

Analisi (classica, Rasch, dei distrattori) di una prova di lettura a scelta multipla della certificazione di italiano per stranieri CILS

PAOLO TORRESAN*
Alma Edizioni

Received 5 October 2014; received in revised form 13 February 2015; accepted 13 March 2015

ABSTRACT

- IT** Nell'articolo che segue svolgiamo un'analisi degli item di una prova di lettura a scelta multipla di livello B1 della certificazione di italiano per stranieri CILS (Centro CILS, 2009), allestita dall'omonimo centro dell'Università per Stranieri di Siena. Operiamo inoltre uno studio della manipolazione del testo autentico avvenuta per mano del certificatore; da tale studio emerge una criticità di rilievo che, a nostro giudizio, pregiudica la coerenza del testo. Nel suo insieme, l'indagine consente al lettore di prendere coscienza delle difficoltà che la progettazione di un'esercitazione ampiamente usata, sia in sede di *testing* che nella didattica di ogni giorno, qual è il quesito a scelta multipla, comporta.
- Parole chiave:** DIDATTICA DELL'ITALIANO COME LINGUA STRANIERA, VALUTAZIONE DELLA COMPrensIONE SCRITTA, QUESITO A SCELTA MULTIPLA, ITEM ANALYSIS.
- EN** In the following article, we carry out an analysis of the items from a multiple choice reading test for CILS (Italian as a Foreign Language Certification) (Centro CILS, 2009), run by the Università per Stranieri di Siena. We also examine the way in which the original text was modified for the test by the certifying body, leading to what we identify as a critical issue which, in our view, detracts from the text's coherence. This study illustrates some of the difficulties in designing an exercise used widely in both testing and teaching context, that is, the multiple choice question.
- Key words:** TEACHING ITALIAN AS A FOREIGN LANGUAGE, ASSESSING READING COMPREHENSION, MULTIPLE CHOICE QUESTIONS, ITEM ANALYSIS.
- ES** En el siguiente artículo hemos llevado a cabo un análisis de los ítems de respuesta múltiple en una prueba de lectura de nivel B1 de la Certificación de Italiano como Lengua Extranjera CILS (Centro CILS, 2009), realizada en la Universidad para Extranjeros de Siena. Asimismo, hemos observado el modo en el que el texto original ha sido manipulado por el examinador; de dicho estudio emerge una criticidad relevante que, a nuestro modo de ver, afecta a la coherencia del texto. En su conjunto, el estudio permite al lector tomar consciencia de la dificultad que supone diseñar un ejercicio que se utiliza ampliamente tanto en la evaluación como en la didáctica de aula; las preguntas de respuesta múltiple.
- Palabras clave:** ENSEÑANZA DEL ITALIANO COMO LENGUA EXTRANJERA, EVALUACIÓN DE LA COMPrensIÓN ESCRITA, PREGUNTAS DE OPCION MÚLTIPLE, ANÁLISIS DE ÍTEMS.

* Contatto: piroclastico@tin.it oppure esh@unive.it

1. La certificazione CILS

La certificazione CILS, *Certificazione di Italiano come Lingua Straniera*, elaborata dall'omonimo centro di ricerca facente capo all'Università per Stranieri di Siena, è una tra le quattro certificazioni di italiano LS riconosciute dal Ministero degli Affari Esteri italiano all'interno del marchio CLIQ (*Certificazione Lingua Italiana di Qualità*), finalizzato a garantire *standard* comuni di valutazione.

Dalle *Linee guida* della stessa certificazione CILS leggiamo: “[...] è rilasciata dall'Università dopo un esame costituito da test sulle abilità comunicative. La competenza in italiano come lingua straniera viene descritta, misurata, valutata e garantita con criteri scientifici. La certificazione CILS verifica la capacità d'uso della lingua in situazioni comunicative e in rapporto alle loro caratteristiche socio-culturali e socio-linguistiche” (Barni *et al.*, 2009, p. 9).

La suddetta certificazione attesta la conoscenza dell'italiano facendo riferimento ai livelli descritti dal *Quadro Comune Europeo di Riferimento per le Lingue* (d'ora in poi CEFR; Council of Europe, 2001; Tabella 1).

Tabella 1
CILS e Quadro Comune Europeo di Riferimento:
corrispondenza dei livelli di competenza

Livelli certificazione CILS	Livelli Quadro Comune Europeo di Riferimento
Livello A1	A1
Livello A2	A2
CILS UNO	B1
CILS DUE	B2
CILS TRE	C1
CILS QUATTRO	C2

Ad ogni livello vengono misurate le quattro abilità linguistiche; inoltre – ad eccezione delle prove di livello A1 per bambini e dei “moduli per l'integrazione”, vale a dire le prove di livello A1 e A2 per apprendenti adulti in Italia, (Barni *et al.* 2009, p. 9) – la certificazione rileva competenze specifiche: morfolessicali sin dal livello A1, sociopragmatiche ai livelli B, testuali ai livelli C.

2. La prova oggetto di studio

Novello descrive in maniera sintetica le caratteristiche di un *test a scelta multipla* (Novello, 2014, p. 63; il corsivo è nostro):

L'item a scelta multipla prevede che lo studente scelga tra più opzioni, in genere tre o quattro, l'alternativa corretta. È particolarmente impiegato nella verifica della comprensione scritta e orale [...] e della competenza morfosintattica. *La scelta multipla* permette di testare molti elementi, tra i quali inferenze e opinioni, in modo più controllato rispetto alle domande aperte. È un tipo di esercizio adatto a tutti i livelli di competenza comunicativa in quanto può essere costruito in molti modi, prevedendo l'utilizzo di immagini, singole parole, frasi brevi o frasi complesse. La costruzione di un *item a scelta multipla* si dimostra particolarmente complessa in quanto necessita della creazione di tre o quattro opzioni adatte al compito, senza possibilità assurde (che non porterebbero a nessuna riflessione, ma solo ad una scelta per eliminazione) o che il candidato possa intuire senza aver letto/ascoltato il testo¹.

¹ Nell'Appendice A rimandiamo a una serie di indicazioni pratiche relative alla confezione di un *test a scelta multipla*.

Il *test a scelta multipla* oggetto di analisi si compone di sette *item*, con quattro opzioni ciascuno, riferiti a un testo scritto relativo a fatti d'attualità. Assieme all'*individuazione di informazioni* (che è una forma elaborata di *vero/falso*; Torresan, 2014) e all'incastro o riordino di testo (noto anche come *jigsaw* o *sequencing task*; Torresan, 2015), il *quesito a scelta multipla* contribuisce a definire il *set* di prove di cui si dota il certificatore per valutare l'abilità di lettura del candidato.

Testo e *item* sono stati confezionati in modo da accertare le competenze di un candidato di livello B1 (CILS UNO). La prova è stata presentata nella sessione estiva del 2009 (Centro CILS, 2009).

La nostra indagine mira ad accertarne la validità. Per validità intendiamo la capacità di misurare effettivamente ciò che la prova si propone di misurare, vale a dire la comprensione di un testo scritto da parte di uno studente di italiano (per un prospetto generale sulla questione della validità, D'Este 2013). In particolare l'abilità di comprensione va rapportata al profilo di uno studente di livello B1, così sintetizzato nelle *Linee guida CILS* (Barni *et al.*, 2009, p. 30) "Il candidato sa comprendere il senso globale e le principali informazioni presenti in testi scritti con scrittura lineare di argomento quotidiano e generale".

Ci concentriamo, in particolare, sui seguenti elementi:

- *istruzioni e input* (paragrafo 2.1.)
- *item* (paragrafo 2.2.)

2.1. Le istruzioni e l'input

Le istruzioni risultano chiare e precise, come si può osservare nel seguente esempio estratto dal test:

Completa le seguenti frasi. Scegli una delle quattro proposte di completamento che ti diamo per ogni frase.

Il testo della prova di lettura presenta consistenti modifiche rispetto all'originale da cui è stato tratto, ossia un articolo del *Corriere della Sera* (Taglietti, 2007).

Le *modifiche* operate sul testo originale (d'ora in poi TO) sono tese a:

- un '*alleggerimento*' del testo di arrivo, con l'*espunzione di informazioni (probabilmente) ritenute accessorie o ridondanti*
- una *semplificazione della lingua del testo di partenza (in termini morfosintattici e lessicali)*

Nella presentazione affiancata dei testi – TO e testo modificato (d'ora in poi TM) – illustriamo le modifiche occorse. Evidenziamo, inoltre, criticità che emergono nel TM.

Tabella 2

Testo originale e testo modificato a confronto

Testo originale	Testo modificato CILS
<u>TENDENZE ALLA FIERA DI BOLOGNA ARRIVANO NARRATORI GIOVANISSIMI. RIGOROSAMENTE UNDER 20</u> «E adesso i libri ce li scriviamo da soli». La nuova generazione di autori adolescenti Ragazzi che (forse) leggono poco, ma certo scrivono molto. Per rendersene conto basterà fare un giro tra gli stand della Fiera del libro per ragazzi che domani apre i battenti a Bologna, tradizionale appuntamento dove si vendono e comprano i diritti e rituale occasione per fare il punto sulla situazione di un mercato sempre sull'orlo della crisi. L'offerta di romanzi per i cosiddetti young adults è sempre ampia, i dati dell'Aie (Associazione italiana editori) quest'anno sono addirittura più confortanti (rilevano un aumento della lettura nella fascia che va dagli 11 ai 17 anni), ma ciò che emerge come una tendenza nuova è la crescita di una giovane leva di autori che sanno parlare ai loro coetanei affrontando, con ambizioni ed esiti diversi, qualunque genere. Così se una delle iniziative di maggior successo della prima edizione del Fantasio festival di Perugia è stato il concorso «G.a.s. al minimo» che ha raccolto le storie brevi di oltre tremila giovani aspiranti scrittori tra i 17 e i 23 anni (37 sono state scelte e pubblicate in un volume), a Bologna autrici come la ventunenne Sara Boero al suo terzo romanzo con <i>Piumo di drago</i> (Piemme) o la ventiseienne Licia Troisi, bestsellerista affermata con la trilogia delle <i>Cronache del mondo emerso</i> (Mondadori), sono ormai delle veterane. L'editoria sforna continuamente giovani autori (anzi più spesso autrici), essi il nuovo caso letterario è la sedicenne Valentina F., autrice di TVUKDB, cioè «Ti voglio un kasino di bene», diario di un anno nella vita di una ragazza romana tra scuola, amicizia, amori. «Io scrivo da anni un diario che faccio leggere solo alla mia amica del cuore, Marta, la ragazza di cui parlo anche nel libro. Insieme abbiamo pensato che era fico farlo pubblicare» spiega Valentina. Il libro non ha nulla di spinto o di pruriginoso, Valentina F. non è Melissa P. e il massimo a cui si può arrivare è un bacio; il cognome di Valentina, il suo aspetto, persino la sua voce sono rigorosamente protetti dalla famiglia che non vuole gettarla nel circo mediatico (al punto da fa pensare che Valentina sia solo un nom de plume), eppure in due settimane il diario ha venduto 12 mila copie costringendo l'editore Fanucci che lo pubblica nella collana Teens a una ristampa. Per Valentina il livello dei libri per ragazzi è abbastanza buono (l'ultimo che ha letto è lo sono di legno di Giulia Carcasi, <i>Tre metri sopra il cielo</i> lo ha letto addirittura tre volte) e si possono «trovare un sacco di storie che parlano di noi, anche se spesso non raccontano come siamo e viviamo veramente». Tuttavia i ragazzi non scrivono soltanto di loro stessi, del loro mondo, dei loro bisogni e desideri, ma si cimentano anche su temi forti, difficili da raccontare. Basti pensare a Randa Ghazy nata a Saronno da genitori egiziani che, nel 2002, appena quindicenne, sconvolse la Fiera con il suo <i>Segnando Palestina</i>, all'indomani dell'11 settembre. Quest'anno, ormai diciannovenne, ripropone un altro titolo forte, <i>Forse oggi non ammazzo nessuno</i>, storia di Jasmine, «una giovane musulmana stranamente non terrorista» che, con l'arma dell'ironia riesce a stare a cavallo tra due mondi, quello musulmano e quello occidentale, senza appartenere pienamente a nessuno. Un tema difficile, soprattutto da raccontare a un pubblico di adolescenti, è	“E ADESSO I LIBRI CE LI SCRIVIAMO DA SOLI”. LA NUOVA GENERAZIONE DI AUTORI ADOLESCENTI <u>Secondo recenti indagini in Italia i</u> ragazzi leggono <u>ancora troppo</u> poco, ma certo scrivono molto. Per rendersene conto basta fare un giro tra gli stand della Fiera del libro per ragazzi a Bologna, aperta dal 27 aprile, dove arrivano scrittori giovanissimi, rigorosamente sotto i 20 anni. <u>La Fiera di Bologna</u> è l'occasione per fare il punto sulla situazione del mercato dell'editoria. L'offerta dei romanzi per i giovani è sempre ampia. I dati dell'Aie - Associazione italiana editori - quest'anno rilevano un aumento della lettura nella fascia che va dagli 11 ai 17 anni, ma ciò che emerge come una tendenza nuova è la crescita di una giovane schiera di autori che parlano ai loro coetanei e affrontano qualunque genere letterario. Una iniziativa di successo per i giovani è il concorso letterario G.A.S.- Giovani Aspiranti Scrittori, che si svolge ormai da due anni all'interno del festival Fantasia di Perugia. Nel 2008 G.A.S. ha raccolto le storie brevi di oltre tremila giovani aspiranti scrittori tra i 17 e i 23 anni. <u>La casa editrice PIM ha pubblicato in un volume 37 di queste storie, molti altri racconti sono invece disponibili, con relativa scheda identificativa dell'autore, sul sito internet www.fantasiafestival.it.</u> L'editoria sforna continuamente giovani autori, più spesso autrici. Il nuovo caso letterario è la sedicenne Valentina F., autrice di [PTM1➔] TVUKDB “Ti voglio un Kasino di bene”, diario di un anno della vita di una ragazza di Roma, tra scuola, amicizia, amori. “Io scrivo da anni un diario che faccio leggere solo alla mia amica del cuore, Marta, la ragazza di cui parlo anche nel libro. Insieme abbiamo pensato di pubblicarlo” spiega Valentina. Valentina non è il suo nome, la

~~anche quello affrontato da Erika Silvestri, ventenne di Ladispoli. Nel suo // commerciante di bottoni (Fabbri), introdotto da una prefazione di Walter Veltroni, racconta la storia di Piero Terracina, sopravvissuto ad Auschwitz, conosciuto quando aveva 14 anni. «Il libro è una raccolta di incontri, di conversazioni quasi private. Non volevo che l'esperienza di Piero, a cui sono legata come se fosse un nonno, andasse perduta». Erika racconta di aver sempre amato scrivere, e anche leggere: «Ma non ho mai letto i cosiddetti libri per ragazzi. Leggo i libri "normali", molti saggi, perché amo i temi storici, in particolare quelli legati all'Olocausto». Erika di Federico Moccia ha soltanto sentito parlare. «Non lo conosco perché non è il mio genere, ma probabilmente il suo successo sta nel fatto che ha trovato un linguaggio adatto per parlare ai ragazzi».~~

Anche Maurizio **Temporin, diciottenne** di Novi Ligure, non legge la cosiddetta letteratura per ragazzi, ~~non considera il suo romanzo, *Il tango delle cattedrali* (che uscirà da Rizzoli il 2 maggio) appartenente al tanto popolare genere fantasy, nonostante la presenza di un gargoyle nelle prime pagine («se proprio dovessi scegliere un genere direi che è fantascienza», spiega): *Il Tango delle cattedrali* è un libro **dall'architettura complessa e dalle grandi ambizioni**, che parla di tango, di arte, **di immortalità**, che **mescola** mondi **paralleli e giochi** letterari, **Oscar Wilde** e **gli** amici dell'autore. Maurizio **ha interrotto gli studi al terzo anno del liceo scientifico**, ~~a 15 anni è andato a vivere con la sua ragazza~~, scrive da quando **ne aveva 12**. Se gli parli di Federico Moccia risponde con Fabrizio De André, Ray Bradbury, Tim Burton e Federico Fellini.~~

*** ERIKA

Erika Silvestri, 20 anni, 1 vive a Ladispoli e studia storia medievale alla Sapienza di Roma. Ama dipingere e disegnare

*** MAURIZIO

Maurizio Temporin, 18 anni, vive tra Novi Ligure e Barcellona. **Ha interrotto gli**

studi al liceo scientifico, ama il tango e gli scacchi

*** VALENTINA

Valentina F., 16 anni, romana, studia al liceo scientifico. **Non vuole rivelare la**

sua vera identità (la foto è stata ritoccata)

*** LA KERMESSE

La Fiera internazionale del libro per ragazzi di **Bologna è aperta da domani al 27**

aprile. Vi prendono parte circa 1200 editori, provenienti da 54 Paesi.

Al Paese ospite, il Belgio, è dedicata una sezione speciale della Mostra degli illustratori che quest'anno propone 85 artisti provenienti da 23 Paesi.

Informazioni e programma si possono trovare al sito

www.bookfair.bolognafiere.it.

famiglia vuole mantenere nascosta la sua vera identità, la sua immagine, persino la sua voce. Il diario in due settimane ha venduto 12 mila copie e l'editore ha dovuto fare anche una ristampa. Per Valentina il livello dei libri per ragazzi è abbastanza buono: "Ci sono molti scrittori grandi che scrivono buone storie che parlano di noi, ma spesso non raccontano come siamo e come viviamo veramente".

[PTM2→] Anche il diciottenne Maurizio T. [PTM3→] di Novi Ligure non legge la letteratura per ragazzi, **ma scrive romanzi**. Il suo ultimo libro, *Il tango delle cattedrali* [PTM4→] è un libro **complesso**, che parla di tango e di arte, che **mischia** mondi **letterari** e amici dell'autore. Maurizio scrive da quando **aveva 12 anni** e ora ha **anche** interrotto **la scuola** **per dedicarsi interamente alla scrittura**.

Legenda. Testo originale: **blu barrato** - Parte soggetta a espunzione (informazioni eliminate); **grassetto grigio** - Parte soggetta a manipolazione significativa. Testo modificato CILS: **blu sottolineato** - Inserzioni (informazioni aggiunte); **PTM** - Problemi nel TM (testo modificato); **Riquadrettato** - Le informazioni equivalenti su porzioni distinte dei testi sono poste all'interno di riquadri.

Le criticità presenti nel TM riguardano, per la maggior parte dei casi, la punteggiatura. In un caso, a nostro giudizio di particolare rilievo, l'espunzione di informazioni presenti nel TO produce incoerenza nel TM (PTM2).

Nella Tabella 3 presentiamo in dettaglio il prospetto delle criticità presenti nel TM (gli emendamenti da noi suggeriti vengono rappresentati, di volta in volta, in grassetto).

Tabella 3
Criticità presenti nel testo modificato

	Testo modificato	Emendamento suggerito	Commento
PTM1	Il nuovo caso letterario è la sedicenne Valentina F., autrice di [➡] TVUKDB “Ti voglio un Kasino di bene”, diario di un anno della vita di una ragazza di Roma, tra scuola, amicizia, amori.	Il nuovo caso letterario è la sedicenne Valentina F., autrice di [➡] “TVUKDB [➡] . Ti voglio un [➡] kasino di bene”, diario di un anno della vita di una ragazza di Roma, tra scuola, amicizia, amori.	Occorre racchiudere il titolo per intero tra virgolette, inserire un punto tra titolo e sottotitolo, rendere in minuscolo l'iniziale della parola “kasino”.
PTM2	[➡] Anche il diciottenne Maurizio T. di Novi Ligure non legge la letteratura per ragazzi, ma scrive romanzi.	Anche il diciottenne Maurizio T. di Novi Ligure [➡] [non legge la letteratura per ragazzi, ma] scrive romanzi.	Il dato di cui si parla (il fatto che Maurizio T. non legga libri per ragazzi) si raccorda con quanto è stato espunto dal TO (il fatto che Erika Silvestri non legga libri per ragazzi) e non con il dato che invece precede l'informazione nel TM (Valentina legge libri per ragazzi). L'emendamento che suggeriamo, posto che sia plausibile alterare il testo come il certificatore ha fatto, è l'espunzione dell'informazione.
PTM3	Anche il diciottenne Maurizio T. [➡] di Novi Ligure non legge la letteratura per ragazzi, ma scrive romanzi.	Anche il diciottenne Maurizio T. [➡] , di Novi Ligure [➡] , non legge la letteratura per ragazzi, ma scrive romanzi.	L'inciso andrebbe compreso tra virgole.
PTM4	Il suo ultimo libro, <i>Il tango delle cattedrali</i> [➡] è un libro complesso, che parla di tango e di arte, che mischia mondi letterari e amici dell'autore.	Il suo ultimo libro, <i>Il tango delle cattedrali</i> [➡], è un libro complesso, che parla di tango e di arte, che mischia mondi letterari e amici dell'autore.	È necessaria una virgola a chiusura dell'inciso.

Riportiamo ora le nostre osservazioni in merito alla scelta del testo (TO) e alle operazioni che su di esso sono state eseguite (TM).

Per quanto riguarda il testo, il certificatore ha agito in conformità alle indicazioni presenti nelle *Linee guida CILS* (Barni *et al.*, 2009). Il testo di riferimento è infatti un articolo informativo-divulgativo e come tale rientra nei generi testuali relativi al livello B1, secondo le suddette *Linee* (Barni *et al.* 2009, p. 30: “lettere, brani narrativi, articoli informativo-divulgativi, regolamenti”).

Nello stesso documento (Barni *et al.*, 2009, p. 30), peraltro, “si presume che i testi”, in una prova di lettura rivolta a un candidato di livello B1, “siano comprensibili ad un italofono con istruzione di base”. Questo secondo aspetto non è, a nostro avviso, evidente nel TO: vi compaiono infatti riferimenti culturali che possono risultare di difficile presa per un italofono con istruzione di base.

Se poi lo confrontiamo con i descrittori del CEFR relativi alle competenze attese da un lettore di livello B1, il TO non risulta adatto. È lungo, sintatticamente complesso, tutt'altro che lineare, e l'argomento di cui tratta (l'editoria giovanile) non è considerabile familiare per un candidato-tipo. Ciò può aver spinto il

certificatore a operare un rimaneggiamento sia a livello micro (operazioni lessicali, morfologiche e sintattiche circoscritte) che a livello macro:

- riorientando il tema: alla Fiera di Bologna viene dato un rapido accenno;
- espungendo parti ritenute accessorie: i resoconti dell'attività letteraria di un paio di autrici vengono eliminati;
- riorganizzando le informazioni: si riprendono dati presenti nei box laterali e li si incorpora nell'articolo.

L'operazione, tuttavia, è da considerarsi non riuscita. Il TM risulta infatti diviso in due blocchi, il blocco A (la Fiera di Bologna e il concorso di Perugia) e il blocco B (due autori, Valentina F. e Maurizio T.), ciascuno dei quali presenta delle difficoltà.

Nel Blocco A:

- gli eventi paiono giustapposti (sia a causa della modifica del layout in virtù della quale le informazioni vengono compattate, sia per la mancanza di dispositivi coesivi che mettano in relazione quanto descritto);
- l'informazione iniziale ("*i ragazzi leggono ancora troppo poco*") mal si accorda con quanto emerge dai dati dell'Aie ("*un aumento della lettura nella fascia che va dagli 11 ai 17 anni*").

Nel Blocco B:

- l'informazione di apertura ("*L'editoria sforna continuamente giovani autori, più spesso autrici*") pare non adeguatamente sviluppata. In sostanza, mentre riceve una giustificazione nel TO (dato che ad essa segue la menzione dell'attività di scrittura da parte di tre ragazze), nel TM pare avere una scarsa tenuta con l'intero blocco (perché seguita dalla descrizione della produzione letteraria di una ragazza e di un ragazzo);
- si dà un cortocircuito logico, come già visto:
 - "*Per Valentina il livello dei libri dei ragazzi è abbastanza buono*";
 - "*Anche il diciottenne Maurizio T. di Novi Ligure non legge letteratura per ragazzi*".

Le *Guidelines for Writers of Reading Test* (d'ora in poi GWRT) allestite dal *British Council* ungherese (Alderson & Cseresznyés, 2002) e punto di riferimento per molti *item writer*² invitano a non manipolare in alcun modo le fonti:

Do not make any changes to the original text. Do not delete words, sentences or paragraphs from the selected body of the text. If text contains any offensive words that you think should be replaced, only change these with great care and always seek the advice of a fellow teacher or a native speaker as to the acceptability of the changes you have made. (paragrafo 3.1.10)³

Una posizione meno intransigente può ammettere operazioni superficiali, puntuali, che non comportino uno stravolgimento della struttura testuale.

Nel caso in esame, come abbiamo visto, le manomissioni sono, invece, consistenti. Il passaggio dal TO al TM presenta i caratteri di una giustapposizione poco sorvegliata, con pesanti ricadute sulla resa del testo di arrivo.

Il problema alla base, secondo noi, sta nella scelta della fonte; è infatti evidente che operazioni micro sono possibili solo se la fonte presenta caratteristiche testuali in accordo con quelle previste dal CEFR per quel dato livello. In altre parole, se il testo di partenza avesse presentato i caratteri di semplicità e linearità e se l'argomento fosse risultato familiare (come previsto per il livello B1), il certificatore non sarebbe dovuto

² Si possono consultare all'interno del progetto *Into Europe* (Alderson & Cseresznyés, 2002).

³ Circa la legittimità/opportunità della semplificazione dei testi in ambito glottodidattico, rimandiamo a Humphris in Derosas, Torresan, 2009; Spinelli, 2000; Urbani, 1989; Wilkins, 1972, 1976.

intervenire così massicciamente e, con grande probabilità, non sarebbe incorso in un risultato così poco soddisfacente.

2.2. Gli item

Nella sezione che segue consideriamo:

- la distribuzione delle corrispondenze delle chiavi nel testo (paragrafo 2.3.1.);
- l'analisi statistica (classica e Rasch) delle risposte agli item da parte di un campione di studenti (paragrafi 2.2.2.; 2.2.3.);
- l'attribuzione di livello degli item da parte di un gruppo di formandi esaminatori (paragrafo 2.2.4.);
- l'analisi dei distrattori (paragrafo 2.2.5.).

2.2.1. La distribuzione delle corrispondenze delle chiavi nel testo

Le corrispondenze delle chiavi paiono adeguatamente distribuite lungo il testo, coerentemente con quanto si legge nelle GWRT: "Items must follow the text sequence" (punto 2.3.2.).

Tabella 4

La distribuzione delle corrispondenze delle chiavi nel testo

Testo modificato	Item
[1] Secondo recenti indagini in Italia i ragazzi leggono ancora troppo poco, ma certo scrivono molto. Per rendersene conto basta fare un giro tra gli stand della Fiera del libro per ragazzi a Bologna, aperta dal 27 aprile, dove arrivano scrittori giovanissimi, rigorosamente sotto i 20 anni. [2] La Fiera di Bologna è l'occasione per fare il punto sulla situazione del mercato dell'editoria. L'offerta dei romanzi per i giovani è sempre ampia. I dati dell'Aie - Associazione italiana editori - quest'anno rilevano un aumento della lettura nella fascia che va dagli 11 ai 17 anni, ma ciò che emerge come una tendenza nuova è la crescita di una giovane schiera di autori che parlano ai loro coetanei e affrontano qualunque genere letterario. [3] Una iniziativa di successo per i giovani è il concorso letterario G.A.S.- Giovani Aspiranti Scrittori, che si svolge ormai da due anni all'interno del festival Fantasia di Perugia. Nel 2008 G.A.S. ha raccolto le storie brevi di oltre tremila giovani aspiranti scrittori tra i 17 e i 23 anni. La casa editrice PIM ha pubblicato in un volume 37 di queste storie, molti altri racconti sono invece disponibili, con relativa scheda identificativa dell'autore, sul sito internet www.fantasiefestival.it. [4] L'editoria sforna continuamente giovani autori, più spesso autrici. Il nuovo caso letterario è la sedicenne Valentina F., autrice di TVUKDB "Ti voglio un Kasino di bene", [5] diario di un anno della vita di una ragazza di Roma, tra scuola, amicizia, amori. "Io scrivo da anni un diario che faccio leggere solo alla mia amica del cuore, Marta, la ragazza di cui parlo anche nel libro. Insieme abbiamo pensato di pubblicarlo" spiega Valentina. Valentina non è il suo nome, la famiglia vuole mantenere nascosta la sua vera identità, la sua immagine, persino la sua voce. Il diario in due settimane ha venduto 12 mila copie e l'editore ha dovuto fare anche una ristampa. Per Valentina il livello dei libri per ragazzi è abbastanza buono: [6] "Ci sono molti scrittori grandi che	1. Ricerche recenti sull'editoria hanno rilevato che in Italia sono in aumento A) i giovani scrittori. B) i lettori di romanzi. C) le pubblicazioni per ragazzi. D) le case editrici di piccole dimensioni. 2. La Fiera del libro di Bologna offre l'occasione per A) intervistare molti autori contemporanei. B) conoscere le nuove tendenze letterarie. C) comprare libri a prezzi scontati. D) acquistare testi rari e originali. 3. Il concorso letterario G.A.S. ha permesso a diversi giovani di A) farsi conoscere dal grande pubblico. B) riuscire a guadagnare qualche soldo. C) partecipare a festival letterari internazionali. D) vedere pubblicati i propri racconti. 4. Un dato che emerge nell'editoria è che A) prevalgono le autrici straniere. B) molti scrittori preferiscono pubblicare su internet. C) gli autori amano diffondere immagini di loro stessi. D) le ragazze scrivono di più dei ragazzi. 5. Valentina F. ha pubblicato un diario A) in cui racconta un anno della sua vita. B) redatto a due mani con la sua amica. C) dove parla del suo rapporto con la famiglia. D) scritto per sfuggire alla depressione. 6. Secondo Valentina F. gli scrittori adulti A) non sanno raccontare il mondo degli adolescenti.

scrivono buone storie che parlano di noi, ma spesso non raccontano come siamo e come viviamo veramente".

Anche il diciottenne Maurizio T. di Novi Ligure non legge la letteratura per ragazzi, ma scrive romanzi. [7] *Il suo ultimo libro, Il tango delle cattedrali è un libro complesso, che parla di tango e di arte, che mischia mondi letterari e amici dell'autore.* Maurizio scrive da quando aveva 12 anni e ora ha anche interrotto la scuola per dedicarsi interamente alla scrittura.

B) usano uno stile troppo elaborato.

C) non conoscono la scrittura abbreviata dei giovani.

D) scrivono romanzi lunghi e noiosi.

7. Maurizio T. ha scritto

A) storie per ragazzini.

B) un romanzo sul tango.

C) un manuale sul ballo.

D) racconti fantastici.

2.2.2. Analisi classica degli item

Un item 'funziona' nella misura in cui discrimina gli esaminandi in base alla competenza, vale a dire distingue i candidati abili da quelli che non lo sono, in riferimento al livello per il quale il test è stato confezionato. Se l'item risulta troppo facile (e quindi non discrimina a sufficienza) o, al contrario, troppo difficile, la prova è in disequilibrio rispetto al livello al quale viene riferita (McNamara, 2000, pp. 60-61):

Where the text purpose is to make distinctions between candidates, to spread them out in terms of their performance on the texts, the items should be neither too easy nor too difficult. If the items are too easy, then people with differing levels of ability or knowledge will all get them right, and the differences in ability or knowledge will not be revealed by the item. Similarly if the items are too hard, then able and less able candidates alike will get them wrong, and the item won't help us in distinguishing between them.

Un indicatore utile, in tal senso, è la percentuale di studenti che hanno risposto correttamente ai singoli item, il cosiddetto *facility value* (d'ora in poi FV). Quale FV considerare per stabilire l'appropriatezza di un item rispetto al livello dichiarato? Per un *proficiency test*, McNamara sostiene quanto segue:

Ideal item facility [i.e. Facility Value, ndt.] is 0.5 [i.e. 50%, ndt.] but of course it is hard to hit this target exactly, and a range of item facilities from 0.33 [i.e. 33%, ndt.] to 0.67 [i.e. 67%, ndt.] is usually accepted. (McNamara, 2000, p. 67)

Poi precisa:

Even though [...] items that are very easy (item with high item facility) do not distinguish between candidates, it may be useful to include some at the beginning of a test in order to ease candidates into the test and to allow them a chance to get over their nerves⁴. (McNamara, 2000, p. 67)

Rita Green riconosce margini più ampi:

The most useful information about a test taker's proficiency comes from facility values which are around 50 per cent (see Popham, 2000) as this value suggests that the item might be discriminating between the test takers (though it is not always the case); 40 to 60 per cent (see Bachman, 2004), which are the values many test developers use when making initial decisions about whether an item is working or not. Facility values between 20 and 80 per cent can also provide useful information provided the items still discriminate and contribute to the test's internal consistency [...]. Facility values of below 20 per cent and above 80 per cent in a proficiency test suggest that most of the test population is either answering the item incorrectly or correctly respectively, which reduces the chance of gaining useful information about the test takers and/or the items. (Green, 2013, pp. 26-27)

Abbiamo somministrato la prova a 179 studenti di livello B1, avvalendoci della disponibilità di colleghi che operano sia in Italia (in prevalenza nelle scuole private) che all'estero (in prevalenza negli Istituti Italiani di Cultura e nelle università).

Nel prospetto che segue abbiamo raccolto gli FV (Tabella 4) ed evidenziato le risposte nulle (omesse o doppie). Posto che il nostro campione di riferimento sia significativo per numero e tipologia (179 studenti,

⁴ Per il ruolo esercitato dal *placebo effect* sulla motivazione, cfr. Geers *et al.*, 2005.

di nazionalità diversa) e soprattutto per livello (B1, stando all'individuazione di livello da parte della ventina di docenti coinvolti)⁵, abbiamo riscontrato la presenza di due item (#1 e #7) molto facili ($FV > 90$), e di un item (#4) particolarmente difficile ($FV = 31,3$). L'item #4 e l'item #6 presentano, inoltre, un elevato numero di risposte nulle.

Tabella 5
FV e risposte nulle

Item	FV	Risposte nulle
1	92,2	0
2	81,6	3
3	57,0	1
4	31,3	7
5	64,2	3
6	81,6	6
7	92,2	0

Il seguente istogramma (Figura 1) illustra la distribuzione dei punteggi complessivi raggiunti dal campione. La maggior parte degli studenti a cui il test è stato somministrato (51) ha risposto correttamente a 6 item su 7 (la colonna in riferimento al #6 sulle ascisse è, in effetti, la più alta). Si tratta di un profilo con disallineamento negativo (*negatively skewed*; in sostanza, la coda della curva a campana punta verso lo zero, quindi idealmente verso i valori negativi): il test è, cioè, relativamente facile per il campione a cui è stato somministrato.

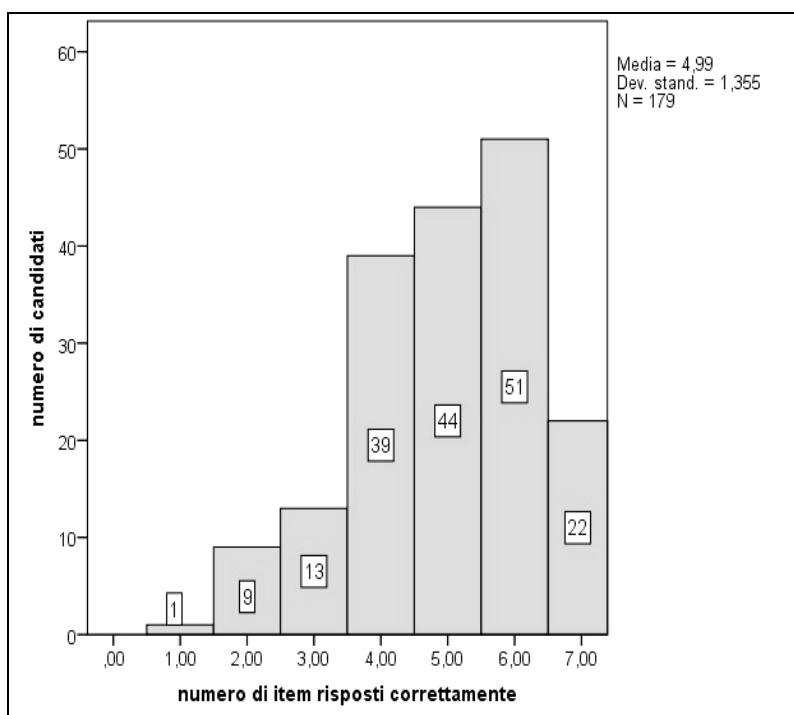


Figura 1. Punteggi dei candidati alla prova di lettura a scelta multipla, Certificazione CILS, Sessione giugno 2009, livello B1.

⁵ L'individuazione del livello delle singole classi è stata operata dagli stessi insegnanti che hanno collaborato alla raccolta dati. È ad essi, dunque, che ci siamo affidati. Nel caso in cui singoli docenti manifestassero perplessità, abbiamo fornito loro le indicazioni del CEFR, e ci siamo sintonizzati del livello attraverso l'analisi del manuale di italiano in adozione.

Per quanto concerne la *discriminazione*, ovvero la capacità del test di distinguere gli studenti abili da quelli meno abili, l'indice di riferimento è la *Correlazione del totale item corretta*⁶.

Questo indice oscilla da +1 a -1. Valori superiori allo 0.30 (o 0.25 a dipendere dal grado di rigore, cfr. Henning, 1987) ci informano che il test discrimina adeguatamente; vale a dire, il candidato che risponde correttamente all'*item* ha maggiore probabilità di rispondere correttamente agli altri *item*. Valori inferiori allo 0.30 (0.25, vedi sopra) o addirittura negativi ci indicano il contrario: l'*item* non discrimina adeguatamente (può capitare che a rispondere correttamente siano i meno abili, per via di strategie di *guessing*).

Un indice di discriminazione basso (inferiore allo 0.30/0.25) può essere dovuto a:

- sotto o sovradimensionamento dell'*item*;
- cattiva qualità dell'*item* (doppia chiave, ambiguità della chiave, ecc.);
- comportamento divergente dell'*item* rispetto agli altri; l'*item* cioè sta verificando qualcosa di diverso rispetto agli altri⁷. Si ha, in tal senso, un caso di *construct-irrelevant variance*: una variabile esterna al costrutto esercita un impatto sul comportamento del candidato (nel nostro caso ciò significherebbe che per rispondere correttamente alla prova di lettura non è coinvolta solo l'abilità di comprensione scritta ma anche altre competenze, come per esempio il *background* culturale).

Un *item* con un basso indice di discriminazione deve essere, in ogni caso, scartato. Nella tabella sottostante riportiamo la *Correlazione del totale item corretta* degli *item* in questione. Da notare il valore basso di tre *item*, #2, #4, #7 (evidenziati in grigio).

Tabella 6
*Correlazione del totale item
corretta*

<i>Item</i>	Correlazione del totale item corretta
1	,258
2	1,69
3	,293
4	,145
5	,265
6	,249
7	,078

Risalendo a quanto detto sopra, e anticipando i risultati dell'analisi degli *item* che verranno discussi nel paragrafo 2.3.5:

- l'*item* #2 presenta caratteristiche di construct-irrelevant variance, le quali inficiano, a nostro dire, la validità della prova;
- l'*item* #4 è sovradimensionato;
- l'*item* #7 è sottodimensionato.

⁶ Non ci occupiamo dell'affidabilità dei singoli *item* (Alfa di Cronbach). Il numero esiguo degli *item* non ci permette, infatti, di pervenire a dati attendibili; imporrebbe indagini fattoriali che esulano dalla ricerca in corso (si veda Green, 2013, pp. 39-40).

⁷ In sostanza, la *Correlazione del totale item corretta* ci restituisce informazioni sulla 'compattezza' della prova: sappiamo cioè se un *item* 'lavora' in sinergia con gli altri o ha, per così dire, un comportamento a sé.

2.2.3. Analisi Rasch sul rapporto tra la difficoltà degli item e l'abilità dei candidati

L'analisi Rasch è un modello di calcolo probabilistico in cui l'abilità dei candidati viene messa in relazione alla difficoltà degli *item*. Il modello si basa sulla premessa secondo cui la possibilità di un candidato di rispondere correttamente a un *item* è una funzione della sua abilità e della difficoltà dell'*item* (Henning, 1987, pp. 107-8).

Riprendiamo da McNamara quanto segue:

Consider the case of a dichotomous scored item [com'è il nostro caso, N.d.A.], that is one in which the responses may be 'correct' or 'incorrect'. If we know how an individual has performed on other items (we have some estimate of his or her ability), and if we know how other individuals have performed on the item in question (we have an estimate of how difficult the item is), then we are in a position to make an estimate of how likely it is that the individual will respond in a particular way –that is, to choose either the 'correct' or the 'incorrect' answer. If an item is easy, and the individual ability is high, there will be a high likelihood of a correct response. On the other hand, if the item is difficult, and the individual ability is low is unlikely that he or she will make a 'correct' response [...]. Note that we are talking about expectancies, either outcome ('correct' or 'incorrect') is theoretically possible for each response to each item, but in proportions that will differ as function of the item and the estimated ability of the candidate. (McNamara, 1996, p. 160)

Un esempio. Immaginiamo che i candidati A e B abbiano ottenuto il medesimo punteggio, avendo risposto correttamente a una batteria di 5 *item*, e quindi totalizzando 5 punti ciascuno, uno per *item*. Consideriamo però che l'*item* #5 sia di diversa difficoltà: più difficile nel caso di B (Figura 2). In un'analisi statistica classica, questa differenza non viene apprezzata. In realtà, nonostante il punteggio sia equivalente, la competenza che B ha modo di dimostrare è maggiore.

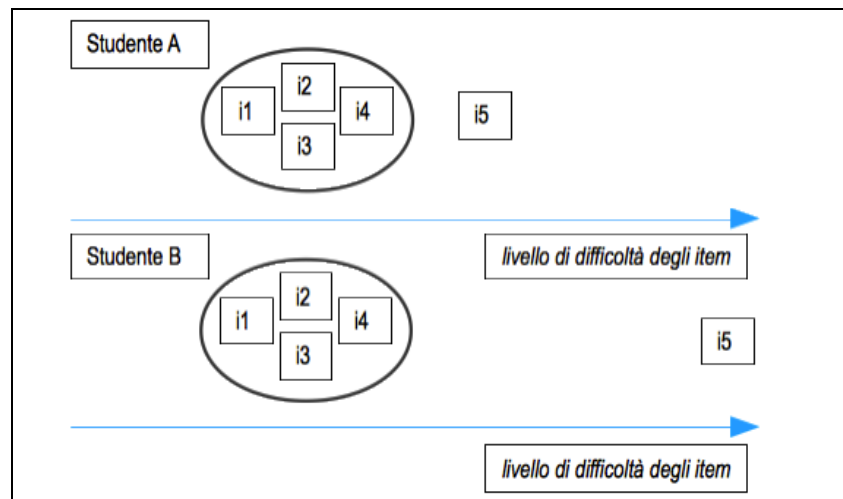


Figura 2. Diverse abilità degli studenti a parità di punteggio (raw score)

Una rilevazione Rasch tiene conto di questi fattori, distinguendo il punteggio osservato (*Obs. Average* nella Tabella 7), che di fatto equivale al FV (si veda Tabella 5, paragrafo precedente), dal punteggio atteso, ovvero determinato secondo l'analisi della probabilità (*Fair Average*, nella Tabella 7) e sensibile al rapporto difficoltà dell'*item*/abilità del candidato.

Nella Tabella 7 emerge la notevole discrepanza tra il punteggio osservato dell'*item* #4 (0.31) e il punteggio atteso (0.15). In sostanza, posto che il nostro campione sia rappresentativo, un candidato di livello B1 ha un 15% di probabilità di rispondere correttamente all'*item* #4, che risulta con evidenza un *item*

sovradimensionato (l'*item* è evidenziato in grigio). Gli *item* più facili, invece (pur'essi evidenziati in grigio), hanno un punteggio atteso che ne accentua il carattere sottostimato (#1 e #7, rispettivamente 0.94 e 0.95)⁸.

La *Fit Statistics*, rappresentata sempre nella Tabella 7, accerta inoltre l'adeguatezza (*fit*) dei dati raccolti a un modello di discriminazione secondo il quale è più probabile che (si veda Bond & Fox, 2007):

- un candidato risponda a un *item* semplice rispetto a uno difficile;
- i candidati più abili rispondano correttamente ai diversi *item*, a differenza di quelli meno abili.

Se non si verificano queste premesse, possiamo dedurre situazioni di divergenza, vale a dire di superuniformità (*overfitting*) o di variabilità eccessiva (*misfitting*).

Per i candidati, tali divergenze sono rappresentabili in termini di:

- *overfitting*, assenza di variabilità nelle risposte (es. un candidato che dia sempre la stessa risposta, del tipo: 1b, 2b, 3b, 4b, 5b, 6b, 7b);
- *misfitting*, variabilità eccessiva delle risposte (es.: candidati abili che rispondono in maniera scorretta a *item* facili, o viceversa, candidati poco abili che rispondono correttamente a *item* difficili).

Per gli *item*, le divergenze si traducono così:

- *overfitting*, assenza di interdipendenza tra gli *item* (i.e. ridondanza⁹);
- *misfitting*, variazioni nella *performance* dei candidati in riferimento a quel singolo *item*, che risultano inesplicite per il modello, cioè inattese (dovute per esempio a *item bias*, a trascrizioni errate delle risposte¹⁰ o a multidimensionalità per via di fattori esterni a quelli previsti¹¹).

Nella nostra indagine, ciò cui prestiamo attenzione non è il comportamento dei candidati. L'oggetto dello studio è la prova; pertanto torna utile concentrarci sul 'funzionamento' degli *item*. I risultati ci confortano: i valori di riferimento (la *ZStd*, *Standardized Fit Statistics*, con un campione $N < 200$, come nel nostro caso) rientrano nella norma, vale a dire tra ± 2 , con zero come valore atteso; come è stato messo in luce da Linacre, 2012b)

⁸ Come già detto, la nostra analisi si basa su una distribuzione con disallineamento negativo. La statistica conferma (significatività pari a zero nel test di normalità secondo il metodo Shapiro-Wilk). Una misurazione Rasch può perdere in affidabilità con una distribuzione non-normale, come sostiene Wright (2001), "In general, skew reduces reliability, so your reliability should have increased a little with the normalized sample" (p. 786). Linacre tuttavia contempla la possibilità di una distribuzione tale e quale appare nel nostro caso (il corsivo è nostro): "*Persons*: we usually expect a normal distribution, or a distribution skewed in a certain way. [...] *Items*: we usually expect the items to be uniformly distributed (like marks on a ruler) or clustered at pass-fail points" (2012a, slide 36, punto 1).

⁹ Scrive Linacre (2000): "For measurement purposes, we want an item that focuses on the same thing as the other items, but produces one item's worth of new information. A useful item is «as similar as possible, but as different as possible». Including the same item twice in a test is nearly always redundant. Such items do not provide two item's worth of new information, since responses to one are overly predictable from responses to the other" (p. 755).

¹⁰ Scrive Linacre (1995): "One item may be cuing another or an item may be miskeyed. A response form may have been unintentionally scanned twice or may have been mis-scanned" (p. 6).

¹¹ Per esempio le strategie di *guessing* costituiscono una *dimensione* 'altra' rispetto all'abilità del candidato, sicché le risposte che derivano da quelle 'inquinano', per così dire, la nostra lettura dei dati; si veda a tal proposito Linacre (1998); o ancora il fatto di disporre di un costruito complesso, com'è il caso di un *test* di abilità integrate, come il riassunto, previsto per esempio nella certificazione ISE del Trinity College o nella Certificazione CELI dell'Università per Stranieri di Perugia.

Tabella 7¹²*Fit Statistics riferita agli item*

Tot score	Tot count	Obsvd Average	Fair (M) Average	Measure	Model S. E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtMea	Corr. PtExp	it
55	179	.31	.15	2.82	.21	1.16	1.4	1.37	1.3	.74	.51	.52	i4
102	179	.57	.50	1.11	.18	.92	-.9	.88	-1.0	1.17	.60	.54	i3
115	179	.64	.61	.66	.19	.95	-.6	.87	-1.0	1.14	.57	.53	i5
145	179	.81	.84	-.52	.22	.92	-.6	1.00	.0	1.07	.47	.44	i6
147	179	.82	.85	-.61	.22	1.06	.4	.92	-.2	.97	.42	.44	i2
164	179	.92	.94	-1.69	.29	.89	-.5	.54	-1.1	1.11	.40	.33	i1
165	179	.92	.95	-1.78	.30	1.13	.6	.96	.0	.92	.26	.32	i7
Tot score	Tot count	Obsvd Average	Fair (M) Average	Measure	Model S. E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtMea	Corr. PtExp	it

Legenda. Model, Populn: RMSE .24 Adj (True) S.D. 1.51 Separation 6.41 Strata 8.88 Reliability .98; Model, Sample: RMSE .24 Adj (True) S.D. 1.63 Separation 6.93 Strata 9.58 Reliability .98; Model, Fixed (all same) chi-square: 288.1 d.f.: 6 significance (probability): .00; Model, Random (normal) chi-square: 5.9 d.f.: 5 significance (probability): .32

Va rilevata comunque la presenza di numerose risposte inattese in corrispondenza all'*item* #4 (Tabella 8). È un'*item* difficile, abbiamo detto, a cui un numero discreto di studenti meno abili (N=6) ha risposto convenientemente (le risposte inattese sono evidenziate in grigio).

Tabella 8
Risposte inattese

Cat	Score	Exp.	Resd	StRes	Num	per	it
1	1	.0	1.0	5.3	110	110	i4
1	1	.0	1.0	5.3	166	166	i4
0	0	1.0	-1.0	-4.8	89	89	i4
0	0	1.0	-1.0	-4.6	50	50	i7
0	0	1.0	-1.0	-4.6	56	56	i7
0	0	1.0	-1.0	-4.6	57	57	i7
0	0	1.0	-1.0	-4.6	75	75	i6
0	0	1.0	-1.0	-4.6	164	164	i7
0	0	1.0	-1.0	-4.6	173	173	i6
0	0	1.0	-1.0	-4.6	179	179	i6
0	0	1.0	-1.0	-4.4	98	98	i1
1	1	.1	.9	3.5	13	13	i4
1	1	.1	.9	3.5	59	59	i4
1	1	.1	.9	3.5	92	92	i4
1	1	.1	.9	3.5	162	162	i4
Cat	Score	Exp.	Resd	StRes	Num	per	it

¹² Per approfondimenti sulle singole voci, si rimanda a Green (2013, pp. 218-219).

Nella Figura 4 (*Vertical Ruler*, pagina 37) è rappresentata una mappatura complessiva del rapporto difficoltà degli *item*/abilità dei candidati, espresso in *logit*. La scala *logit* è determinata da intervalli costanti, attorno ai quali sono rapportati, da un lato, gli *item* secondo il loro grado di difficoltà, dall'altro, i candidati, secondo la loro abilità.

Per rendere più immediata la comprensione, nella Figura 3, qui sotto, abbiamo immaginato di costruire una scala analoga, a intervalli regolari, tratta da Mellinson (2001). Su questa scala si è messo in relazione l'impegno che un'attività fisica comporta con il grado di energia di vari soggetti. Una persona allo stremo delle forze (con un grado di energia pari a 1) è *assai probabile* sia in grado di svolgere compiti leggeri, come alzarsi dal letto o farsi la doccia; *meno probabile* abbia energia in quantità sufficiente per curare il giardino, e ancora meno per giocare a pallone. Al contrario, sarà *quasi scontato* che la persona con il grado di energia maggiore (11) possa alzarsi dal letto e farsi una doccia; potrà *forse* aver la forza di radere l'erba e *presumibilmente* –anche se è *meno evidente* –sarà in grado di stare in campo per l'intera durata di una partita di calcio.

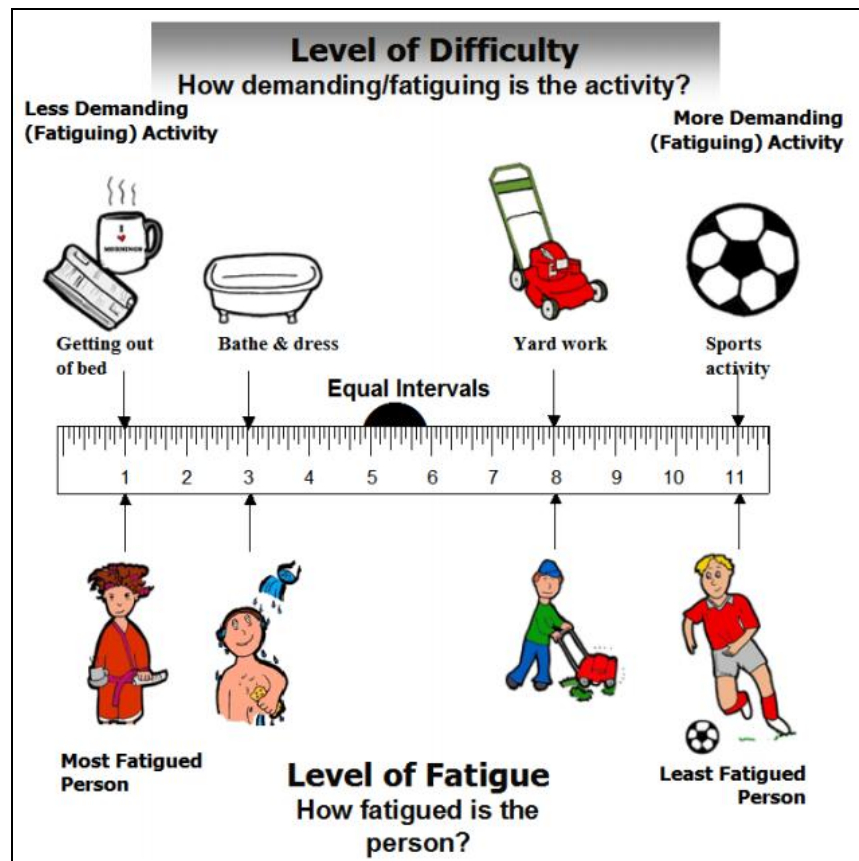


Figura 3. Mappa del rapporto tra il grado di impegno di attività fisiche e i livelli di energia della persona (tratta da Mallinson, 2001). Riprodotta con il gentile permesso di Trudy Mallinson.

Una scala a intervalli regolari costruita secondo una logica Rasch definisce pari a 50% la probabilità di una persona di svolgere con successo un compito che ha un *logit* equivalente al suo grado di energia, abilità, intelligenza, ecc., a dipendere dal parametro che si osserva. Venendo a noi, la scala cui facciamo riferimento (Figura 4), è disposta verticalmente (*Vertical Ruler*):

- nella prima colonna abbiamo *la scala logit*;
- nella seconda sono distribuiti i candidati a seconda dell'abilità (quelli più abili sono disposti nella parte superiore mentre quelli meno abili nella parte inferiore);

- la terza si riferisce agli *item*, a seconda del grado di difficoltà (gli *item* più difficili sono distribuiti nella parte superiore mentre quelli più facili sono distribuiti nella parte inferiore).

Seguendo la logica espressa sopra (in riferimento al rapporto impegno fisico/grado di energia del soggetto), candidati allineati all'*item* #6 (cioè che condividono lo stesso *logit* con l'*item* #6), evidenziati dalla linea rossa - sempre nella Figura 4 - hanno il 50% di probabilità di rispondere correttamente a quell'*item* (e a *item* di pari difficoltà): quell'*item*, in altre parole, costituisce una sfida possibile, non troppo facile né troppo difficile; la probabilità aumenta in riferimento all'*item* #1 (1 *logit* circa di differenza; 75% di probabilità; è quindi molto più probabile che i candidati in questione rispondano correttamente ad esso), mentre diminuisce in maniera drastica in riferimento all'*item* #4, spostato verso l'estremità superiore della scala (3,5 *logit* di scarto; 3% di probabilità). Queste percentuali sono calcolabili mediante tabelle di convergenza (Green 2013, p. 165), che riportiamo qui sotto (Tabella 9).

Tabella 9
Tabelle di convergenza

Valori positivi		Valori negativi	
Differenza tra l'abilità del candidato e la difficoltà dell' <i>item</i>	Probabilità di risposta corretta	Differenza tra l'abilità del candidato e la difficoltà dell' <i>item</i>	Probabilità di risposta corretta
5.0	99%	-5.0	1%
4.6	99%	-4.6	1%
4.0	98%	-4.0	2%
3.0	95%	-3.0	5%
2.2	90%	-2.2	10%
2.0	88%	-2.0	12%
1.4	80%	-1.4	20%
1.1	75%	-1.1	25%
1.0	73%	-1.0	27%
0.8	70%	-0.8	30%
0.5	62%	-0.5	38%
0.4	60%	-0.4	40%
0.2	55%	-0.2	45%
0.1	52%	-0.1	48%
0	50%	0	50%

Il dato che emerge dalla rappresentazione nella Figura 4 è un *mismatching* tra l'abilità dei candidati e la difficoltà degli *item*: la maggioranza dei *test taker* è spostata verso la parte superiore della scala (sopra lo zero), mentre quattro *item* su sette risultano concentrati, invece, nella parte inferiore. Si tratta di una conferma della relativa facilità della prova rispetto all'abilità dei candidati, di cui abbiamo parlato nella sezione precedente. Spicca, ancora una volta, la situazione atipica dell'*item* #4, al limite dell'accessibilità, così come si nota il sottodimensionamento degli *item* #1 e #7.

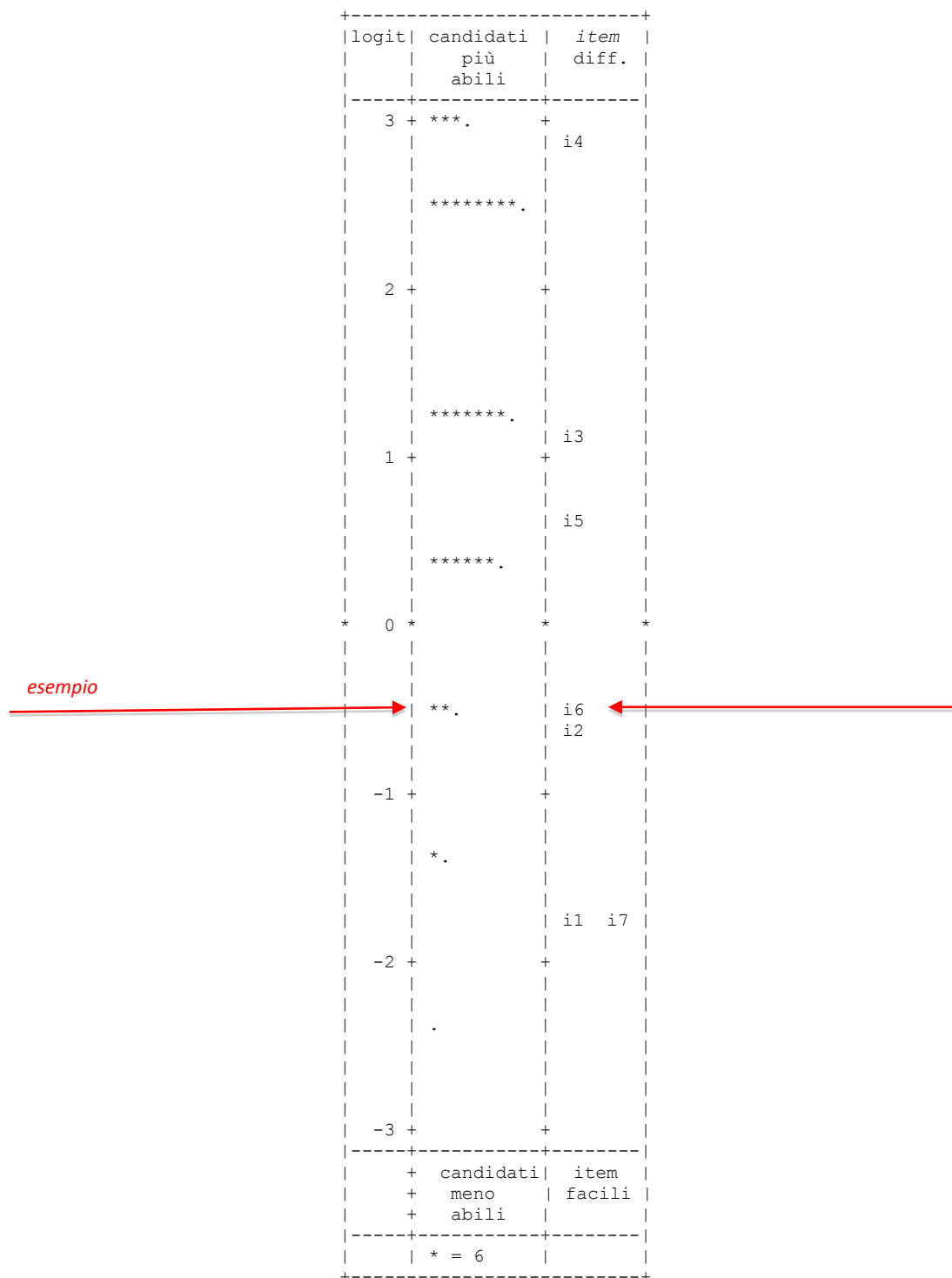


Figura 4. Vertical Ruler

2.2.4. Attribuzione di livello degli item

Abbiamo sottoposto la prova a 36 formandi esaminatori, chiedendo loro di ragionare sul livello dei singoli *item*, con il CEFR alla mano.

L'attribuzione è avvenuta considerando paritempo il testo di riferimento e i singoli *item*, come indicato dal Manuale collegato al CEFR:

It is essential to consider both the question of the **level of the source text** and the **difficulty of the individual item(s)** associated with it. A text does not have a "level". It is the competence of the test takers as demonstrated by their responses to the items that can be related to a CEFR level. The most that can be said about a text is that it is suitable for inclusion in a test aimed at a particular level. (Council of Europe, 2009, p. 61)

L'attribuzione di livello costituisce l'azione principale di un'operazione di *Standard Setting*, vale a dire di calibrazione degli *item*. La calibrazione concorre a garantire la qualità di una prova: un test sotto- o sovradimensionato non consente di formulare giudizi attendibili sulla competenza del candidato (si vedano Bejar, 2008; Bejar, Braun & Tannenbaum, 2007; Cizek & Bunch, 2007; Hambleton & Pitoniak, 2006; Näsström & Nyström, 2008; Zieky, Perie, & Livingstone, 2008). Nella Figura 5, tratta da Green e Spoetl (2011), illustriamo come la calibrazione si inserisca all'interno del processo di costruzione di un test.

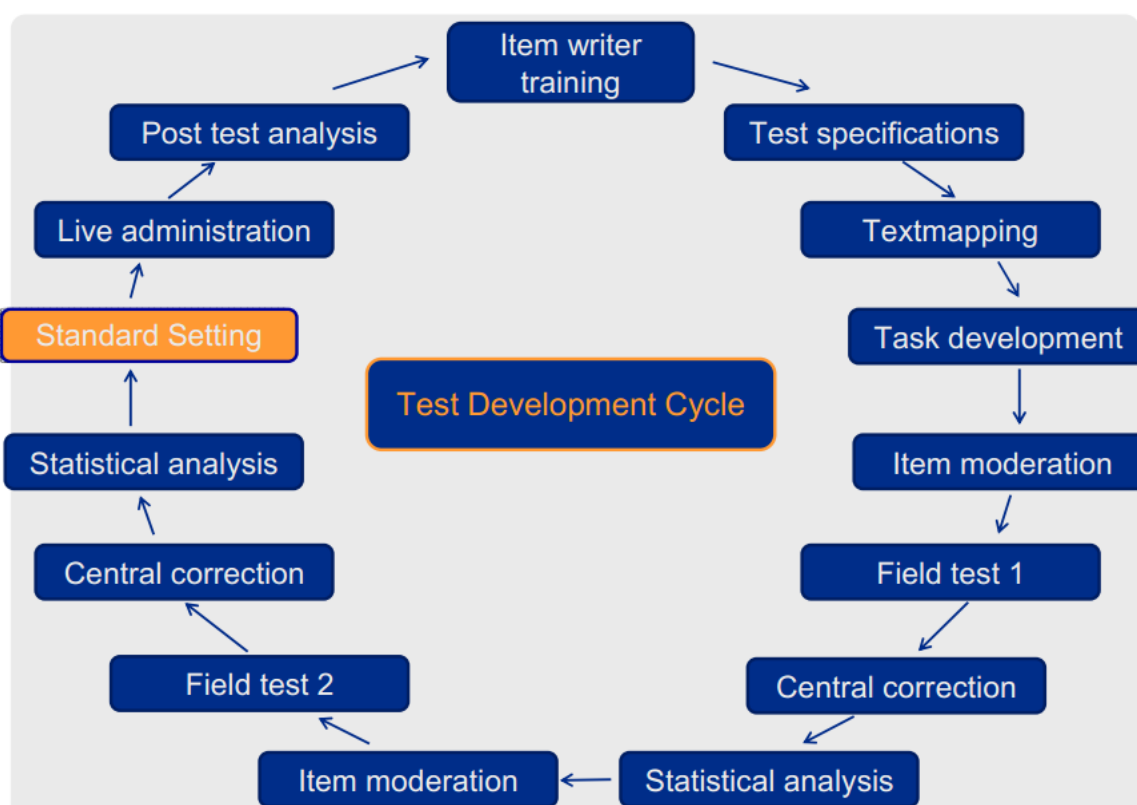


Figura 5. L'attribuzione di livello all'interno di un processo di calibrazione degli *item* (Green & Spoetl 2011). Riprodotta con permesso di Rita Green e Carol Spoetl.

Durante una sessione di *Standard Setting*, i giudici sono tenuti a (Freunberger, Bazinger, & Itzlinger-Brunefort, 2014):

- familiarizzare con i parametri di riferimento (nel nostro caso, descrittori del CEFR relativi all'abilità di lettura);
- svolgere la prova e confrontare le proprie risposte con le chiavi;
- attribuire un livello a ciascun *item* sulla base dei parametri di riferimento (cfr. punto a);

- d. confrontare il proprio giudizio con quello degli altri giudici e prendere visione dei risultati dell'analisi statistica classica;
- e. confermare il proprio giudizio o attribuire un nuovo livello sulla base di quanto emerso nella fase precedente.

Nella nostra esperienza ci siamo limitati alle prime tre azioni (a-c). Il giudizio che ne deriva manca quindi di quel raffinamento che si sarebbe potuti avere grazie al confronto collettivo.

Cionondimeno, riteniamo utile condividere con il lettore i risultati dal momento che costituiscono un dato ulteriore da triangolare con quelli già in possesso. Nell'Appendice B riportiamo i giudizi occorsi durante la sessione.

Nella Tabella 9 illustriamo una sintesi delle occorrenze dei giudizi degli esaminatori. In grassetto sono evidenziati i soli *item* riguardo al livello dei quali il giudizio degli esaminatori collima con i dati raccolti.

Tabella 10

Occorrenze delle attribuzioni di livello operate dagli esaminatori

<i>Item</i>	A2	B1	B2	FV
1	8	23	5	92,2
2	2	16	18	81,6
3	1	16	19	57,0
4	12	19	5	31,3
5	26	10	0	64,2
6	4	29	3	81,6
7	16	18	2	92,2

Solo per gli *item* #3, #6, #7 il giudizio degli esaminatori conferma i dati che provengono dall'*Item Analysis* (FV). Diversamente:

- l'*item* #1 è ritenuto calibrato, mentre dalle analisi statistiche risulta sottodimensionato
- l'*item* #4 è considerato calibrato, mentre dalle analisi statistiche risulta sovradimensionato
- l'*item* #5 è reputato sottodimensionato, mentre dalle analisi statistiche risulta calibrato

È assai probabile che se i *rater* avessero condiviso i giudizi e avessero avuto modo di confrontarsi tra loro, per procedere poi a una seconda valutazione, le divergenze in merito al livello di alcuni *item* (#2, #3, #4, #5, #7) si sarebbero ridotte. La questione viene ripresa nella sezione che segue.

2.2.5. Analisi dei distrattori

Consideriamo in dettaglio il comportamento dei distrattori durante la somministrazione della prova al campione di studenti.

Item #1. È tra gli *item* più facili, non discrimina a sufficienza. Il distrattore (d) non assolve al suo ruolo, non ‘distrare’ (un solo studente l’ha scelto); rappresenta, in effetti, un dato di cui non si ha notizia nel testo.

Tabella 11
Analisi dei distrattori dell’item #1

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A = Chiave	165	92,2%	92,2%	92,2%
B	7	3,9%	3,9%	96,1%
C	6	3,4%	3,4%	99,4%
D	1	0,6%	0,6%	100,0%
Totale	179	100,0%	100,0%	

Legenda. Item #1. Ricerche recenti sull’editoria hanno rilevato che in Italia sono in aumento: A) i giovani scrittori. B) i lettori di romanzi. C) le pubblicazioni per ragazzi. D) le case editrici di piccole dimensioni.

Rileviamo un’imprecisione nella formulazione dello *stem*. Nel TM si dice che ci sono molti scrittori giovani: “*Secondo recenti indagini in Italia i ragazzi leggono ancora troppo poco, ma certo scrivono molto*”.

Questo però non significa che ci sia un “*aumento*” rispetto al passato, come si sostiene invece nello *stem* “*Ricerche recenti sull’editoria hanno rilevato che in Italia sono in aumento i giovani scrittori*”. (Su questo genere di problemi, si veda Peirce, 1992).

In merito ai dati che provengono dall’attribuzione di livello, l’*item* è stato reputato, dalla maggioranza degli esaminatori, come adeguato al livello B1.

Tabella 12
Giudizio dei rater in merito al livello dell’item #1

Item	A2	B1	B2	FV
1	8	23	5	92,2

In effetti, se scorriamo i riferimenti al *Quadro*, alla voce “Leggere per orientarsi” troviamo una rappresentazione delle strategie richieste e del tipo di testo che si confà alla prova oggetto di studio.

B1	LEGGERE PER ORIENTARSI
	È in grado di scorrere testi di una certa lunghezza alla ricerca di informazioni specifiche e di reperire in punti diversi di un testo [...] le informazioni necessarie per portare a termine un compito specifico.

La facilità dell’*item* (FV=92,2) si può allora giustificare, a nostro avviso, per:

- la debolezza dei distrattori;
- l’ancoraggio *stem*/testo mediante la ripresa del lessico “*ricerche recenti*”/“*recenti indagini*”; il lettore è, quindi, condotto alla risposta corretta.

Item #2. È un *item* relativamente facile per il campione cui è stato somministrato (FV=82).

Tabella 13
Analisi dei distrattori dell'item #2

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A	10	5,6%	5,7%	5,7%
B = Chiave	146	81,6%	83,0%	88,6%
C	17	9,5%	9,7%	98,3%
D	3	1,7%	1,7%	100,0%
Totale	176	98,3%	100,0%	
Mancanti				
Nessuna risposta	3	1,7%		
Totale	179	100,0%		

Legenda. Item #2. La Fiera del libro di Bologna offre l'occasione per A) intervistare molti autori contemporanei. B) conoscere le nuove tendenze letterarie. C) comprare libri a prezzi scontati. D) acquistare testi rari e originali.

Si dà comunque, a nostro avviso, un problema di logica nella costruzione.

La chiave recita: “La Fiera del libro di Bologna offre l'occasione per conoscere le nuove tendenze letterarie”. Nel TM si legge: “La Fiera di Bologna è l'occasione per fare il punto sulla situazione del mercato dell'editoria”. Il fatto di “conoscere nuove tendenze” (chiave) è un possibile risultato del fare “il punto sulla situazione del mercato dell'editoria” (TM).

Scrivo Giacomo Pierini, responsabile *marketing* ALMA Edizioni (comunicazione personale, gennaio 2014): “Fare il punto [in un contesto editoriale] è un calderone e vuol dire: tendenze; vendite; canali distributivi; prezzi; *ebook* [...]; nuovi generi, come il *comic*”.

Chiave e rimando nel testo non sono, in sostanza, perfettamente sovrapponibili. E in ogni caso, anche ammesso lo siano, l'associazione delle informazioni rischia di dipendere da conoscenze specifiche, che non è detto un candidato possieda (*construct-irrelevant variance*, cui abbiamo accennato al paragrafo 2.2.2). Del resto, proprio la difficoltà legata a questo processo di inferenza può aver guidato i formandi a ritenere questo *item* collocabile tra B1/B2.

L'alto FV (82) fa supporre comunque che, nonostante la criticità della chiave, l'*item* non abbia comportato difficoltà significative al campione di studio. A ciò ha contribuito, molto plausibilmente, la scarsa attrattività dei distrattori, (a) e (d) in particolar modo.

Item #3. L'item si presenta come impegnativo (FV=57); così emerge anche nel giudizio degli esaminatori.

Tabella 14
Analisi dei distrattori dell'item #3

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A	44	24,6%	24,7%	24,7%
B	3	1,7%	1,7%	26,4%
C	29	16,2%	16,3%	42,7%
D = Chiave	102	57,0%	57,3%	100,0%
Totale	178	99,4%	100,0%	
Mancanti				
Doppia risposta	1	0,6%		
Totale	179	100,0%		

Legenda. Item #3. Il concorso letterario G.A.S. ha permesso a diversi giovani di A) farsi conoscere dal grande pubblico. B) riuscire a guadagnare qualche soldo. C) partecipare a festival letterari internazionali. D) vedere pubblicati i propri racconti.

Balza all'occhio il comportamento del distrattore (a) che raccoglie molti consensi (quasi il 25%).

Si può pensare che il *test taker* possa esser stato confuso da questo distrattore, dal momento che esso si sovrappone, in termini di logica, alla chiave (d). Dire cioè che la Fiera di Bologna consente "a diversi giovani" di "farsi conoscere dal grande pubblico" (distrattore a) è una naturale conseguenza del fatto che i loro racconti siano stati oggetto di pubblicazione (chiave).

Oltretutto c'è da riconoscere che abbiamo due tipi di pubblicazione:

- una a stampa, di cui si fa carico l'editore PIM (e di cui trae vantaggio un gruppo ristretto di giovani scrittori);
- una *online* sul sito del Festival (a favore di un gruppo più ampio).

Noi non sappiamo se il *test taker* che ha scelto il distrattore (a) avesse in mente gli aspiranti scrittori il cui racconto è stato pubblicato *online*, e i quali, grazie a ciò, hanno avuto modo di "farsi conoscere dal grande pubblico":

Una iniziativa di successo per i giovani è il concorso letterario G.A.S.- Giovani Aspiranti Scrittori, che si svolge ormai da due anni all'interno del festival Fantasia di Perugia. Nel 2008 G.A.S. ha raccolto le storie brevi di oltre tremila giovani aspiranti scrittori tra i 17 e i 23 anni. La casa editrice PIM ha pubblicato in un volume 37 di queste storie, molti altri racconti sono invece disponibili, con relativa scheda identificativa dell'autore, sul sito [internet www.fantasiafestival.it](http://www.fantasiafestival.it). (Centro CILS, 2009)

Rupp *et al.* (2006) mettono in evidenza, la correlazione negativa tra la somiglianza delle opzioni e il FV (maggiore è la prima, minore il secondo).

Carr, da parte sua, ammonisce sulla necessità di disporre di chiavi precise e incontrovertibili (2011, p. 90): "There should be one option that is clearly correct, not two or more that are partially correct".

Item #4. L'item si presenta come molto difficile (FV=31,3).

Due terzi dei volontari hanno optato per i distrattori (b) e (c). Si registra un numero significativo di omissioni (3) e di risposte doppie (4). Dall'analisi Rasch emerge un numero discreto di comportamenti disattesi: candidati poco abili che indovinano la risposta.

L'intero quadro ci informa di una generale confusione tra i lettori.

Tabella 15
Analisi dei distrattori dell'item #4

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A	16	8,9%	9,3%	9,3%
B	49	27,4%	28,5%	37,8%
C	50	27,9%	29,1%	66,9%
D = Chiave	56	31,3%	32,6%	99,4%
M	1	0,6%	0,6%	100,0%
Totale	172	96,1%	100,0%	
Mancanti				
Nessuna risposta	3	1,7%		
Doppia risposta	4	2,2%		
Totale	7	3,9%		
Totale	179	100,0%		

Legenda. Item #4. Un dato che emerge nell'editoria è che A) prevalgono le autrici straniere. B) molti scrittori preferiscono pubblicare su internet C) gli autori amano diffondere immagini di loro stessi. D) le ragazze scrivono di più dei ragazzi.

Va segnalato, *in primis*, che il passaggio del TM cui l'item fa riferimento è espresso con un lessico di bassa frequenza: "L'editoria sforna continuamente giovani autori, più spesso autrici"; il verbo "sfornare" rimanda a una competenza avanzata (livello C).

Si potrebbe ribattere che il fatto di non conoscere la parola "sforna" non inficia la comprensione del passaggio, posto che un allievo di livello intermedio riesca a dedurre il significato di una parola sconosciuta, facendo leva su indizi contestuali.

È vero però che, da un lato, si registra un co-testo povero, dall'altro, se ci rifacciamo al CEFR, una lettura di tipo *top-down*, a un livello B1, è legata alla familiarità dell'argomento (il corsivo è nostro).

	INDIVIDUARE INDIZI E FARE INFERENZE
B1	È in grado di identificare in base al contesto parole sconosciute, relativamente ad argomenti che si riferiscono al suo campo di interesse. È in grado di estrapolare dal contesto il significato di una parola sconosciuta e ricostruire il significato della frase, a condizione di avere familiarità con l'argomento in questione

Collegli che insegnano italiano ad Augsburg ci hanno fatto notare, poi, che la chiave e il rimando del TM non si sovrappongono perfettamente; dire che le ragazze scrivono più dei ragazzi (chiave) non significa che i loro scritti siano i più pubblicati (TM). In effetti, scrivere è condizione necessaria, ma non sufficiente, per la pubblicazione.

I collegli segnalano, inoltre, la possibilità che uno studente di livello B1, in virtù di una lettura inferenziale, sia portato a dedurre che se è vero (TM) che le donne pubblicano di più sul cartaceo ("editoria") potrebbe essere, allora, che i maschi pubblicano di più sull'*online*, come appunto vuole il distrattore (b): "molti scrittori preferiscono pubblicare su internet".

Quest'ultima osservazione andrebbe verificata sottoponendo un questionario metacognitivo a ulteriori candidati, attraverso il quale indagare i processi che li hanno guidati nella scelta dell'opzione. Questo, infatti, non è confermabile con i dati che abbiamo a disposizione (difficoltà del lessico nel TM; possibilità di letture inferenziali, per un livello B1, legata alla familiarità dell'argomento; risposte inattese da parte degli studenti meno abili e non da parte dei più abili). Va comunque riconosciuto il grande rischio legato a 'finestre' di plausibilità che un distrattore può comportare, e tale eventualità va ridotta al minimo. A tal proposito Buck annota che "it is difficult to say with certainty that any inference is completely wrong. After all, we are free to infer what we want" (2001, p. 148).

Da parte nostra siamo convinti che l'attrattività del distrattore (b) sia stata determinata dal fatto che i lettori si siano lasciati 'distrarre' da un dato riportato in precedenza:

La casa editrice PIM ha pubblicato in un volume 37 di queste storie, molti altri racconti sono invece disponibili, con relativa scheda identificativa dell'autore, sul sito internet www.fantasiafestival.it. L'editoria sforna continuamente giovani autori, più spesso autrici. (Centro CILS, 2009)

Una lettura attenta avrebbe fatto cogliere che non si tratta di una preferenza (da parte degli autori), quanto di un'opportunità (o condizione, a seconda dei punti di vista) riservata a coloro le cui storie non hanno avuto modo di essere pubblicate a stampa.

Per quanto riguarda, invece, il distrattore (c), "*gli autori amano diffondere immagini di loro stessi*" la scelta dei candidati può essere giustificata dal fatto che, immediatamente dopo la corrispondenza della chiave nel TM ("*l'editoria sforna continuamente giovani autori, più spesso autrici*"), si fa riferimento a una ragazza che parla di sé attraverso un diario:

L'editoria sforna continuamente giovani autori, più spesso autrici. Il nuovo caso letterario è la sedicenne Valentina F., autrice di TVUKDB «Ti voglio un Kasino di bene», diario di un anno della vita di una ragazza di Roma, tra scuola, amicizia, amori. (Centro CILS, 2009)

C'è da riconoscere che si tratta di una circostanza particolare, relativa a una specifica autrice, e non di un "*dato che emerge dall'editoria*" (come recita lo *stem*). Sono pochi gli esaminatori (N=5) che hanno ravvisato la difficoltà dell'*item*. L'*item* è stato giudicato, dalla maggioranza, come collocabile tra il livello A2 e il livello B1.

Tabella 16
Giudizio dei rater in merito al livello dell'*item* #4

Item	A2	B1	B2	FV
4	12	19	5	31,3

L'attribuzione di livello, in tal caso, non considera la scarsa familiarità dei *test taker* con l'argomento in questione né dà conto dell'esiguità del co-testo nel quale l'informazione è contenuta; infine alla maggioranza degli esaminatori sfugge la non sovrapposizione tra chiave e rimando nel TM.

Assai probabile, come già detto, che se l'attribuzione di livello si fosse completata con un confronto tra i giudici, come vuole una sessione di *Standard Setting*, tali difficoltà sarebbero emerse, e quindi probabilmente si sarebbe raggiunto un giudizio finale complessivo diverso.

Item #5. L'item ha un FV pari a 64,2.

Tabella 17

Analisi dei distrattori dell'item #5

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A = Chiave	115	64,2%	65,3%	65,3%
B	55	30,7%	31,3%	96,6%
C	5	2,8%	2,8%	99,4%
D	1	0,6%	0,6%	100,0%
Totale	176	98,3%	100,0%	
Mancanti				
Nessuna risposta	2	1,1%		
Doppia risposta	1	0,6%		
Totale	3	1,7%		
Totale	179	100,0%		

Legenda. Item #5. Valentina F. ha pubblicato un diario A) in cui racconta un anno della sua vita. B) redatto a due mani con la sua amica. C) dove parla del suo rapporto con la famiglia. D) scritto per sfuggire alla depressione.

Il distrattore (b) è stato scelto da un terzo del campione. L'attrattività può essere dovuta alla bassa frequenza del lessico con cui è stato formulato: *"redatto a due mani"*. Tale *chunk* può essere confuso con *"insieme abbiamo pensato di pubblicarlo"* (TM), riferito alla complicità con l'amica, da parte di Valentina F., nell'inoltare il diario a una casa editrice (il corsivo è nostro).

Il nuovo caso letterario è la sedicenne Valentina F., autrice di TVUKDB «Ti voglio un Kasino di bene», diario di un anno della vita di una ragazza di Roma, tra scuola, amicizia, amori. «Io scrivo da anni un diario che faccio leggere solo alla mia amica del cuore, Marta, la ragazza di cui parlo anche nel libro. *Insieme abbiamo pensato di pubblicarlo*» spiega Valentina.

C'è da chiedersi se sia legittimo servirsi di opzioni formulate con un lessico sovradimensionato rispetto a quello che si suppone di certificare. La nostra risposta è negativa, dal momento che l'obiettivo è la comprensione del testo e non dell'item.

Segnaliamo, inoltre, il fatto che il distrattore (d), il quale allude a un'informazione di cui non si ha traccia nel testo, sia stato scelto da un solo allievo.

Molti esaminatori hanno sottostimato l'item; è parso loro corrispondere a un livello A2. Si sono avvalsi delle seguenti scale.

A2	COMPRESIONE GENERALE
	È in grado di comprendere testi brevi e semplici di contenuto familiare e di tipo concreto, formulati nel linguaggio che ricorre frequentemente nella vita di tutti i giorni o sul lavoro.
A2	LEGGERE PER ORIENTARSI
	È in grado di trovare informazione specifiche in semplice materiale scritto di uso corrente che gli/le può capitare tra le mani, quali lettere, opuscoli e brevi articoli di cronaca.

In realtà, dal nostro punto di vista, il testo cui l'item fa riferimento non è né "breve" né "semplice", non costituisce un "materiale scritto di uso corrente" né si riscontra l'impiego di un "linguaggio che ricorre frequentemente nella vita di tutti i giorni".

Item #6. L'item è risultato piuttosto facile (FV=81,6). Il distrattore (d) ha costituito un'opzione plausibile solo per un paio di candidati. Da notare, un numero significativo di risposte nulle.

Tabella 18
Analisi dei distrattori dell'item #6

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A = Chiave	146	81,6%	84,4%	84,4%
B	15	8,4%	8,7%	93,1%
C	10	5,6%	5,8%	98,8%
D	2	1,1%	1,2%	100,0%
Totale	173	96,6%	100,0%	
Mancanti				
Nessuna risposta	5	2,8%		
Doppia risposta	1	0,6%		
Totale	6	3,4%		
Totale	179	100,0%		

Legenda. Item #6. Secondo Valentina F. gli scrittori adulti A) non sanno raccontare il mondo degli adolescenti. B) usano uno stile troppo elaborato. C) non conoscono la scrittura abbreviata dei giovani. D) scrivono romanzi lunghi e noiosi.

Item #7. L'item si rivela molto facile (FV=92). Il distrattore (c) non è stato scelto da nessuno. Scarse sono state pure le convergenze sugli altri distrattori.

Tabella 19
Analisi dei distrattori dell'item #7

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi				
A	8	4,5	4,5	4,5
B = Chiave	165	92,2	92,2	96,6
C	0	0	0	0
D	6	3,4	3,4	100,0
Totale	179	100,0	100,0	

Legenda. Item #7. Maurizio T. ha scritto A) storie per ragazzini. B) un romanzo sul tango. C) un manuale sul ballo. D) racconti fantastici

3. Bilancio conclusivo

La nostra indagine si è concentrata sul testo e sugli *item* della prova. La natura della ricerca, abbiamo visto, è illustrativa: lo scopo è quello di indurre il lettore a considerare possibili insidie durante le fasi di confezione di un *quesito a scelta multipla*. La prova della certificazione CILS, in tal senso, ha valso da 'pretesto' per una riflessione ampia sulla validazione di una prova oggettiva di comprensione.

L'analisi è articolata in diverse fasi e, oltre al nostro personale giudizio, conta su una serie di informazioni raccolte mediante strumenti di vario genere.

Quale premessa delle riflessioni cui giungiamo alla fine, vanno riconosciuti margini di approssimazione durante:

- la raccolta dei dati: l'attribuzione del livello delle classi-campione è stata affidata a colleghi; da parte nostra, abbiamo operato un controllo esterno piuttosto limitato, come dichiarato in precedenza; un volume maggiore di dati avrebbe inoltre garantito una maggiore affidabilità alle indagini;
- l'attribuzione di livello degli *item*: una procedura più precisa e significativa sarebbe stata garantita se avessimo contemplato l'intero arco di azioni cui si compone lo *Standard Setting*; la stessa sarebbe stata ottimizzata, peraltro, se avessimo disposto, al massimo, della metà di giudici convocati.

Riconoscere tali limiti, significa per noi, contestualizzare i dati raccolti, definire un perimetro all'interno del quale attribuire un valore alla ricerca. Ad ogni modo, tali limiti, non invalidano le conclusioni cui siamo pervenuti e che presentiamo a seguire, distinte a seconda dell'oggetto: il *testo* e gli *item*.

Testo. Si rileva che, per quanto riguarda il testo, sono state operate:

- una scelta poco avveduta della fonte (TO) sia per quanto riguarda il tema – che risulta poco familiare - che in termini di struttura, poiché il testo è lungo e complesso;
- una manipolazione poco sorvegliata. Il testo cui giunge il certificatore (TM) presenta disequilibri e 'salti logici': la coesione e la coerenza risultano, in parte, compromesse.

Item. Dalle indagini condotte emergono criticità diffuse nella costruzione degli *item*. Tali criticità sono riassunte nella Tabella 20 in corrispondenza dei rispettivi *item*.

Tabella 20
Criticità dei singoli item

Item	Criticità	Distrattori inefficaci
#1	Imprecisione nella formulazione dello <i>stem</i> .	d)
#2	Non sovrapposizione tra chiave e informazione presente nel testo: la chiave rappresenta una "informazione parziale" rispetto ai contenuti corrispondenti nel brano e inferibile in virtù di conoscenze specifiche.	d)
#3	Sovrapposizione tra chiave e distrattore (i.e. distrattore ambiguo).	b)
#4	Complessità del testo cui la chiave rimanda: lessico sovradimensionato per il livello <i>target</i> . Non sovrapposizione tra chiave e informazione presente nel testo.	
#5	Complessità di un distrattore: lessico sovradimensionato per il livello <i>target</i> .	d)
#6	-	d)
#7	-	c)

Numerosi sono i distrattori inefficaci. Carr, al riguardo, sentenzia: "Implausible distractors are a waste of space on the test" (2011, p. 92). Haladyna *et al.* (2002) sono dell'idea che questa situazione si presenti spesso nei *test a scelta multipla* con quattro opzioni e concludono che tre opzioni sono più che sufficienti nella maggiore parte dei casi.

Per quanto riguarda il dimensionamento della prova, segnaliamo, in particolare, la facilità degli *item* #1 e #7 e la difficoltà dell'*item* #4.

A uno sguardo conclusivo ci risulta opportuno, anziché un rimodellamento (i.e. intervento sul TM e riscrittura degli *item*), ripensare la prova da capo. Mettendoci nei panni del certificatore, adatteremmo un testo che presenti caratteristiche di livello B1, secondo quanto indicato dal CEFR, quali fattualità, semplicità, linearità, familiarità dell'argomento da parte del candidato. Insomma, pur considerando il triplo ordine di problemi – afferenti rispettivamente a TO, TM, *item* – ci pare poco proficuo ragionare su come 'riparare' le conseguenze degli ultimi due, se la scelta di base, abbiamo detto, non è adeguata.

Riferimenti bibliografici

- Alderson, Charles (2000). *Assessing reading*. Cambridge, Regno Unito: Cambridge University Press.
- Alderson, Charles (1990). Testing reading comprehension skills. *Reading in a Foreign Language*, 7(1), 465-503.
- Alderson, Charles, Clapham, Caroline, & Wall, Diane (1995). *Language test construction and evaluation*. Cambridge, Regno Unito: Cambridge University Press.
- Alderson, J. Charles, & Cseresnyés, Mária (2002) Appendice Guidelines for Writers of Reading and Use of English Tests del volume *Into Europe Project*, British Council, Hungary, 2002. Consultato su: http://www.lancs.ac.uk/fass/projects/examreform/Media/GL_Reding&Use.pdf
- Bachman, Lyle F. (2004). *Statistical analyses for language assessment*. Cambridge, Regno Unito: Cambridge University Press.
- Bachman, Lyle F. (1990). *Fundamental considerations in language testing*. Oxford, Regno Unito: Oxford University Press.
- Balboni, Paolo E. (1999). *Tecniche didattiche per l'educazione linguistica*. Torino, Italia: Utet.
- Barni, Monica, Bandini, Anna, Sprugnoli, Laura, Lucarelli, Silvia, Scaglioso, Anna Maria, Strambi, Beatrice, Fusi, Chiara, & Arruffoli, Anna Maria (2009). *Linee guida CILS. Certificazione di italiano come LS Università per stranieri di Siena*. Perugia, Italia: Guerra.
- Bejar, Isaac I. (2008). Standard setting: What is it? Why is it important? *R&D Connections*, 7, 1-6. Consultato su: http://www.ets.org/Media/Research/pdf/RD_Connections7.pdf
- Bejar, Isaac I., Braun, Henry I., & Tannenbaum, Richard (2007). A prospective, predictive and progressive approach to standard setting. In Robert W. Lissitz (Ed.). *Assessing and modeling cognitive development in school: intellectual growth and standard setting* (pp. 1-30). Maple Grove, Minnesota: Jam Press.
- Bond, Trevor G., & Fox, Christine M. (2007). *Applying the Rasch model. fundamental measurement in the human sciences*. Mahwah, New Jersey: Lawrence Erlbaum Assoc.
- Buck, Gary (2001). *Assessing listening*. Cambridge, Regno Unito: Cambridge University Press.
- Cangelosi, James S. (1990). *Designing tests for evaluating student achievement*. White Plains, NY: Longman.
- Carr, Nathan T. (2011). *Designing and analyzing language tests*. Oxford, Regno Unito: Oxford University Press.
- Centro CILS (2009). *Prova d'esame. Giugno 2009 Livello UNO - B1*. Università per Stranieri di Siena, Siena, Italia. Consultato su: <http://www.borg2700.at/unterricht/CILS/downloads/CILS%20B1%20quaderno%20del%20candidato%202009.06.pdf>
- Cizek, J. Gregory, & Bunch, Michael B. (2007). *Standard setting: a guide to establishing and evaluating performance standards on tests*. Thousand Oaks, California: Sage Publications.
- Council of Europe, Language Policy Division (2009). *Relating language examination to the common European framework of reference for languages: learning, teaching, assessment*. Consultato su http://www.coe.int/t/dg4/linguistic/source/manualrevision-proofread-final_en.pdf
- Davidson, Fred, & Lynch, Brian (2001). *Testcraft: A teacher's guide to writing and using language test specification*. New Haven, Connecticut: Yale University Press.
- Council of Europe (2001). *The common European framework of reference for languages: learning, teaching, assessment*. Cambridge: CUP (edizione italiana: *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*. Firenze: La Nuova Italia).
- Derosas, Manuela, & Torresan, Paolo (2009). Circa l'uso dei testi autentici. A colloquio con Christopher Humphris. *Bollettino Itals*, 7(29)
Consultato su: www.itals.it
- D'Este, Claudia (2013). New views of validity in language testing. *EL.LE. Educazione Linguistica*, 1(1), 49-63.
- Freunberger, Roman, Bazinger, Claudia, & Itzlinger-Bruneft, Ursula (2014). *Linking the Austrian standards-based test for English to the CEFR: the standard-setting process*. Salzburg, Austria: BIFIE.

- Geers, Andrew L., Weiland, Paul E., Kosbab, Kristin, Landry, Sarah J., & Helfer, Suzanne G. (2005). Goal activation, expectations, and the placebo effect. *Journal of Personality and Social Psychology*, 89(2), 143-159.
- Grabowski, Kirby C., & Dakin, Joe, W. (2014). Test development literacy. In Antony J. Kunnan (Ed.), *The companion to language assessment*. Hoboken, New Jersey: John Wiley and Sons, 751-768.
- Green, Rita (2013). *Statistical analyses for language teachers*. Basingtoke, Regno Unito: Palgrave MacMillan.
- Green, Rita, & Spoetl, Carol (2011, May). Building up a pool of standard setting judges: problems. solutions and insights. Eight European Association of Language Testing and Assessment Conference. Consultato su: http://www.ealta.eu.org/conference/2011/friday/EALTA2011_Green_Spoetl.pdf.
- Haladyna, Thomas M., Downing, Steven M., & Rodriguez, Michael C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Hambleton, Ronald K., & Pitoniak, Mary (2006). Setting performance standards. In Robert L. Brennan (Ed.). *Educational measurement* (pp. 433-470). Westport, Connecticut: American Council on Education & Praeger Publishers.
- Harding, Luke (2011a). *Specific guidelines for constructing test item/tasks*, in Id. *Modulo Master in Applied Linguistics*. Lancaster, Regno Unito: Lancaster University.
- Harding, Luke (2011b). *Guidelines for preparing listening tests*. In Id. *Modulo Master in Applied Linguistics*. Lancaster, Regno Unito: Lancaster University.
- Heaton, John B. (1988). *Writing English language tests*. London: Longman.
- Henning, Grant (1987). *A Guide to language testing: Developing, evaluation and research*. Cambridge, Massachussets: Newbury House.
- Hughes, Arthur (2003). *Testing for language teachers*. Cambridge, Regno Unito: Cambridge University Press.
- Hulstijn, Jan H. (2011). Language proficiency in native and non-native speakers: an agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8(3), 229-249.
- Kirschner, Michal, Wexler, Carol, & Spector-Cohen, Elana (1992). *Avoiding obstacles to student comprehension of test questions*. *TESOL Quarterly*, 26, 537-556.
- Linacre, John Michael (2012a). *Winsteps tutorial*. (Presentazione powerpoint). Consultato su: <http://www.winsteps.com/a/winsteps-tutorial-2.pdf>
- Linacre, John Michael (2012b). *Practical Rasch measurement*. Consultato su: www.winsteps.com/tutorial.
- Linacre, John Michael (2000). Redundant items, overfit and measure bias. *Rasch Measurement Transactions* 14(3), 755.
- Linacre, John M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of Outcome Measurement* 2(3), 266-283.
- Linacre, John M. (1995, April). *The effect of misfit on measurement*. Paper presented at the Eighth Annual Meeting of the International Objective Measurement Workshop, Berkeley, CA.
- Mallinson, Trudy (2001, October). *Measuring the impact of fatigue on everyday activities during chemotherapy*. Paper presented at the International Conference on Objective Measurement. University of Illinois, October 19-20, 2001, Chicago, IL. (Presentazione powerpoint).
- McNamara, Tim (2000). *Language testing*. Oxford, Regno Unito: Oxford University Press.
- McNamara, Tim (1996). *Measuring second language performance*. London, Regno Unito: Longman.
- Näsström, Gunilla, & Nyström, Peter (2008). A comparison of two different methods for setting performance standards for a test with constructed-response items. *Practical Assessment Research & Evaluation*, 13(9), 1-12. Consultato su: <http://pareonline.net/getvn.asp?v=13&n=9>
- Novello, Alberta (2015). Testare l'abilità di ascolto. *Officina.it*, 24. Consultato su: www.almaedizioni.it
- Novello, Alberta (2014). *La valutazione delle lingue straniere e secondo nella scuola. Dalla teoria alla pratica*. Venezia, Italia: Edizioni Ca' Foscari.
- Novello, Alberta (2009). *Le certificazioni della competenza linguistica*. Università Ca' Foscari, Dipartimento di Scienze del Linguaggio. Tesi di dottorato: Venezia, Italia.
- Pierce, B. Norton (1992). Demystifying the TOEFL reading test. *TESOL Quarterly*, 26(4), 665-689.

- Popham, W. James (2000). *Modern educational measurement*. Boston, Massachusetts: Allyn & Bacon.
- Powers, Donald E., & Wilson, Susan T. (1993). Passage dependence on the new SAT reading comprehension questions. *College Board Report*, 93(3), 1-18. Consultato su: <http://research.collegeboard.org/>
- Purpura James E. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.
- Rupp, André A., Ferne, Tracy, & Choi, Hyeran (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441-471.
- Schedl, Mary, & Malloy, Jeanne (2014). *Writing items and tasks*. In Antony J. Kunnan (Ed.). *The companion to language assessment* (pp. 1-18). Hoboken, New Jersey: John Wiley and Sons.
- Spinelli, Barbara (2000). L'utilizzo dei materiali autentici nell'insegnamento dell'italiano LS. In Roberto Dolci & Paola Celentin (Eds.). *La formazione di base dell'insegnante di italiano a stranieri* (pp. 133-147). Roma, Italia: Bonacci.
- Taglietti, Cristina (2007, 23 aprile). 'E adesso i libri ce li scriviamo da soli'. La nuova generazione di autori adolescenti. *Corriere della Sera*. p. 29. Consultato su:
http://archiviostorico.corriere.it/2007/aprile/23/adesso_libri_scriviamo_soli_nuova_co_9_070423056.shtml e
http://www.fantasiofestival.it/fantasio_2007/documenti/rassegna/Corriere_della_sera_23_04_2007.pdf
- Thorndike, Robert L., & Hagen, Tracy (1969). *Measurement and evaluation in psychology and education*. Hoboken, New Jersey: John Wiley and Sons.
- Torresan, Paolo (2014). Il test a individuazione di informazioni della certificazione di italiano per stranieri CILS: aspetti critici. *Rassegna italiana di valutazione*, 18(59), 104-123.
- Torresan, Paolo (2015). Insidie nella confezione di un sequencing task quale test di lettura: uno studio a partire da una prova della certificazione CILS. *EL.LE Educazione Linguistica* [di prossima pubblicazione].
- Urbani, Stefano (1989). Testo, autenticità, grammatica. In Christopher Humphris (Ed.), *Atti del I seminario internazionale per insegnanti di lingua* (pp. 9-26). Roma, Italia: Dilit.
- Weir, Cyril (1993). *Understanding and developing language tests*. New York, New York: Prentice Hall.
- Wright, Benjamin D. (2001). Separation, reliability and skewed distributions: statistically different levels of performance. *Rasch Measurement Transactions*, 14(4), 786.
- Wilkins, David (1976). *Notional syllabuses*. Oxford, Regno Unito: Oxford University Press.
- Wilkins, David (1972). *Linguistics and language teaching*. London, Regno Unito: Arnold.
- Zieky, Michael J. Perie, Marianne, & Livingstone, Samuel A. (2008). *Cut-scores: A manual for setting standards on educational and occupational tests*. Princeton, New Jersey: ETS.
 Consultato su: https://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf

Appendice A

Indicazioni per la confezione di un quesito a scelta multipla ai fini della valutazione delle abilità ricettive

Riportiamo le caratteristiche relative al quesito a scelta multipla e una serie di indicazioni pratiche relative alla confezione.

Tabella 21

Quesito a scelta multipla: vantaggi, limiti, attenzioni richieste

		Commenti
Vantaggi	rapidità di correzione	La tabulazione dei dati e la correzione dei test sono rapide da eseguirsi (Balboni, 1999, p. 181).
	flessibilità	L'esercizio si adatta a diverse strategie di comprensione (<i>for a gist, detail-mapping, ecc.</i>).
	duttilità	L'esercizio è confezionabile in diversi formati: esercizi elettronici (per esempio attraverso il software <i>Hotpotatoes</i>), a video (attraverso <i>Blubbr</i> ; <i>Eduuzzle</i> ; <i>Educanon</i>), ecc.
	non è richiesta la scrittura	A differenza delle domande aperte o delle attività di completamento, una scelta multipla non prevede attività di scrittura (quindi si limitano questioni che possono insorgere, in tal senso, in termini di <i>construct-irrelevant variance</i> , vale a dire di interposizione di fattori esterni al costrutto che ci si propone di misurare).
Limiti	difficoltà di costruzione	I distrattori devono essere, idealmente, di pari plausibilità: un distrattore che non viene scelto da nessuno è inutile (Alderson <i>et al.</i> , 1995).
	abbinamento a certi tipi di testo	La scelta multipla si accompagna a testi densi, dettagliati, non a testi lineari e/o che si riducono a una lista di informazioni (orientativamente, quindi, vale per testi dal livello B2 in su). Nei livelli più alti un quesito a scelta multipla può prestarsi a rilevare la comprensione di contenuti impliciti, che il lettore può cogliere attraverso processi inferenziali (Weir, 1993). Va mantenuta però un'attenzione particolare, in questo caso, all'appropriatezza dell' <i>item</i> , specie nelle prove di ascolto, dato che "it is difficult to say with certainty that any inference is completely wrong. After all, we are free to infer what we want" (Buck, 2001, p. 148).
	effetto ombra	Come per tutte le prove oggettive (compreso il vero/falso), gli <i>item</i> non consentono di indagare i processi che hanno guidato il <i>test taker</i> nella risoluzione del compito (Alderson, 1990). Si tratta di una caratteristica inversamente proporzionale alla rapidità di correzione; scrive Alderson (2000, p. 212): "If candidates were required to give their reasons for making their choice [...], the problem [i.e. l'effetto ombra] would be mitigated, but then the practical advantage of multiple-choice questions in terms of marking would be vitiated".
Attenzioni richieste	linearità	Occorre far in modo che la progressione degli <i>item</i> rispetti la progressione del testo.
	logica	Si evitino opzioni che si autoescludono (chiave vs distrattore, o distrattore vs distrattore).
	attinenza al costrutto	Gli item devono misurare l'abilità come descritta nelle <i>test specifications</i> . Nei test di ascolto, bisogna evitare che gli <i>item</i> misurino la memoria anziché la comprensione del testo. Più in generale, se la comprensione è intesa come un processo interpretativo e non di puro riconoscimento, si evitino chiavi che ricalcano tale e quale il contenuto corrispondente nel testo di comprensione; si procederà piuttosto attraverso una parafrasi del contenuto.

immediatezza	Evitare di scrivere <i>item</i> più difficili da comprendere del testo stesso (Thorndike & Hagen, 1969; Kirschner <i>et al.</i> , 1992). Occorre evitare l'uso di negazioni (del tipo: "scegli l'opzione non corretta", o formulazioni di <i>item</i> o <i>stem</i> al negativo), che costringono a contorsioni logiche (Haladyna <i>et al.</i> , 2002). Occorre ricordarsi che si sta misurando la capacità dell'allievo di comprendere un testo, non quella di comprendere l' <i>item</i> .
brevità	È meglio se <i>stem</i> e opzioni siano succinti, specie per i livelli bassi (A1-B1).
indipendenza	È bene evitare <i>item</i> la cui risposta dipende da quella a <i>item</i> precedenti. L'interdipendenza degli <i>item</i> sortisce un effetto negativo sull'affidabilità della prova (Hughes, 2003).
esclusività	Due <i>item</i> non devono prevedere una risposta simile (Carr, 2011): in caso di risposta scorretta, uno studente verrebbe penalizzato due volte.
precisione	La chiave deve essere precisa (totalmente sovrapponibile, e non solo in parte, all'informazione del testo cui fa riferimento) e unica (non ci devono essere due chiavi). Occorre, in questo senso, controllare che a un distrattore non corrispondano informazioni presenti in una porzione del testo a cui non si aveva pensato. Allo stesso modo, occorre che lo <i>stem</i> , nel modo in cui è formulato, non anticipi la chiave.
riferimento al testo	Vanno evitati <i>item</i> a cui si può rispondere correttamente con la sola logica o in virtù di conoscenze previe (senza far riferimento al testo oggetto di comprensione; i.e. <i>background knowledge</i> ; cfr. Powers & Wilson, 1993). Allo stesso modo non sono ammissibili <i>item</i> che facciano capo a informazioni non presenti nel testo (Schedl & Malloy, 2014). E infine occorre essere molto cauti nella scrittura di distrattori che sono falsi, stando al testo, ma veri nella realtà delle cose (Cangelosi, 1990).
uniformità	Occorre gli <i>item</i> siano di pari lunghezza. "No options should 'stick out' from the others. Item options should look like a coherent set" (Grabowski & Dakin, 2014). Si consideri, a tal proposito, che gli studenti tendono a scegliere l'opzione più lunga, la quale spesso, nel caso di <i>item writer</i> poco esperti, coincide con la chiave (cfr. Heaton, 1988).
anticipazione	Vanno anticipate nello <i>stem</i> le informazioni comuni alle opzioni.
cautela nell'uso delle immagini	L'uso delle immagini, benché intuitivamente (in quanto a codice più universale della lingua) si potrebbe supporre faciliti il lavoro dello studente, va limitato il più possibile nei test di comprensione. Le immagini possono essere percettivamente poco chiare (specie quelle pensate per la didattica e la valutazione), avere margini di ambiguità, presentare <i>bias</i> culturali, sociali, di genere, di religione, ecc., non essere percepite come utili dagli studenti né risultare un codice familiare. La regola d'oro quindi è: "Do not include visual materials unless it makes real contribution. Make sure that it is clear and informative, and does not cause more problems than it is worth" (Harding, 2011a).
layout	I distrattori vanno disposti verticalmente, non in blocchi orizzontali, al fine di facilitare il lettore (Haladyna <i>et al.</i> , 2002; Purpura, 2004).
calibrazione	<i>Item</i> poco calibrati non consentono di misurare ciò che ci si propone di misurare; di conseguenza compromettono la validità della prova. Occorre, a tal proposito, considerare che più sono sottili le differenze tra le opzioni, più l' <i>item</i> si configura come arduo (Rupp <i>et al.</i> , 2006).
familiarità	Come ogni strumento di indagine, il <i>quesito a scelta multipla</i> deve risultare familiare ai candidati. In caso contrario, la non familiarità con il <i>format</i> può incidere sul risultato in quanto a <i>construct-irrelevant variance</i> .

	positività del punteggio e uniformità dei valori	Si evitino punteggi negativi in corrispondenza di risposte errate, al fine di non compromettere la validità e l'affidabilità della prova (Torresan 2015a). Si assegni ad ogni <i>item</i> lo stesso punteggio (GWRT).
	validazione esterna	È bene presentare il test a un collega o un nativo esperto che non ha presieduto alla confezione e considerare il suo <i>feedback</i> (Davidson & Lynch, 2001; Hulstijn, 2011).
	dispersione	Specie nei test di ascolto, è bene evitare <i>item</i> relativi a porzioni di testo troppo ravvicinate. A uno studente può sfuggire la risposta, perché impegnato nello scegliere l'opzione corretta dell' <i>item</i> precedente (Novello, 2015). Allo stesso modo, sempre nei test di ascolto, va evitato che ci siano pause troppo lunghe, tra un <i>item</i> e l'altro: lo studente può smarrirsi e pensare che la risposta all' <i>item</i> sia stata fornita durante quella pausa (cfr. Harding, 2011b).
	fattualità	Specie in riferimento a test di ascolto, "prestare attenzione a non testare le emozioni (in quanto si interpretano in modo diverso anche a seconda della cultura)" (Novello, 2015). Fanno eco le <i>Guidelines for Writers of Listening Test</i> (GWLT) allestite dal <i>British Council</i> (Alderson & Cseresznyés, 2002), le quali invitano a sollecitare "factual information, and not [...] subjective judgment".
	disponibilità di tempo	Si calcolino esattamente i tempi, onde evitare che lo studente, per la fretta, sia portato a tirare a indovinare. Per l'ascolto, le GWLT raccomandano il rispetto dei seguenti tempi: - prima della prova: "Allow enough time for students to familiarise themselves with items before listening to the text. Intermediate: 4 seconds per item; advanced: 3 seconds per item". - dopo la prova: "Allow enough time for students to finalise their answer after the text is finished: 3 seconds per item".

Appendice B

Risultati della sessione di attribuzione di livello

Riportiamo l'esito finale della sessione di attribuzione di livello condotta con 36 formandi esaminatori, in merito degli *item* della prova oggetto di studio.

Tabella 22
Giudizi dei rater

item

formandi esaminatori

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	FV
1	B	B	A	B	B	C	B	B	B	B	B	B	C	A	B	B	B	B	B	B	B	A	B	C	B	B	C	C	B	A	A	B	B	A	A	A	92,2
2	B	A	B	C	B	B	B	B	C	C	C	B	B	C	C	B	B	B	A	C	C	B	C	C	B	C	C	C	C	B	C	C	C	B	B	C	81,6
3	C	C	B	B	C	C	C	C	B	B	B	C	B	B	C	C	C	B	C	B	B	B	C	C	B	C	B	C	C	C	B	A	C	C	B	B	57
4	B	A	A	A	B	B	B	B	C	C	A	B	A	A	B	B	B	A	A	B	B	B	C	B	B	B	C	B	B	A	C	A	B	A	A	B	31,3
5	A	A	A	A	A	B	A	A	A	A	A	A	A	A	B	B	B	B	A	A	A	A	A	B	B	A	B	A	A	A	B	B	A	A	A	A	64,2
6	B	B	B	B	A	A	B	B	B	B	B	B	B	B	B	C	C	B	B	B	A	B	B	B	B	A	C	B	B	B	B	B	B	B	B	B	81,6
7	A	A	B	A	A	B	B	B	B	B	B	B	A	A	A	C	B	B	A	A	B	B	A	B	B	B	B	C	B	A	A	A	B	A	A	A	92,2

Legenda. A= A2; B= B1; C= B2.

Paolo Torresan, Alma Edizioni
 piroclastico@tin.it; esh@unive.it

- IT** | **Paolo Torresan** attualmente insegna Didattica delle lingue moderne e Didattica dell'italiano a stranieri presso l'Università di Catania (Struttura Didattica Speciale di Lingue e Letterature Straniere, sede di Ragusa). Ha svolto attività di ricerca presso l'Università Ca' Foscari di Venezia, le università Complutense e Autónoma di Madrid, la Lancaster University. È stato Fulbright Visiting Scholar presso il Santa Monica College (California) e Visiting Professor presso l'Università dello Stato di Rio de Janeiro. È autore e formatore per ALMA Edizioni. Cura la redazione delle riviste *Officina.it* e *Bollettino Itals*. Ha scritto: *Intelligenze e didattica delle lingue* (EMI, Bologna 2008); *Didáctica de las lenguas culturas. Nuevas perspectivas* (con Manuela Derosas, a cura di; Sb International, Buenos Aires, 2011); *Il noticing comparativo* (con Francesco Della Valle; Lincom, München, 2013), *Studio su cloze mirati della certificazione CILS, livello avanzato* (Dialogarts, Rio de Janeiro, 2015).
- EN** | **Paolo Torresan** teaches courses in Modern Language Pedagogy and Teaching Italian to Speakers of Other Languages at the University of Catania (Special School of Foreign Languages and Literatures based in Ragusa). He has conducted research at the University Ca' Foscari in Venice, Complutense University and Autónoma University in Madrid, and at Lancaster University in England. He was Fulbright Visiting Scholar at Santa Monica College, California, USA and Visiting Professor at Rio De Janeiro State University, Brazil. He is author and trainer for the Italian publishing company ALMA Edizioni and editor of the journals *Officina.it* and *Bollettino ITALS*. He co/authored: *Intelligenze e didattica delle lingue* (Bologna: EMI, 2008); *Il noticing comparativo* (with Francesco Della Valle. München: Lincom, 2013), and *Studio su cloze mirati della certificazione CILS, livello avanzato* (Rio de Janeiro: Dialogarts, 2015). He also edited with Manuela Derosas, *Didáctica de las lenguas culturas. Nuevas perspectivas* (Buenos Aires: Sb International, 2011).
- ES** | **Paolo Torresan** enseña actualmente Didáctica de las lenguas modernas y Didáctica del italiano como lengua extranjera en la Universidad de Catania (Centro educativo especializado de Lenguas y Literaturas extranjeras, sede de Ragusa). Ha llevado a cabo labores de investigación en la Universidad Ca' Foscari de Venecia, las universidades Complutense y Autónoma de Madrid y Lancaster University. Ha sido becario Fulbright en Santa Mónica College (California) y profesor visitante en la Universidad Estatal de Rio de Janeiro. Es autor y formador para ALMA Edizioni. Supervisa la redacción de las revistas *Officina.it* y *Bolletino Itals*. Ha escrito *Intelligenze e didattica delle lingue* (EMI, Bolonia 2008); *Didáctica de las lenguas culturas. Nuevas perspectivas* (con Manuela Derosas, editado por Sb International, Buenos Aires, 2011); *Il noticing comparativo* (con Francesco Della Valle; Lincom, München, 2013), *Studio su cloze mirati della certificazione CILS, livello avanzato* (Dialogarts, Rio de Janeiro, 2015).