

Percorsi secondari di una prova adattativa multilivello e valutazione formativa

Secondary paths of a multilevel adaptive test and formative assessment

Emanuela Botta*

Riassunto

In questo articolo si illustra il processo di analisi delle risposte fornite dagli studenti in una prova adattativa multilivello *computer based* per la stima delle abilità in matematica nel grado 10 del sistema di istruzione italiano. La stima è stata effettuata tramite una prova adattativa multilivello 1 - 3 - 3 costruita nell'ambito di una ricerca di dottorato in Psicologia Sociale, dello Sviluppo e della Ricerca Educativa. Nella prima parte dell'articolo si descrivono la struttura della prova e il suo funzionamento, oltre ai vantaggi che essa offre rispetto a una classica prova lineare, quali una maggiore precisione nella stima delle abilità e la possibilità di confrontare direttamente i risultati ottenuti da studenti che hanno svolto percorsi differenti. In seguito, si analizza l'andamento degli studenti in due dei percorsi secondari della prova sia per mettere in luce in quali quesiti essi hanno incontrato maggiori difficoltà, nell'ottica di una valutazione formativa, sia per indagare i motivi per cui in una data fase della prova gli studenti hanno mostrato un inatteso calo nel rendimento. Nell'analisi dei risultati di una prova adattativa multilivello lo studio dei percorsi secondari è di particolare interesse perché in essi ricadono gli studenti per i quali la stima dell'abilità varia nel passaggio da un livello all'altro ed è dunque ipotizzabile la parziale acquisizione di alcune conoscenze e abilità.

Parole chiave: test adattativi, test multilivello, matematica, valutazione formativa, *computer based*

Abstract

In this paper we illustrate the process of analyzing the responses provided by students in a computer-based multilevel adaptive test for the estimation of

* Università degli Studi di Roma "La Sapienza", Dipartimento di Psicologia dei Processi di Sviluppo e Socializzazione.

mathematics skills in grade 10 of the Italian education system. The estimation was carried out by means of a multilevel adaptive test 1 - 3 - 3 constructed within the framework of a doctoral research in Social Psychology, Developmental Psychology and Educational Research. In the first part of the article we describe the structure of the test and how it works, as well as the advantages it offers over a classic linear test, such as greater precision in the estimation of abilities and the possibility of directly comparing the results obtained by students who have followed different paths. Then, we analyze the students' performance in one of the secondary pathways of the test both to highlight in which questions they encountered greater difficulties, with a view to a formative assessment, and to investigate the reasons why in a given phase of the test the students showed an unexpected drop in performance. In the analysis of the results of a multilevel adaptive test, the study of secondary paths is of particular interest because they include students for whom the estimate of ability varies in the passage from one level to another and it is therefore possible to hypothesize the partial acquisition of some knowledge and skills.

Keywords: adaptive test, multistage test, mathematics skills, formative assessment, computer-based, MST

Articolo sottomesso: 30/09/2021; accettato: 25/11/2021

Disponibile online: 23/12/2021

1. Il grado 10 nel sistema scolastico italiano

In questo articolo si riportano alcuni risultati che provengono dall'approfondimento di una ricerca che per la prima volta ha realizzato in Italia un modello adattativo multilivello per la stima delle abilità in matematica degli studenti del grado 10 (Botta, 2021).

Il grado 10 del sistema di istruzione italiano, secondo anno della scuola secondaria di secondo grado, è caratterizzato da un bacino di utenza ampio e variegato, distribuito su tre percorsi di studio: liceale, tecnico e professionale.

Facendo riferimento alle relative Linee guida e Indicazioni nazionali per l'insegnamento della matematica è evidente che gli obiettivi di apprendimento in termini di conoscenze, abilità e competenze si differenziano solo in minima parte fra un percorso e l'altro, per cui è sensato, ad esempio, che l'INVALSI, definisca un Quadro di Riferimento unico per il grado 10 e lo utilizzi annualmente per le rilevazioni nazionali. Ma a prescindere dalla previsione normativa è chiaro che la popolazione di riferimento raccoglie studenti con livelli di abilità che coprono un esteso intervallo di riferimento.

A titolo esemplificativo si può considerare che nella popolazione studentesca ci sono studenti con Bisogni Educativi Speciali (BES), cioè con disabilità, disturbi specifici dell'apprendimento (DSA) o svantaggio socioeconomico, linguistico, culturale.

Secondo i dati del Ministero dell'istruzione (MI, 2020) nell'anno scolastico 2020/21 gli studenti della scuola secondaria di II grado erano 2.635.110 di cui circa il 2,96% con disabilità e circa il 7,67%, con cittadinanza non italiana¹. Il Ministero dell'istruzione (MIUR, 2019) fornisce anche i dati relativi agli studenti con DSA, relativamente però all'anno scolastico 2017/18: nella scuola secondaria di secondo grado erano circa il 4,68% del totale. Complessivamente, al netto delle variazioni intercorse fra le due annualità considerate, circa il 16% degli studenti del grado 10 ha bisogni educativi speciali, considerando che da questa stima restano esclusi parte degli studenti con svantaggio socioeconomico, linguistico, culturale.

La varietà della popolazione di riferimento ha portato alla considerazione che per stimare accuratamente le abilità degli studenti era opportuno costruire una prova che superasse la tradizionale struttura lineare che non consente di stimare l'abilità del singolo alunno con una precisione soddisfacente per quella parte della popolazione che si discosta significativamente dalla media nazionale. Le prove lineari standardizzate, utilizzate spesso per la valutazione sommativa o di sistema, hanno il limite di mettere tutti gli alunni di fronte alle stesse domande e di valutare il risultato ottenuto in relazione a un valore di riferimento, inadatto per definizione, a misurare ciò che si discosta dalla norma.

In generale una prova lineare ha lunghezza fissa e tutti i soggetti rispondono alle medesime domande. Questo comporta che molti soggetti si trovano a rispondere ad item che hanno un livello di difficoltà troppo alto o troppo basso rispetto al loro livello di abilità e che sono dunque poco informativi dal punto di vista psicometrico e contribuiscono poco alla precisione della stima dell'abilità del soggetto. In letteratura le prove lineari sono distinte in due tipologie, le prove centrate, in cui la prova è formata da un insieme di item la cui difficoltà è prossima a un certo livello di difficoltà stabilito a priori e le prove rettangolari, formate dallo stesso numero di item per ciascun livello di difficoltà scelto opportunamente all'interno di un intervallo dato (Weiss, 1985). Ci si trova allora a dover scegliere se utilizzare le prove centrate, misurando molto bene attorno al valore centrale dell'abilità e progressivamente e rapidamente sempre peggio a mano a mano che ci si allontana da esso, o utilizzare le prove rettangolari, misurando tutti i soggetti con un grado di precisione piuttosto basso. Inoltre, la lunghezza fissa della prova limita le possibilità di scelta degli item, riducendone

¹ Valore atteso, calcolo previsionale basato sui dati delle rilevazioni sulle scuole riferiti agli anni scolastici precedenti.

lo spettro di variabilità in relazione al parametro di difficoltà e al contenuto. Questo si rivela molto problematico quando i risultati necessitano di essere interpretati su scala individuale e la popolazione risulti molto variegata in termini di abilità, coprendo un ampio intervallo dei valori del tratto latente, come nel caso di nostro interesse. Ci sono numerose evidenze che i test adattativi, che regolano la difficoltà della prova in base al rendimento degli studenti durante la prova stessa, soddisfano queste due esigenze meglio rispetto a una prova lineare (Hambleton, Zaal e Pieters, 1991; Sands, Waters e McBride, 1997; Sireci, 2004; Wainer, 2000).

Dalla ricerca internazionale è emerso dunque che il modello adattativo risolve i problemi illustrati con la costruzione dinamica di una prova su misura per il soggetto, contenente item che hanno un livello di difficoltà adeguato al suo livello di abilità e che garantiscono quindi la massima precisione della misura. Gli item da somministrare a ciascun individuo vengono infatti selezionati sulla base di stime progressive della sua abilità, in modo che egli abbia sempre circa il 50% di probabilità di rispondere correttamente. In una prova adattativa classica la stima dell'abilità del soggetto viene ricalcolata ogni volta che egli risponde a una domanda considerando l'insieme di tutte le domande a cui ha risposto in precedenza, e tale stima viene utilizzata per la selezione della domanda successiva. Nel corso della prova se lo studente risponde correttamente gli verrà somministrato un item più difficile, viceversa, un item più facile. Il limite di questo tipo di prove è la difficoltà nel controllo della validità di contenuto, può accadere infatti che ad un soggetto vengano poste in prevalenza domande relative ad una certa area del costrutto di riferimento piuttosto che ad altre, limitando la possibilità di verificare l'abilità su più ambiti nel caso di costrutti complessi come quello della matematica. Una possibile soluzione a questo problema la offrono le prove adattative multilivello che si pongono come obiettivo proprio quello di coniugare la precisione della misura con la validità di contenuto (Weiss, 1985; Luecht, Brumfield e Breithaupt, 2006).

Di seguito si descrivono brevemente il modello di prova adattativa scelto e il costrutto della matematica adottato, indispensabili per comprendere i risultati ottenuti, che costituiscono la parte principale di questo lavoro.

2. Il disegno della prova multilivello

Un test multilivello (*Multistage Test*, MST) è caratterizzato da una struttura a livelli e moduli. Ogni test è suddiviso in alcuni livelli, ciascuno dei quali è costituito da uno o più moduli. I moduli dello stesso livello sono centrati su differenti valori del parametro di difficoltà. Ad ogni livello al soggetto viene somministrato un insieme di item, detto appunto modulo, pre-assemblato e centrato attorno a un valore della difficoltà adeguato all'abilità stimata del

soggetto. La stima dell'abilità del soggetto viene effettuata in uscita da ogni livello sulla base delle risposte che ha fornito in tutti i livelli precedenti. L'insieme di moduli che un dato soggetto riceve è detto percorso.

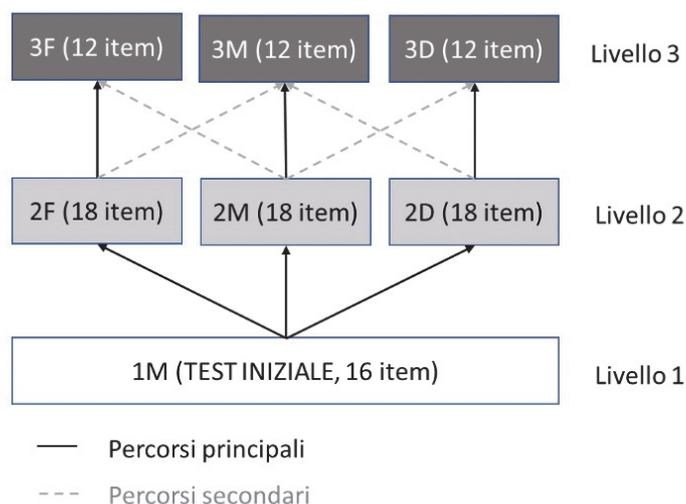
Il disegno della prova può variare in relazione al numero dei livelli, al numero di moduli che costituiscono ciascun livello, al numero di item per modulo e ai percorsi che è concesso di fare passando da un livello all'altro.

La realizzazione di questo tipo di prova richiede prioritariamente la costruzione di una banca di item (Choppin, 1976) che può influenzarne significativamente la struttura, in particolare in relazione alla disponibilità di item con date caratteristiche psicometriche o di contenuto. La banca di item è generalmente calibrata in relazione ad un modello di *Item Response Theory* (IRT) a uno o più parametri.

Considerando la composizione della banca di item costruita nel corso della ricerca², in particolare la disponibilità di item in relazione al parametro di difficoltà, si è ritenuto che fosse opportuno suddividere l'insieme degli item in tre classi di difficoltà, soddisfacenti per ottenere tre profili distinti del livello di abilità.

Il tempo a disposizione per la somministrazione della prova (1 ora e 30 minuti) e il livello di precisione della stima dell'abilità che si voleva ottenere sono stati i principali elementi tenuti in considerazione per la definizione del numero di livelli.

Figura 1 - Modello MST implementato



² La banca è stata calibrata secondo il modello IRT a un parametro, la difficoltà dell'item.

Si è inoltre preso in considerazione lo studio di Luecht e Nungester (1998) nel quale si verifica che per una serie di ragioni operative, psicometriche e di sicurezza, una delle configurazioni più efficienti è il modello MST a 3 livelli. Si è optato quindi per la realizzazione di un modello MST 1 – 3 – 3, come quello illustrato in figura 1, in cui ciascun percorso fosse composto da 46 item e bilanciato rispetto ai contenuti.

Le lettere F, M e D indicano la difficoltà media di ogni modulo (F: bassa, M: media, D: alta).

In termini di navigazione all'interno della prova è stato stabilito che nel passaggio da un livello all'altro non fosse possibile fare passi avanti o indietro di lunghezza superiore ad un unico livello di difficoltà e che all'interno di ciascun livello fosse possibile navigare rivedendo gli item e modificando le risposte, funzionalità inibita invece tra un livello e l'altro.

I percorsi che si possono fare all'interno della prova sono indicati dalle frecce. Ogni soggetto compie un solo percorso composto dal modulo iniziale, un modulo del livello 2 e un modulo del livello 3. Come si può osservare in figura sono previsti due tipi di percorso, i percorsi principali, indicati dalle linee continue, per i quali nel passaggio dal livello 2 al livello 3 la classe di assegnazione del soggetto non cambia, e quelli secondari, per i quali invece la classe di assegnazione cambia passando dal livello 2 al livello 3.

Complessivamente sono possibili 7 diversi percorsi, ciascuno dei quali ha come esito la stima dell'abilità dello studente all'interno di un dato intervallo del continuo dell'abilità. Per l'ottimizzazione delle proprietà misuratorie si è scelto di massimizzare la funzione informativa di ogni modulo. Il numero di item in ciascun modulo è stato definito in modo da garantire una reale differenziazione dei percorsi all'interno della prova ed evitare il rischio che la stima dell'abilità fosse predeterminata dal modulo iniziale.

3. Il costrutto

Il costrutto è stato definito a partire dai principali riferimenti internazionali, come l'OCSE PISA, e nazionali, come il quadro di riferimento dell'INVALSI, le Linee guida e le Indicazioni nazionali per il grado 10, ed è stato adattato alle necessità di produzione di una prova adattativa *computer based*. È stato necessario considerare i vincoli posti dalla natura e dal formato della prova, pur essendo essa strettamente collegata agli obiettivi di apprendimento ordinamentali.

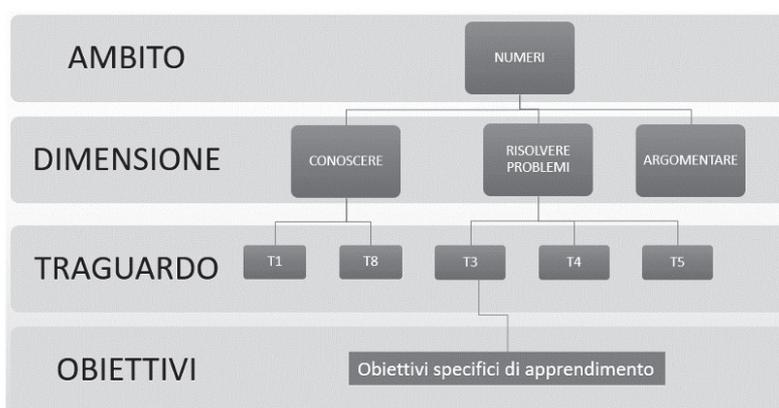
Una prova strutturata, contenente quesiti a scelta multipla, a risposta univoca e *cloze*, non è adatta a misurare gli aspetti metacognitivi e motivazionali propri della competenza matematica, così come i caratteri dell'autonomia e della responsabilità dello studente. Questa prova misura dunque conoscenze e abilità

in matematica e la capacità di utilizzarle nel risolvere problemi e nell'argomentare. Ad esempio, si richiede allo studente di scegliere l'argomentazione più adeguata a sostenere una determinata tesi, di stabilire il valore di verità di una proposizione o di completare una dimostrazione inserendo alcune parole. Relativamente alla formalizzazione si richiede di individuare la modellizzazione più adatta a rappresentare una situazione problematica e nella risoluzione di problemi non si richiede di riportare i passaggi effettuati ma solamente la soluzione trovata.

Il costrutto è articolato in 4 Ambiti di contenuto, Numeri (NU), Dati e previsioni (DP), Relazioni e funzioni (RF), Spazio e figure (SF), in tre dimensioni Conoscere, Risolvere Problemi e Argomentare, in 12 Obiettivi di apprendimento generali, detti Traguardi, e in numerosi obiettivi specifici di apprendimento.

In figura 2 è rappresentato un esempio di articolazione del costrutto.

Figura 2 - Esempio di articolazione del costrutto dell'abilità matematica



4. La somministrazione della prova sul campo e i risultati

4.1. Il campione

La somministrazione della prova sul campo ha coinvolto un campione di 4132 studenti, di cui 278 con BES, frequentati istituti tecnici, istituti professionali e licei.

Il campione è stato selezionato con un metodo a due stadi. Al primo stadio, in ognuna delle tre grandi aree geografiche italiane, Nord, Centro, Sud e Isole, è stato individuato un campione di giudizio di 82 scuole. Al secondo stadio, in

ciascuna scuola è stato selezionato un campione di classi, almeno due per ogni scuola. La tabella che segue mostra la ripartizione del campione nelle macroaree geografiche.

Tabella 1 - Distribuzione del campione nelle macroaree geografiche

Macroarea geografica	Distribuzione percentuale	Numerosità del campione
Nord	34	1408
Centro	34	1406
Sud - Isole	32	1318
Totale Italia	100	4132

La tabella 2 mostra la ripartizione del campione sui vari percorsi di studio.

Tabella 2 - Distribuzione del campione fra i percorsi di studio

Percorso di studio	Distribuzione percentuale	Numerosità del campione
Istituti Professionali	30,3	1253
Istituti Tecnici	31,1	1286
Licei	38,6	1593
Totale	100,0	4132

4.2. Studio dei percorsi secondari

I percorsi secondari, quelli in cui lo studente varia la classe di assegnazione dell'abilità durante lo svolgimento della prova, sono stati svolti da un numero considerevolmente ridotto di studenti. Si è ritenuto comunque che fosse utile studiarli nel dettaglio per cercare di comprendere più a fondo il funzionamento della prova e il comportamento degli studenti in un test adattativo MST.

Si è scelto inoltre di utilizzarli come esempio di un uso formativo della valutazione, inteso come individuazione di errori ricorrenti o misconcezioni diffuse per identificare specifiche aree di intervento da condividere con gli studenti affinché possano trarne beneficio in modo diretto. L'accettazione degli errori come fonti essenziali di regolazione e di avanzamento (Astolfi, 1997) si realizza, infatti, solo a condizione che essi vengano analizzati e compresi. La restituzione del *feedback* (Hattie e Timperley, 2007, Black e Wiliam, 2009) è dunque intesa come parte integrante del processo di valutazione e può essere realizzata dal docente a posteriori oppure, sulla base di quanto rilevato nelle fasi di pre-test o nelle precedenti somministrazioni del test, progettata all'interno del modello come restituzione automatica allo studente al termine dello svolgimento della prova.

In particolare, sono stati analizzati il Percorso 3, 1M – 2M – 3F, e il Percorso 6, 1M – 2D – 3M. In entrambi i casi, nel passaggio dal livello 2 al livello 3, gli studenti subiscono una riduzione dell'abilità stimata. L'analisi del Percorso 6 è un'integrazione della ricerca iniziale che ha confermato le osservazioni fatte in

prima istanza. Il Percorso 3 è stato svolto da 162 studenti, mentre il Percorso 6 è stato svolto da 371 studenti.

Prioritariamente è stata osservata la distribuzione degli studenti in relazione al numero di risposte corrette fornite in ciascun modulo del percorso, in relazione alle soglie stabilite nel processo di navigazione per muoversi all'interno di esso. Di seguito si illustrano i risultati ottenuti.

Nel Percorso 3 le soglie di riferimento per accedere a 2M in uscita da 1M sono di un minimo di 7 e un massimo di 10 risposte corrette, su un totale di 16 item.

Tabella 3 - Distribuzione degli studenti del Percorso 3 in relazione al numero di risposte corrette nel modulo 1M

Numero Risposte Corrette	Frequenza	Percentuale
7	100	61,7
8	37	22,8
9	23	14,2
10	2	1,2
Totale	162	100,0

Come si può osservare la maggior parte degli studenti accede al Percorso 3 con il numero minimo di risposte corrette, e questo connota tali studenti come soggetti al limite inferiore del livello di abilità per affrontare un percorso di difficoltà media. Solo una percentuale molto piccola di studenti, l'1,2%, si colloca all'estremo opposto.

Nel Percorso 6 le soglie di riferimento per accedere a 2D in uscita da 1M sono di un minimo di 11 e un massimo di 16 risposte corrette, sempre su un totale di 16 item.

Tabella 4 - Distribuzione degli studenti del Percorso 6 in relazione al numero di risposte corrette nel modulo 1M

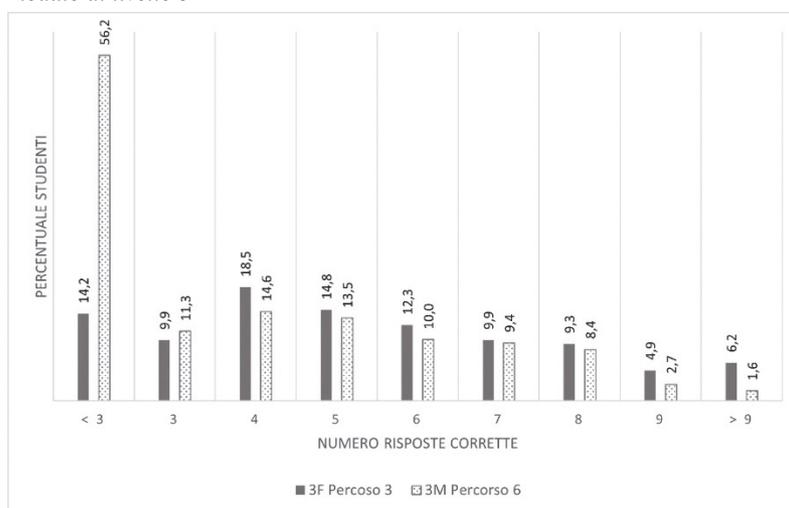
Numero Risposte Corrette	Frequenza	Percentuale
11	191	51,5
12	119	32,1
13	48	12,9
14	10	2,7
15	3	0,8
16	0	0
Totale	371	100,0

Anche in questo caso la maggior parte degli studenti accede al Percorso 6 con il numero minimo di risposte corrette, quindi con un livello di abilità appena sufficiente ad affrontare poi un modulo di difficoltà alta. La percentuale di studenti che si colloca all'estremo opposto è estremamente ridotta e pari allo 0,8%.

Analogamente si è proceduto per i moduli di livello 2, nel Percorso 3 il modulo 2M e nel Percorso 6 il modulo 2D.

Le soglie di riferimento per accedere a 3F in uscita da 2M sono rispettivamente di un minimo di 0 e un massimo di 5 risposte corrette. Osservando la distribuzione degli studenti in relazione al numero di risposte corrette fornite nel modulo 2M si riscontra che una piccola parte degli studenti fornisce da 0 a 2 risposte corrette, mentre tutti gli altri si suddividono in maniera abbastanza uniforme fra le 3 e le 5 risposte corrette.

Figura 2: Distribuzione degli studenti dei Percorsi 3 e 6 in relazione al numero di risposte corrette nel modulo di livello 3



Questo dato va interpretato ricordando che il modulo 2M è composto da 18 item e che le soglie indicate mettono in evidenza che lo studente di fatto non è in grado di affrontare item di difficoltà media, se non in minima parte, 5 risposte corrette su 18 rappresentano infatti meno del 28% del totale.

Le soglie di riferimento per accedere a 3M in uscita da 2D sono rispettivamente di un minimo di 0 e un massimo di 6 risposte corrette. Osservando la distribuzione degli studenti in relazione al numero di risposte corrette fornite nel modulo 2D si rileva una situazione molto simile alla precedente, infatti la percentuale di studenti che fornisce da 0 a 2 risposte corrette è pari a circa il 25%, c'è poi una distribuzione abbastanza uniforme fra le 3 e le 5 risposte corrette e solo pochi studenti, il 9,4%, forniscono 6 risposte corrette.

Particolarmente interessante si è rivelato osservare la distribuzione degli studenti in relazione al numero di risposte corrette nei moduli di terzo livello, 3F nel Percorso 3 e 3M nel Percorso 6. A questo punto infatti l'adattatività della

prova MST ci consente di correggere il tiro e di riavvicinare la difficoltà degli item al livello di abilità dello studente, ottenendo il massimo dell'accuratezza nella misura. I dati sono riportati nel grafico in figura 2.

È opportuno ricordare che essendo all'ultimo livello della prova il numero di risposte corrette può variare fra 0 e 12, il numero totale di item che compongono il modulo.

I grafici nelle figure 3 e 4 illustrano i risultati in termini di numero medio di risposte corrette in ciascun modulo dei due percorsi. Dalla loro analisi, si può osservare che in entrambi i casi il numero medio di risposte corrette nel livello 3 aumenta rispetto a quello del modulo di livello 2. Nel caso del Percorso 3 il valore medio torna quasi a quello del livello 1, mentre nel Percorso 6 si mantiene decisamente più basso. In entrambi i casi occorre osservare, come detto all'inizio, che nel passaggio dal livello 2 al livello 3 la difficoltà dei moduli si riduce, nel Percorso 3 è persino inferiore a quella del modulo iniziale, teoricamente dunque ci saremmo attesi un miglioramento nel rendimento degli studenti nell'ultima parte del percorso che invece non c'è stato.

Si ritiene che le ragioni del verificarsi di questo fenomeno potrebbero essere ricondotte, almeno in parte, agli studi di Kahneman (2012) relativi allo sforzo cognitivo e alla deplezione dell'io.

Rispondere ad item con indice di difficoltà mediamente più alto del livello di abilità dello studente richiede concentrazione e controllo dell'attenzione che generano stanchezza.

Figura 3 - Media del numero di risposte corrette in ciascuno dei moduli del Percorso 3

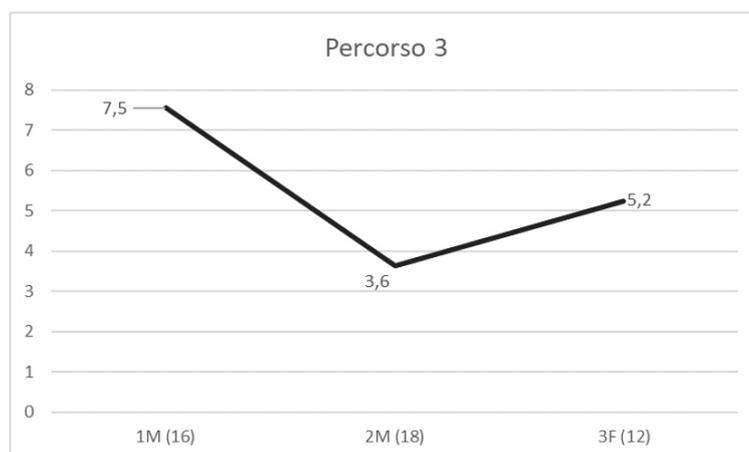
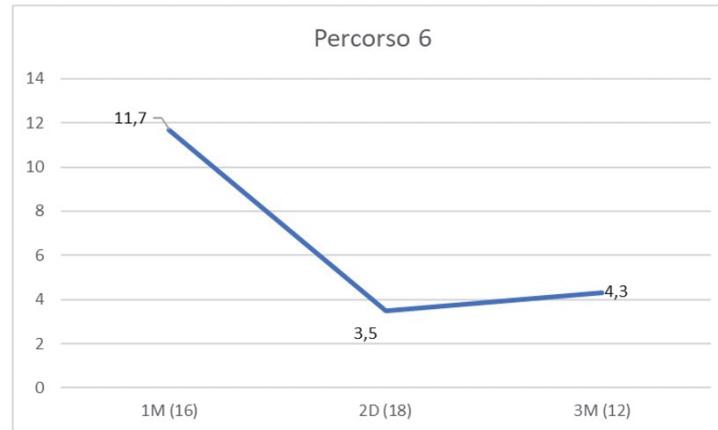


Figura 4 - Media del numero di risposte corrette in ciascuno dei moduli del Percorso 6



A questo punto si è meno disposti o meno capaci di esercitare l'autocontrollo di fronte all'esigenza di affrontare un nuovo problema.

Fra i sintomi della deplezione dell'io Kahneman individua l'ottenere risultati mediocri in compiti cognitivi e processi decisionali logici. Egli afferma inoltre che

[...] attività molto impegnative per il sistema 2³ richiedono autocontrollo, e l'esercizio dell'autocontrollo è sempre spiacevole e "svuotante". Diversamente dal carico cognitivo, la deplezione dell'io è, almeno in parte, una perdita di motivazione. Dopo aver esercitato l'autocontrollo per affrontare un compito, non abbiamo più voglia di sottoporci a sforzo per affrontarne un altro, anche se ne saremmo in grado ove fosse proprio necessario. In diversi esperimenti, i soggetti resistettero agli effetti della deplezione quando fu dato loro un forte incentivo (p. 55).

Come osservato nell'ambito della ricerca sul Percorso 3 (Botta, 2021) possiamo ipotizzare che la lunghezza dei moduli di livello 2 e la difficoltà dei loro item,

³ Kahneman (Kahneman, 2012) adotta per le due principali modalità del pensiero i termini Sistema 1 e Sistema 2.

- Sistema 1. Opera in fretta e automaticamente, con poco o nessuno sforzo e nessun senso di controllo volontario.
- Sistema 2. Indirizza l'attenzione verso le attività mentali impegnative che richiedono focalizzazione, come i calcoli complessi. Le operazioni del sistema 2 sono molto spesso associate all'esperienza soggettiva dell'azione, della scelta, della concentrazione.

abbiano portato al manifestarsi negli studenti del fenomeno della deplezione dell'io, o anche solo sforzo cognitivo, con una conseguente perdita di motivazione, non compensata, nel caso specifico, da alcun incentivo a resistere alla deplezione per ottenere risultati migliori (p.149).

Un'altra riflessione nasce dal considerare il processo di valutazione come parte integrante dell'apprendimento. In quest'ottica le osservazioni fatte portano a supporre che l'esperienza di svolgimento della prova per questi studenti non sia stata un'esperienza educativa nel senso di Dewey (1949). Affinché lo diventi e sia parte feconda e creativa delle esperienze future è fondamentale che il processo si espliciti nelle modalità della valutazione formativa.

4.3. Analisi degli item in ottica di valutazione formativa

Dal punto di vista dell'analisi degli errori e dell'individuazione delle difficoltà in vista della restituzione di un costruttivo *feedback* agli studenti sono stati individuati e verranno descritti gli item dei due percorsi che hanno ottenuto il minor numero di risposte corrette.

Tralasciando le domande più difficili si può osservare che fra le domande di IM che mettono maggiormente in difficoltà gli studenti con livello di abilità più scarso (Percorso 3), ci sono tre domande con indice di difficoltà piuttosto basso, la nona ($b = -0,409$), l'undicesima ($b = -0,542$) e la tredicesima ($b = -0,678$) su 16, rispetto all'ordinamento decrescente in relazione ai valori del parametro di difficoltà b . Due appartengono all'ambito Dati e previsioni e una all'ambito Numeri.

Entrando nel merito si evidenziano difficoltà nel riconoscere la modalità corretta per il calcolo di una media ponderata, la probabilità dell'evento contrario di un evento la cui probabilità è nota e la trasformazione di semplici formule o espressioni algebriche, più precisamente riconoscere l'espressione algebrica equivalente al prodotto di due binomi di primo grado. I primi due, facendo ancora riferimento a Kahneman, potrebbero essere ricondotti all'idea che generalmente non siamo dei buoni statistici intuitivi, il che richiede che gli studenti utilizzino una forma di pensiero più lenta, riflessiva e impegnativa. Un ulteriore elemento che può aver influenzato la risposta ai primi due item è una scarsa euristica della disponibilità, il che rende difficile la ricerca nella memoria di problemi simili o associabili a quello da risolvere, considerando che la statistica ancora stenta ad entrare diffusamente nelle prassi didattiche consolidate della scuola. Le difficoltà nel calcolo algebrico sono invece abbastanza sorprendenti, considerando che l'argomento è ampiamente trattato sui libri di testo più diffusi e probabilmente affrontato nelle classi con notevole frequenza. Per gli studenti del Percorso 6 il quesito che ha avuto il minor numero di risposte corrette è il

quinto in ordine di difficoltà ($b = -0,208$). In questo quesito si richiede allo studente di individuare, fra quelli proposti, l'insieme delle soluzioni di una disequazione. In questo caso la difficoltà potrebbe risiedere nel fatto che il polinomio di secondo grado di cui si richiede l'intervallo di positività non è scomponibile nel campo dei numeri reali e questo impedisce una risoluzione esclusivamente procedurale della disequazione.

A livello 2 si osserva che nel modulo 2M (Percorso 3) due delle domande con indice difficoltà più basso sono fra quelle con uno scarso numero di risposte corrette. Una ($b = -0,444$) relativa alla risoluzione di una equazione lineare in un'incognita e l'altra ($b = -0,411$) che richiede di individuare l'espressione simbolica che descrive una relazione illustrata a parole. Il primo quesito si riferisce a una tematica ampiamente trattata già al termine della scuola secondaria di primo grado, mentre la seconda sembra essere rappresentativa della difficoltà di passare da un registro di rappresentazione ad un altro, benché in un caso piuttosto semplice nel quale allo studente è richiesto di scegliere la formulazione simbolica corretta fra le quattro proposte e non di procedere autonomamente alla sua scrittura. Nel modulo 2D (Percorso 6) le percentuali di risposte corrette variano fra il 9,7% e il 29,9%, laddove i quesiti di difficoltà confrontabile del modulo di ingresso, 1M, raggiungono percentuali di risposte corrette superiori al 50%. Ad esempio, uno degli item più facili del modulo 2D ($b = 0,379$), richiede allo studente di desumere, a partire dal grafico, il valore di una funzione a gradini noto il valore della variabile dipendente.

A livello 3 si ha che nel modulo 3F (Percorso 3) nessun item ha superato il 56% di risposte corrette, e due dei più facili sono fra quelli che hanno messo gli studenti maggiormente in difficoltà. Entrambi gli item sono dell'ambito Relazioni e funzioni e contestualizzati. Nel primo ($b = -1,774$) si richiede di comprendere e applicare la relazione che lega il consumo di una certa quantità di materiale al tempo: sapendo che la decrescita è uniforme e che il consumo in due ore è pari a quattro unità di materiale, si chiede qual è il consumo in un'ora. Nel secondo quesito ($b = -1,888$), in analogia a quanto già osservato nel modulo 2D, è dato il grafico che descrive un fenomeno e si chiede di individuare a partire da esso il valore della variabile indipendente noto quello della variabile dipendente. In entrambi i casi l'indice di difficoltà è significativamente più basso di quello degli altri esempi del Percorso 3, ma le percentuali di risposte corrette sono confrontabili con quelle degli item del modulo 1M e decisamente superiori a quelle degli item del modulo 2M.

Nel modulo 3M (Percorso 6) le difficoltà si concentrano su tre item di Spazio e Figure, con difficoltà che va da $b = -0,395$ fino a $b = 0,166$. Anche in questo caso i tre item hanno un indice di difficoltà più basso di tutti quelli del modulo 2D e anche di parte di quelli del modulo 1M. Le richieste sono sempre differenti: riconoscimento dell'immagine di una figura rispetto a una

trasformazione geometrica data, risoluzione di un problema sulla determinazione del perimetro e individuazione delle caratteristiche geometriche di una figura piana.

5. Conclusioni

L'articolo illustra la struttura e il funzionamento di una prova MST e il fatto che essa risulta essere significativamente più informativa di una prova lineare, consentendo di stimare con maggiore precisione l'abilità degli studenti entro un ampio intervallo del continuo. Di particolare interesse è l'analisi del Percorso 6, che conferma le osservazioni fatte nella ricerca originale (Botta, 2021) a partire dal Percorso 3: gli studenti a cui, nel livello 2 della prova, vengono somministrati item con un livello di difficoltà più alto del loro livello di abilità mostrano poi nel livello 3 un calo significativo del rendimento, hanno infatti difficoltà a rispondere anche a domande di difficoltà pari o inferiore sia rispetto al livello di abilità stimato sia rispetto ai quesiti cui hanno risposto nella prima parte della prova. Resta da verificare l'ipotesi che alla base di questo comportamento ci siano la deplezione dell'io o lo sforzo cognitivo (Kahneman, 2012). Nel complesso la riflessione potrebbe portare a mettere in discussione anche alcune procedure usualmente utilizzate nella strutturazione delle prove lineari, come l'alternarsi di blocchi di domande di diversa difficoltà. Dagli esempi illustrati nel paragrafo 4.3 si evince che dall'analisi degli item e degli errori degli studenti è possibile individuare gli obiettivi di apprendimento difficili da raggiungere per un dato gruppo di studenti, in questo caso individuato sulla base del percorso effettuato all'interno della prova MST.

Riferimenti bibliografici

- Astolfi, J. P. (1997). *L'erreur, un outil pour enseigner*. Paris, ESF éditeur.
- Black, P. & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21.1, 5-31.
- Botta, E. (2021). *Sperimentazione di un modello adattativo multilivello per la stima delle abilità in matematica nelle rilevazioni su larga scala*, Roma: Nuova Cultura
- Choppin, B. (1976). *Developments in item banking*. Windsor U.K.: Monitoring National Standards of Attainment in School, 1976.
- Dewey, J. (1949). *Esperienza e educazione*, tr. it. Firenze, La Nuova Italia.
- Hambleton, R. K., Zaal, J. N., & Pieters, P. (1991). Computerized adaptive testing: Theory, applications, and standards. In Hambleton, R. K. e Zaal, J. N. (Eds.). *Advances in educational and psychological testing*, Norwell, MA: Kluwer, pp. 341-366.

- Hattie J. & Timperley H. (2007). The Power of Feedback. *Review of Educational Research*, 77.1, 81-112.
- Kahneman, D. (2012). *Pensieri lenti e veloci*. Edizioni Mondadori, Milano.
- Luecht, R. M., & Nungester, R. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*, 35, pp. 229-249.
- Luecht, R., Brumfield, T., & Breithaupt, K. (2006). A testlet assembly design for adaptive multistage tests. *Applied Measurement in Education*, 19.3, 189-202.
- MI (2021). Focus Principali dati della scuola – Avvio Anno Scolastico 2020/2021, Direzione Generale per i sistemi informativi e la statistica.
- MIUR (2019). I principali dati relative agli alunni con DSA anno scolastico 2017/2018, Ufficio Statistica e Studi.
- Sands, W. A., Waters, B. K., & McBride, J. R. (Eds.) (1997). *Computerized adaptive testing: From inquiry to operation*. Washington, DC: American Psychological Association
- Sireci, S. G. (2004). Computerized-adaptive testing: An introduction. In Wall, J. e Walz, G. (Eds). *Measuring up: Assessment issues for teachers, counselors, and administrators*, Greensboro, NC: CAPS Press, pp. 685-694.
- Veldkamp, B. P. (2014). Item pool design and maintenance for multistage testing. In Yan, D., von Davier, A.A. & Lewis C., *Computerized multistage testing: Theory and applications*, New York: CRC Press, pp. 39-54.
- Wainer, H. (Ed.) (2000). *Computerized adaptive testing: A primer* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.
- Weiss, D. J. (1985). Adaptive testing by computer. *Journal of consulting and clinical psychology*, 53.6, 774.