedentemente indicato.

⁹ Dal punto di vista propriamente statistico un determinato valore di sintesi, in questo caso il coefficiente di correlazione punto-biseriale, è statisticamente significativo se il valore $t(p)$ ha una probabilità, indicata tra parentesi, di prodursi per solo effetto del caso inferiore allo 0.05 (5%). Più precisamente t rappresenta il valore della t di Student e il numero indicato tra parentesi il relativo p-value, ossia la probabilità di ottenere un valore di t maggiore o uguale a quello ottenuto.

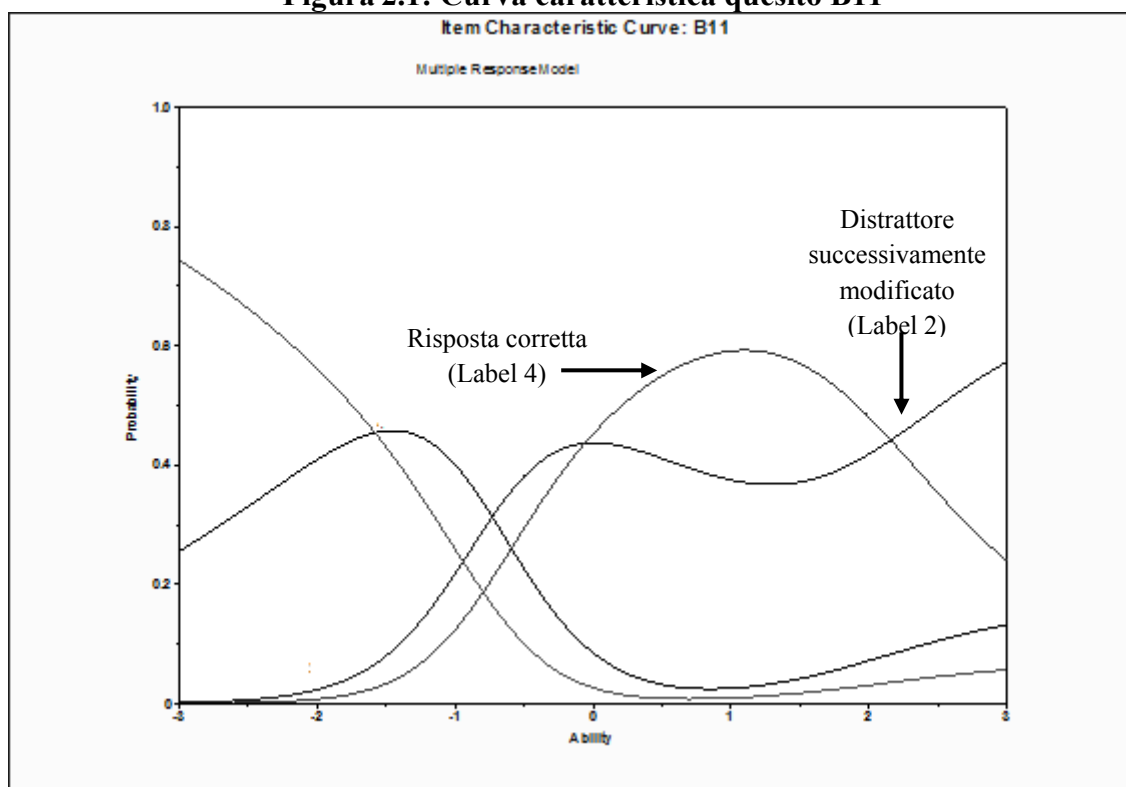
probabilità di $t(p) = (.444)$ è un valore molto superiore allo 0.05). Questo significa che l'opzione di risposta 2 viene scelta anche da soggetti con abilità non troppo basse.

Infine, la colonna 7 permette di valutare su una scala di *Rasch* il livello medio di prestazione dei rispondenti che scelgono una determinata risposta, mentre la colonna 8 fornisce una misura della variabilità di ciascuno dei risultati riportati in colonna 7. È importante notare che, in media, il livello di preparazione dei rispondenti che scelgono l'opzione corretta di risposta è più elevato di quello di coloro che scelgono le altre opzioni e ciò fornisce un'ulteriore indicazione positiva circa l'adeguatezza della domanda presa in esame.

Si studiano anche le Curve Caratteristiche per confermare il funzionamento degli item per intervenire laddove necessario. Al crescere del livello di abilità θ_i , cresce la probabilità di rispondere correttamente all'item; maggiore è il valore del parametro β_i maggiore deve essere l'abilità del soggetto per riuscire a rispondere correttamente all'item.

In riferimento al quesito B11, preso come esempio, anche l'elaborazione delle curve caratteristiche conferma la necessità di modificare il secondo distrattore. L'andamento della curva della risposta corretta risente del fatto che il secondo distrattore ha attirato anche studenti con abilità non troppo basse.

Figura 2.1: Curva caratteristica quesito B11



22

In questo caso, quindi, si procede alla modifica della opzione di risposta 2 come si può vedere nel Box 2 in cui è stato anche modificato l'ordine con cui sono presentati i distrattori.

Per cui l'opzione di risposta modificata (nel pretest opzione 2) la ritroviamo come opzione 1, la risposta corretta come opzione 3. Una piccolissima modifica è stata apportata all'opzione di risposta 2 (che nel pretest era 1), sostituendo il verbo "iniziare" con "cominciare" in modo che le prime due opzioni avessero una continuità linguistica.

Box 2.2: Quesito nella versione definitiva SNV 2011-2012

Testo: La gara di barche

B8. In quale momento Piddo ha detto che la barca della squadra avversaria aveva un forellino?

- A. Quando la gara non era ancora cominciata
- B. Quando la gara era appena cominciata
- C. Quando la gara era in pieno svolgimento
- D. Quando la gara era finita

Il quesito con le modifiche apportate ha fatto parte della versione definitiva del fascicolo¹⁰ come quesito B8. I dati elaborati per tale quesito mostrano che la correzione ha in effetti ottimizzato le sue caratteristiche strutturali come si osserva dai dati che seguono (Tabella 2.2).

**Tabella 2.2: Analisi IRT
Tabella dati: quesito B8**

Label	Score	Count	% of tot	Pt Bis	t	(p)	PV1Avg:1	PV1 SD:1
1	0.00	2022	6.39	-0.23	-41.16	(.000)	-0.74	0.86
2	0.00	5045	15.94	-0.20	-36.03	(.000)	-0.41	0.89
3	1.00	21515	67.98	0.49	99.26	(.000)	0.29	0.89
4	0.00	2233	7.06	-0.24	-44.69	(.000)	-0.76	0.77
7	0.00	68	0.21	-0.08	-13.85	(.000)	-1.48	1.29
9	0.00	766	2.42	-0.22	-39.81	(.000)	-1.21	0.98

¹⁰ http://www.invalsi.it/snv2012/documenti/Ril_apprendimenti/Italiano_SECONDA_primaria_SNV2012.pdf

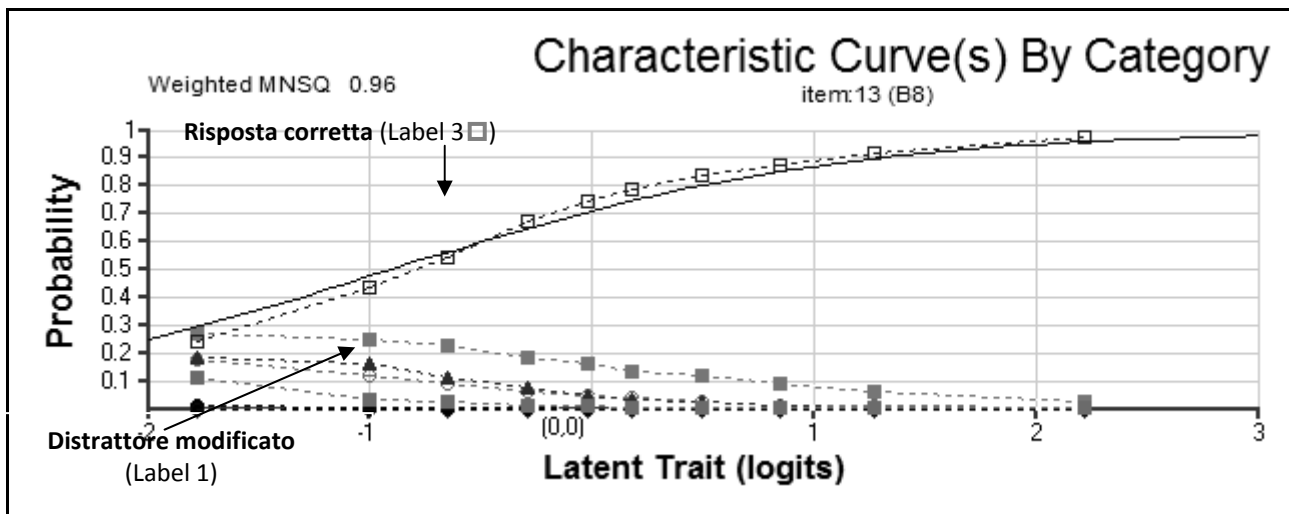


Figura 2.2: Curva caratteristica quesito B8

Le curve relative al quesito B8 (Figura 2.2) mostrano chiaramente che la modifica apportata all'opzione di risposta 1 (nel pretest opzione 2) ha funzionato; l'andamento della curva relativa al primo distrattore si allinea a quella degli altri e la curva della opzione di risposta corretta si allinea a quella prevista dal modello.

ESEMPIO 2: QUESITO DI V PRIMARIA - ITALIANO

Box 2.3: Quesito nella versione pre-test

Testo: La magia della bidella

A9. Qual è stata la reazione della mamma al racconto del figlio?

- A. Si è inquietata
- B. Si è divertita
- C. Si è offesa
- D. Si è interessata

Tabella 2.3: Analisi IRT
Tabella dati: quesito A9

```

Item 9
-----
item:9 (A9)
Cases for this item      409      Discrimination  0.18
Item Threshold(s):     -0.79      Weighted MNSQ  1.22
Item Delta(s):         -0.79
-----

```

Label	Score	Count	% of tot	Pt Bis	t (p)	PVIAvg:1	PVI SD:1
1	1.00	272	66.50	0.18	3.77 (.000)	0.07	0.87
2	0.00	6	1.47	-0.20	-4.09 (.000)	-1.14	0.69
3	0.00	15	3.67	-0.23	-4.87 (.000)	-0.95	0.73
4	0.00	106	25.92	-0.01	-0.11 (.914)	-0.03	0.82
7	0.00	2	0.49	-0.18	-3.79 (.000)	-1.83	0.48
9	0.00	8	1.96	-0.03	-0.51 (.613)	-0.04	0.69

La prima indicazione che troviamo nella Tabella 2.3 è il numero di rispondenti ai quali è stato sottoposto l'item. Nel caso in questione, l'item 9 (quesito A9) è stato somministrato a 409 individui. L'informazione immediatamente successiva, che troviamo sulla stessa riga, è l'indice di discriminazione (*Discrimination*) che consente di valutare in termini quantitativi la capacità di una domanda di individuare gruppi di rispondenti in funzione del loro livello di preparazione. In altri termini, quanto più una domanda è *discriminativa*, tanto più essa è in grado di misurare la variazione di probabilità di fornire la risposta corretta anche per piccole variazioni di abilità del rispondente. Gli standard in letteratura indicano che l'indice di discriminazione deve raggiungere almeno il valore 0.20 e può considerarsi buono quando supera il valore di 0.25. La domanda in questione risulta avere una discriminazione ai limiti della discriminatività (0.18).

Nella prima colonna (Label) sono indicate le quattro alternative di risposta (1, 2, 3, 4), le mancate risposte (9) e le risposte non valide (7). Nella colonna 2 (Score) viene indicata l'opzione di risposta corretta con il codice 1.00, in questo caso la risposta corretta è l'opzione 1. Le colonne 3 (Count) e 4 (% of tot) riportano rispettivamente le frequenze assolute e percentuali delle risposte registrate. Questa informazione viene letta assieme ai dati della colonna 5 (Pt Bis) che riporta i valori della correlazione punto-biseriale (la correlazione tra la probabilità di scegliere una data opzione e l'abilità complessiva del rispondente). Tale correlazione deve essere negativa per le opzioni di risposta non corrette e positiva per quella esatta. Una domanda a scelta multipla è ben formulata se, in media, coloro che rispondono correttamente a quella domanda ottengono un punteggio elevato

nella prova complessivamente considerata¹¹. Nell'esempio presentato, la correlazione punto-biseriali dell'opzione corretta è +0.18.

Per le altre opzioni di risposta, la 2 e la 3 hanno correlazioni punto-biseriali decisamente negative, mentre la 4 ha una correzione punto-biseriali negativa vicina allo zero (-0.01).

I dati della colonna 6 consentono di valutare la significatività della correlazione punto-biseriali: come si può vedere nell'esempio, i valori delle opzioni di risposta 1, 2, 3 (colonna 1) registrati nella colonna 5 sono significativi in senso propriamente statistico¹², mentre il valore relativo all'opzione 4 registra una correlazione punto biseriali negativa ma troppo vicina allo 0, non significativa da un punto di vista statistico (la probabilità di $t(p) = (.914)$ è un valore molto superiore allo 0.05). Questo significa che l'opzione di risposta 4 viene scelta anche da soggetti con abilità non troppo basse.

Infine, la colonna 7 permette di valutare su una scala di *Rasch* il livello medio di prestazione dei rispondenti che scelgono una determinata risposta, mentre la colonna 8 fornisce una misura della variabilità di ciascuno dei risultati riportati in colonna 7. È importante notare che, in media, il livello di preparazione dei rispondenti che scelgono l'opzione corretta di risposta è più elevato di quello di coloro che scelgono le altre opzioni e ciò fornisce un'ulteriore indicazione positiva circa l'adeguatezza della domanda presa in esame.

26

Si studiano anche le Curve Caratteristiche per confermare il funzionamento degli item per intervenire laddove necessario. Al crescere del livello di abilità θ , cresce la probabilità di rispondere correttamente all'item; maggiore è il valore del parametro β_i maggiore deve essere l'abilità del soggetto per riuscire a rispondere correttamente all'item.

In riferimento al quesito A9, preso come esempio, anche l'elaborazione delle curve caratteristiche conferma la necessità di modificare il quarto distrattore che non segue l'andamento degli altri due attirando anche studenti con abilità non troppo basse.

¹¹ cfr. nota 9.

¹² cfr. nota 10.

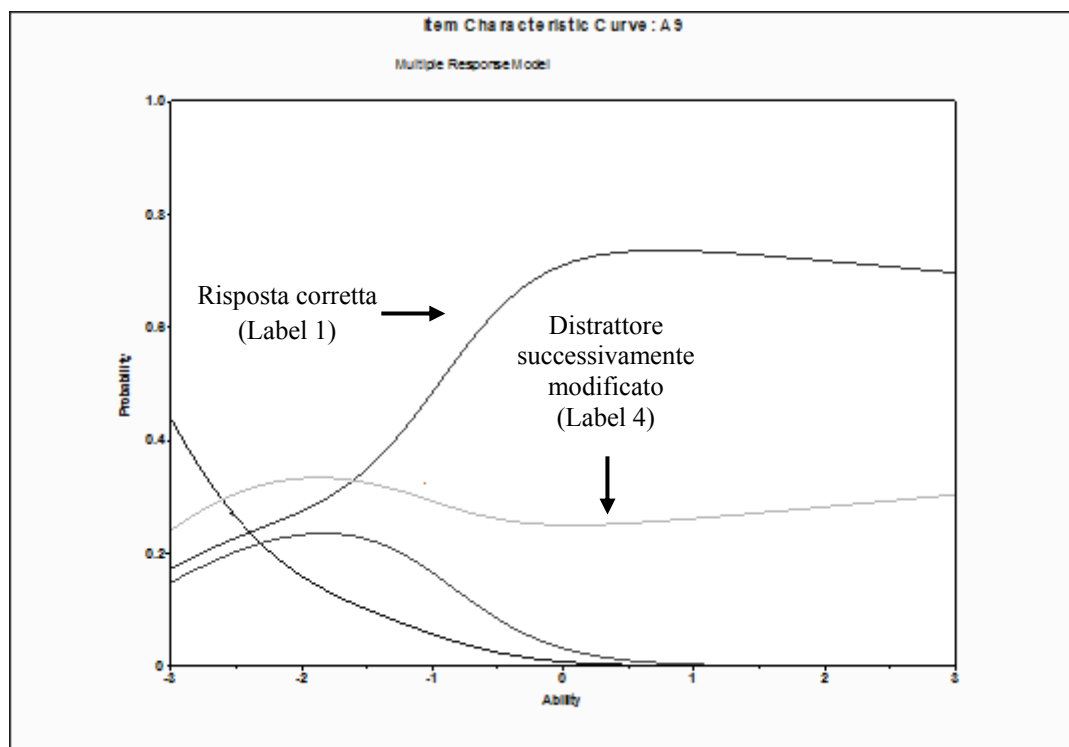


Figura 2.3: Curva caratteristica quesito A9

In questo caso, quindi, si procede alla modifica della opzione di risposta 4 come si può vedere nel Box 2.4.

Box 2.4: Quesito nella versione definitiva SNV 2011-2012

Testo: Le magie della bidella

A8. Qual è stata la reazione della mamma al racconto del figlio?

- A. Si è inquietata
- B. Si è divertita
- C. Si è offesa
- D. Si è tranquillizzata

Il quesito con le modifiche apportate ha fatto parte della versione definitiva del fascicolo¹³ come quesito A8. I dati elaborati per tale quesito mostrano che la correzione ha in effetti ottimizzato le sue caratteristiche strutturali come si osserva dai dati che seguono (Tabella 2.3).

Tabella 2.3: Analisi IRT
Tabella dati: quesito A8

```

Item 8
-----
item:8 (A8)
Cases for this item 30877 Discrimination 0.38
Item Threshold(s): -2.20 Weighted MNSQ 0.96
Item Delta(s): -2.19
    
```

Label	Score	Count	% of tot	Pt Bis	t (p)	PVIAvg:1	PV1 SD:1
1	1.00	26883	87.06	0.38	72.18 (.000)	0.12	0.85
2	0.00	341	1.10	-0.14	-24.23 (.000)	-1.05	0.86
3	0.00	1760	5.70	-0.18	-32.37 (.000)	-0.61	0.91
4	0.00	1715	5.55	-0.28	-51.37 (.000)	-0.94	0.86
7	0.00	17	0.06	-0.04	-6.44 (.000)	-1.06	1.04
9	0.00	161	0.52	-0.08	-14.96 (.000)	-0.93	0.92

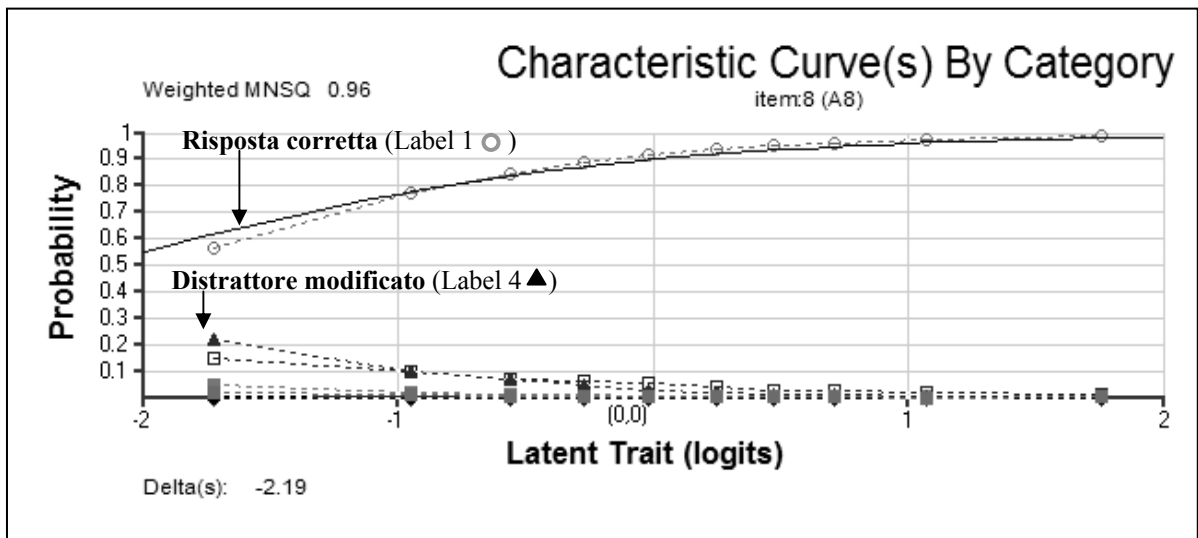


Figura 2.4: Curva caratteristica quesito A8

Le curve relative al quesito A8 (Figura 2.4) mostrano chiaramente che la modifica apportata all'opzione di risposta 4 ha funzionato; l'andamento della curva relativa al quarto distrattore si allinea a quella degli altri e la curva della opzione di risposta corretta si allinea a quella prevista dal modello.

¹³ http://www.invalsi.it/snv2012/documenti/Ril_apprendimenti/Italiano_QUINTA_primaria_SNV2012.pdf

Esempio di modifica del quesito

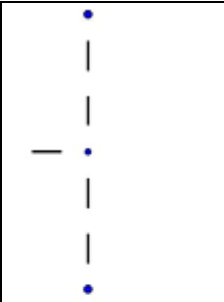
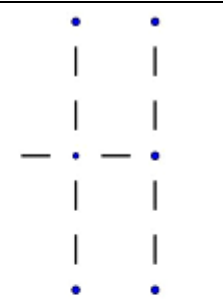
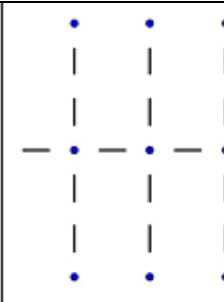

In generale il quesito di seguito riportato è risultato al pretest semplice (ad eccezione dell’item d.). Infatti, come si può osservare dalle curve caratteristiche del pretest, anche studenti con abilità molto basse rispondono correttamente, e in alcuni casi i distrattori non hanno svolto il ruolo per cui sono stati costruiti. Nello specifico l’item a. da item a scelta multipla è stato trasformato in item a risposta aperta univoca. Gli item b. c. e d. sono stati modificati rispetto all’oggetto della richiesta (nel pretest si chiedeva di ragionare sui punti mentre nella versione definitiva si è chiesto di ragionare sui segmenti) per cui l’item b. è stato modificato nella domanda e quindi nelle relative alternative di risposta, l’item c. e d. nella versione definitiva sono stati considerati come un unico item. Tutte le modifiche hanno portato alla costruzione di item più discriminativi in grado di cogliere in modo più puntuale le differenze tra le abilità degli studenti.

ESEMPIO 1: QUESITO DI V PRIMARIA – MATEMATICA

Box 2.5: Quesito nella versione pretest

Quesito A8

Osserva la sequenza:

			
Figura 1	Figura 2	Figura 3	Figura 4
Figura 5	Figura 6	Figura 7	Figura 8

Box 2.6: Item nella versione pre-test

Item a.

a. Immagina di continuare la sequenza. Da quanti segmenti sarà composta la figura 5?

- A. 25
- B. 30
- C. 20
- D. 35

Tabella 2.4: Analisi IRT

Tabella dati: item a.

```

Item 10
-----
item:10 (A8.a)
Cases for this item   395   Discrimination  0.39
Item Threshold(s):   -1.02   Weighted MNSQ  1.04
Item Delta(s):       -1.02
    
```

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	1.00	275	69.62	0.39	8.44 (.000)	0.23	0.93
2	0.00	19	4.81	-0.17	-3.52 (.000)	-0.63	0.74
3	0.00	62	15.70	-0.20	-3.96 (.000)	-0.49	0.77
4	0.00	21	5.32	-0.18	-3.73 (.000)	-0.55	0.80
7	0.00	2	0.51	0.00	0.10 (.924)	-0.14	0.63
9	0.00	16	4.05	-0.15	-3.10 (.002)	-0.79	0.95

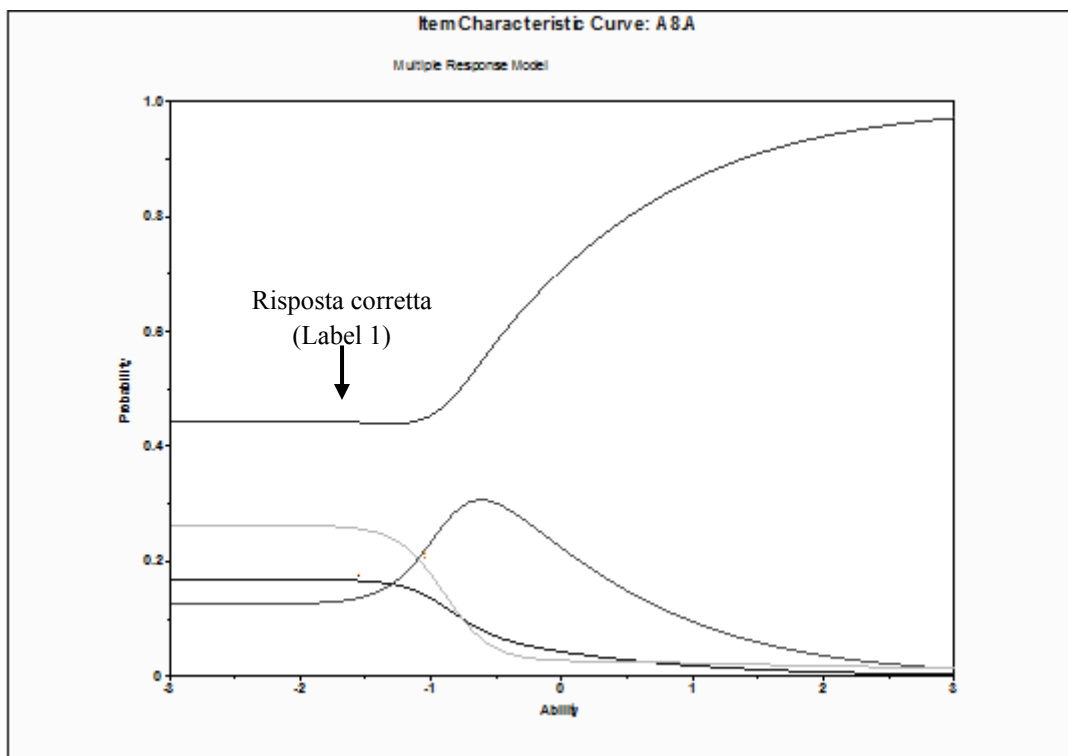


Figura 2.5: Curva caratteristica item a.

Box 2.7: Item nella versione pre-test

Item b.

b. Sempre immaginando di continuare la sequenza, quanti punti ci saranno nella figura 6?

A. 18

B. 15

C. 12

D. 20

Tabella 2.5: Analisi IRT
Tabella dati: item b

```

Item 11
-----
item:11 (A8.b)
Cases for this item   395   Discrimination   0.35
Item Threshold(s):   -0.53   Weighted MNSQ   1.06
Item Delta(s):       -0.53
    
```

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	1.00	238	60.25	0.35	7.44 (.000)	0.27	0.95
2	0.00	43	10.89	-0.07	-1.42 (.157)	-0.26	0.80
3	0.00	28	7.09	-0.16	-3.22 (.001)	-0.59	0.82
4	0.00	49	12.41	-0.22	-4.43 (.000)	-0.57	0.79
9	0.00	37	9.37	-0.13	-2.52 (.012)	-0.38	0.84

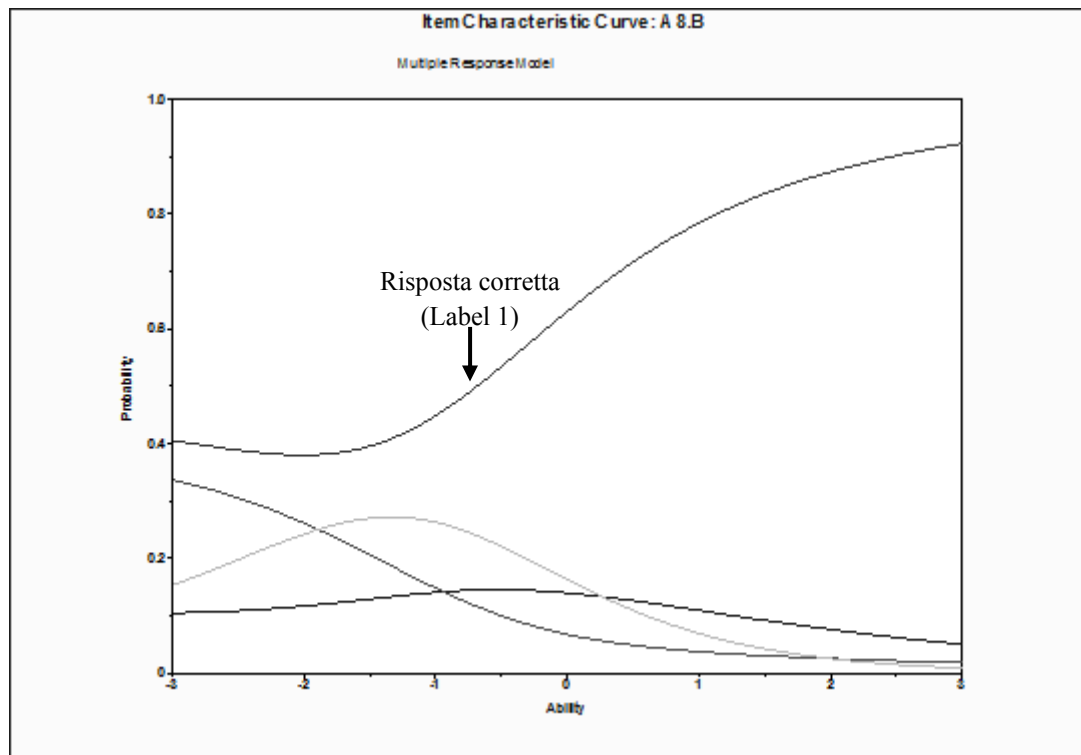


Figura 2.6: Curva caratteristica item b.

Box 2.8: Item nella versione pre-test

Item c.
c. È possibile costruire una figura con 32 punti?

A. Sì

B. No

Tabella 2.6: Analisi IRT
Tabella dati: item c.

```
Item 12
-----
item:12 (A8.c)
Cases for this item   395   Discrimination  0.33
Item Threshold(s):   -0.13   Weighted MNSQ  1.09
Item Delta(s):       -0.12
-----
```

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	153	38.73	-0.28	-5.83 (.000)	-0.30	0.83
2	1.00	205	51.90	0.33	6.85 (.000)	0.26	0.97
9	0.00	37	9.37	-0.09	-1.75 (.081)	-0.33	0.97

32

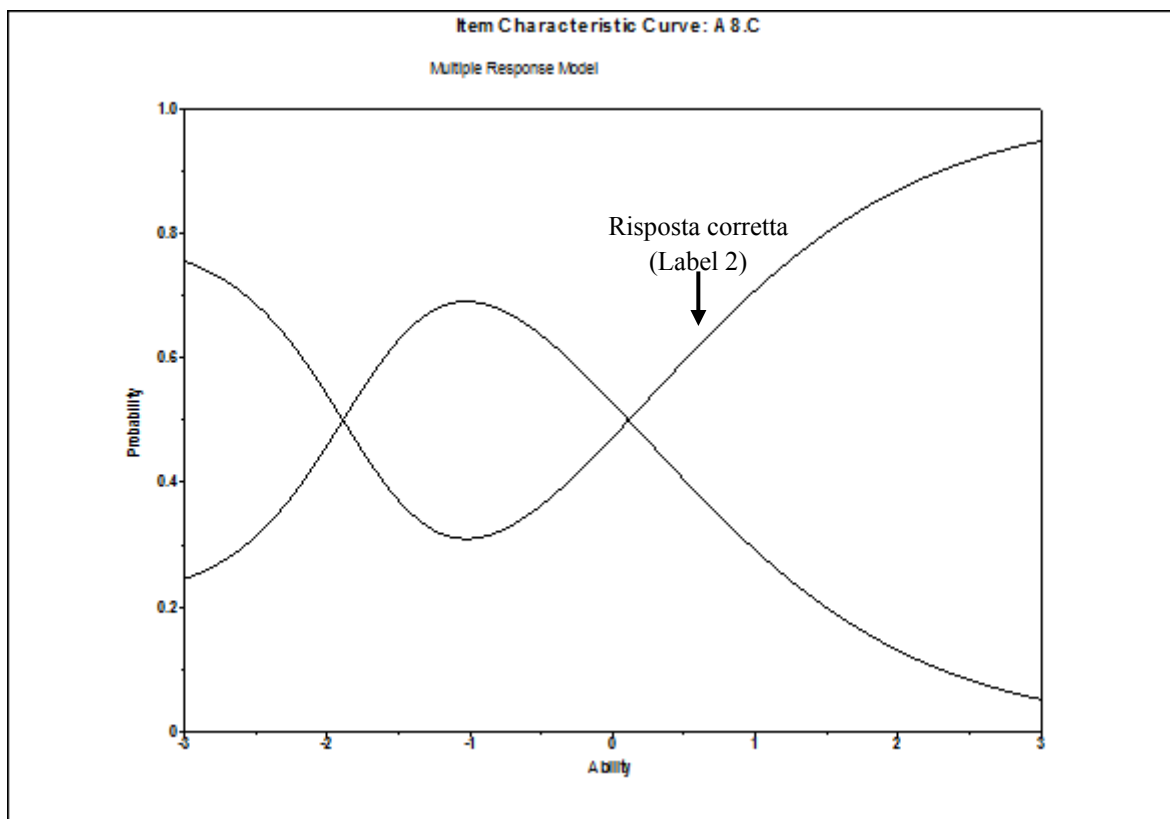


Figura 2.7: Curva caratteristica item c.

Box 2.9: Item nella versione pretest

Item d.

d. Giustifica la tua risposta?

.....

Tabella 2.7: Analisi IRT

Tabella dati: item d.

```

Item 13
-----
item:13 (A8.d)
Cases for this item      395  Discrimination  0.41
Item Threshold(s):      1.67  Weighted MNSQ  0.99
Item Delta(s):          1.67
    
```

Label	Score	Count	% of tot	Pt Bis	t	(p)	PVIAvg:1	PV1 SD:1
1	1.00	77	19.49	0.41	8.89	(.000)	0.73	0.87
2	0.00	229	57.97	-0.13	-2.59	(.010)	-0.10	0.89
7	0.00	2	0.51	0.06	1.18	(.238)	0.58	1.32
9	0.00	87	22.03	-0.25	-5.06	(.000)	-0.46	0.86

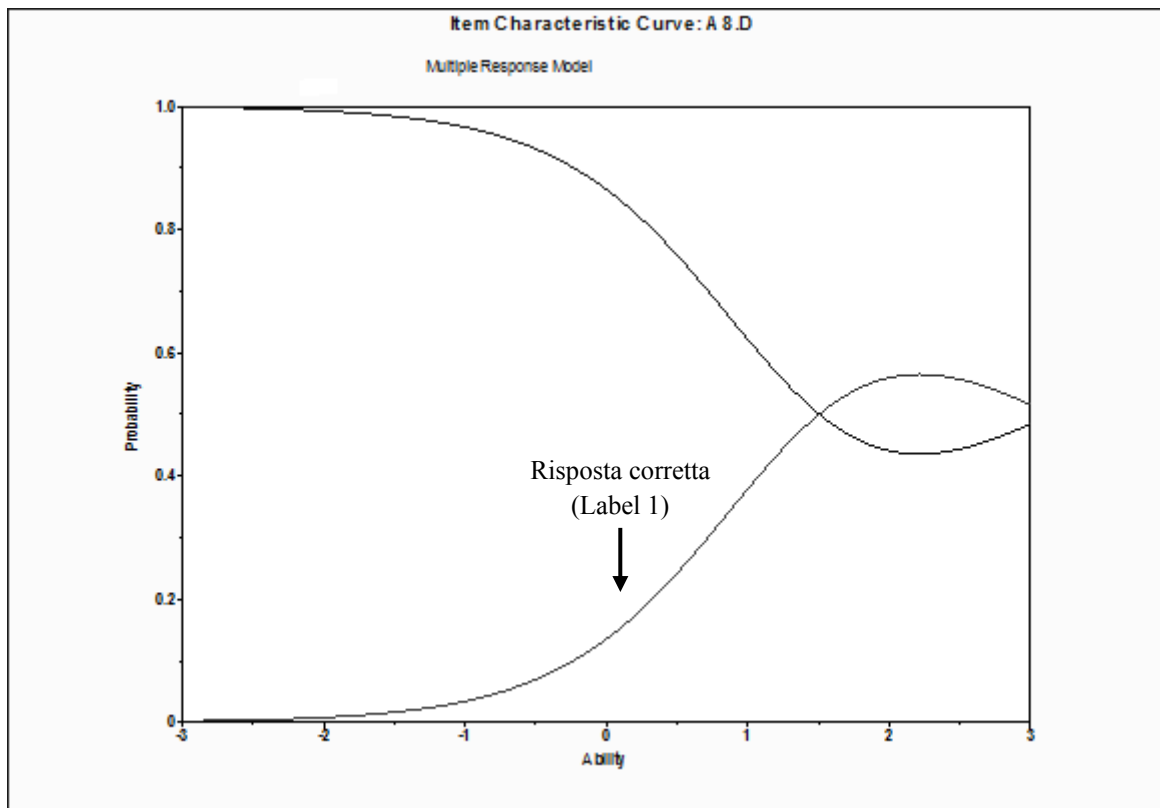
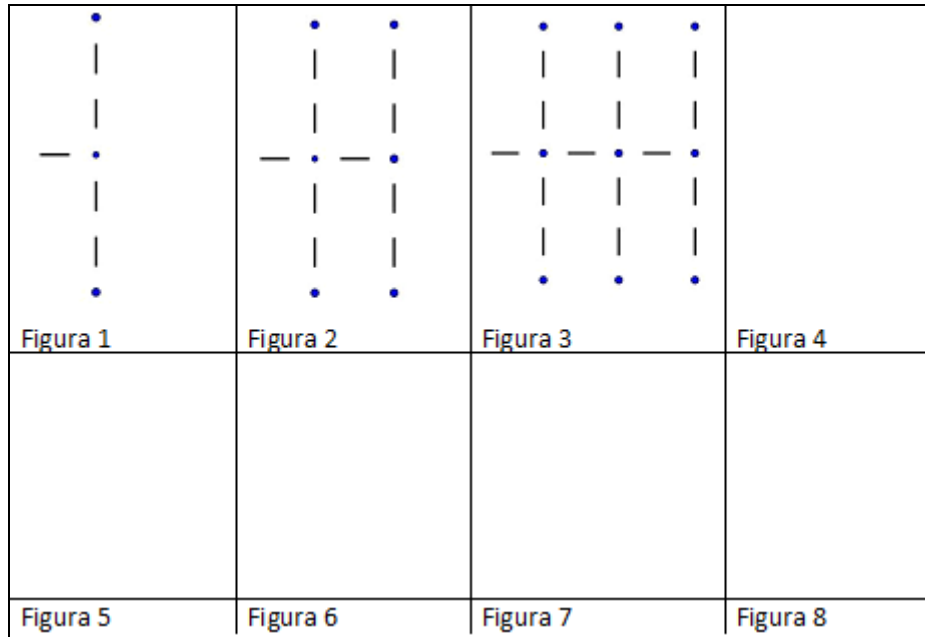


Figura 2.8: Curva caratteristica item d.

Box 2.10: Quesito nella versione definitiva SNV 2011-2012

Quesito D17

Osserva la sequenza:



Box 2.11: Item nella versione definitiva SNV 2011-2012

Item a.

a. Immagina di continuare la sequenza. Da quanti segmenti sarà composta la figura 5?

Risposta:

Tabella 2.8: Analisi IRT
Tabella dati: quesito d17a.

```

Item 21
-----
item:21 (D17_a)
Cases for this item 30869 Discrimination 0.50
Item Threshold(s): -0.31 Weighted MNSQ 0.96
Item Delta(s): -0.31
    
```

Label	Score	Count	% of tot	Pt Bis	t (p)	PVIAvg:1	PV1 SD:1
0	0.00	12526	40.58	-0.44	-86.38 (.000)	-0.55	0.92
1	1.00	17266	55.93	0.50	102.59 (.000)	0.47	1.00
7	0.00	46	0.15	-0.02	-3.70 (.000)	-0.60	0.89
9	0.00	1031	3.34	-0.18	-32.70 (.000)	-1.06	0.90

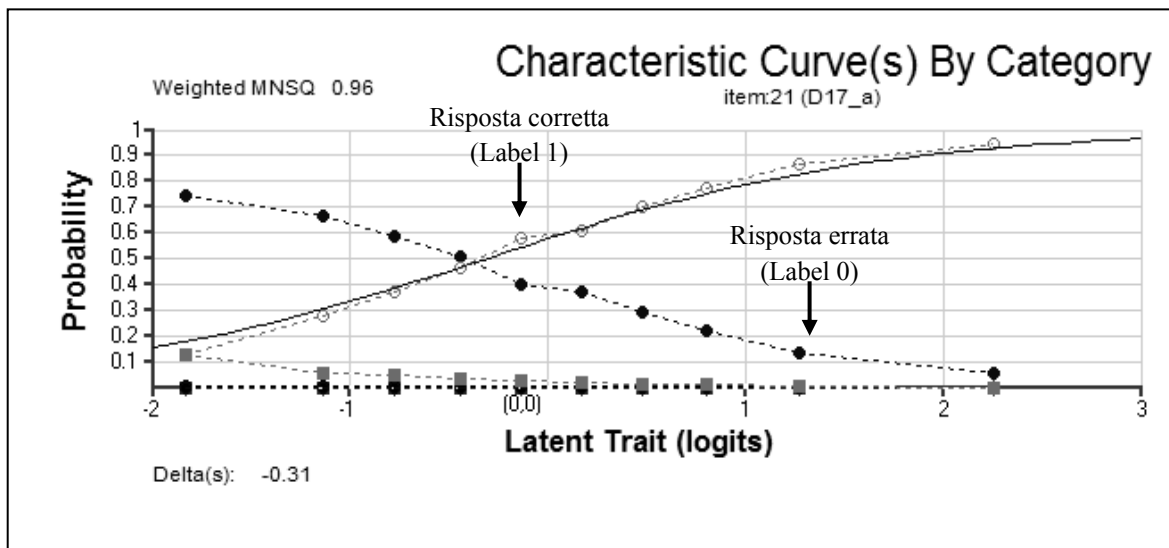


Figura 2.9: Curva caratteristica quesito D17a.

Box 2.12: Item nella versione definitiva SNV 2011-2012

Item b.

b. Sempre immaginando di continuare la sequenza, quale figura sarà formata da 40 segmenti?

- A. La figura 7
- B. La figura 8
- C. La figura 9
- D. La figura 10

**Tabella 2.9: Analisi IRT
Tabella dati: quesito d17b.**

```

Item 22
-----
item:22 (D17_b)
Cases for this item 30869 Discrimination 0.44
Item Threshold(s): -0.63 Weighted MNSQ 1.02
Item Delta(s): -0.62
    
```

Label	Score	Count	% of tot	Pt Bis	t (p)	PVIAvg:1	PV1 SD:1
1	0.00	2255	7.31	-0.17	-29.65 (.000)	-0.60	0.90
2	1.00	19157	62.06	0.44	85.85 (.000)	0.36	1.04
3	0.00	2021	6.55	-0.13	-22.24 (.000)	-0.50	0.94
4	0.00	6289	20.37	-0.29	-52.74 (.000)	-0.59	0.92
7	0.00	87	0.28	-0.02	-4.24 (.000)	-0.40	0.95
9	0.00	1060	3.43	-0.12	-20.98 (.000)	-0.67	0.96

36

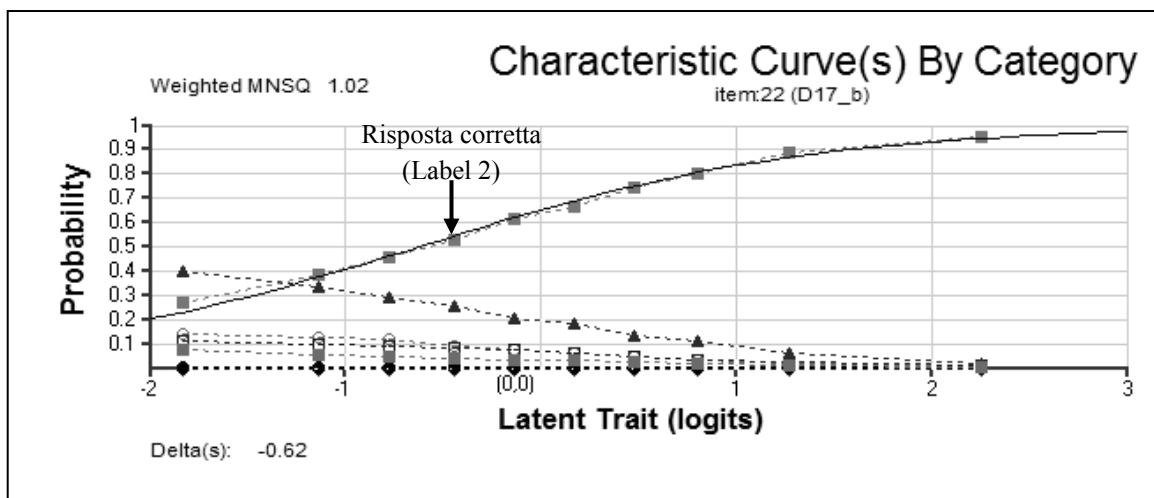


Figura 2.10: Curva caratteristica quesito D17b.

Box 2.13: – Item nella versione definitiva SNV 2011-2012

Item c.

**c. La sequenza potrebbe comprendere una figura con 32 segmenti?
Scegli una delle due risposte e completa la frase.**

Sì, perché

.....
.....

No, perché

.....
.....

**Tabella 2.10: Analisi IRT
Tabella dati: quesito d17c.**

```
Item 23
-----
item:23 (D17_c)
Cases for this item 30869 Discrimination 0.55
Item Threshold(s): 0.25 Weighted MNSQ 0.91
Item Delta(s): 0.25
-----
```

Label	Score	Count	% of tot	Pt Bis	τ (p)	PVIAvg:1	PV1 SD:1
0	0.00	13300	43.09	-0.39	-74.92 (.000)	-0.46	0.92
1	1.00	13830	44.80	0.55	115.11 (.000)	0.63	0.96
7	0.00	443	1.44	-0.05	-9.61 (.000)	-0.43	0.96
9	0.00	3296	10.68	-0.23	-41.99 (.000)	-0.70	0.92

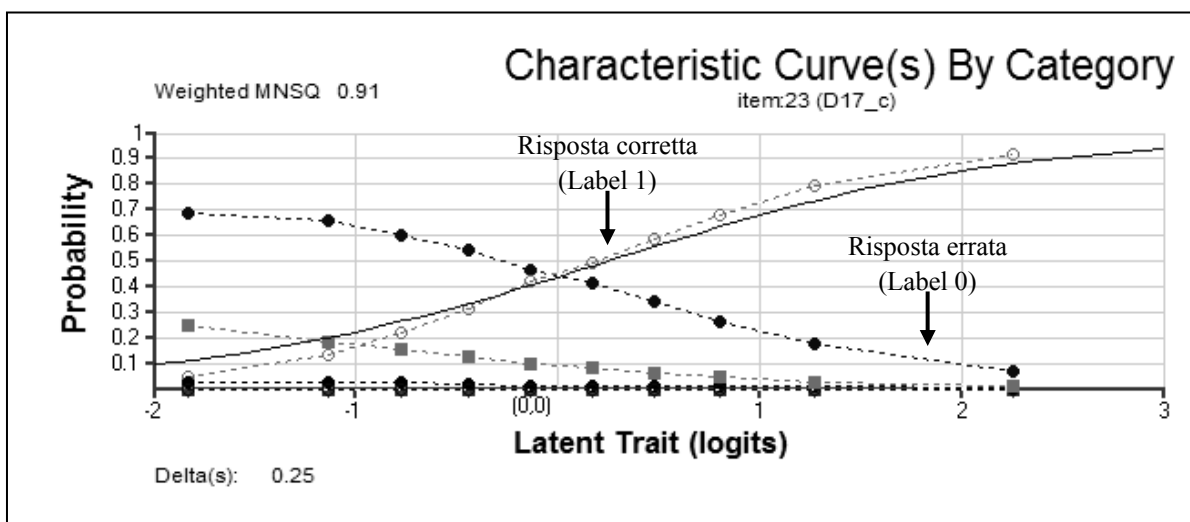


Figura 2.11: Curva caratteristica quesito D17c.

CAPITOLO 3 - La valutazione delle prove standardizzate del Servizio nazionale di valutazione

3.1 L'analisi degli item

La valutazione delle caratteristiche di una prova standardizzata può avvenire su diversi piani, uno di questi, ovvero quello concernente l'analisi dei contenuti coperti dalla prova stessa unita all'analisi dei processi cognitivi che ciascun item mira a rilevare, è stato descritto nel capitolo 3 del Rapporto nazionale.

Un secondo piano è invece relativo alle caratteristiche psicometriche degli item stessi e implica, pertanto, un insieme di analisi di tipo statistico sulle risposte degli studenti volte, in sostanza, a identificare eventuali item che non svolgono nel modo dovuto la loro funzione misuratoria.

Le informazioni di base da cui si parte per l'analisi del funzionamento di un item sono sostanzialmente tre:

- a) quanti alunni hanno risposto correttamente;
- b) quanti hanno scelto ciascuno dei possibili distrattori;
- c) quanti hanno omesso la risposta.

A partire da queste informazioni è possibile rispondere a tre domande:

- 1) quanto è difficile l'item?
- 2) L'item è in grado di discriminare fra gli alunni più abili e quelli meno abili?
- 3) Gli alunni hanno scelto tutte le opzioni? O ci sono alcune opzioni che non sono state scelte da nessuno?¹⁴

La valutazione della difficoltà di una prova è, *in primis*, effettuata basandosi sul giudizio di esperti e su premesse teoriche. Successivamente, dal punto di vista statistico e secondo la teoria classica dei test, viene valutata la percentuale di risposte corrette fornite da un numero adeguato di alunni, tale cioè da coprire una varietà di comportamenti sufficientemente ampia. L'elevato numero di risposte è, in altre parole, un elemento cruciale per ottenere informazioni tecnicamente fondate.

¹⁴ Auspicabilmente, l'opzione scelta il maggior numero di volte dovrebbe essere quella corretta, ma tutti i distrattori dovrebbero essere scelti un certo numero di volte; in caso contrario, essi non svolgono correttamente la loro funzione.

A partire da questi dati di base, la teoria dei test nella sua forma più avanzata che va sotto il nome di *Item Response Theory* e di *Rasch Analysis*¹⁵, permette di effettuare sofisticate e robuste valutazioni sulla difficoltà complessiva di una prova, sulla sua coerenza generale e sull'adeguatezza delle domande che la compongono.

Analogamente a quanto avviene in altre rilevazioni quantitative sugli apprendimenti, sia a livello nazionale sia a livello internazionale, l'INVALSI utilizza metodi e tecniche la cui validità trova riscontro in tutta la letteratura di settore. In particolare, per ogni parte delle prove SNV¹⁶ (Italiano o Matematica), sono state effettuate diverse analisi basate sui dati forniti da campioni di non meno di 1250 classi, come nel caso della Prova nazionale, fino a raggiungere le 2250 classi, come è avvenuto per la scuola secondaria di secondo grado.

Gli strumenti tecnico-statistici principali con i quali comunemente vengono valutate le caratteristiche misuratorie di una prova standardizzata, e che sono stati utilizzati anche per le prove SNV 2011-2012, sono:

1. valutazione generale di affidabilità della prova mediante l'indice α di *Cronbach*;
2. la *Rasch Analysis* per la stima della difficoltà delle domande e dell'abilità degli studenti;
3. *placement* relativo delle domande rispetto al punteggio di Rasch assegnato a ciascun allievo;
4. l'indice di discriminazione statistica (coefficiente r punto-biseriale).

3.2 L'indice α di Cronbach

Oltre alla validità di una prova e degli item che la compongono, occorre valutarne anche l'attendibilità¹⁷.

L'indice α di *Cronbach*, permette di esprimere una valutazione sintetica e generale sulla coerenza complessiva della prova e quindi sull'attendibilità delle informazioni che da essa si possono desumere. Più precisamente, l' α di *Cronbach* consente di comprendere se le domande che

¹⁵ Esula dallo scopo del presente lavoro qualsiasi approfondimento circa le differenze metodologiche e teoriche tra *Item Response Theory* e *Rasch Analysis*. A tal riguardo si veda, ad esempio, BARBARANELLI C., NATALI E. (2005), *I test psicologici: teorie e modelli psicometrici*, Carrocci Editore, Roma.

¹⁶ Inclusa la Prova nazionale.

¹⁷ La validità ha a che fare con la capacità dello strumento di misurare ciò che intende misurare; l'attendibilità riguarda, invece, quanto tale misurazione sia coerente e precisa. È importante ricordare che validità e attendibilità sono strettamente legate: uno strumento non può essere valido senza essere attendibile. Tuttavia, l'attendibilità di uno strumento non dipende dalla sua validità.

compongono la prova sono tra loro coerenti e volte alla misurazione dello stesso oggetto (Cronbach, 1951). Come è noto in letteratura, l' α di *Cronbach* non è sufficiente per poter esprimere un giudizio sicuro e fondato su questo aspetto, ma consente di effettuare un'analisi preliminare che deve essere ulteriormente approfondita con altri strumenti. A tal fine, la *Rasch Analysis* consente di effettuare gli approfondimenti necessari per valutare se e in che misura la prova proposta agli studenti sia caratterizzata da un grado di affidabilità sufficiente per ottenere misurazioni precise dei livelli di apprendimento raggiunti.

L' α di *Cronbach* è un indice espresso da un numero compreso fra 0 e 1 e tanto più si avvicina a 1, tanto maggiore è la coerenza complessiva della prova oggetto d'interesse. La letteratura psicometrica fornisce valori di riferimento che definiscono degli standard generalmente accettati per esprimere un primo giudizio orientativo sulla coerenza generale di una prova.¹⁸

In termini tecnici, l' α esprime la correlazione fra gli item che compongono la prova con il costrutto latente (il livello di apprendimento) oggetto di misurazione: maggiore sarà l'intercorrelazione fra gli item, maggiore sarà l'indice α . Tale indice, tuttavia, è anche influenzato dal numero complessivo di item: se il test è troppo breve, l' α ne risulterà ridotto.

Tavola 3.1: Soglie di α per prove su larga scala

Valori di α	Attendibilità della prova
0,80 o superiore	Molto buona
da 0,70 a 0,80	Buona
da 0,50 a 0,70	Modesta
inferiore a 0,50	Inadeguata

I valori soglia riportati nella tavola 3.1 hanno naturalmente una valenza orientativa e dipendono dalla tipologia di prova presa in esame. Se, come nel caso delle prove SNV, inclusa la Prova nazionale (PN), i rispondenti sono tutti gli allievi che frequentano una data classe, allora possono essere considerati adeguati valori di α di almeno 0,70.

La tabella 3.1 riporta i valori dell'indice α di *Cronbach* per i fascicoli di Italiano e Matematica delle prove SNV 2011-2012 e della Prova nazionale 2012.

¹⁸ Nunnally (1978), in un testo classico della psicometria, ha indicato 0,7 quale valore minimo accettabile per l' α di Cronbach, ma in letteratura è possibile trovare anche indicazioni di soglie più basse (ad es. Kline, 1999).

Tabella 3.1: Valori di α delle prove SNV-PN

Livello	Italiano	Matematica
II primaria	0,82	0,85
V primaria	0,86	0,91
I sec. di I gr.	0,88	0,89
III sec. di I gr.	0,87	0,84
II sec. di II gr.	0,91	0,89

Come si può vedere dalla tabella 3.1, per tutti i livelli scolastici coinvolti l'indice α di *Cronbach* supera nettamente la soglia di 0,70 sia in Italiano sia in Matematica.

Gli elevati valori di α consentono di ritenere che le due prove (Italiano e Matematica) sono internamente coerenti dal punto di vista psicometrico, ovvero sono costruite in modo che l'oggetto di misurazione abbia una sua unitarietà prevalente. Si tratta, quindi, di un primo indicatore di affidabilità delle misure ottenute tramite i fascicoli, intesa come ottima consistenza interna della prova di Italiano e della prova di Matematica; in altre parole, la prova, complessivamente intesa, non sembra presentare problemi di incoerenza o di ambiguità macroscopiche tali da mettere in difficoltà i rispondenti, anche se, come sarà mostrato in seguito, sono necessarie analisi ulteriori per giungere a un giudizio maggiormente fondato.

3.3 La Rasch Analysis per la stima della difficoltà delle domande e delle “abilità” in italiano e in matematica degli studenti

Dopo una prima valutazione sintetica sull'attendibilità generale delle prove SNV e della Prova nazionale, l'analisi è stata approfondita applicando i metodi della cosiddetta *Rasch Analysis* (RA).

Il modello di Rasch è riconducibile dal punto di vista operativo alla famiglia dei modelli afferenti all'*Item Response Theory* (IRT). Esso consente di ottenere una scala a intervalli sulla quale vengono collocati gli item in base alla loro difficoltà¹⁹; sulla medesima scala, successivamente, vengono

¹⁹ Quando chi risponde ha un livello di abilità pari a quello della difficoltà dell'item, allora c'è il 50% di probabilità che quel soggetto risponda correttamente all'item. Nel caso in cui il livello di abilità sia inferiore al livello di difficoltà dell'item, allora tale probabilità è minore del 50%; viceversa, nel caso in cui il livello di abilità del soggetto sia superiore al livello di difficoltà dell'item, la probabilità è superiore al 50%.

posizionati gli alunni, o meglio, il livello della loro *performance*²⁰, in funzione delle risposte fornite agli item stessi²¹.

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 alla difficoltà di una determinata domanda o di un insieme di quesiti, e di conseguenza comprendere ciò che questi allievi sono in grado di fare o che conoscono. Risulta quindi evidente l'enorme potenzialità di questa metodologia anche dal punto di vista didattico e per la promozione di azioni volte al miglioramento che possono essere realizzate dalle singole scuole.

Nei modelli IRT è possibile rappresentare ciascun item tramite una "curva caratteristica" che esprime l'idea della difficoltà intesa come un indice di posizione.

La Figura II.1 mostra la curva caratteristica di tre item caratterizzati dal medesimo potere di discriminazione, ma con differenti livelli di difficoltà. La curva a sinistra rappresenta un item facile in quanto la probabilità di rispondere correttamente - $P(\Theta)$ - è piuttosto elevata anche per allievi con un basso livello di abilità; la curva centrale, invece, rappresenta un item mediamente difficile poiché la probabilità di dare la risposta corretta è bassa ai livelli inferiori di abilità dei rispondenti, intorno a 0,50 al centro della scala di abilità e vicino a 1 ai livelli più elevati di abilità. Infine, la curva a destra illustra un item difficile: la probabilità di fornire la risposta esatta è bassa per quasi tutti i livelli della scala di abilità e aumenta solo in corrispondenza di livelli molto alti di abilità, non superando comunque 0,80 perfino al livello di abilità maggiore fra quelli illustrati nel grafico (+3)²².

²⁰ Poiché gran parte delle ricerche in questo ambito hanno riguardato variabili cognitive quali abilità matematiche, scientifiche, di lettura, nell'IRT viene utilizzato il termine generico "abilità" per riferirsi al tratto latente oggetto della misurazione.

²¹ In altre parole, sulla base della difficoltà degli item che compongono una prova e delle risposte date dagli studenti all'insieme di tali item, si ottengono misure quantitative, riferite alle singole persone, del costrutto latente che si intende misurare. Ci si basa, cioè, sull'assunzione che ciascun alunno possieda una certa quantità di abilità sottostante, espressa da un valore numerico, ovvero un punteggio, che lo posiziona sulla scala di abilità.

²² Baker, F.B. (2001), *The Basics of Item Response Theory*; Baker, F.B., Kim S., (2004), *Item Response Theory. Parameter Estimation Techniques*.

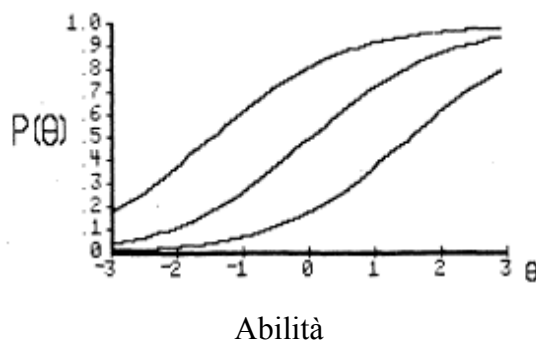


Figura 3.1: Curva caratteristica di tre item con lo stesso potere discriminante ma con difficoltà differente

Le tavole del presente paragrafo riportano i cosiddetti *parametri di difficoltà* di ciascuna domanda delle prove (fascicolo d'Italiano e di Matematica) SNV²³ per la classe II della scuola primaria. Tali parametri sono definiti su una scala convenzionale centrata sul valore 0 e compresa tra i valori -4 e +4. Ciò significa che domande con parametri di difficoltà inferiori a 0 sono generalmente facili e sono tanto più facili quanto più tali parametri assumono valori vicino a -4, mentre quelle con parametri positivi sono generalmente difficili, diventando più difficili quanto più ci si avvicina a +4.

3.4 L'indice di discriminazione statistica

43

Un ulteriore elemento fondamentale per esprimere un giudizio complessivo adeguatamente approfondito sulle prove e su ciascuna domanda che le compone è l'indice di discriminazione ovvero il **coefficiente r punto-biseriale**²⁴. È necessario, infatti, valutare se le diverse domande contengano o meno delle ambiguità che possano aver tratto in inganno i rispondenti²⁵: in una domanda formulata in modo adeguato e priva di ambiguità, solo l'opzione corretta si deve associare positivamente con il risultato complessivo della prova, nel senso che coloro che rispondono correttamente a una data domanda devono, in media, ottenere un risultato complessivo nella prova migliore di coloro che rispondono in modo errato o non rispondono affatto.

In termini più tecnici, il quesito deve essere in grado di *discriminare* positivamente fra allievi più abili e allievi meno abili. L'indice di discriminazione consente di valutare in termini quantitativi la

²³ I risultati della prova preliminare di lettura della II primaria sarà oggetto, insieme ai risultati del questionario studente somministrato in V primaria e nella I secondaria di primo grado, di un apposito approfondimento che l'INVALSI pubblicherà nel corso dell'anno scolastico 2012-13.

²⁴ Cfr. anche Capitolo I.

²⁵ Per ogni domanda è stata valutata sia in sede di pre-test sia *ex post*, ovvero sui risultati della rilevazione vera e propria, la capacità misuratoria di ciascuna opzione (cfr. Capitolo I).

capacità di una domanda di individuare gruppi di rispondenti in funzione del loro livello di preparazione. In altri termini, tanto più una domanda è *discriminante*, tanto più essa è in grado di misurare la variazione di probabilità di fornire la risposta corretta anche per piccole variazioni di abilità del rispondente. Gli standard di letteratura indicano che l'indice di discriminazione deve raggiungere almeno il valore 0,20 e può considerarsi buono quando supera il valore di 0,25²⁶.

Osservando le tabelle 4.1-4.10 si può notare che, in generale, i quesiti delle prove relative ai vari livelli di scolarità mostrano una correlazione punto-biseriala positiva con l'esito complessivo della prova medesima; solo in un numero limitato di casi il coefficiente di correlazione punto-biseriala assume valori inferiori a 0,20 e ciò consente di eliminare, su una base empirica, la preoccupazione che siano presenti ambiguità o passaggi poco chiari nelle domande che costituiscono le prove SNV.

²⁶ Si veda OECD, *Pisa 2006 Technical Report*, Parigi, OECD, 2009, pag. 147.

Capitolo 4 - I dati sulle prove per i vari livelli coinvolti

Di seguito sono presentati i dati relativi alle caratteristiche degli item componenti le prove di Italiano e Matematica dei livelli II e V primaria, I e III secondaria di primo grado, II secondaria di secondo grado.

4.1 La scuola primaria

Le tabelle 4.1 e 4.2 riportano i parametri di difficoltà e i coefficienti di correlazione punto-biseriale delle prove di Italiano e Matematica della classe II primaria; le figure 4.1 e 4.2 illustrano l'informazione veicolata dall'intera prova di Italiano e Matematica.

Tabella 4.1: Parametri di difficoltà delle domande del fascicolo di Italiano (II primaria) e loro afferenza a parti del testo e a processi

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto-biseriale
1	A1	Narrativo	Ricostruire il significato del testo	-0,83	0,40
2	A2	Narrativo	Ricostruire il significato del testo	-0,80	0,42
3	A3	Narrativo	Ricostruire il significato del testo	-1,32	0,44
4	A4	Narrativo	Ricostruire il significato del testo	-1,54	0,26
5	A5	Narrativo	Ricostruire il significato del testo	-1,15	0,49
6	B1	Narrativo	Individuare informazioni	-2,11	0,43
7	B2	Narrativo	Ricostruire il significato del testo	-0,47	0,41
8	B3	Narrativo	Individuare informazioni	-0,75	0,44
9	B4	Narrativo	Individuare informazioni	-0,06	0,31
10	B5	Narrativo	Ricostruire il significato del testo	0,05	0,45
11	B6	Narrativo	Ricostruire il significato del testo	1,15	0,20
12	B7	Narrativo	Ricostruire il significato del testo	-0,20	0,43
13	B8	Narrativo	Individuare informazioni	-0,90	0,49
14	B9	Narrativo	Ricostruire il significato del testo	-0,53	0,43
15	B10	Narrativo	Ricostruire il significato del testo	-0,92	0,43
16	B11	Narrativo	Ricostruire il significato del testo	-0,41	0,55
17	B12	Narrativo	Ricostruire il significato del testo	-0,30	0,51
18	B13	Narrativo	Ricostruire il significato del testo	0,84	0,32
19	B14	Narrativo	Ricostruire il significato del testo	-0,92	0,46
20	C1	Riordino frasi	Riflessione sulla lingua	-2,17	0,45
21	C2	Riordino frasi	Riflessione sulla lingua	-0,63	0,50
22	C3	Riordino frasi	Riflessione sulla lingua	-2,47	0,46
23	C4	Riordino frasi	Riflessione sulla lingua	-0,15	0,53
24	C5	Riordino frasi	Riflessione sulla lingua	-0,21	0,43
25	C6	Riordino frasi	Riflessione sulla lingua	0,27	0,51

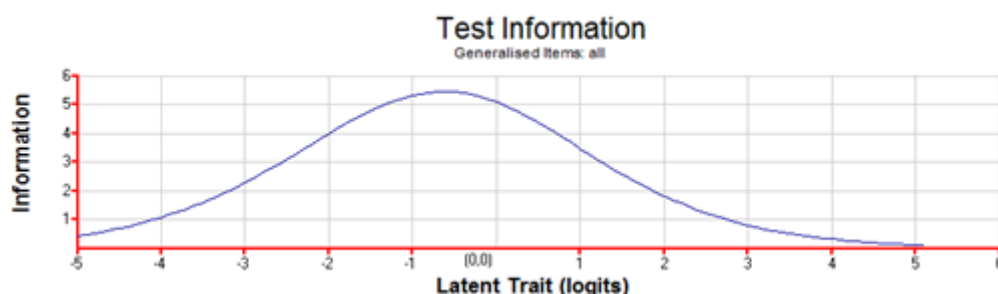


Figura 4.1: Informazione totale sulla prova di Italiano – II primaria

Tabella 4.2: Parametri di difficoltà delle domande del fascicolo di Matematica (II primaria) e loro afferenza ad ambiti e a macro processi

N.	Item	Ambiti	Macro Processi	Difficoltà	Correlazione punto-biseriale
1	D1	Spazio e figure	Rappresentazioni	-1,88	0,30
2	D2	Numeri	Argomentazione	-0,57	0,36
3	D3_a	Numeri	Modellizzazione	0,12	0,52
4	D3_b	Numeri	Modellizzazione	0,28	0,54
5	D4_a	Numeri	Rappresentazioni	-0,65	0,43
6	D4_b	Numeri	Rappresentazioni	-0,43	0,46
7	D5_a	Dati e previsioni	Modellizzazione	-2,06	0,30
8	D5_b	Dati e previsioni	Modellizzazione	-2,58	0,34
9	D5_c	Dati e previsioni	Modellizzazione	-1,57	0,49
10	D5_d	Dati e previsioni	Modellizzazione	-2	0,42
11	D6	Numeri	Concetti e procedure	0,7	0,45
12	D7_a	Spazio e figure	Rappresentazioni	0,02	0,32
13	D7_b	Spazio e figure	Rappresentazioni	-0,38	0,38
14	D8	Spazio e figure	Rappresentazioni	-0,25	0,37
15	D9_a	Dati e previsioni	Modellizzazione	-1,39	0,49
16	D9_b	Dati e previsioni	Modellizzazione	-0,07	0,47
17	D9_c	Dati e previsioni	Modellizzazione	-0,35	0,58
18	D10	Numeri	Concetti e procedure	0,75	0,46
19	D11	Numeri	Modellizzazione	0,03	0,09
20	D12	Numeri	Concetti e procedure	-0,01	0,26
21	D13	Spazio e figure	Rappresentazioni	-0,36	0,38
22	D14	Numeri	Concetti e procedure	0,42	0,45
23	D15	Numeri	Modellizzazione	0,3	0,30
24	D16	Numeri	Modellizzazione	0,37	0,52
25	D17	Numeri	Concetti e procedure	-0,33	0,44
26	D18_a	Spazio e figure	Modellizzazione	-0,94	0,25
27	D18_b	Spazio e figure	Modellizzazione	-0,14	0,35

N.	Item	Ambiti	Macro Processi	Difficoltà	Correlazione punto-biseriale
28	D19	Numeri	Modellizzazione	-0,06	0,46
29	D20_a	Numeri	Concetti e procedure	-1,06	0,57
30	D20_b	Numeri	Concetti e procedure	-0,22	0,61
31	D20_c	Numeri	Concetti e procedure	-0,48	0,61
32	D20_d	Numeri	Concetti e procedure	-0,25	0,63

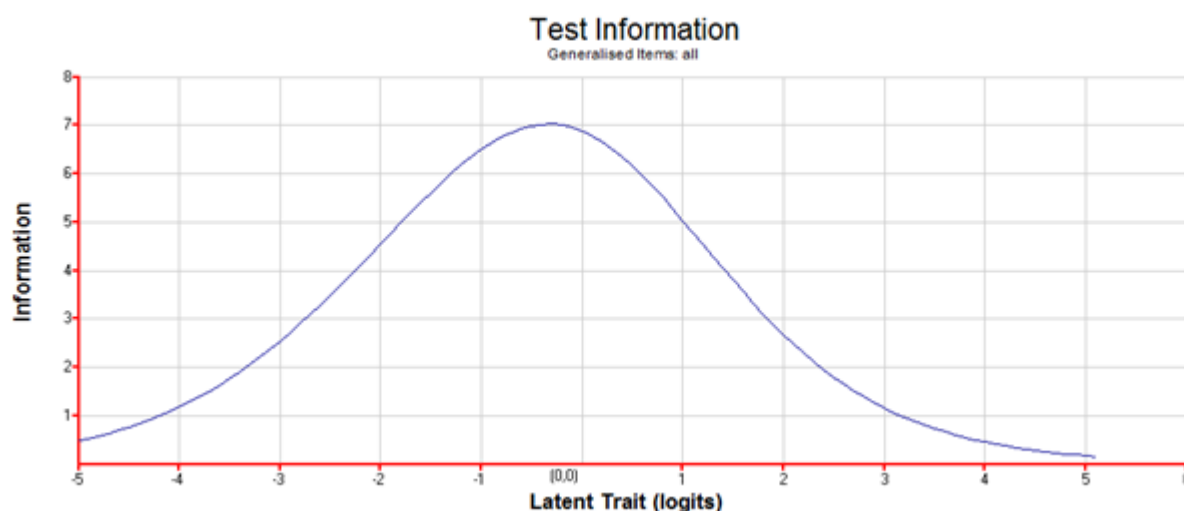


Figura 4.2: Informazione totale sulla prova di Matematica – II primaria

La lettura dei dati della tabella 4.1 mostra come la prova di Italiano della seconda primaria possa considerarsi, in base alle risposte effettivamente espresse dagli allievi, mediamente facile, confermando quindi tutte le ipotesi emerse durante la fase di pre-test delle prove medesime. L’item più difficile (B6), infatti, risulta avere un valore di difficoltà pari a 1,15, ben lontano dal valore massimo di 4; soltanto 4 item risultano avere un parametro di difficoltà superiore a 0, mentre tutti gli altri (21 item) riportano un valore inferiore a 0; occorre anche notare che le diverse parti del testo e i diversi processi sono presenti lungo tutta la scala di difficoltà della prova. La difficoltà contenuta delle prove della seconda primaria è coerente con le scelte effettuate dall’INVALSI in sede di progettazione degli strumenti di misurazione per bambini di 7 anni al termine del loro secondo anno di scuola. È importante, come indicato dalla letteratura del settore²⁷, che bambini così piccoli affrontino prove standardizzate che non creino loro senso di frustrazione per l’eccesso di

²⁷ Hambleton, R.K., Swaminathan, H., Rogers, H.J., *Fundamentals of Item Response Theory*, Amherst, University of Massachusetts, SAGE Publications, 1991.

difficoltà, ma che, al contempo, contengano domande sufficientemente stimolanti per evitare fenomeni di disinteresse e di calo di attenzione durante lo svolgimento della prova stessa. Tuttavia, è al contempo opportuno inserire nelle prove proposte dall'INVALSI anche domande in grado di rilevare livelli di abilità più alti. Naturalmente ciò deve avvenire nel pieno rispetto della peculiarità delle prove proposte a bambini di soli 7 anni, senza creare quindi nei rispondenti la sensazione di richieste troppo complesse che possano risultare scoraggianti e influire negativamente sulla *performance*.

Occorre ricordare che i parametri di difficoltà degli item sono calcolati sulla base delle reali risposte fornite da diverse centinaia di migliaia di allievi, ossia tutti i bambini che frequentano la seconda classe della scuola primaria; pertanto, i dati mostrano chiaramente che le prove di Italiano sono abbastanza agevolmente risolte da bambini di 7 anni. Accade frequentemente che le capacità di comprensione della lettura degli alunni venga sottostimata, con la conseguenza che non vengono proposti loro testi ritenuti complessi, ma che in realtà non lo sono realmente per i bambini stessi. Anche in questo senso, le prove standardizzate proposte a tutti gli alunni della classe seconda primaria sono molto utili poiché consentono di fornire una misura solida e comparativamente informativa sulla reale comprensione della lettura dei bambini di 7 anni che frequentano le scuole italiane, statali e paritarie.

Allo stesso modo, la prova di Matematica (Tabella 4.2) risulta essere di difficoltà medio-bassa; 9 dei 32 item che compongono la prova risultano avere un parametro di difficoltà superiore allo 0, mentre i restanti 23 item hanno un valore di difficoltà al di sotto dello 0. La prova di Matematica, pertanto, risulta maggiormente in grado di rappresentare anche i livelli più alti della scala di abilità dei rispondenti rispetto alla prova di Italiano; tuttavia, per la Matematica come per l'Italiano, sarà necessario prevedere per il futuro un maggior numero di quesiti in grado di permettere una stima più accurata dei livelli elevati di abilità dei rispondenti. In questa direzione occorre considerare anche il progetto *Misurazione dei progressi e degli apprendimenti nelle scuole* (<http://www.invalsi.it/invalsi/ri/sis/misurazione.php>), finalizzato allo studio diacronico sui dati relativi agli studenti italiani e alla messa a punto di un sistema di prove ancorate fra loro che consenta la comparabilità diretta delle rilevazioni effettuate nei diversi anni scolastici. Inoltre, la prova di Matematica pare meglio articolata in termini di distribuzione delle difficoltà delle domande, ma è necessario rimarcare la maggiore semplicità di realizzare ciò con i quesiti di matematica anziché con quelli di italiano. Infatti, questi ultimi sono legati a un testo e quindi i margini di libertà nella scelta delle domande sono decisamente minori. Tuttavia, proprio in questa direzione si concretizza

una sfida per l'INVALSI, ossia quella di trovare testi adatti a bambini di 7 anni, ma sufficientemente ricchi e densi per formulare domande di diversi livelli di difficoltà, in grado, quindi, di stimare accuratamente tutti i livelli di abilità dei rispondenti, da quelli più modesti a quelli più elevati.

Le tabelle 4.3 e 4.4 riportano i parametri di difficoltà e i coefficienti di correlazione punto-biseriali delle prove di Italiano e Matematica della classe V primaria; le figure 4.3 e 4.4 illustrano l'informazione che le prove di Italiano e di Matematica forniscono sulla distribuzione di abilità.

Tabella 4.3: Parametri di difficoltà delle domande del fascicolo di Italiano (V primaria) e loro afferenza a parti del testo e a processi

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto-biseriali
1	A1	narrativo	Ricostruire il significato del testo	-1,29	0,35
2	A2	narrativo	Ricostruire il significato del testo	-2,10	0,47
3	A3	narrativo	Ricostruire il significato del testo	1,58	0,2
4	A4	narrativo	Ricostruire il significato del testo	-1,34	0,29
5	A5	narrativo	Ricostruire il significato del testo	-1,19	0,22
6	A6	narrativo	Ricostruire il significato del testo	-1,14	0,28
7	A7	narrativo	Ricostruire il significato del testo	-1,16	0,28
8	A8	narrativo	Ricostruire il significato del testo	-2,19	0,38
9	A9	narrativo	Ricostruire il significato del testo	-2,48	0,38
10	A10	narrativo	Ricostruire il significato del testo	-1,91	0,27
11	A11	narrativo	Ricostruire il significato del testo	-1,64	0,45
12	A12	narrativo	Ricostruire il significato del testo	-0,78	0,44
13	A13	narrativo	Ricostruire il significato del testo	-0,66	0,41
14	A14	narrativo	Ricostruire il significato del testo	-1,45	0,33
15	A15	narrativo	Ricostruire il significato del testo	-1,90	0,31
16	A16	narrativo	Ricostruire il significato del testo	-0,26	0,42
17	A17	narrativo	Ricostruire il significato del testo	1,09	0,18
18	A18	narrativo	Ricostruire il significato del testo	-1,76	0,41
19	A19	narrativo	Ricostruire il significato del testo	-0,48	0,31
20	A20	narrativo	Interpretare e valutare	0,28	0,35
21	B1	espositivo	Individuare informazioni	-1,93	0,33
22	B2	espositivo	Individuare informazioni	-1,38	0,51
23	B3	espositivo	Individuare informazioni	-0,34	0,48
24	B4	espositivo	Individuare informazioni	0,78	0,26
25	B5	espositivo	Ricostruire il significato del testo	0,83	0,4
26	B6	espositivo	Individuare informazioni	-1,62	0,38
27	B7	espositivo	Ricostruire il significato del testo	-1,42	0,4
28	B8	espositivo	Individuare informazioni	0,47	0,32

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto-biserial
29	B9	espositivo	Individuare informazioni	-0,73	0,56
30	B10	espositivo	Ricostruire il significato del testo	-0,60	0,44
31	B11	espositivo	Ricostruire il significato del testo	-0,63	0,41
32	B12	espositivo	Ricostruire il significato del testo	-1,59	0,4
33	C1_a	grammatica	Riflessione sulla lingua	0,16	0,42
34	C1_b	grammatica	Riflessione sulla lingua	-0,02	0,46
35	C2	grammatica	Riflessione sulla lingua	-1,62	0,42
36	C3	grammatica	Riflessione sulla lingua	-1,49	0,47
37	C4	grammatica	Riflessione sulla lingua	-0,23	0,43
38	C5	grammatica	Riflessione sulla lingua	-0,25	0,49
39	C6	grammatica	Riflessione sulla lingua	-1,94	0,42
40	C7	grammatica	Riflessione sulla lingua	-2,04	0,34
41	C8	grammatica	Riflessione sulla lingua	0,21	0,51
42	C9	grammatica	Riflessione sulla lingua	-0,98	0,49
43	C10	grammatica	Riflessione sulla lingua	-2,82	0,37

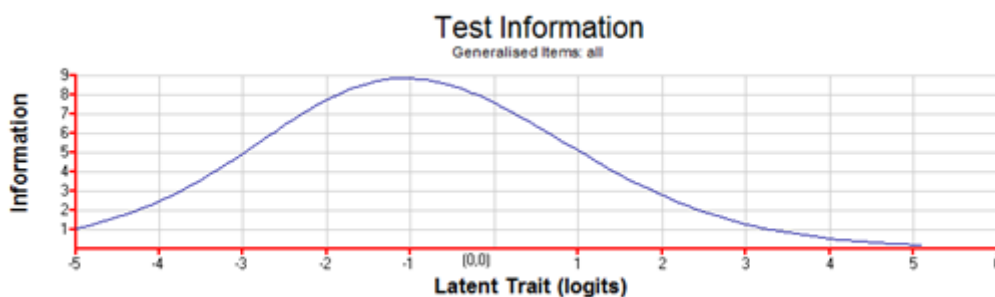


Figura 4.3: Informazione totale sulla prova di Italiano – V primaria

Tabella 4.4: Parametri di difficoltà delle domande del fascicolo di Matematica (V primaria) e loro afferenza ad ambiti e a macro processi

N.	Item	Ambiti	Macro Processi	Difficoltà	Correlazione punto-biseriale
1	D1	Numeri	Concetti e procedure	-1,52	0,42
2	D2	Dati e previsioni	Modellizzazione	0,89	0,39
3	D3	Numeri	Rappresentazioni	-1	0,37
4	D4	Relazioni e funzioni	Modellizzazione	0,59	0,42
5	D5_a	Spazio e figure	Concetti e procedure	-0,36	0,48
6	D5_b	Spazio e figure	Argomentazione	0,34	0,49
7	D6	Dati e previsioni	Modellizzazione	-0,26	0,37
8	D7_a	Numeri	Concetti e procedure	0,34	0,45
9	D7_b	Numeri	Concetti e procedure	-0,01	0,56
10	D8_a	Relazioni e funzioni	Modellizzazione	1,62	0,46
11	D8_b	Relazioni e funzioni	Modellizzazione	1,88	0,46
12	D9	Dati e previsioni	Modellizzazione	0,53	0,46
13	D10	Spazio e figure	Rappresentazioni	-0,06	0,45
14	D11	Numeri	Concetti e procedure	0,44	0,55
15	D12_a	Dati e previsioni	Modellizzazione	-0,97	0,43
16	D12_b	Dati e previsioni	Concetti e procedure	0,54	0,54
17	D13	Relazioni e funzioni	Concetti e procedure	-0,19	0,49
18	D14	Spazio e figure	Modellizzazione	-0,89	0,35
19	D15	Numeri	Concetti e procedure	-1,58	0,37
20	D16	Spazio e figure	Rappresentazioni	0,08	0,45
21	D17_a	Relazioni e funzioni	Argomentazione	-0,31	0,5
22	D17_b	Relazioni e funzioni	Argomentazione	-0,62	0,44
23	D17_c	Relazioni e funzioni	Argomentazione	0,25	0,55
24	D18	Relazioni e funzioni	Concetti e procedure	0,16	0,6
25	D19_a	Dati e previsioni	Modellizzazione	-0,32	0,54
26	D19_b	Dati e previsioni	Modellizzazione	-0,17	0,46
27	D19_c	Dati e previsioni	Modellizzazione	0,74	0,49
28	D19_d	Dati e previsioni	Modellizzazione	-0,19	0,57
29	D20	Dati e previsioni	Concetti e procedure	-1,48	0,4
30	D21_a	Numeri	Concetti e procedure	0,63	0,55
31	D21_b	Numeri	Concetti e procedure	-0,88	0,46
32	D22	Spazio e figure	Modellizzazione	0,54	0,3
33	D23	Relazioni e funzioni	Modellizzazione	0,77	0,46
34	D24	Numeri	Concetti e procedure	-0,06	0,53
35	D25_a	Numeri	Rappresentazioni	-0,07	0,43
36	D25_b	Numeri	Concetti e procedure	1,06	0,37

N.	Item	Ambiti	Macro Processi	Difficoltà	Correlazione punto-biseriale
37	D26	Spazio e figure	Argomentazione	1,08	0,42
38	D27	Spazio e figure	Rappresentazioni	-0,29	0,44
39	D28	Numeri	Concetti e procedure	0,66	0,44
40	D29	Relazioni e funzioni	Rappresentazioni	1,35	0,24
41	D30	Numeri	Modellizzazione	0,75	0,51
42	D31	Spazio e figure	Rappresentazioni	-0,47	0,37
43	D32	Numeri	Concetti e procedure	0,3	0,42
44	D33	Dati e previsioni	Modellizzazione	-0,81	0,36

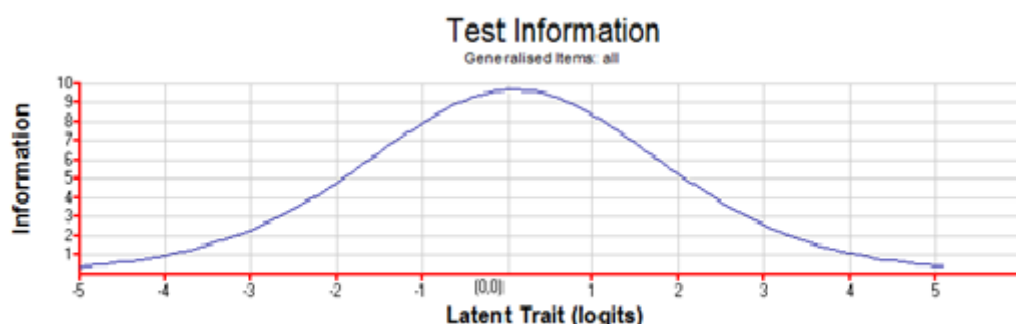


Figura 4.4: Informazione totale sulla prova di Matematica – V primaria

Anche per le prove delle V primaria si confermano i risultati dei pre-test: le ipotesi formulate in quella sede, in relazione alla difficoltà complessiva della prova, escono corroborate dai risultati della somministrazione a tutti gli allievi della quinta classe della scuola primaria.

I valori dei parametri di difficoltà della prova di Italiano riportati nella tabella 4.3 mostrano che essa è formata da quesiti adeguati per gli allievi della V primaria poiché i livelli di difficoltà delle domande, pur prevalendo quelle medio-facili, si distribuiscono lungo la scala delle abilità. Inoltre, i differenti livelli di difficoltà sono abbastanza equamente distribuiti per le diverse parti del testo e i diversi processi, sebbene il testo espositivo, come è lecito attendersi, sia maggiormente presente nella parte più alta della scala di difficoltà.

Se si confronta la prova di Italiano di V primaria con quella della classe II primaria, si osserva che nella prima sono presenti più quesiti di maggiore difficoltà, anche se non particolarmente elevata. Tale fenomeno è ancora più evidente se, come si vedrà in seguito, il confronto è effettuato con la prova di Italiano della I secondaria di primo grado. Ciò, almeno in parte, si spiega con la maggiore

articolazione della prova. Infatti, non bisogna dimenticare che, a differenza della prova di Italiano della II primaria, le altre prove di Italiano sono composte da due o più testi.

La tabella 4.4 riporta i parametri di difficoltà della prova di Matematica. Anche in questo caso si riscontra ciò che è già stato evidenziato per la prova di Matematica della II primaria, ossia una sostanziale adeguatezza in termini di distribuzione delle difficoltà; inoltre, rispetto allo scorso anno, la prova consente di stimare con maggiore accuratezza i livelli moderatamente elevati della scala di abilità.

4.2 La scuola secondaria di primo grado

Le tabelle 4.5 e 4.6 riportano i parametri di difficoltà delle prove di Italiano e Matematica della classe I della scuola secondaria di primo grado; le figure 4.5 e 4.6 illustrano l'informazione che le prove di Italiano e di Matematica forniscono sulla distribuzione di abilità.

Tabella 4.5: Parametri di difficoltà delle domande del fascicolo di Italiano (I sec. I gr.) e loro afferenza a parti del testo e processi

N.	Item	Parte del testo	Processi	Difficoltà	Correlazione punto biseriale
1	A1	Narrativo	Ricostruire il significato del testo	0,59	0,39
2	A2	Narrativo	Ricostruire il significato del testo	-1,13	0,37
3	A3	Narrativo	Ricostruire il significato del testo	-1,76	0,48
4	A4	Narrativo	Ricostruire il significato del testo	-1,94	0,41
5	A5	Narrativo	Ricostruire il significato del testo	-1,19	0,27
6	A6	Narrativo	Ricostruire il significato del testo	-1,97	0,41
7	A7	Narrativo	Ricostruire il significato del testo	-0,69	0,42
8	A8	Narrativo	Ricostruire il significato del testo	-0,52	0,36
9	A9	Narrativo	Ricostruire il significato del testo	-0,95	0,44
10	A10	Narrativo	Interpretare e valutare	-1,64	0,41
11	A11	Narrativo	Ricostruire il significato del testo	-1,77	0,51
12	A12	Narrativo	Individuare informazioni	-0,62	0,33
13	A13	Narrativo	Ricostruire il significato del testo	-2,10	0,43
14	A14	Narrativo	Ricostruire il significato del testo	-0,86	0,51
15	A15	Narrativo	Ricostruire il significato del testo	-2,72	0,43
16	A16	Narrativo	Ricostruire il significato del testo	-1,19	0,24
17	A17	Narrativo	Ricostruire il significato del testo	-0,76	0,43
18	A18	Narrativo	Interpretare e valutare	1,10	0,35
19	B1	Espositivo	Ricostruire il significato del testo	-0,02	0,37
20	B2	Espositivo	Ricostruire il significato del testo	0,83	0,45
21	B3	Espositivo	Ricostruire il significato del testo	-0,01	0,33
22	B4	Espositivo	Ricostruire il significato del testo	-0,10	0,35
23	B5	Espositivo	Ricostruire il significato del testo	-0,45	0,28
24	B6	Espositivo	Ricostruire il significato del testo	0,23	0,36
25	B7	Espositivo	Individuare informazioni	-1,07	0,48
26	B8	Espositivo	Ricostruire il significato del testo	-0,73	0,52
27	B9	Espositivo	Ricostruire il significato del testo	1,37	0,38
28	B10	Espositivo	Individuare informazioni	-1,15	0,45
29	B11	Espositivo	Individuare informazioni	0,75	0,17
30	B12	Espositivo	Ricostruire il significato del testo	-0,04	0,54
31	B13	Espositivo	Ricostruire il significato del testo	-1,19	0,41
32	B14	Espositivo	Individuare informazioni	-0,09	0,47
33	B15	Espositivo	Ricostruire il significato del testo	0,35	0,45
34	B16	Espositivo	Individuare informazioni	0,23	0,53
35	B17	Espositivo	Ricostruire il significato del testo	-0,09	0,39
36	B18	Espositivo	Ricostruire il significato del testo	1,03	0,49

N.	Item	Parte del testo	Processi	Difficoltà	Correlazione punto biseriale
37	C1	Grammatica	Riflessione sulla lingua	0,04	0,50
38	C2	Grammatica	Riflessione sulla lingua	-1,21	0,38
39	C3	Grammatica	Riflessione sulla lingua	1,43	0,29
40	C4	Grammatica	Riflessione sulla lingua	-1,04	0,26
41	C5	Grammatica	Riflessione sulla lingua	0,00	0,25
42	C6	Grammatica	Riflessione sulla lingua	0,10	0,49
43	C7	Grammatica	Riflessione sulla lingua	-0,78	0,38
44	C8	Grammatica	Riflessione sulla lingua	1,15	0,23
45	C9	Grammatica	Riflessione sulla lingua	1,13	0,40

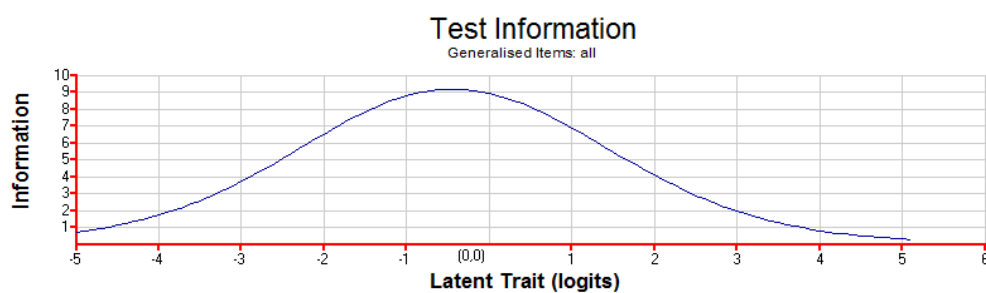


Figura 4.5: Informazione totale sulla prova di Italiano – I sec. I gr.

Tabella 4.6: Parametri di difficoltà delle domande del fascicolo di Matematica (I sec. I gr.) e loro afferenza ad ambiti e a macro processi

N.	Item	Ambito	Macro processi	Difficoltà	Correlazione punto-biseriale
1	D1_a	Spazio e figure	Rappresentazioni	-2,7	0,23
2	D1_b	Spazio e figure	Rappresentazioni	-1,6	0,33
3	D2_a	Dati e previsioni	Modellizzazione	-0,06	0,48
4	D2_b	Dati e previsioni	Argomentazione	0,95	0,43
5	D3_a	Relazioni e funzioni	Argomentazione	-2,19	0,34
6	D3_b	Relazioni e funzioni	Concetti e procedure	0,66	0,52
7	D4	Numeri	Concetti e procedure	-0,46	0,46
8	D5_a	Dati e previsioni	Modellizzazione	-0,24	0,42
9	D5_b	Dati e previsioni	Modellizzazione	1,48	0,49
10	D5_c	Dati e previsioni	Modellizzazione	1,22	0,51
11	D6_a	Numeri	Modellizzazione	-2,19	0,27
12	D6_b	Numeri	Concetti e procedure	1	0,19
13	D6_c	Numeri	Modellizzazione	-0,15	0,42
14	D7	Spazio e figure	Modellizzazione	2,22	0,22
15	D8	Relazioni e funzioni	Rappresentazioni	0,1	0,36

N.	Item	Ambito	Macro processi	Difficoltà	Correlazione punto-biseriale
16	D9	Numeri	Modellizzazione	1,1	0,14
17	D10	Spazio e figure	Concetti e procedure	0,1	0,42
18	D11_a	Numeri	Modellizzazione	-1,54	0,47
19	D11_b	Numeri	Modellizzazione	1,24	0,38
20	D12	Dati e previsioni	Rappresentazioni	-0,66	0,54
21	D13_a	Relazioni e funzioni	Modellizzazione	-0,05	0,53
22	D13_b	Relazioni e funzioni	Modellizzazione	0,6	0,3
23	D13_c	Relazioni e funzioni	Modellizzazione	1,46	0,43
24	D14	Spazio e figure	Argomentazione	1,37	0,35
25	D15	Spazio e figure	Concetti e procedure	2,14	0,01
26	D16	Relazioni e funzioni	Modellizzazione	1,88	0,27
27	D17_a	Relazioni e funzioni	Modellizzazione	-0,48	0,49
28	D17_b	Relazioni e funzioni	Modellizzazione	1,69	0,5
29	D17_c	Relazioni e funzioni	Modellizzazione	2,14	0,48
30	D18_a	Numeri	Concetti e procedure	0,08	0,51
31	D18_b	Numeri	Concetti e procedure	0,45	0,51
32	D19_a	Spazio e figure	Modellizzazione	-1,16	0,38
33	D19_b	Spazio e figure	Modellizzazione	0,28	0,38
34	D20	Numeri	Concetti e procedure	2,61	0,28
35	D21_a	Dati e previsioni	Modellizzazione	-1,44	0,42
36	D21_b	Dati e previsioni	Modellizzazione	-1,97	0,34
37	D21_c	Dati e previsioni	Modellizzazione	1,21	0,39
38	D22	Numeri	Concetti e procedure	0,02	0,45
39	D23	Numeri	Concetti e procedure	2,44	0,27
40	D24	Spazio e figure	Argomentazione	0,53	0,43
41	D25	Spazio e figure	Concetti e procedure	-0,06	0,35
42	D26	Numeri	Argomentazione	0,82	0,34
43	D27_a	Spazio e figure	Concetti e procedure	-0,89	0,51
44	D27_b	Spazio e figure	Concetti e procedure	0,1	0,53
45	D28_a	Relazioni e funzioni	Modellizzazione	0,23	0,47
46	D28_b	Relazioni e funzioni	Modellizzazione	0,97	0,51
47	D29	Relazioni e funzioni	Modellizzazione	0,22	0,33
48	D30	Spazio e figure	Concetti e procedure	-0,48	0,41
49	D31	Dati e previsioni	Modellizzazione	-0,76	0,44

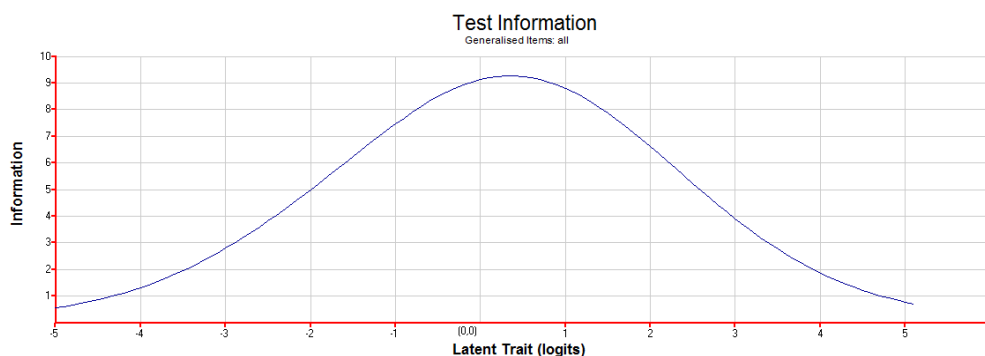


Figura 4.6: Informazione totale sulla prova di Matematica – I sec. I gr.

I valori riportati nelle tabelle 4.5 e 4.6 confermano la tendenza emersa nell'analisi comparativa delle prove della scuola primaria. Le prove rivolte alla prima classe della scuola secondaria di primo grado, quindi ad allievi che di norma hanno 11-12 anni, si caratterizzano, infatti, per la distribuzione dei quesiti lungo tutta la scala di abilità, sia per Italiano sia per Matematica, con una maggiore rappresentazione, rispetto alle prove di II e V primaria, dei livelli più elevati di difficoltà, in particolare nella prova di Matematica. La distribuzione dei quesiti lungo l'intera scala consente di rappresentare con precisione i livelli di preparazione raggiunti da tutti gli allievi, sia quelli che hanno raggiunto traguardi formativi elevati sia quelli che si trovano in condizioni di minore vantaggio.

Anche in questo caso, inoltre, risultano ben distribuiti tutte le parti del testo e i processi, per la prova di Italiano, e tutti gli ambiti e i macro processi per la prova di Matematica; unica eccezione in tal senso, il macro processo "Rappresentazioni" nella Matematica, rappresentato solo ai livelli più bassi di difficoltà.

4.2.1. La Prova nazionale

A partire dal 2008, al termine della scuola secondaria di primo grado e all'interno dell'esame di Stato, si svolge la cosiddetta Prova nazionale. Sebbene risponda a finalità in parte diverse rispetto a quelle del Servizio nazionale di valutazione, dal punto di vista della costruzione delle prove non vi sono differenze sostanziali, sia dal punto di vista strettamente tecnico sia per quanto riguarda il quadro di riferimento. In ogni caso, poiché la Prova nazionale contribuisce, anche se solo nella misura di un settimo rispetto al totale, al voto finale di licenza, una particolare attenzione è rivolta alla scelta della difficoltà dei quesiti, proprio per non alterare lo spirito generale che ispira l'esame di Stato conclusivo del primo ciclo di istruzione.

Le tabelle 4.7 e 4.8 riportano i parametri di difficoltà e i coefficienti di correlazione punto-biseriale delle prove di Italiano (denominata in sede di esame Fascicolo 2) e Matematica (denominata in sede di esame Fascicolo 1) della classe III della scuola secondaria di primo grado; le figure 4.7 e 4.8 illustrano l'informazione che le prove di Italiano e di Matematica forniscono sulla distribuzione di abilità.

Tabella 4.7: Parametri di difficoltà dei quesiti della prova di Italiano (III sec. I gr.) e loro afferenza a parti del testo e processi

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto-biseriale
1	A1	Narrativo	Ricostruire il significato del testo	-0,92	0,37
2	A2	Narrativo	Ricostruire il significato del testo	0,23	0,27
3	A3	Narrativo	Ricostruire il significato del testo	-2,21	0,36
4	A4	Narrativo	Ricostruire il significato del testo	-0,14	0,34
5	A5	Narrativo	Ricostruire il significato del testo	-2,40	0,44
6	A6	Narrativo	Ricostruire il significato del testo	-1,84	0,32
7	A7	Narrativo	Ricostruire il significato del testo	-3,89	0,28
8	A8	Narrativo	Ricostruire il significato del testo	-0,53	0,38
9	A9	Narrativo	Ricostruire il significato del testo	-2,23	0,43
10	A10	Narrativo	Ricostruire il significato del testo	-0,82	0,37
11	A11	Narrativo	Interpretare e valutare	-2,42	0,46
12	A12	Narrativo	Ricostruire il significato del testo	-1,31	0,43
13	A13	Narrativo	Ricostruire il significato del testo	-0,98	0,4
14	A14	Narrativo	Ricostruire il significato del testo	-0,03	0,35
15	A15	Narrativo	Ricostruire il significato del testo	-1,01	0,42
16	A16	Narrativo	Ricostruire il significato del testo	-1,18	0,31

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto-biserial
17	A17	Narrativo	Ricostruire il significato del testo	-0,58	0,48
18	A18	Narrativo	Individuare informazioni	-0,18	0,39
19	A19	Narrativo	Ricostruire il significato del testo	-1,53	0,44
20	A20	Narrativo	Ricostruire il significato del testo	-0,94	0,15
21	A21	Narrativo	Interpretare e valutare	-1,53	0,29
22	A22	Narrativo	Ricostruire il significato del testo	0,36	0,49
23	B1	Espositivo	Ricostruire il significato del testo	-1,07	0,36
24	B2	Espositivo	Individuare informazioni	-0,60	0,45
25	B3	Espositivo	Individuare informazioni	-0,95	0,34
26	B4	Espositivo	Individuare informazioni	0,10	0,44
27	B5	Espositivo	Ricostruire il significato del testo	0,40	0,48
28	B6	Espositivo	Individuare informazioni	-2,11	0,42
29	B7	Espositivo	Individuare informazioni	-0,45	0,41
30	B8	Espositivo	Interpretare e valutare	-0,25	0,3
31	C1	Non continuo	Individuare informazioni	-2,74	0,28
32	C2	Non continuo	Ricostruire il significato del testo	-2,34	0,46
33	C3	Non continuo	Ricostruire il significato del testo	-1,03	0,32
34	C4	Non continuo	Ricostruire il significato del testo	-0,98	0,3
35	C5	Non continuo	Ricostruire il significato del testo	-1,40	0,37
36	C6	Non continuo	Individuare informazioni	-2,53	0,4
37	C7	Non continuo	Ricostruire il significato del testo	2,14	0,24
38	C8	Non continuo	Ricostruire il significato del testo	-0,01	0,08
39	D1	Grammatica	Riflessione sulla lingua	-1,37	0,42
40	D2	Grammatica	Riflessione sulla lingua	-1,98	0,45
41	D3	Grammatica	Riflessione sulla lingua	-0,60	0,25
42	D4	Grammatica	Riflessione sulla lingua	-1,29	0,47
43	D5	Grammatica	Riflessione sulla lingua	-1,13	0,49
44	D6	Grammatica	Riflessione sulla lingua	-0,21	0,43
45	D7	Grammatica	Riflessione sulla lingua	-1,23	0,41
46	D8	Grammatica	Riflessione sulla lingua	-2,42	0,35
47	D9	Grammatica	Riflessione sulla lingua	0,07	0,33
48	D10	Grammatica	Riflessione sulla lingua	-0,27	0,48
49	D11	Grammatica	Riflessione sulla lingua	-0,66	0,44

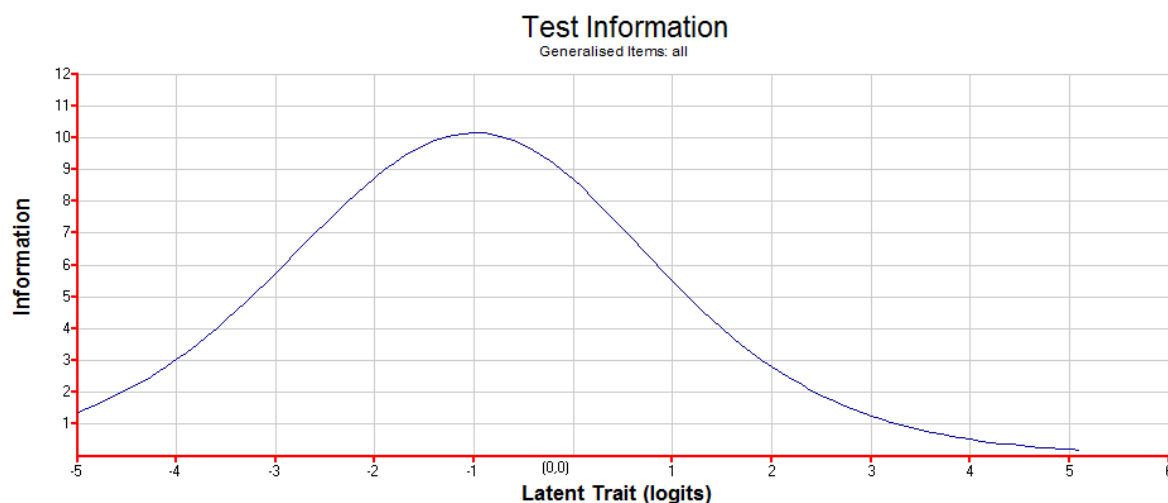


Figura 4.7: Informazione totale sulla prova di Italiano – III sec. I gr.

Tabella 4.8: Parametri di difficoltà delle domande del fascicolo di Matematica (III sec. I gr.) e loro afferenza ad ambiti e macro processi

N.	Item	Ambiti	Macro processi	Difficoltà	Correlazione punto-biseriale
1	E1	Spazio e figure	Rappresentazioni	-4,24	0,11
2	E2	Numeri	Concetti e procedure	-0,84	0,31
3	E3_a	Dati e previsioni	Concetti e procedure	0,55	0,31
4	E3_b	Dati e previsioni	Argomentazione	-0,81	0,36
5	E4_a	Relazioni e funzioni	Concetti e procedure	-0,27	0,3
6	E4_b	Relazioni e funzioni	Modellizzazione	-1,09	0,33
7	E4_c	Relazioni e funzioni	Modellizzazione	-1,6	0,32
8	E5	Relazioni e funzioni	Argomentazione	-0,27	0,4
9	E6	Spazio e figure	Argomentazione	1,5	0,18
10	E7	Numeri	Concetti e procedure	-0,16	0,47
11	E8_a	Spazio e figure	Modellizzazione	0,55	0,26
12	E8_b	Spazio e figure	Modellizzazione	1	0,45
13	E9_a	Relazioni e funzioni	Modellizzazione	0,87	0,42
14	E9_b	Relazioni e funzioni	Argomentazione	0,37	0,23
15	E10_a	Dati e previsioni	Modellizzazione	-0,11	0,4
16	E10_b	Dati e previsioni	Concetti e procedure	1,05	0,27
17	E11	Numeri	Rappresentazioni	1,53	0,4
18	E12_a	Spazio e figure	Rappresentazioni	-0,51	0,43
19	E12_b	Spazio e figure	Argomentazione	0,26	0,43
20	E13	Numeri	Argomentazione	0,89	0,31
21	E14_a	Spazio e figure	Rappresentazioni	1,6	0,42
22	E14_b	Spazio e figure	Rappresentazioni	1	0,3

N.	Item	Ambiti	Macro processi	Difficoltà	Correlazione punto-biserial
23	E15	Numeri	Modellizzazione	0,37	0,49
24	E16_a	Spazio e figure	Modellizzazione	0,06	0,45
25	E16_b	Spazio e figure	Argomentazione	0,79	0,51
26	E17_a	Relazioni e funzioni	Concetti e procedure	-0,44	0,47
27	E17_b	Relazioni e funzioni	Modellizzazione	-0,61	0,45
28	E17_c	Relazioni e funzioni	Rappresentazioni	-0,11	0,41
29	E18	Numeri	Concetti e procedure	0,88	0,34
30	E19_a	Dati e previsioni	Concetti e procedure	-1,79	0,26
31	E19_b	Dati e previsioni	Concetti e procedure	0,05	0,31
32	E20_a	Numeri	Modellizzazione	-1,11	0,35
33	E20_b	Numeri	Modellizzazione	0,33	0,41
34	E20_c	Numeri	Modellizzazione	0,23	0,37
35	E21	Numeri	Concetti e procedure	-0,78	0,47
36	E22_a	Dati e previsioni	Modellizzazione	-0,77	0,3
37	E22_b	Dati e previsioni	Modellizzazione	0,11	0,41
38	E23	Relazioni e funzioni	Modellizzazione	-1,08	0,37
39	E24	Dati e previsioni	Concetti e procedure	-0,14	0,52
40	E25	Spazio e figure	Concetti e procedure	-1,02	0,34

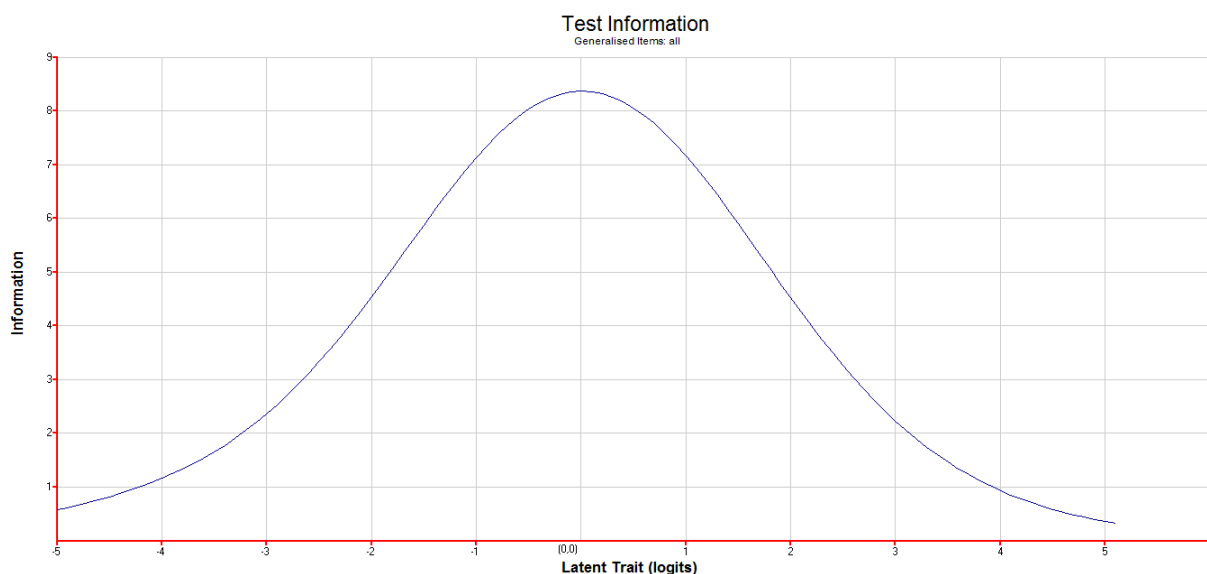


Figura 4.8: Informazione totale sulla prova di Matematica – III sec. I gr.

I dati delle tabelle 4.7 e 4.8 confermano, soprattutto per la prova di Italiano, una certa prevalenza di domande di difficoltà media e medio-bassa, mentre per la prova di Matematica sono presenti alcuni quesiti di difficoltà medio-alta, sebbene restino prevalenti domande mediamente difficili o facili.

4.3. La scuola secondaria di secondo grado

Il presente anno scolastico vede coinvolta, per la seconda volta, la scuola secondaria di secondo grado. Gli studenti interessati dalla rilevazione sono quelli frequentanti le classi seconde.

Prima di riproporre le analisi mostrate in precedenza per le altre classi interessate dalla rilevazione del SNV, sono necessarie alcune considerazioni di carattere generale. In primo luogo, le rilevazioni del SNV si collocano alla fine dell'obbligo di istruzione, quando gli studenti hanno, di norma, frequentato dieci anni di istruzione. In altre parole, queste prove sono proposte a studenti che, indipendentemente dal tipo di scuola e dall'indirizzo di studio frequentato, dovrebbero possedere conoscenze e competenze di base, ovvero quelle conoscenze e competenze utili sia se continueranno il loro percorso di studi sia se usciranno dalla scuola per inserirsi nel mondo del lavoro.

Si è pertanto ritenuto opportuno predisporre prove uniche per tutti i tipi di scuola (licei, istituti tecnici e professionali), volte alla misurazione degli apprendimenti di base relativi a due ambiti fondamentali: 1) la comprensione di testi scritti di varia natura e la riflessione sulla lingua e 2) la matematica. La scelta di sottoporre agli studenti prove non differenziate rispetto agli indirizzi di studio implica necessariamente che esse facciano riferimento a contenuti, abilità e competenze comuni a tutti i tipi di scuola e che i quesiti proposti siano in grado di coprire un'ampia gamma di livelli di difficoltà.

Ciò premesso, è fondamentale prendere in considerazione gli stessi parametri proposti per le scuole del primo ciclo di istruzione anche per le prove della scuola secondaria di secondo grado.

Le tabelle 4.9 e 4.10 riportano i parametri di difficoltà e i coefficienti di correlazione punto-biseriale delle prove di Italiano e Matematica della classe II della scuola secondaria di secondo grado; le figure 4.9 e 4.10 illustrano l'informazione che le prove di Italiano e di Matematica forniscono sulla distribuzione di abilità.

Tabella 4.9: Parametri di difficoltà delle domande del fascicolo di Italiano (II sec. II gr.) e loro afferenza a parti del testo e processi

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto- biseriale
1	A1	Regolativo	Ricostruire il significato del testo	-2,66	0,31
2	A2	Regolativo	Individuare informazioni	0,60	0,33
3	A3	Regolativo	Individuare informazioni	-2,75	0,30
4	A4	Regolativo	Individuare informazioni	-3,34	0,21
5	A5	Regolativo	Individuare informazioni	-3,39	0,29
6	B1	Narrativo	Ricostruire il significato del testo	-1,28	0,28
7	B2	Narrativo	Ricostruire il significato del testo	-1,50	0,40
8	B3	Narrativo	Ricostruire il significato del testo	-1,89	0,31
9	B4	Narrativo	Ricostruire il significato del testo	-0,26	0,38
10	B5	Narrativo	Ricostruire il significato del testo	-0,93	0,20
11	B6	Narrativo	Ricostruire il significato del testo	-1,01	0,32
12	B7	Narrativo	Ricostruire il significato del testo	-0,48	0,37
13	B8	Narrativo	Individuare informazioni	0,44	0,43
14	B9	Narrativo	Individuare informazioni	-0,85	0,33
15	B10	Narrativo	Ricostruire il significato del testo	-0,92	0,40
16	B11	Narrativo	Ricostruire il significato del testo	-1,58	0,45
17	B12	Narrativo	Ricostruire il significato del testo	-0,94	0,50
18	B13	Narrativo	Interpretare e valutare	0,08	0,34
19	B14 a	Narrativo	Ricostruire il significato del testo	-1,17	0,43
20	B14 b	Narrativo	Ricostruire il significato del testo	0,53	0,39
21	B15	Narrativo	Ricostruire il significato del testo	2,12	0,13
22	B16	Narrativo	Interpretare e valutare	-1,52	0,41
23	C1	Poetico	Ricostruire il significato del testo	-1,86	0,39
24	C2a	Poetico	Ricostruire il significato del testo	0,26	0,36
25	C2b	Poetico	Ricostruire il significato del testo	-0,81	0,45
26	C3	Poetico	Ricostruire il significato del testo	1,31	0,24
27	C4	Poetico	Individuare informazioni	-1,47	0,42
28	C5	Poetico	Ricostruire il significato del testo	-1,37	0,41
29	C6	Poetico	Individuare informazioni	-2,12	0,50
30	C7	Poetico	Ricostruire il significato del testo	0,18	0,36
31	C8	Poetico	Ricostruire il significato del testo	-0,22	0,31
32	C9	Poetico	Ricostruire il significato del testo	0,88	0,28
33	C10	Poetico	Interpretare e valutare	-0,97	0,51
34	C11	Poetico	Ricostruire il significato del testo	-0,38	0,40
35	C12	Poetico	Ricostruire il significato del testo	2,24	0,17
36	C13	Poetico	Ricostruire il significato del testo	-1,56	0,38
37	D1	Espositivo_Misto	Individuare informazioni	-0,79	0,50
38	D2	Espositivo_Misto	Ricostruire il significato del testo	-0,15	0,41

N.	Item	Parti del testo	Processi	Difficoltà	Correlazione punto- biseriale
39	D3	Espositivo_Misto	Ricostruire il significato del testo	0,08	0,41
40	D4	Espositivo_Misto	Interpretare e valutare	-0,61	0,36
41	D5	Espositivo_Misto	Individuare informazioni	-2,96	0,35
42	D6	Espositivo_Misto	Ricostruire il significato del testo	-0,33	0,52
43	D7	Espositivo_Misto	Ricostruire il significato del testo	-1,10	0,41
44	D8	Espositivo_Misto	Individuare informazioni	-0,52	0,47
45	D9	Espositivo_Misto	Ricostruire il significato del testo	-0,42	0,43
46	D10	Espositivo_Misto	Individuare informazioni	-0,73	0,54
47	E1	Espositivo_Argomentativo	Individuare informazioni	0,54	0,35
48	E2	Espositivo_Argomentativo	Interpretare e valutare	-0,53	0,52
49	E3	Espositivo_Argomentativo	Ricostruire il significato del testo	-0,61	0,43
50	E4	Espositivo_Argomentativo	Individuare informazioni	-1,21	0,47
51	E5	Espositivo_Argomentativo	Ricostruire il significato del testo	-1,62	0,48
52	F1	Grammatica	Riflessione sulla lingua	-1,70	0,49
53	F2	Grammatica	Riflessione sulla lingua	-1,07	0,33
54	F3	Grammatica	Riflessione sulla lingua	-1,20	0,41
55	F4	Grammatica	Riflessione sulla lingua	0,80	0,22
56	F5	Grammatica	Riflessione sulla lingua	-0,01	0,50
57	F6	Grammatica	Riflessione sulla lingua	-1,28	0,47
58	F7	Grammatica	Riflessione sulla lingua	-0,84	0,49
59	F8	Grammatica	Riflessione sulla lingua	-0,37	0,37
60	F9	Grammatica	Riflessione sulla lingua	-1,03	0,47
61	F10	Grammatica	Riflessione sulla lingua	-0,55	0,58

64

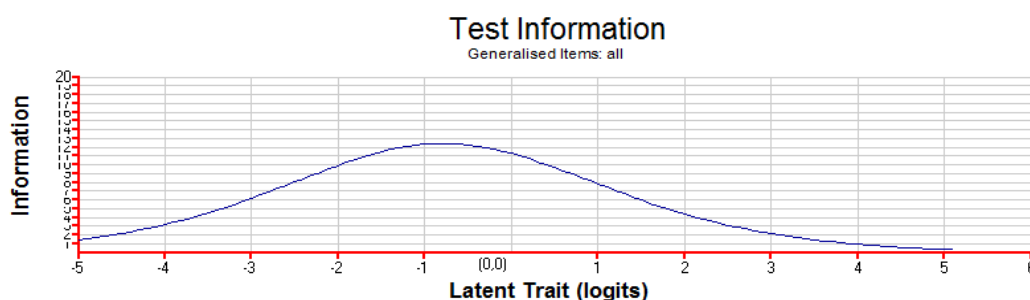


Figura 4.9: Informazione totale sulla prova di Italiano – II sec. II gr.

Tabella 4.10: Parametri di difficoltà delle domande del fascicolo di Matematica (II sec. II gr.) e loro afferenza ad ambiti e macro processi

N.	Item	Ambiti	Processi	Difficoltà	Correlazione punto-biseriale
1	M1	Numeri	Concetti e procedure	1,08	0,35
2	M2_a	Relazioni e funzioni	Rappresentazioni	-1,90	0,27
3	M2_b	Relazioni e funzioni	Concetti e procedure	-2,16	0,34
4	M2_c	Relazioni e funzioni	Modellizzazione	-1,97	0,35
5	M2_d	Relazioni e funzioni	Rappresentazioni	0,05	0,59
6	M2_e	Relazioni e funzioni	Rappresentazioni	2,26	0,49
7	M3	Spazio e figure	Argomentazione	1,11	0,38
8	M4	Dati e previsioni	Modellizzazione	-0,01	0,43
9	M5	Numeri	Concetti e procedure	0,78	0,39
10	M6_a	Dati e previsioni	Rappresentazioni	-1,83	0,35
11	M6_b	Dati e previsioni	Concetti e procedure	-0,42	0,40
12	M6_c	Dati e previsioni	Concetti e procedure	2,06	0,49
13	M7_a	Relazioni e funzioni	Modellizzazione	-0,15	0,41
14	M7_b	Relazioni e funzioni	Modellizzazione	-1,32	0,30
15	M8	Spazio e figure	Argomentazione	0,11	0,54
16	M9	Dati e previsioni	Rappresentazioni	0,81	0,36
17	M10_a	Relazioni e funzioni	Concetti e procedure	-0,29	0,40
18	M10_b	Relazioni e funzioni	Rappresentazioni	0,37	0,38
19	M11_a	Numeri	Concetti e procedure	-2,14	0,37
20	M11_b	Numeri	Argomentazione	1,71	0,34
21	M11_c	Numeri	Argomentazione	1,63	0,46
22	M12	Numeri	Concetti e procedure	2,01	0,50
23	M13	Spazio e figure	Concetti e procedure	0,74	0,44
24	M14_a	Dati e previsioni	Concetti e procedure	1,62	0,18
25	M14_b	Dati e previsioni	Concetti e procedure	-1,20	0,41
26	M14_c	Dati e previsioni	Concetti e procedure	1,69	0,43
27	M15	Numeri	Concetti e procedure	0,65	0,46
28	M16	Dati e previsioni	Argomentazione	1,18	0,42
29	M17	Spazio e figure	Modellizzazione	2,12	0,39
30	M18	Numeri	Rappresentazioni	-0,59	0,53
31	M19	Numeri	Argomentazione	1,26	0,21
32	M20	Numeri	Modellizzazione	1,09	0,52
33	M21	Numeri	Concetti e procedure	0,76	0,50
34	M22	Spazio e figure	Rappresentazioni	0,47	0,28
35	M23_a	Numeri	Concetti e procedure	-0,92	0,45
36	M23_b	Numeri	Argomentazione	1,85	0,42
37	M24	Spazio e figure	Modellizzazione	1,15	0,59
38	M25	Numeri	Concetti e procedure	1,36	0,43

N.	Item	Ambiti	Processi	Difficoltà	Correlazione punto-biseriale
39	M26_a	Spazio e figure	Concetti e procedure	-0,53	0,45
40	M26_b	Spazio e figure	Concetti e procedure	-0,37	0,44
41	M26_c	Spazio e figure	Concetti e procedure	0,14	0,45
42	M27	Relazioni e funzioni	Modellizzazione	-0,14	0,48
43	M28	Numeri	Concetti e procedure	-0,05	0,47
44	M29	Spazio e figure	Concetti e procedure	1,78	0,30
45	M30	Relazioni e funzioni	Argomentazione	1,57	0,21

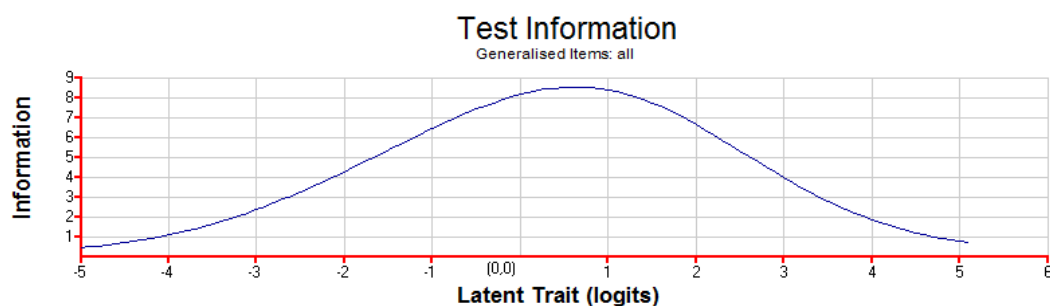


Figura 4.10: Informazione totale sulla prova di Matematica – II sec. II gr.

I dati della tabella 4.9 mostrano che la prova di Italiano della seconda classe della scuola secondaria di secondo grado si caratterizza per la prevalenza di domande con un livello di difficoltà medio e medio-basso; solo due item (C12 e B15), infatti, superano il valore 2. Le diverse parti del testo e i diversi processi si distribuiscono piuttosto equamente lungo tutta la scala di difficoltà coperta dagli item.

Invece, per la Matematica (cfr. tabella 4.10), la prova nel suo complesso risulta rappresentare meglio anche i livelli di difficoltà (e, di conseguenza, di abilità) più elevati. La maggior parte degli item, infatti, registra un valore di difficoltà superiore allo 0.

La differenza nella difficoltà delle due prove si giustifica con la volontà di introdurre gradualmente le prove standardizzate del SNV nella scuola secondaria superiore; a tal fine, è stata predisposta, per la comprensione della lettura e la grammatica, una prova non troppo complessa. Per la Matematica, invece, essendo la prova caratterizzata dalla netta prevalenza di domande singole, non legate da un macro-stimolo unico, come nel caso dei testi (narrativo, espositivo, ecc.) della prova di

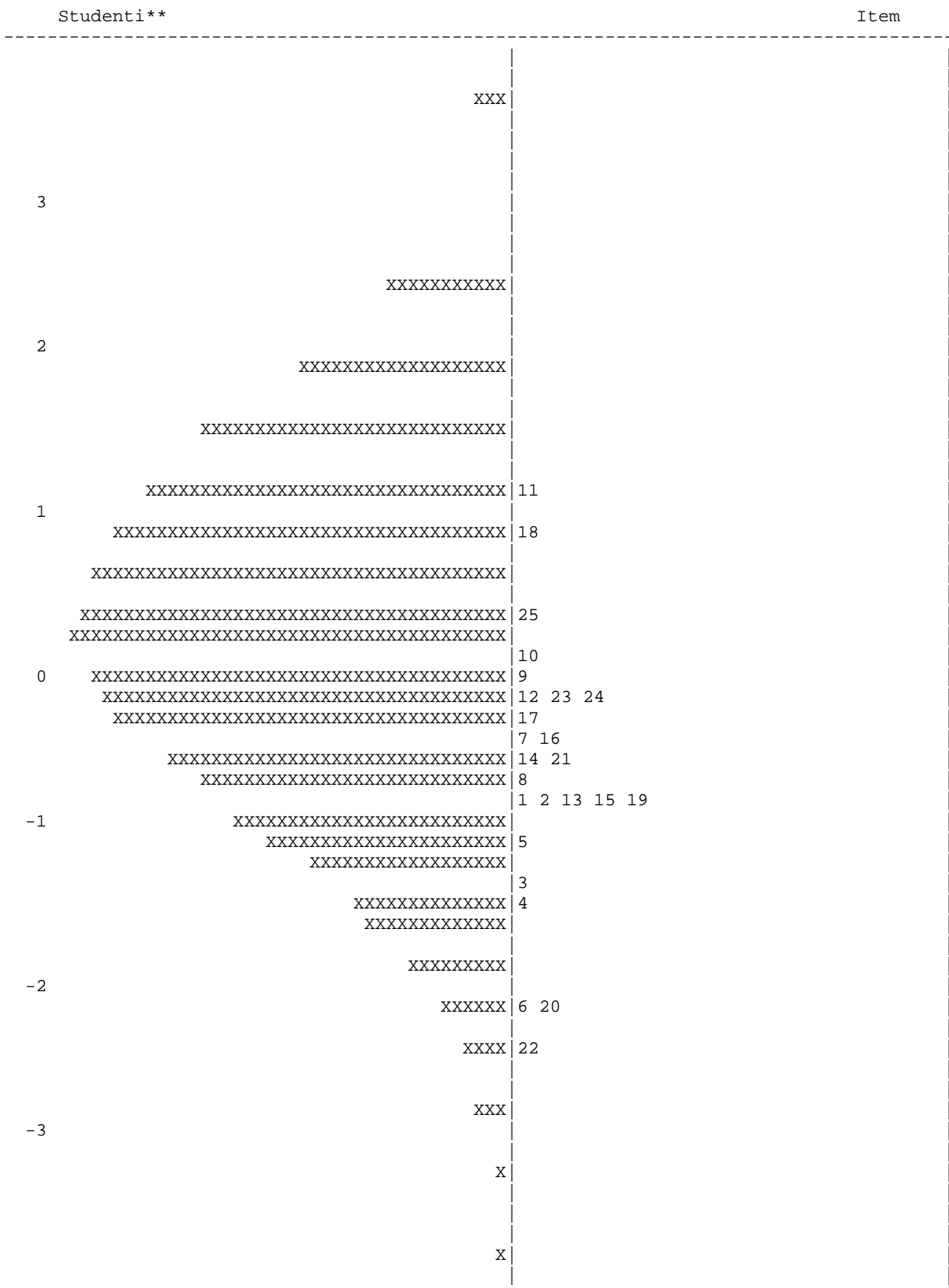
comprensione della lettura, è stato possibile, nell'insieme, articolare meglio la prova al fine di rappresentare una scala di difficoltà/abilità più ampia.

Capitolo 5 – Il *placement* relativo delle domande rispetto al punteggio di Rasch dei rispondenti

L'ultimo passo per la valutazione complessiva di una prova è lo studio del cosiddetto *placement* relativo tra difficoltà delle domande e livelli di apprendimento raggiunti dai rispondenti. Dal punto di vista teorico, il *placement* ideale è dato da una distribuzione della difficoltà delle domande in grado di coprire tutto l'intervallo della scala su cui insistono i livelli di apprendimento degli allievi. Nelle situazioni concrete ciò non è perfettamente realizzabile, specie nelle prove come quelle di comprensione della lettura caratterizzate da un insieme di domande legate a uno o più testi. È però importante che il *placement* sia soddisfacente, ovvero che non vi siano intervalli, soprattutto quelli intermedi, totalmente privi di domande o di allievi con livelli di prestazione che si collocano proprio in quegli intervalli. Inoltre, come già accennato nel commento alle tavole del paragrafo precedente, la scelta della proporzione di domande lungo la scala delle abilità deve tenere conto dell'età degli allievi ai quali la prova è presentata. In altre parole, quando ci si rivolge a bambini molto piccoli è opportuno che le domande tendano a popolare maggiormente i livelli medi e medio-bassi della scala di difficoltà, proprio per creare nel rispondente un senso di adesione alla verifica, tanto più importante, quanto più quest'ultimo si trova a frequentare i primi anni della scuola primaria.

Le figure da 5.1 a 5.10 mostrano il posizionamento relativo delle domande delle prove SNV e della Prova nazionale lungo la stessa scala sulla quale si rappresentano i risultati conseguiti dagli allievi.

Figura 5.1: *Placement* della prova di Italiano (II primaria)*



* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta - ** Ciascuna 'X' rappresenta 59,4 casi

I grafici delle figure 5.1 e 5.2 confermano i dati contenuti nelle tabelle 4.1 e 4.2. Nella parte sinistra dei grafici si trova la distribuzione dei risultati degli allievi che hanno partecipato alla prova SNV (dati campionari) della seconda classe della scuola primaria, mentre nella parte destra la collocazione sulla stessa scala (asse verticale tratteggiato) della difficoltà delle domande²⁸. Più precisamente, tutti gli studenti rappresentati dalle crocette che si trovano in una posizione inferiore a quella occupata da una data domanda hanno una scarsa probabilità (inferiore al 50 per cento) di rispondere correttamente alla domanda medesima, mentre coloro che sono posizionati rispetto all'asse verticale al di sopra di una domanda hanno una probabilità alta (superiore al 50 per cento) di rispondere correttamente e tale probabilità aumenta all'aumentare della distanza studente-domanda.

In questa sede è bene rimarcare la differenza tra una valutazione scolastica tradizionale e una prova standardizzata a carattere nazionale. Quest'ultima si pone come obiettivo principale quello di misurare i livelli di apprendimento raggiunti dagli studenti che frequentano le scuole italiane, ovvero si propone di misurare i risultati raggiunti in un determinato ambito da allievi con diversi livelli di abilità potenzialmente posizionati su ogni punto dell'asse verticale, come quello rappresentato nelle figure 5.1 e 5.2. In altre parole, è assolutamente normale che in una prova somministrata su larga scala vi siano alcuni quesiti ai quali solo una piccola minoranza di studenti è in grado di rispondere correttamente, senza che da questo si debba dedurre che la preparazione dei rispondenti sia inadeguata. Infatti, queste particolari domande sono finalizzate all'individuazione di livelli di apprendimento particolarmente elevati. In assenza di queste domande più difficili la prova non sarebbe in grado di dare indicazioni circa i livelli di apprendimento degli studenti più abili.

Le figure da 5.3 a 5.10 illustrano il *placement* relativo delle prove di Italiano e Matematica della V primaria, della I e III secondaria di primo grado e della II secondaria di secondo grado.

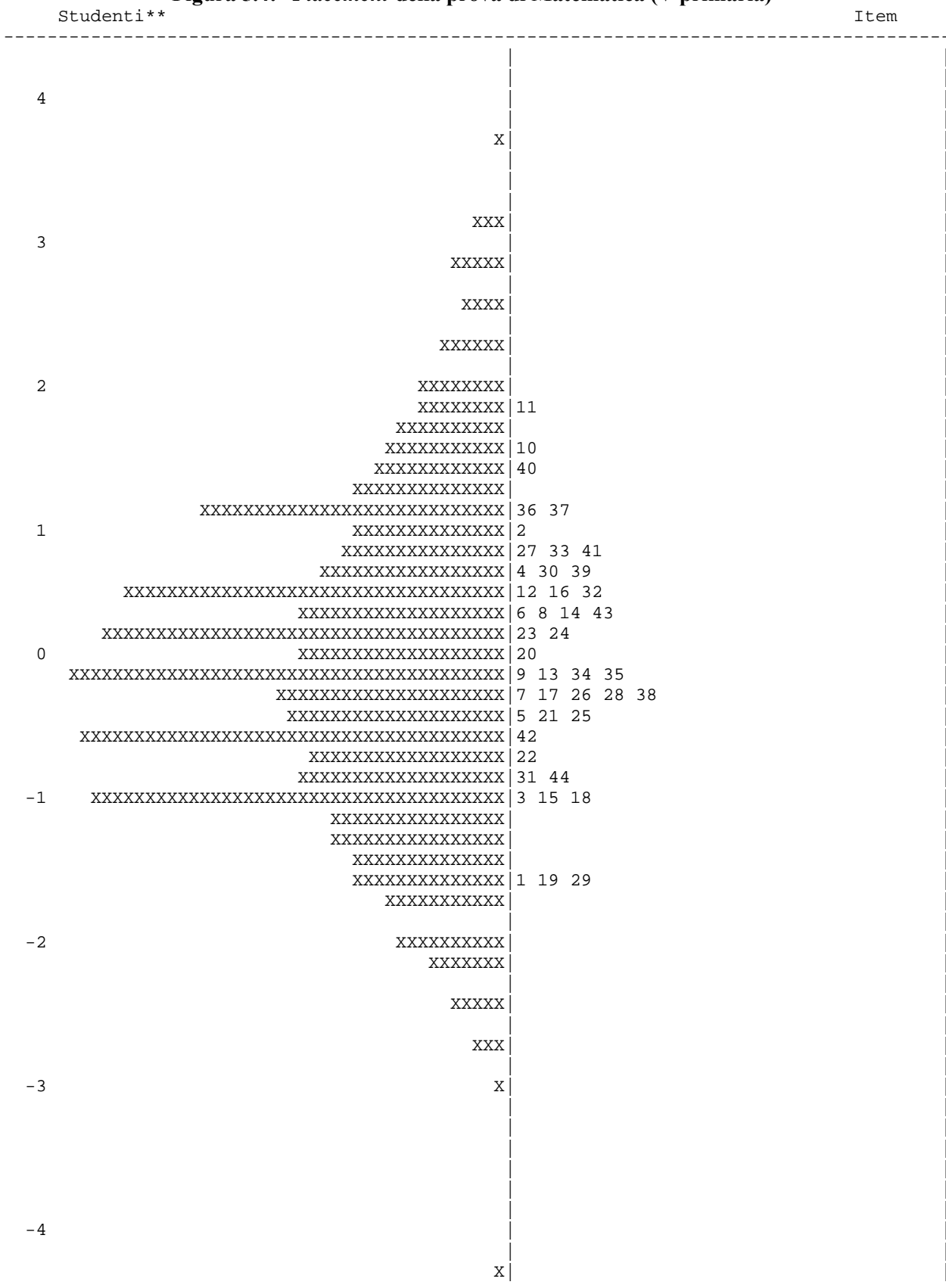
²⁸ Le domande sono etichettate in base al numero progressivo come riportato nelle tavole II.3-II.12.

Figura 5.3: *Placement* della prova di Italiano (V primaria)*



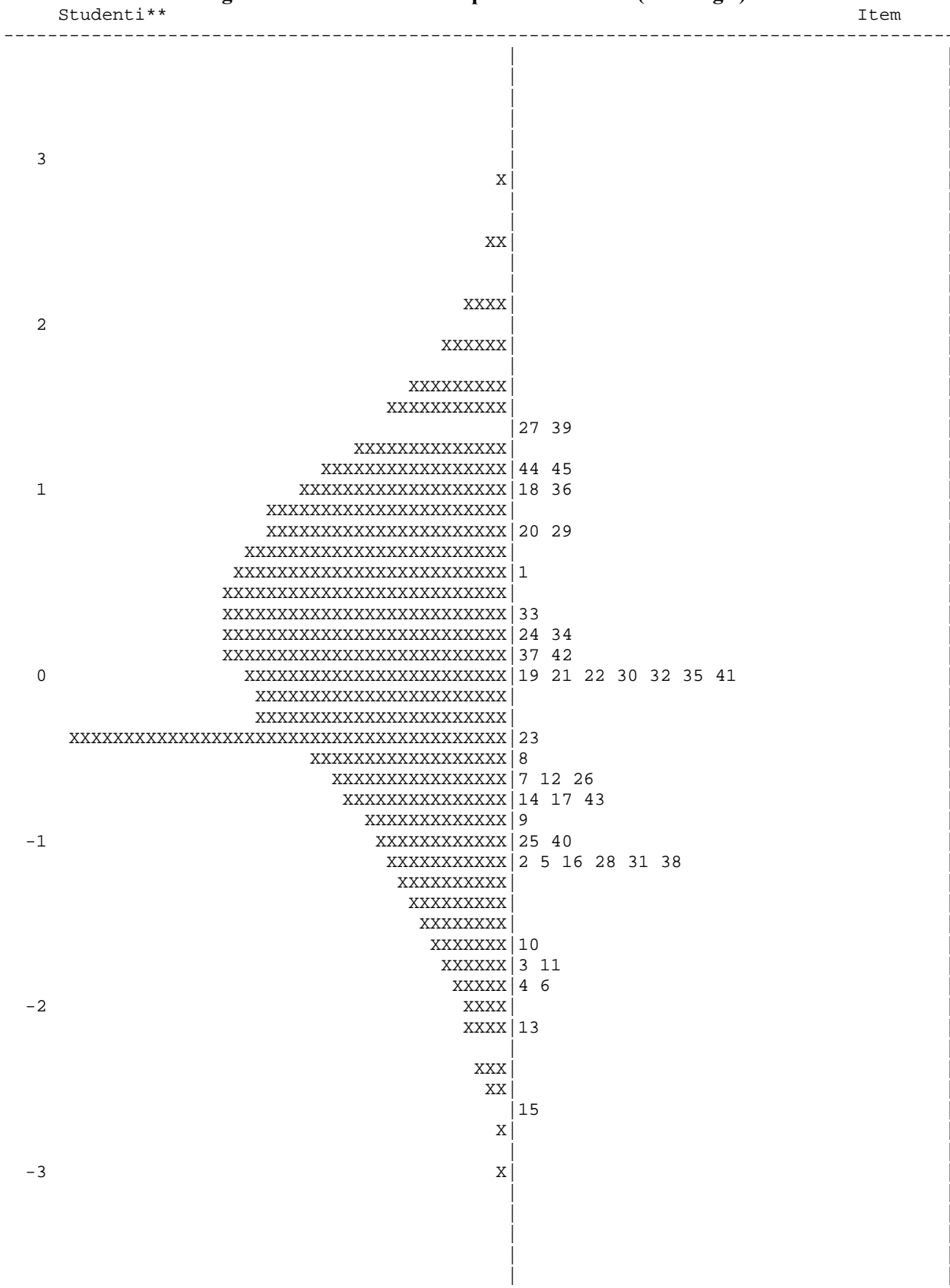
* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta - ** Ciascuna 'X' rappresenta 47,5 casi

Figura 5.4: *Placement* della prova di Matematica (V primaria)*



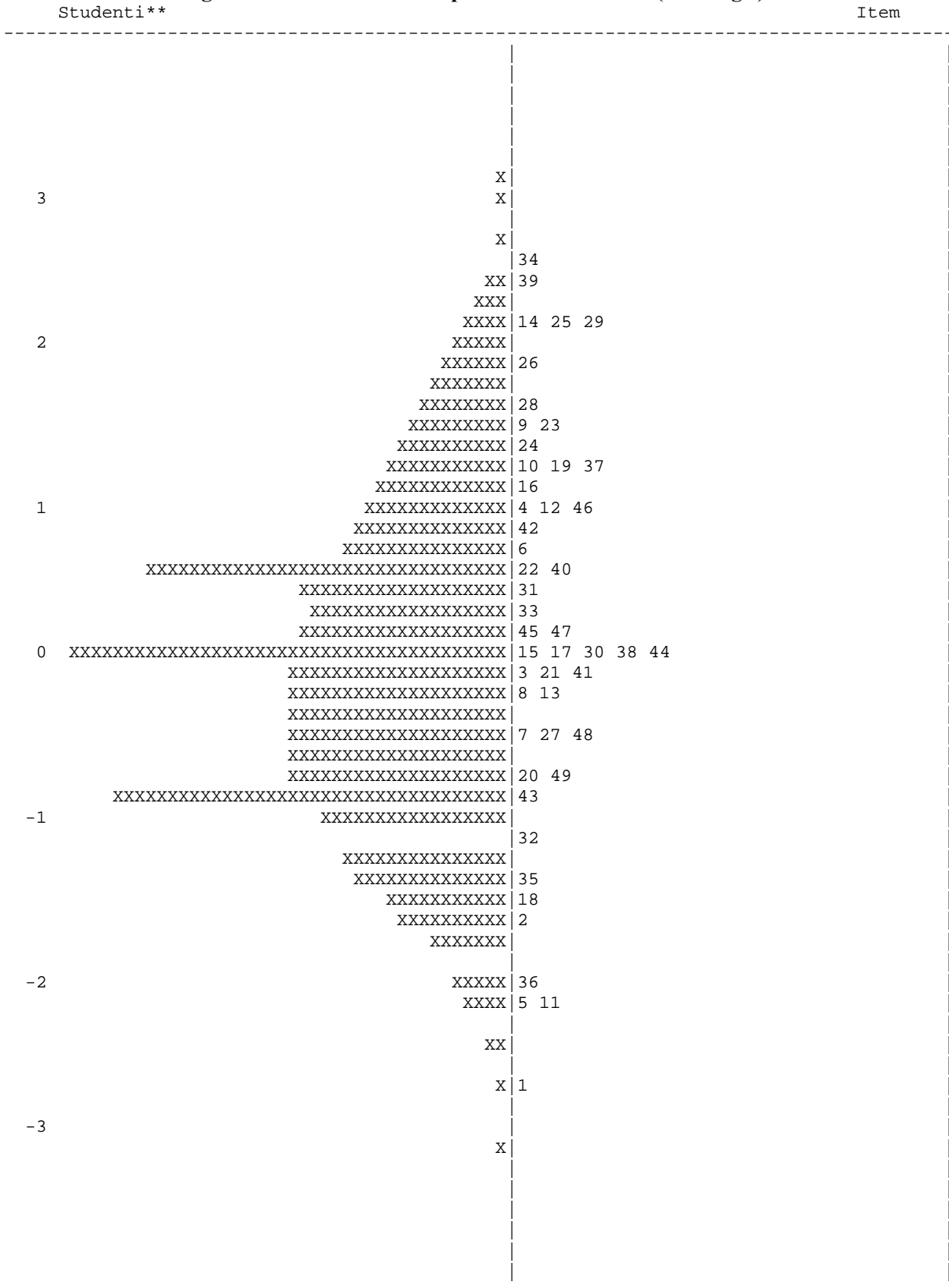
* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta - ** Ciascuna 'X' rappresenta 55,3 casi

Figura 5.5: *Placement* della prova di Italiano (I sec. I gr.)*



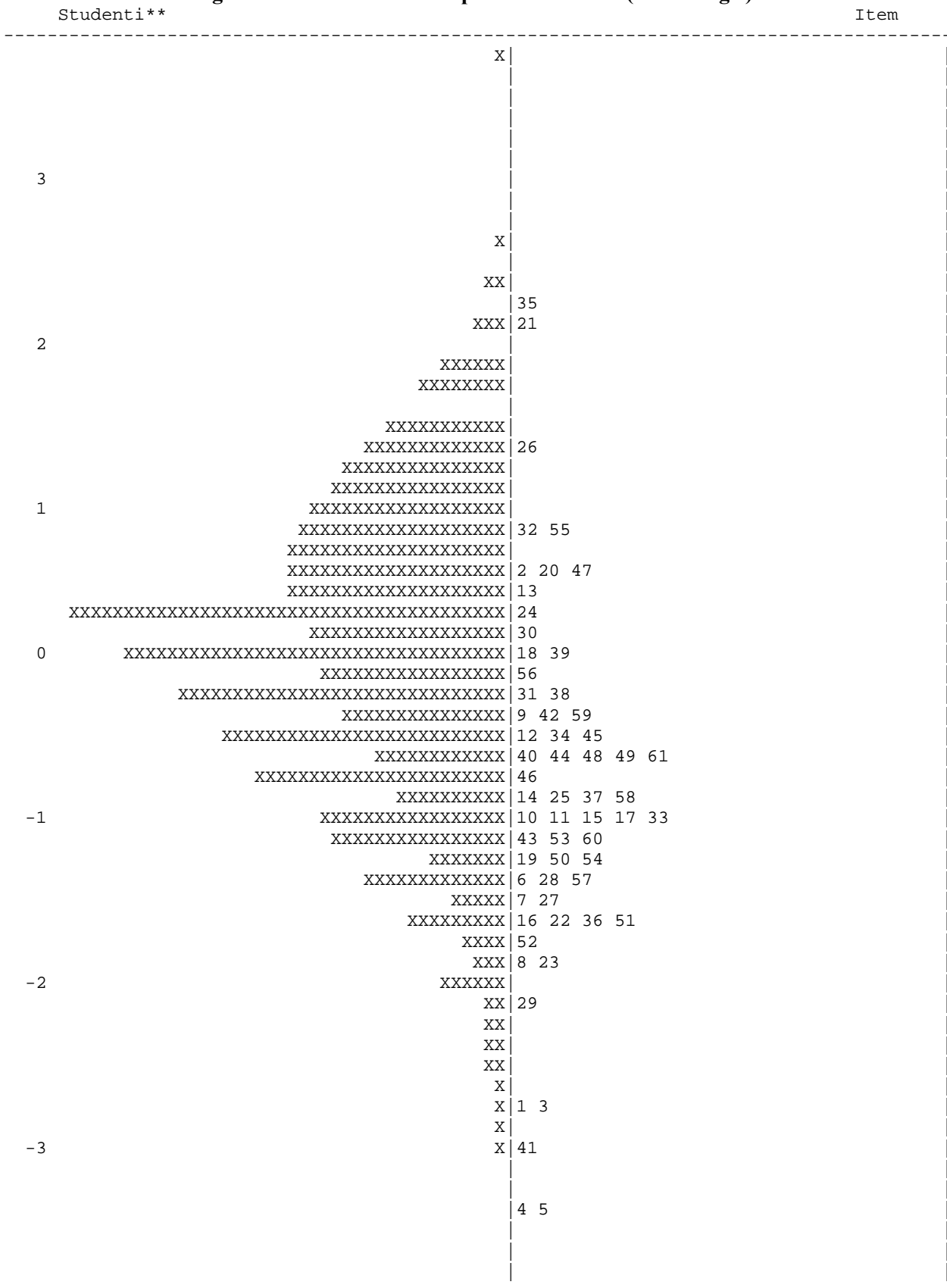
* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta - ** Ciascuna 'X' rappresenta 73,8 casi

Figura 5.6: *Placement* della prova di Matematica (I sec. I gr.)*



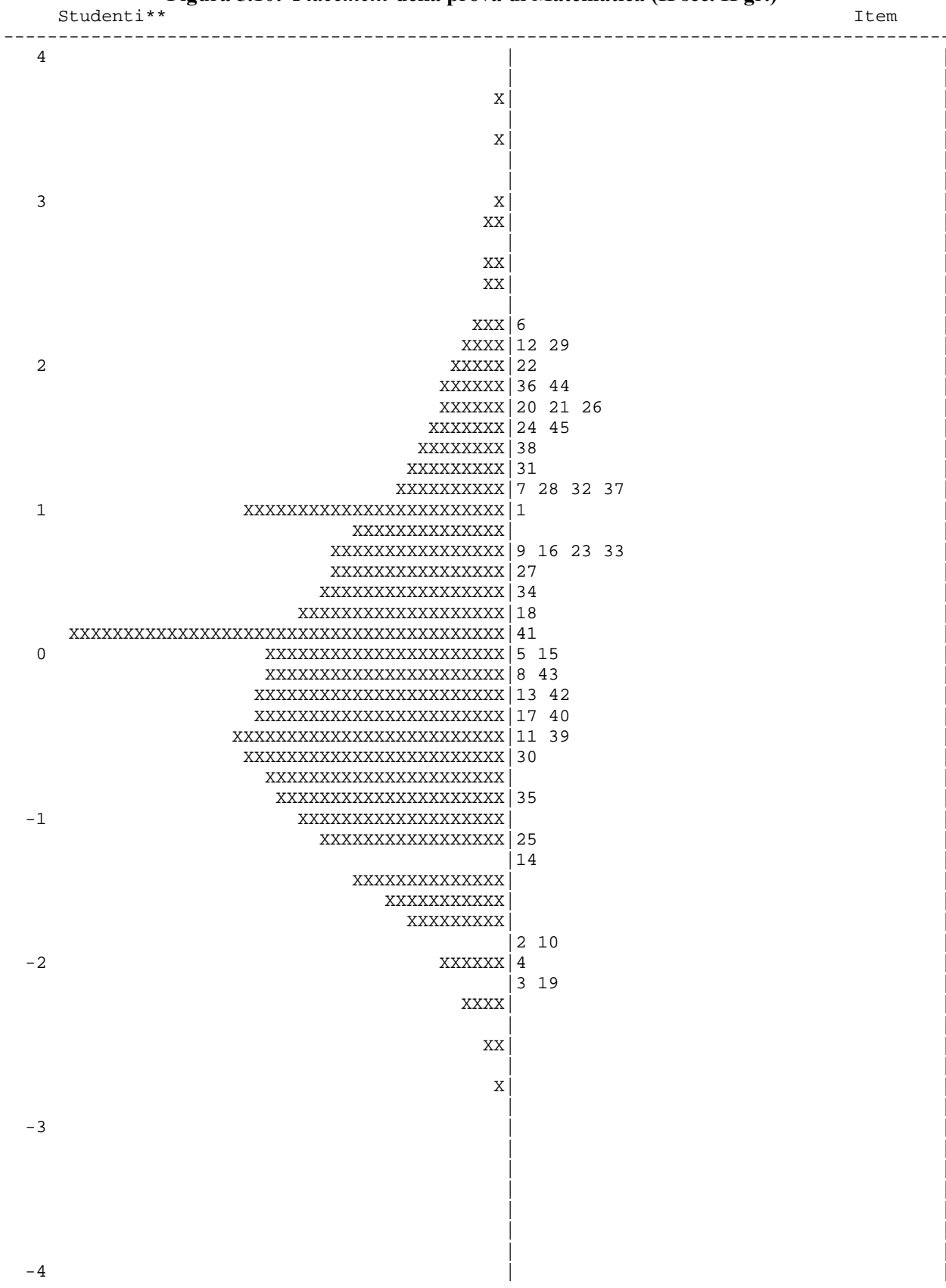
* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta ** Ciascuna 'X' rappresenta 80,3 casi

Figura 5.9: *Placement* della prova di Italiano (II sec. II gr.)*



* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta - ** Ciascuna 'X' rappresenta 85 casi

Figura 5.10: *Placement* della prova di Matematica (II sec. II gr.)*



* Il grafico rappresenta la mappa delle stime WLE e delle stime dei parametri del modello di risposta - ** Ciascuna 'X' rappresenta 86,4 casi

I grafici riportati nelle figure 5.1-5.10 mostrano come la maggior parte degli allievi si collochi, come prevedibile, ai livelli intermedi della scala di difficoltà/abilità, ma anche come, sebbene in minoranza, siano presenti allievi sia ai livelli più bassi sia ai livelli più alti. Una prova standardizzata e finalizzata alla misurazione degli apprendimenti a livello di sistema, pertanto, deve essere in grado di coprire, nel modo più completo possibile, l'intera scala di difficoltà, in modo tale da fornire informazioni sia sugli studenti meno o mediamente abili, sia su quelli più abili.

Se si osservano in sequenza le figure 5.1-5.10 si può notare, come già accennato in precedenza, la differente articolazione della difficoltà delle prove nei diversi livelli di scolarità, al fine di realizzare quanto suggerito dalla letteratura specifica nel campo della costruzione delle prove standardizzate, ovvero costruire prove adatte all'età degli alunni, non solo in termini di contenuti ma anche in termini di difficoltà.

Riferimenti bibliografici

- Baker, F.B. (2001), *The Basics of Item Response Theory*, ERIC Clearinghouse on Assessment and Evaluation.
- Baker, F.B., Kim S., (2004), *Item Response Theory. Parameter Estimation Techniques*, New York, Marcel Dekker.
- Barbaranelli, C. & Natali, E. (2005). *I test psicologici: teorie e modelli psicometrici*. Roma: Carrocci Editore.
- Boomsma, A., van Duijn M.A.J, Snijders, T.A.B.(1991), *Essays on Item Response Theory*, New York, Springer.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Falocci N., Gnaldi, M., Matteucci, M. & Mignani, S. (2010). *La validazione statistica di test standardizzati di profitto: principali aspetti di metodo e due casi di studio sulla valutazione degli apprendimenti nella scuola primaria*. Collana Working Paper INVALSI, 9. http://www.invalsi.it/download/wp/wp09_Falocci.pdf
- Fischer, G.H., Molenaar, I.W.(1995), *Rasch Models: Foundation, Recent Developments, and Application*, New York, Springer.
- Haladyna, T. M. (1994), *Developing and validating multiple-choice test items*, Hillsdale, N.J: L. Erlbaum Associates.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J.(1991), *Fundamentals of Item Response Theory*, Newbury Park, CA, Sage.
- Hambleton, R.K., Swaminathan,H., Rogers, H.J., (1991), *Fundamentals of Item Response Theory*, Sage.
- INVALSI (2011a). *Quadro di riferimento di Italiano*. (<http://www.invalsi.it/snv1011/>)
- INVALSI (2011b). *Quadro di riferimento di Matematica*. (<http://www.invalsi.it/snv1011/>)
- Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., & Erberber, E. (2005).
- OCSE (2007). *Valutare le competenze in scienze, lettura e matematica*. Roma: Armando Editore.
- OECD, (2009), *PISA 2006 Technical Report*, Parigi, OECD.
- Osterlind, S. J. (1983), *Test item bias*, Beverly Hills, Sage Publications.
- Osterlind, S. J. (1998), *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*, Boston, Kluwer Academic Publishers.
- *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.



