

Logical Intelligence Enhancement Test: uno strumento per la valutazione delle abilità logiche nella scuola primaria

Logical Intelligence Enhancement Test: a measure for the evaluation of logical skills in primary school

Valeria Di Martino

Università di Catania, valeria.dimartino@unict.it

Marta Pellegrini

Università di Firenze, marta.pellegrini@unifi.it



Ricerche

This work is part of a pilot study conducted during 2018-2019 to evaluate the effectiveness of the Logical Intelligence Enhancement Programme (LIEP, Calvani, Peru & Zanaboni, 2019), a program which consists of logical-verbal and numeric exercises to enhance logical reasoning and metacognitive skills of fifth grade students. During this study the Logical Intelligence Enhancement test was developed to measure the effects of the training on the logical-verbal skills. This paper describes the development of the LIEP test, its psychometric features and validation process. For the validation the measure was administered to a sample of 173 fifth grade students in Palermo and Florence. The final version of the test consists of 18 items and includes eight dimensions of logical exercises, such as tables to be completed, numeric series, problem solving and logical inferences.

Keywords: cognitive enhancement; Logical Intelligence Enhancement Test (LIEP test); logical-verbal skills; test validation; primary school.

Il presente lavoro si inserisce in uno studio pilota condotto nell'anno scolastico 2018-2019 per valutare l'efficacia del Logical Intelligence Enhancement Programme (LIEP, Calvani, Peru, & Zanaboni, 2019), un programma composto da tavole di logica verbale e numerica volte a sviluppare il ragionamento logico e le capacità metacognitive di studenti di quinta primaria. All'interno di questo studio è stato sviluppato il Logical Intelligence Enhancement test per stimare gli effetti del training sulle abilità logico-verbali. Il contributo presenta la costruzione del test, le sue caratteristiche psicometriche e la validazione. A tal fine, lo strumento è stato somministrato a un campione di 173 studenti di classe quinta primaria nell'area di Palermo e Firenze. Il LIEP Test si compone di 22 item e include otto dimensioni di esercizi logici come il completamento di tabelle o serie di numeri, la risoluzione di problemi numerici o geometrici e tavole di inferenza logica.

Parole chiave: potenziamento cognitivo; Logical Intelligence Enhancement Test (LIEP test); abilità logico-verbali; validazione; scuola primaria.

All'interno di una impostazione condivisa, i paragrafi 1, 5 e 6 sono da attribuire a Valeria Di Martino, i paragrafi 2, 3, 4 e 7 a Marta Pellegrini.

Le autrici desiderano ringraziare il Prof. Antonio Calvani per averle seguite nel corso dello studio pilota insieme con il Prof. Andrea Peru. Desiderano inoltre ringraziare il Prof. Santo Di Nuovo per i suggerimenti forniti sulla parte metodologica di questo lavoro.

Logical Intelligence Enhancement Test: uno strumento per la valutazione delle abilità logiche nella scuola primaria

1. Introduzione

Il presente lavoro si inserisce nell'ambito degli studi condotti dall'associazione SApIE (Società per l'Apprendimento e l'Istruzione informati da Evidenza) per valutare l'efficacia di attività di *cognitive enhancement* o potenziamento cognitivo. Due sono i programmi creati per sviluppare le capacità logiche degli studenti dalla scuola dell'infanzia alla scuola secondaria di primo grado: il primo è il Visual Intelligence Enhancement Programme (VIEP) (Calvani & Zanaboni, 2018) composto da tavole logico-visive volte a sviluppare l'intelligenza visiva e le capacità metacognitive di studenti dai 2 ai 12 anni di età; il secondo è il Logical Intelligence Enhancement Program (LIEP) (Calvani, Peru, & Zanaboni, 2019) composto principalmente da tavole di logica verbale e numerica volte a sviluppare il ragionamento logico e le capacità metacognitive di studenti dai 6 ai 12 anni.

Approcci come il VIEP, il LIEP e altri programmi simili creati a livello internazionale¹ mirano a sviluppare le funzioni esecutive come il ragionamento, la memoria di lavoro e l'autoregolazione, anche grazie alle modalità didattiche utilizzate per la loro applicazione (Diamond, 2012; Diamond & Lee, 2011). Le strategie didattiche impiegate per l'applicazione delle tavole VIEP e LIEP sono principalmente il modellamento, più o meno guidato, attraverso cui l'insegnante mostra come si ragiona ad alta voce e fornisce esempi di risoluzione del compito, e il lavoro a coppie fra studenti volto a promuovere la verbalizzazione delle procedure attuate e la consapevolezza sui processi cognitivi coinvolti. Queste strategie, collocabili nel modello didattico della Direct



1 Programmi di potenziamento cognitivo internazionali sono ad esempio: Processing And Cognitive Enhancement Program (PACE) (<http://www.accomplished.ca/programs/pace/>); the Arrowsmith Cognitive Enhancement Program (<https://arrowsmithschool.org/enhancementprogram/>); Brain Enhancement Program (<http://www.autismbep.com>); Tools of the Mind (<https://toolsofthemind.org>), Instrumental Enrichment di Feuerstein, Rand, Hoffman, e Miller (1980). Per una panoramica di ricerca sull'efficacia di diverse modalità di potenziamento cognitivo consultare Diamond e Lee (2011).

Instruction (Rosenshine, 2012), sono parte integrante del programma LIEP. Esso, infatti, non si limita a una semplice applicazione di esercizi logici, ma mira a sviluppare capacità e competenze trasversali – come la metacognizione e la collaborazione fra pari – utili all'interno e all'esterno del contesto scolastico.

La sistematicità del percorso di potenziamento cognitivo, data dalla frequenza di utilizzo (due sessioni a settimana di 1 ora e 30 minuti ciascuna per un totale di 24 ore), e la gradualità delle sessioni sulla base della difficoltà, offrono agli studenti la possibilità di consolidare le proprie conoscenze e allo stesso tempo di compiere nuove riflessioni su un uso più approfondito delle strategie risolutive individuate (Di Martino & Pellegrini, 2019).

Nell'ambito della valutazione empirica sull'efficacia del programma LIEP, che si è da poco conclusa, si è reso necessario lo sviluppo di un test interno – il Logical Intelligence Enhancement test – per stimare gli effetti del training sul ragionamento logico-verbale. Il contributo, dopo una breve rassegna sui test internazionali per misurare le capacità logiche, presenta la costruzione del test e la sua validazione su un cam



2. Test standardizzati per misurare le abilità logiche

Numerosi sono i test standardizzati a livello internazionale per valutare le abilità logiche dei bambini in età scolare. Alcuni presentano item a carattere visivo, come le note matrici di Raven, altri comprendono item di natura visiva e verbale, come il Wechsler Preschool and Primary Scale of Intelligence (WPPSI), il Woodcock-Johnson Tests of Cognitive Abilities (WJ-COG) e il test Stanford-Binet. Di seguito si presentano brevemente ciascuno di questi test ampiamente utilizzati nei Paesi anglosassoni.

Le matrici di Raven, dette anche matrici progressive, sono state pubblicate dallo psicologo inglese John Carlyle Raven nel 1938 per misurare il ragionamento non verbale, analogico e l'intelligenza generale (Raven, 2000). Esse sono composte da varie figure che possono essere ruotate, capovolte o avere elementi aggiunti o sottratti da completare con la figura mancante. Il soggetto, sottoposto al test, deve selezionare la figura mancante tra sei o otto opzioni. Grazie al loro carattere visivo possono essere risolte da soggetti di ogni età e nazione, limitando così gli svantaggi linguistici tipici delle prove verbali. Tre sono le forme di matrici progressive utilizzate a seconda dell'età anagrafica del soggetto e del livello di difficoltà: (i) Matrici Colorate (CPM), formate da tre serie colorate rivolte ai bambini nella fascia d'età 3-11 e ai soggetti con

disabilità cognitiva; (ii) Matrici Standard (SPM), formate da 60 elementi suddivisi in cinque serie composti da 12 item ciascuno scritti in inchiostro nero su sfondo bianco di difficoltà crescente; (iii) Matrici avanzate (APM), suddivise in due set di complessità crescente, sono utilizzate con giovani e adulti con capacità intellettiva superiore alla media.

Le scale di Wechsler sono fra i test più comuni per misurare le abilità cognitive generali di bambini della scuola dell'infanzia e primaria. Tali test sono stati ideati da David Wechsler alla fine degli anni Trenta con l'obiettivo di ottenere informazioni cliniche dinamiche a partire dalla conduzione di alcuni compiti da parte dell'individuo. La Wechsler-Bellevue Intelligence Scale (Wechsler, 1939) fu la prima scala a essere pubblicata, mentre il test Wechsler Intelligence Scale for Children (WISC) per misurare l'intelligenza dei bambini fra i 5 e i 15 anni fu ideata nel 1949 (Wechsler, 1949). La WISC ha come principale vantaggio quello di comprendere categorie di item verbali e visivi allo scopo di misurare diverse abilità cognitive. La quarta edizione del test (Wechsler, 2003) si compone di 15 subtest organizzati in cinque indici: (i) visuo-spaziale (ii) ragionamento fluido; (iii) memoria di lavoro; (iv) velocità di processamento; (v) comprensione verbale.

Woodcock-Johnson Tests of Cognitive Abilities (WJ-COG) è stato sviluppato da Richard Woodcock e Mary Bonner Johnson nel 1977 e riadattato in successive edizioni, fra le quali la più recente è stata pubblicata nel 2014 (McGrew, LaForte, & Schrank, 2014). Il test è utilizzato principalmente per misurare le capacità necessarie all'apprendimento, al linguaggio orale e più in generale per valutare le abilità cognitive. Il test si compone di una batteria standard, formata da 7 scale, e di una batteria estesa costituita da 11 scale. Le scale della batteria standard, la più diffusa a livello internazionale, sono: (i) vocabolario orale; (ii) serie di numeri; (iii) attenzione verbale; (iv) corrispondenza lettera-modello; (v) processamento fonologico; (vi) ricordo di una storia; (vii) visualizzazione (Reynolds & Niileksela, 2015).

Simile alle scale WISC e WJ-COG, è il test Stanford-Binet (prima edizione 1916) sviluppato per misurare l'intelligenza generale di un individuo sulla base di cinque abilità cognitive: (i) la memoria; (ii) l'elaborazione visuale-spaziale; (iii) il ragionamento quantitativo; (iv) la conoscenza; (v) il ragionamento fluido. Il test è rivolto a individui dai due anni fino all'età adulta (Roid & Pomplun, 2012).



3. Il Logical Intelligence Enhancement Test

Il Logical Intelligence Enhancement Test (LIEP test) è stato elaborato con l'obiettivo di misurare le abilità logico-verbali sviluppate nello specifico dal programma LIEP per la classe quinta primaria. Il test è composto da una selezione di tavole di tipologie simili a quelle proposte durante il training, può essere pertanto definito una *proximal measure* (Fuchs et al., 2014). Le proximal measure si riferiscono a misure allineate con l'intervento attuato durante un esperimento, mentre le *distal measure* presentano materiale non inerente all'intervento con l'obiettivo di misurare la capacità degli studenti di trasferire gli apprendimenti acquisiti a contesti e problemi diversi. Ad esempio nella valutazione empirica del LIEP che si è appena conclusa (a.s. 2018-2019) e nella precedente condotta nell'a.s. 2017-2018 (Pellegrini, Nepi, & Peru, 2018) le Matrici di Raven, test di tipo visivo, sono state utilizzate come distal measure. Il LIEP test, invece, non si propone di misurare le abilità logiche del soggetto in generale, *culture free* come le Matrici di Raven, ma abilità a forte valenza logica strettamente in raccordo con il percorso curricolare della scuola primaria. A tal fine, il processo di costruzione della prova si è avvalso della collaborazione di docenti di scuola primaria e di esperti universitari.

Questo strumento può essere utilizzato anche a fini didattici; gli insegnanti possono ad esempio somministrare il LIEP test per misurare le capacità logiche dei bambini e valutare l'eventuale opportunità di intraprendere un percorso didattico di allenamento al ragionamento logico.

La prima versione del Logical Intelligence Enhancement Test per la quinta primaria, costituita da 22 item a scelta multipla con quattro opzioni di risposta di cui solo una corretta, si compone delle stesse dimensioni che costituiscono il programma LIEP per la classe quinta primaria (Tab. 1).

Tipologie	n. di item
Inferenze numeriche	3
Problem solving geometrico	3
Problem solving numerico	3
Equazioni logiche	3
Tabelle	2
Grafici	3
Combinatoria	2
Inferenze logiche	3
Totale n. item	22

Tab. 1. Tipologie e numerosità degli item del LIEP test.



Di seguito sono descritte le tipologie di item che compongono lo strumento, si rimanda all'appendice per una loro esemplificazione.

Le inferenze numeriche richiedono allo studente di individuare il dato mancante all'interno della sequenza logico-numerica presentata. Tre sono gli item di inferenza numerica usati nel test e presentati in ordine di difficoltà: il primo presenta una sequenza semplice di numeri interi, il secondo una sequenza più complessa di numeri interi, il terzo una sequenza di numeri decimali.

Il problem solving geometrico si compone di esercizi che per la loro risoluzione richiedono la conoscenza del calcolo dell'area e della nozione di frazione. Nel test sono presenti tre tipi di problem solving geometrico del programma LIEP: (i) la superficie, in questi item occorre calcolare l'area della figura composta presentata; (ii) le frazioni geometriche, in cui occorre individuare la frazione numerica corrispondente alla parte colorata della figura; (iii) le composizioni geometriche, in cui occorre confrontare due figure composte e rispondere a domande relative alla loro superficie.

Le tavole di problem solving numerico richiedono la risoluzione di quesiti numerici o l'individuazione dei dati mancanti di un problema. Il test include due quesiti di difficoltà crescente e un item in cui individuare i dati mancanti del problema presentato.

Le equazioni logiche di tipo verbale e visivo attivano il ragionamento che sarà poi formalizzato nei successivi gradi scolastici con una equazione. Il test presenta tre item di cui uno verbale e due visivi.

Le tabelle si compongono di esercizi di due tipi: completare i dati mancanti di una tabella oppure rispondere alle domande data una tabella completa. Il test si compone di un esercizio di ciascuna tipologia.

Le tavole sui grafici presentano areogrammi, istogrammi o diagrammi di Gantt e alcune affermazioni sulla cui veridicità viene richiesto di esprimersi o in cui si chiede di individuare il grafico corretto sulla base dei dati riportati in tabella. Il test include tre item, uno per ciascuna tipologia di grafico.

Le combinatorie sono tavole di ragionamento logico verbale in cui individuare attraverso la combinazione delle proposizioni presentate la risposta corretta. Tali esercizi sono ispirati al *Flag test* e sono presentati in ordine di difficoltà crescente. Il test presenta due tavole di questa tipologia.

Le inferenze logiche sono esercizi in cui gli studenti devono stabilire quale delle alternative può essere inferita sulla base dell'affermazione presentata. La difficoltà maggiore per risolvere l'esercizio sta nella scelta fra alternative verosimili ma non direttamente deducibili dall'affermazione. Il test si compone di tre item progettati con gradi di difficoltà crescente.



4. Campione e modalità di somministrazione del questionario

Per la validazione del LIEP test è stato utilizzato un campione non probabilistico costituito da studenti della classe quinta primaria selezionato sulla base della disponibilità data dal dirigente scolastico e dagli insegnanti. La somministrazione dello strumento ha coinvolto 173 studenti delle province di Firenze e Palermo. La Tab. 2 mostra le caratteristiche del campione rispetto alla provincia e al genere.

	Maschi	Femmine	Totale
Provincia			
Firenze	38	38	76
Palermo	42	55	97
Totale	80	93	173

Tab.2. Distribuzione del campione rispetto alla provincia e al genere.



La somministrazione dello strumento è stata effettuata nelle due province a marzo 2019 su supporto cartaceo e sotto la supervisione dei ricercatori o di loro collaboratori. L'anonimato è stato garantito attraverso l'attribuzione di un codice identificativo per ciascuno studente. Il tempo complessivo per la somministrazione dello strumento è stato di 40 minuti, di cui 30 per l'effettivo svolgimento della prova da parte degli studenti e 10 minuti di presentazione dello stesso prima del suo svolgimento.

5. Principali caratteristiche psicometriche

Le analisi condotte e di seguito riportate sono finalizzate ad evidenziare le caratteristiche psicometriche della prova ed eventuali possibili problemi legati sia alla chiarezza e alla comprensione del testo del quesito, sia alla ragionevolezza delle possibili risposte, sia al livello di difficoltà di un item e alla coerenza del questionario nel suo complesso al fine di modificare e correggere problemi nelle domande e giungere alla definizione di un test con elevato livello di attendibilità misuratoria. Questi aspetti, delicati e cruciali, sono stati affrontati tramite la Classical Test Theory (CTT), che permette un'analisi descrittiva immediata e di facile interpretazione dei risultati, e l'Item Response Theory (IRT) che offre un approfondimento sulle caratteristiche degli item in relazione al costrutto latente. Quest'ultima ha richiesto la verifica empirica dell'uni-

dimensionalità della prova, condotta tramite analisi fattoriale esplorativa con approccio delle variabili soggiacenti (UVA, Underlying Variable Approach, Moustaki, 2000). Di seguito sono riportate le analisi effettuate seguendo la CTT con riferimento all'analisi dei distrattori e agli indici di discriminatività e affidabilità (descritti nei paragrafi 5.1, 5.2 e 5.3), la verifica dell'unidimensionalità (paragrafo 5.4) nonché i parametri di difficoltà, input e output secondo il modello di Rasch (paragrafo 5.5).

5.1 Analisi dei distrattori

La validazione del test attraverso i modelli e le procedure tradizionalmente impiegate nell'ambito della CTT richiede l'elaborazione di una serie di indicatori di natura descrittiva che consentono di individuare sia gli item che funzionano in una prova sia quelli che, invece, devono essere modificati. A tal fine risulta utile anche indagare i distrattori che lo compongono.

Analizzando nello specifico i distrattori del LIEP test (Tab. 3) si nota che andrebbero modificati o sostituiti, poiché scelti da una percentuale di rispondenti inferiore al 5% (Lucisano & Salerni, 2002) l'opzione A dell'item 4, l'opzione C dell'item 5 e l'opzione D degli item 1, 12 e 17. Risulta particolare la situazione dell'item 12, in cui le percentuali maggiori di risposte (70%) si concentrano sull'opzione A (errata). Adeguato risulta invece l'item 2, con un buon indice di difficoltà (0,69) e un perfetto equilibrio dei distrattori, che raccolgono quasi la stessa percentuale di scelte. Un andamento simile si riscontra anche negli item 16, 21 e 22. Ulteriori riflessioni riguardano invece l'andamento della distribuzione delle scelte nell'item 6, in cui le maggiori percentuali di risposte si concentrano sul distrattore D (44%) rispetto all'opzione corretta (33%). Ugualmente attrattive risultano invece le opzioni C e D dell'item 10.



Item	A	B	C	D
1	0,30	0,06	0,61	0,03*
2	0,69	0,10	0,11	0,10
3	0,10	0,71	0,13	0,06
4	0,04*	0,83	0,05	0,09
5	0,14	0,13	0,03*	0,70
6	0,11	0,12	0,33	0,44
7	0,06	0,16	0,09	0,69
8	0,10	0,10	0,40	0,40

9	0,14	0,05	0,70	0,09
10	0,06	0,05	0,44	0,44
11	0,49	0,19	0,15	0,075
12	0,70	0,10	0,19	0,01*
13	0,08	0,06	0,25	0,61
14	0,20	0,09	0,61	0,09
15	0,18	0,08	0,075	0,58
16	0,18	0,19	0,51	0,11
17	0,05	0,80	0,09	0,04*
18	0,10	0,11	0,18	0,58
19	0,68	0,15	0,08	0,08
20	0,18	0,60	0,15	0,08
21	0,64	0,11	0,11	0,14
22	0,73	0,08	0,08	0,08

Tab. 3 Opzioni di scelta per singolo item LIEP test (in grassetto opzioni corrette)



5.2 Indice di discriminatività

L'indice di discriminatività corrisponde al coefficiente di correlazione punto-biseriale del singolo punteggio con quello totale del test, calcolato escludendo dal totale l'item stesso. Fornisce un'indicazione della capacità del singolo item di differenziare allievi con diversi livelli di rendimento nella prova. Sono considerati adeguati gli item con coefficiente di correlazione item-totale uguale o superiore a 0,25.

Dall'analisi degli indici di discriminazione riportati in Tab. 4, si evince che l'item 4 discrimina pochissimo tra gli studenti che ottengono i risultati migliori e peggiori nella prova, risulta pertanto da rigettare. Gli item 10, 11 e 12 discriminano poco, mentre i restanti item discriminano i livelli di abilità in modo adeguato, differenziando i rispondenti coerentemente al punteggio totale al test.

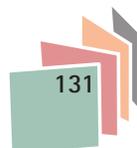
Item	Indice di discriminatività
1	0,40
2	0,47
3	0,41
4	0,02*
5	0,55
6	0,35
7	0,53
8	0,50
9	0,60
10	0,16*
11	0,16*
12	0,19*
13	0,39
14	0,57
15	0,35

16	0,33
17	0,56
18	0,40
19	0,58
20	0,38
21	0,57
22	0,46

Tab. 4 Indici di discriminatività degli item del LIEP test

5.3 L'affidabilità del test

L'affidabilità di un test attiene alla sua accuratezza e coerenza. Obiettivo dell'analisi di affidabilità è verificare che il test fornisca misurazioni precise, stabili e oggettive. Esistono diversi coefficienti che stimano l'attendibilità di un test. Seguendo le raccomandazioni degli *Standard for Educational and Psychological Testing* (AERA, 1999), abbiamo deciso di riportare in Tab. 5 più di un coefficiente per la stima della coerenza interna, come anche il valore dell'errore standard di misurazione (SEM) e l'intervallo di confidenza. In particolare sono riportati: il coefficiente alfa, che si basa sul confronto tra la somma delle varianze dei singoli item e la varianza del punteggio totale nel test, ottenuto sommando gli item che lo compongono; il coefficiente L_2 di Guttman (1945), che presenta nel calcolo anche le covarianze tra gli item e il coefficiente Feldt-Brennan (Feldt & Brennan, 1989) che misura la coerenza interna secondo l'assunto che gli item siano congenerici.



Method	Estimate	95% Conf. Int.	SEM
Guttman's L2	0,852	(0,801-0,896)	1,912
Coefficient Alpha	0,843	(0,789-0,889)	1,972
Feldt-Brennan	0,847	(0,7937-0,8915)	1,948

Tab. 5 Misure di affidabilità della prova.

Dalla Tab. 5 si evince che la prova presenta invece un buon valore del coefficiente di affidabilità (Nunnally & Bernstein, 1994) poiché è compreso tra 0,843 e 0,852.

Nell'ambito della CTT, la qualità di un singolo item viene valutata, oltre che in relazione alla sua affidabilità, calcolando la misura dell'affidabilità globale del test se escludiamo quell'item dal computo del coefficiente. Se l'indice "item omissso" risulta maggiore del coefficiente calcolato sull'intera prova, allora l'omissione dell'item aumenta l'affidabilità globale del test; si rende dunque necessaria la valutazione dell'opportunità di tenere l'item nel test o eliminarlo.

La Tab. 6 mostra i valori dei diversi coefficienti ricalcolati escludendo il singolo item della prova³. Si nota come l'esclusione della maggior parte degli item non provocherebbe un incremento dei valori dei coefficienti globali. La loro eliminazione, dunque, non contribuirebbe ad aumentare l'affidabilità globale del test. Tuttavia, l'omissione degli item 4, 10, 11 e 12 comporterebbe un leggero incremento dei diversi indici.

Item	Guttman's lambda-2	Coefficient alpha	Feldt-Brennan coefficient
<i>PROVA</i>	0,852	0,843	0,847
1	0,847	0,837	0,841
2	0,844	0,834	0,838
3	0,846	0,836	0,840
4	0,857*	0,850*	0,852*
5	0,841	0,831	0,835
6	0,849	0,839	0,843
7	0,842	0,832	0,836
8	0,843	0,833	0,837
9	0,839	0,829	0,833
10	0,856*	0,847*	0,851*
11	0,859*	0,847*	0,851*
12	0,853*	0,844*	0,848*
13	0,847	0,837	0,841
14	0,840	0,830	0,833
15	0,849	0,839	0,843
16	0,850	0,840	0,844
17	0,842	0,832	0,836
18	0,847	0,837	0,841
19	0,840	0,830	0,833
20	0,848	0,838	0,842
21	0,840	0,830	0,834
22	0,845	0,835	0,839

Tab. 6 Coefficienti di affidabilità "item escluso" LIEP test.

5.5 Verifica della dimensionalità

La verifica dell'unidimensionalità è stata verificata tramite l'analisi fattoriale non lineare, dal momento che le prove sono costituite da item dicotomici. La validità interna è stata dunque verificata empiricamente attraverso l'analisi fattoriale con approccio delle variabili soggiacenti (UVA; Moustaki, 2000) tramite il software statistico MPLUS (Muthén & Muthén, 2010) con metodo di stima dei Minimi Quadrati Ponderati (Weighted Least Square, WLS). Si tratta di verificare l'ipotesi secondo cui le risposte agli item sono influenzate da una sola variabile

2 Gli item con * sono quelli che presentano un valore superiore rispetto a quello globale del test.

latente dominante, presupposto rilevante per l'applicazione del modello di Rasch.

In particolare, nella valutazione della bontà della soluzione a un fattore sono stati considerati: l'indice di adattamento RMSEA (Root Mean Square Error of Approximation, Steiger & Lind, 1980), il CFI (Comparative Fit Index, Bentler, 1990), il TLI (Tucker and Lewis Index, Tucker & Lewis, 1973) e SRMR (Standardized Root Mean Square Residual, Barbaranelli, 2007). Oltre a tali indici di adattamento, sono stati presi in considerazione: il rapporto tra primo e secondo autovalore; lo *scree-test* degli autovalori; l'ampiezza delle saturazioni fattoriali.

Suggeriscono un buon adattamento del modello unidimensionale ai dati sia il valore dell'indice RMSEA, pari a 0,041, dunque inferiore al valore soglia di 0,05 (Joreskog, Sorbom, du Toit, & du Toit, 2000), sia l'indice SRMR, pari a 0,142. Il CFI e il TLI risultano maggiori di 0,9, rispettivamente 0,928 e 0,920, sembra quindi che ci sia un'adeguatezza del modello ipotizzato nella popolazione, anche se valori superiori a 0,95 erano sicuramente preferibili (Hu & Bentler, 1999). Il rapporto tra primo e secondo autovalore, pari a 3,342 (7,614/2,278), risulta maggiore di 3, da cui consegue che la prima dimensione riproduce una porzione di variabilità maggiore di quella riprodotta dal secondo fattore estratto. Lo *scree-test* degli autovalori (Fig. 1) mostra graficamente una curva decrescente in funzione del valore estratto e, seppur da considerare con cautela, risulta coerente con l'ipotesi di una unica dimensione dominante sottesa ai dati. Il legame tra domande e dimensione latente, espresso dalle saturazioni, appare globalmente adeguato: il valore delle saturazioni è in gran parte dei casi (per 19 su 22 domande) superiore a |0,30|. I soli item ad avere una saturazione inferiore sono 4, 10 e 11.

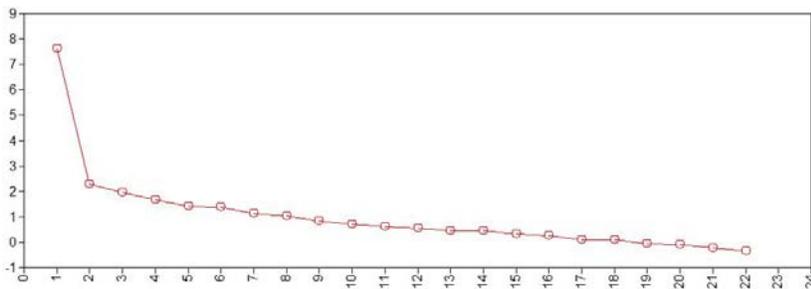


Fig. 1 Scree-plot degli autovalori del LIEP Test.

5.5 Stima dei parametri tramite il modello di Rasch

Nell'ambito dell'IRT le proprietà delle domande sono investigate attraverso l'uso di modelli che esprimono la relazione tra la probabilità di risposta corretta e l'abilità, attraverso un insieme di parametri che derivano dalle caratteristiche degli item risposta (Hambleton, Swaminathan, & Rogers, 1991). Nel modello logistico ad un parametro (*Rasch Model*) si assume che tale probabilità dipenda dalla difficoltà dell'item.

La stima dei parametri è avvenuta tramite il software statistico JMetrik che utilizza la joint maximum likelihood (JML) per stimare i parametri del modello di Rasch e un algoritmo proporzionale di *curve-fitting* che risulta più appropriato per i campioni di piccole dimensioni (Linacre, 2012; Meyer, 2014; Meyer & Hailey, 2012).

Il software utilizzato restituisce la mappa degli item che riepiloga la distribuzione dell'abilità della persona e la distribuzione della difficoltà degli item (Fig. 2). La maggior parte degli item si trova nella regione con la maggior parte degli esaminati. Tuttavia, dal momento che nella fascia alta e bassa della scala si collocano pochi item, i soggetti con un'abilità elevata o molto bassa, verranno misurati con meno precisione.

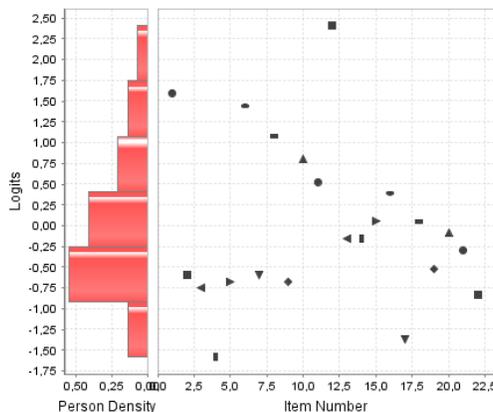


Fig. 2 Mappa degli item del LIEP test.

Per verificare l'adattamento degli item e dei soggetti al modello il software restituisce i parametri di infit (WMS) ed outfit (UMS), sulla base della stima delle differenze tra i valori osservati e attesi di una risposta all'item, una differenza nota come residuo (Meyer, 2014). I valori di infit e outfit sono sempre positivi e hanno un valore atteso di 1; valori vicini a 1 indicano un buon adattamento dei dati al modello

mentre i valori divergono da 1 all'aumentare del grado di disadattamento (Meyer, 2014). Per Bond e Fox (2007) i valori superiori a 1 indicano una variazione maggiore di quanto previsto dal modello di Rasch e dovrebbero, pertanto, essere rivisti (Meyer, 2014). Al contrario, anche valori di infit e outfit inferiori a 1 vanno considerati con attenzione, anche se indicano una tipologia diversa di problema: riflettono la tendenza delle risposte a essere troppo coerenti con ciò che ci si aspetta dal modello (Meyer, 2014).

Bond e Fox (2007) individuano una fascia ottimale di valori di infit e outfit compresa tra 0,8 e 1,2, mentre Linacre (2012) considera accettabili ai fini misurativi anche i valori compresi tra 0,5 e 1,5. Vanno invece prese dovute precauzioni per i valori di infit e outfit superiori a 2,0 perché degradano la misura. Meno preoccupanti sono i valori di infit e outfit più bassi, ma la qualità della misura risulterà gonfiata artificialmente con valori inferiori a 0,5 (Linacre, 2012).

Il software restituisce anche valori di infit e outfit standardizzati che assumono valori positivi e negativi e hanno un valore atteso di 0. Item o soggetti con valori standardizzati di infit e outfit superiori a $|3|$ indicano un problema di adattamento e giustificano un controllo aggiuntivo. Si tratta di valori sensibili alle dimensioni del campione e risultano utili solo per campioni di dimensioni inferiori a 300 soggetti (Linacre, 2012).



Item		Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
1	Problem solving geometrico	1,6	0,28	0,99	-0,04	1,07	0,37
2	Problem solving numerico	-0,6	0,28	0,92	-0,63	1,05	0,28
3	Tabelle	-0,75	0,28	0,97	-0,15	1,03	0,2
4	Inferenze logiche	-1,58	0,33	1,31*	1,52	2,79*	2,98
5	Inferenze numeriche	-0,68	0,28	0,81	-1,51	0,8	-0,75
6	Problem solving geometrico	1,44	0,28	1,01	0,1	1,13	0,6
7	Combinatoria	-0,6	0,28	0,86	-1,12	0,69	-1,32
8	Equazioni logiche	1,07	0,27	0,86	-1,18	0,8	-1,09
9	Equazioni logiche	-0,68	0,28	0,77*	-1,9	0,61*	-1,66
10	Problem solving numerico	0,8	0,26	1,31*	2,59	1,61*	3,09
11	Tabelle	0,52	0,26	1,36*	2,99	1,82*	4,08
12	Inferenze logiche	2,41	0,33	1,1	0,57	1,57*	1,36
13	Inferenze numeriche	-0,16	0,27	1,04	0,39	0,95	-0,2
14	Problem solving geometrico	-0,16	0,27	0,81	-1,72	0,73	-1,48
15	Grafici	0,05	0,26	1,1	0,89	1,09	0,55
16	Inferenze logiche	0,39	0,26	1,13	1,15	1,05	0,38
17	Grafici	-1,37	0,32	0,78*	-1,31	0,51*	-1,42
18	Problem solving numerico	0,05	0,26	1,03	0,32	0,95	-0,24
19	Equazioni logiche	-0,52	0,27	0,79*	-1,76	0,75	-1,06
20	Inferenze numeriche	-0,09	0,26	1,05	0,46	1,06	0,38
21	Combinatoria	-0,3	0,27	0,82	-1,63	0,69	-1,57
22	Grafici	-0,84	0,29	0,95	-0,31	0,77	-0,78

Tab. 7 Parametri del Modello di Rasch del LIEP test.

Come si evince dalla Tab. 7, tra i 22 item della prova la difficoltà varia da un minimo di -1,58 (item 4) a un massimo di 2,41 (item 12). Nel complesso comunque il test risulta leggermente sbilanciato a favore di item facili (gli item più semplici, di segno negativo, sono 13 su 22). Analizzando più nello specifico l'indice di difficoltà degli item relativi a ciascuna tipologia di quesiti utilizzati nella prova, si può notare come in ognuno di essi gli indici di difficoltà siano variegati. Ciò risponde ad uno dei principi seguiti nella costruzione della prova, di prevedere per ciascuna tipologia, quesiti di difficoltà differenti, in grado di cogliere i diversi livelli di abilità.

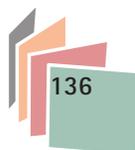
I parametri di infit al quadrato (WMS) variano da un minimo di 0,77 a 1,36 e hanno un valore medio pari a 0,99. Dal momento che il campione è costituito da meno di 300 soggetti, risulta particolarmente utile analizzare anche i valori di infit standardizzati (WMS st.) che non risultano problematici (minimo -1,90; massimo 2,99 e media -0,10), anche se l'item 11 risulta essere al limite (WMS St. = 2,99).

Per quanto riguarda i parametri di outfit, dalla Tab. 9 si evince che essi variano da un minimo di 0,52 a un massimo di 2,79, con un valore medio pari a 1,07. L'analisi evidenzia dunque la criticità dell'item 4 con un valore UMS pari a 2,79. Non sorprende che questo item sia anche il più facile della prova. Dall'analisi dei valori di outfit standardizzati (UMS St.) emergono valori che variano da -1,66 a 4,08 con un valore medio pari a 0,12. Un'attenta analisi di questi parametri rivela che l'item 11 è il più problematico (UMS St. = 4,08. Un comportamento simile è evidente anche nel caso dell'item 10 che presenta un valore UMS St. pari a 3,09. La criticità di tali item era stata già evidenziata anche dai valori di discriminatività e affidabilità secondo la Teoria Classica dei Test.

6. Considerazioni di sintesi sugli item e validazione dello strumento

A conclusione delle analisi svolte, si può sinteticamente affermare che, dalla stima dei parametri di difficoltà secondo il modello di Rasch, non si evidenziano item estremamente facili o estremamente difficili. Solo l'item 4 risulta del tutto non discriminativo, mentre l'attendibilità globale della prova migliora ipotizzando l'eliminazione degli item 4, 10, 11 e 12. La criticità di questi ultimi quattro item è confermata anche dal valore di UMS superiore a 1,5 e dal valore di UMS St. superiore a 3 in valore assoluto per gli item 10 e 11.

Le analisi concordano quindi sulla non appropriatezza degli item 4, 10, 11 e 12 che, pertanto, sono stati eliminati.



È stata quindi effettuata un'analisi fattoriale confermativa non lineare, utilizzando l'UVA (Moustaki, 2000) tramite il software statistico MPLUS (Muthén & Muthén, 2010). I fattori sono stati estratti mediante il metodo dei Minimi Quadrati Ponderati Robusti (WLSMV) che fornisce stime corrette dei parametri e degli errori standard in presenza di variabili categoriali. I risultati sono riportati in Tab. 8.

Chi-Square	RMSEA	CFI	TFI	WRMR
167178 gdl 135 $p=0,0314$	0,037	0,965	0,960	0,808

Tab. 8 Model fit della CFA condotta sul LIEP test revisionato

Gli indici di fit del modello inerenti la struttura fattoriale del test risultano essere buoni. Il Chi-quadrato non è significativo ($p=0,03$), ciò consente di accettare l'ipotesi nulla secondo cui la matrice che contiene le correlazioni tra le variabili osservate è ben riprodotta dalla matrice delle saturazioni che si ottiene dopo l'analisi fattoriale. Tuttavia, poiché in campioni molto piccoli, tale valore tende a non essere significativo anche in presenza di differenze realmente significative (Barbaranelli, 2007), si è deciso di prendere comunque in considerazione anche gli altri indici di fit.

Dai risultati emerge che gli indici CFI e TLI, essendo maggiori di 0,95, indicano un buon fit. L'RMSEA è pari a 0,037, quindi l'errore di approssimazione è minimo. Anche l'indice WRMR (radice della media dei residui al quadrato ponderata) è buono, poiché inferiore al cut-off di 0,90.

La nuova versione della prova (LIEP_{T5} in appendice) risulta migliore anche in termini di attendibilità, come si evince dai coefficienti riportati in Tab. 9.

Method	Estimate	95% Conf. Int.	SEM
Guttman's L2	0,866	(0,819-0,905)	1,703
Coefficient Alpha	0,861	(0,813-0,902)	1,732
Feldt-Brennan	0,862	(0,814-0,903)	1,725

Tab. 9 Misure di affidabilità LIEP_{T5}.

7. Conclusioni

La predisposizione del test è avvenuta attraverso un processo complesso, iniziato con la delimitazione dell'oggetto di misurazione, proseguito



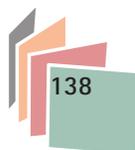
con la costruzione dello strumento di misurazione e la verifica della sua appropriatezza per giungere, successivamente alla sua validazione. Il controllo di appropriatezza è stato condotto attraverso procedure riconducibili alla CTT e IRT, che hanno consentito, da un lato, una riflessione sugli indicatori di natura descrittiva diretti a verificare la validità e l'affidabilità dell'intero test e la bontà dei singoli item in termini di difficoltà, capacità di discriminazione e affidabilità, e dall'altro anche di prevedere in modo probabilistico le risposte ai singoli item, date le caratteristiche degli item e le abilità degli studenti. I risultati delle analisi, condotte congiuntamente seguendo questi diversi approcci metodologici, hanno permesso di esprimere un giudizio sulla bontà dei singoli item e del test nel suo complesso. È stato così possibile selezionare gli item da eliminare in modo da giungere ad una versione definitiva del test, che ha mostrato una buona affidabilità.

Gli esiti presentati sono stati fondamentali per una prima rimodulazione dello strumento che sarà successivamente somministrato a livello nazionale.

Pur trattandosi di una prova nata dall'esigenza di valutazione empirica dell'efficacia del programma LIEP (Calvani, Peru, & Zanaboni, 2019) elaborato nell'ambito degli studi condotti dall'Associazione SApIE, il LIEP_{T5} potrebbe anche essere utilizzato come uno strumento valutativo per indagare le infrastrutture logico-deduttive strettamente interconnesse con il curriculum di quinta primaria. Lo strumento potrebbe essere pertanto somministrato dai docenti per valutare un eventuale inserimento di un percorso di logica nel regolare curriculum scolastico con il fine di sviluppare capacità trasversali, quali la metacognizione, il ragionamento logico e la collaborazione fra pari. In tal senso, la standardizzazione dello strumento su un più corposo campione nazionale, potrebbe consentire il suo utilizzo in un'ottica formativa, segnalando la necessità di eventuali interventi personalizzati e mirati che consentano di affrontare adeguatamente il proseguo degli studi, in particolare in ambito scientifico.

Riferimenti bibliografici

- AERA. American Educational Research Association. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Arrowsmith Cognitive Enhancement Program. <https://arrowsmithschool.org/enhancementprogram/> (ver. 17.12.2019).
- Barbaranelli, C. (2007). *Analisi dei dati. Tecniche multivariate per la ricerca*



- psicologica e sociale* (2nd ed.). Milano: LED.
- Bentler, P.M. (1990). Comparative fit indexes in structural equations models. *Psychological Bulletin*, 107, 238-246.
- Binet, A., & Simon, T. (1916). *The development of intelligence in children* (E. Kit, Trans.). Baltimore, MD: Williams & Wilkins.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Brain Enhancement Program. <http://www.autismbep.com> (ver. 17.12.2019).
- Calvani, A., & Zanaboni, B. (2018). *Tavole logico-visive. Incastri, rotazioni, sviluppi, inferenze, tabelle, punti di vista (3-12 anni)*. Firenze: SApIE.
- Calvani, A., Peru, A., & Zanaboni, B. (2019). *Potenziamento logico (6-12 anni). Deduzioni, equazioni logiche, combinatoria, problem solving, grafici e tabelle*. Firenze: SApIE.
- Di Martino, V., & Pellegrini, M. (2019). Sperimentazione nella quinta primaria. In A. Calvani, A. Peru, & B. Zanaboni (Eds.), *Potenziamento logico (6-12 anni) Deduzioni, equazioni logiche, combinatoria, problem solving, grafici e tabelle* (pp. 182-192). Firenze: SApIE.
- Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, XXI(5), 335-341.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, XXXIII (6045), 959-964.
- Feldt, L. S., & Brennan, R. L. (1999). Reliability. In R. L. Linn (Ed.), *Educational measurement* (pp. 105-146). 3rd ed. New York: American Council on Education and MacMillan.
- Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental Enrichment. An intervention for cognitive modifiability*. Baltimore, MD: University Park Press.
- Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., & Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to prealgebraic knowledge? *Journal of educational psychology*, 106(4), 990-1006.
- Gallucci, M., & Leone, L. (2012). *Modelli statistici per le Scienze Sociali*. Pearson Italia.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Hu, L. T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal*, 6 (1), 1-55.
- Jöreskog, K. G., Sörbom, D., Du Toit, S., & Du Toit, M. (2000). *LISREL 8: New statistical features*. Chicago, IL: Scientific Software International.



- Linacre, J. M. (2012). *A user's guide to WINSTEPS and ministep Rasch model computer programs: Program manual 3.75*.
- Lucisano, P., & Salerni, A. (2002). *Metodologia della ricerca in educazione e formazione*. Roma: Carocci.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.
- Meyer, J. P. (2014). *Applied measurement with jMetrik*. Routledge.
- Meyer, J. P., & Hailey, E. (2012). A study of Rasch partial credit, and rating scale model parameter recovery in WINSTEPS and jMetrik. *Journal of Applied Measurement*, 13(3), 248–258.
- Moustaki, I. (2000). A latent variable model for ordinal variables. *Applied Psychological Measurement*, 24 (3), 211–223.
- Muthén, L. K., & Muthén, B. O. (2010). *MPLUS user's guide: Statistical Analysis with Latent Variables*. Los Angeles, CA: Muthén & Muthén.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory*. New York: McGraw-Hill.
- Pellegrini, M., Nepi, L., & Peru, A. (2018). Effects of logical verbal training on abstract reasoning: evidence from a pilot study. *Educational Cultural and Psychological Studies (ECPS) Journal*, 18, 449–458.
- Processing And Cognitive Enhancement Program (PACE). <http://www.accomplished.ca/programs/pace/> (ver. 17.12.2019).
- Raven, J. C. (1938). *Guide to Using Progressive Matrices (1938): The Design and Use of the Scale-Instructions-Records and Marking-Key-Normal Score Composition-Norms for Children and for Adults Up to the Age of 65-Bibliography*. HK Lewis.
- Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cognitive psychology*, 41(1), 1–48.
- Reynolds, M. R., & Niileksela, C. R. (2015). Test Review: Schrank F.A. McGrew, K. S., & N. Mather (2014). Woodcock-Johnson IV Tests of Cognitive Abilities. *Journal of Psychoeducational Assessment*, 33(4), 381–390.
- Roid, G. H., & Pomplun, M. (2012). The Stanford-Binet intelligence scales. Fifth Edition. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 249-268). New York, NY: The Guilford Press.
- Rosenshine, B. (2012). Principles of Instruction: Research-Based Strategies That All Teachers Should Know. *American educator*, 36(1), 12.
- Steiger, J.H., & Lind, J.M. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.
- Tools of the Mind. <https://toolsofthemind.org> (ver. 17.12.2019).
- Tucker, L.R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- Wechsler, D. (1939). *Wechsler-Bellevue intelligence scale*. New York, NY: Psychological Corporation.
- Wechsler, D. (1949). *Wechsler intelligence scale for children*. San Antonio, TX:

Psychological Corporation.
Wechsler, D. (2003). *Wechsler Intelligence Scale for Children*. 4th ed. San Antonio, TX: The Psychological Corporation.
Woodcock, R. W. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Technical Report.

