

## ANALISI DI PROVE DI ASCOLTO IN ESAMI DI CERTIFICAZIONE ITALIANI (LIVELLO B2)

Italia weist bezüglich Sprachzertifizierung eine Besonderheit auf: Gleich vier Zertifikatsprüfungen sind vom italienischen Bildungsministerium anerkannt: CELI, CILS, PLIDA und .IT. Diese Vielfalt bietet Gelegenheit, die Prüfungen zu vergleichen und die einzelnen Items auf ihre Validität hin zu überprüfen.

161 Maturanden aus Schweizer und Südtiroler Gymnasien haben die vier Hörverständnisübungen auf dem Niveau B2 durchgeführt und zudem die Prüfungen bezüglich ihrer Attraktivität bewertet. Die folgende, eigens für Babylonia durchgeführte Untersuchung legt die Resultate der vier Prüfungen dar und zeigt die Schwachstellen der jeweiligen Prüfungen auf. Die Lektüre dieses Beitrags empfiehlt sich besonders für jeden, der selbst Multiple-Choice-Prüfungen durchführt und um angemessene und unmissverständliche Fragestellungen und Distraktoren ringt.

Die Prüfungen sind auf der Babylonia-Homepage abrufbar:

[babylonia.ch](http://babylonia.ch) > Nummer 2/2016 > Artikel Torresan

### ● Paolo Torresan | Venezia

Paolo Torresan ha svolto attività di insegnamento presso l'Università dello Stato di Rio de Janeiro (*Visiting Professor*), il Santa Monica College (*Fulbright Visiting Scholar*) e l'Università di Catania (sede di Ragusa). Ha svolto attività di ricerca presso l'Università Ca' Foscari di Venezia e la Lancaster University.



#### 1. Il sistema plurale delle certificazioni italiane<sup>1</sup>

I *proficiency test* nell'ambito dell'italiano come lingua straniera riconosciuti dal Ministero degli Affari Esteri italiano sono rappresentati dalle 4 certificazioni ufficiali:

- > CELI, Università per Stranieri di Perugia
- > CILS, Università per Stranieri di Siena
- > PLIDA, Società Dante Alighieri
- > .IT, Università degli Studi Roma Tre

Altre certificazioni, come gli esami AIL dell'Accademia di Firenze<sup>2</sup>, o gli esami CLIDA<sup>3</sup> del Centro Linguistico Italiano Dante Alighieri (con sedi a Firenze e Roma), pur essendo implementate in alcuni contesti (es. in Svizzera nel primo caso; in vari paesi, nel secondo), ricadono fuori dal riconoscimento ministeriale.

Dinanzi a questa pluralità di proposte, domande ricorrenti sono:

- > Perché quattro certificazioni e non una, per una lingua, qual è l'italiano, che non riscuote un interesse internazionale al pari dell'inglese?
- > Esistono dei confronti tra una certificazione e l'altra?

Alla prima domanda non siamo in grado di rispondere, visto che comporta uno sguardo di tipo storico e sociologico esterno alle nostre competenze. Alla seconda rispondiamo che non esistono studi comparativi in termini di risultati, quanto descrizioni, alcune delle quali molto raffinate (Novello, 2014). C'è da riconoscere, ad ogni modo, che una comparazione in termini di risultati tra le 4 certificazioni non è praticabile in sé e di per sé. Si dovrebbe supporre che uno stesso campione si sottoponga alle varie prove; i tempi di gestione di una impresa del genere sono enormi (almeno un giorno per certificazione); inoltre esiste un 'effetto-accumulo', tale per cui, passando dalle prove della prima certificazione alle prove della quarta, il candidato acquisirebbe maggiore competenza (o perlomeno guadagnerebbe in *expertise* nello svolgimento di esercizi simili) e tale fattore inciderebbe a favore della certificazione somministrata per ultima.

In luogo di una comparazione 'in blocco', esiste però la possibilità di un'analisi parallela di singole prove omologhe. Tale indagine ci risulta non essere mai sta-

PIÙ ARTICOLI SU QUESTO TEMA:  
[WWW.BABYLONIA.CH](http://WWW.BABYLONIA.CH) >  
 ARCHIVIO TEMATICO > [SCHEDA 6](#)

ta eseguita nell'ambito dell'italiano ed è l'oggetto del nostro saggio.

## 2. Item Analysis parallelo: presentazione della ricerca

In seno a tre corsi di formazione, svoltisi in tempi assai ravvicinati (dicembre 2015-gennaio 2016), tra la provincia di Bolzano (corso promosso dall'Ispettorato delle scuole di lingua tedesca) e la Svizzera (corsi tenuti rispettivamente presso l'Università di Friburgo e l'Alta Scuola Pedagogica di Zurigo), abbiamo raccolto, grazie alla disponibilità di colleghi, le risposte di un campione omogeneo (studenti maturandi) a prove omolo-

ghe (ascolto, livello B2, quesiti a scelta multipla)<sup>4</sup>, tratte dalle 4 certificazioni di italiano, liberamente consultabili *online*<sup>5</sup>. Il lettore può scaricare le prove dai siti riportati nella sitografia. Ai candidati è stato inoltre somministrato un questionario attraverso il quale, oltre a segnare le preferenze per le singole prove attraverso una scala Likert, ciascuno era libero di esprimere i propri giudizi in merito alla sperimentazione (§ 2.4).

Nella tab. 1 rappresentiamo le caratteristiche delle prove.

Tab. 1. Caratteristiche generali delle prove

	CELI	CILS	PLIDA	.IT
Sessione	estiva 2007	estiva 2012	non dich.	non dich.
Minutaggio soundfile, senza istruzioni	9'12"-12'46"	10'56"-14'09"	28'03"-31'42"	1'05"-3'45"
Durata	3'34"	2'13"	3'45"	2'50"
Numero di ascolti	2	2	2	2
Tipologia	intervista	intervista	monologo	intervista
Dominio	sport	un'offerta formativa	storia dell'urbanistica	un'attività commerciale
Numero item	5	7	5	5
Introduzione al tema	✓			
Item-esempio	✓			✓
Layout chiaro <sup>6</sup>	✓	✓		
Testo autentico <sup>7</sup>	✓		✓	✓
Citazione della fonte			✓	✓
Ortoepia naturale	✓	✓	✓	✓
Testo interessante	✓		✓	✓
Testo chiaro	✓		✓	✓
Punteggio solo positivo		✓		non dich.
Trascrizione		✓		
Soluzioni	✓	✓	✓	✓

1 Ringraziamo Erika Saccuti e Maria Simona Morosin per la lettura e il confronto. Ringraziamo i colleghi che ci hanno aiutato nella raccolta dei dati.

2 Gli esami ALL, nell'ambito della valutazione delle abilità di comprensione, constano soprattutto di prove di *vero/falso*, senza peraltro provvedere ad alcuna misura che consenta di arginare l'arbitrarietà che il *vero/falso* comporta. Si tratta, quindi, di un esame che presenta margini di criticità sin dalla scelta dei *format*.

3 Gli esami CLIDA, nell'ambito della valutazione delle abilità di comprensione, presentano soprattutto *domande aperte*, con numerose ingenuità nella confezione degli *item*, come l'*interdipendenza*, tale per cui si può rispondere a una domanda leggendo una successiva (es. prova CLIDA P3 2014: domanda 1 – "Dove è stato il Campionato Mondiale di Ciclismo 2013?"; Domanda 4 – "Quanti «fan» sono stati presenti lungo le strade tra Firenze e Fiesole?") o domande a cui è facile rispondere grazie a preconcoscienze (es. prova CLIDA P3 2011: domanda 3 – "Dov'è nato il melodramma?").

4 Una leggera differenza è data dal fatto che la prova CILS consta di 7 *item* anziché 5, come le altre; ma ciò non crea difficoltà in termini di comparazione.

5 La prova tratta dal PLIDA corrisponde al vecchio formato; il centro certificatore sta rivedendo l'impianto della certificazione; per ora sono uscite le specifiche del nuovo formato del B1 (Cardillo, Vecchio & Menzinger, 2015), nel corso del 2016 usciranno quelle relative al B2 e al C1.

6 Si tratta di un giudizio soggettivo. Nel caso del PLIDA, la distribuzione della prova su due fogli (fronte/retro) rende poco agevole, secondo noi, l'esecuzione. Nel caso di .IT l'*item*-esempio è stato inteso, da una buona parte del campione, come facente parte della prova.

7 Anche questo è un nostro giudizio: il testo pare essere studiato a tavolino.

Nella tab. 2 riportiamo il tempo a disposizione per leggere i quesiti secondo le indicazioni fornite nei file.

Le *research question* che ci siamo posti sono state:

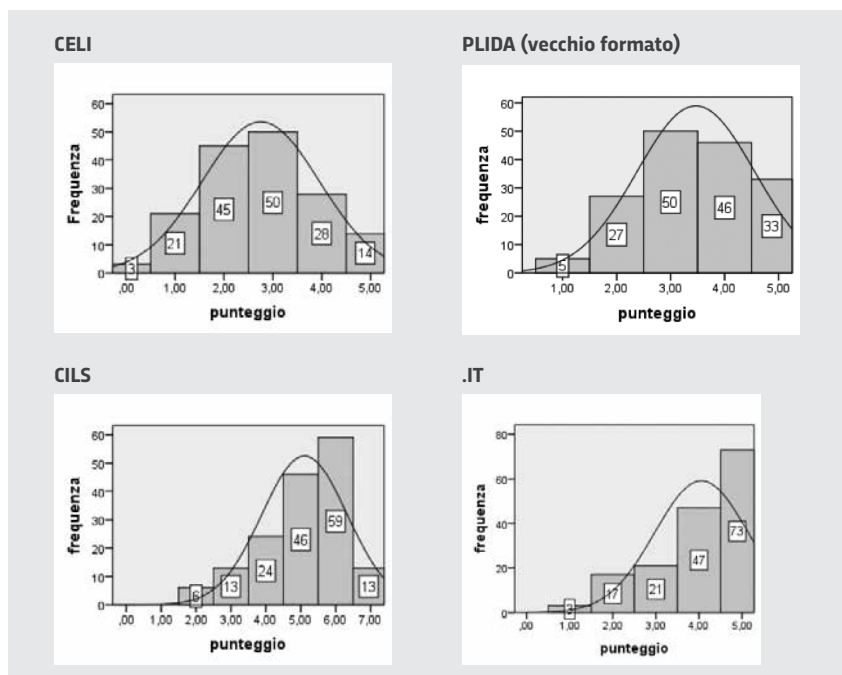
- > I risultati del campione variano *significativamente* da prova a prova?
- > Esistono casi evidenti di sotto- e/o sovracalibrazione?
- > Esistono casi di cattiva scrittura degli item?
- > Qual è il *feedback* dei candidati?

Tab. 2. Tempi di lettura dichiarati nei soundfile prima, tra e dopo gli ascolti

Consideriamo ciascuna nei paragrafi a seguire.

	prima del primo ascolto	tra il primo e il secondo ascolto	dopo il secondo ascolto	tot.
<b>CELI</b>	20" (60" per tutte e tre le prove di ascolto)	25"	35"	1'20"
<b>CILS</b>	-	60"	120"	3'
<b>PLIDA</b>	60"	-	60"	2'
<b>.IT</b>	indicazioni non presenti nell'audio			

Fig. 1. Frequenza dei punteggi



## 2.1. Variazione tra le prove

Da un'analisi degli istogrammi alla fig. 1. si evince che i risultati variano significativamente da prova a prova.

Agli estremi abbiamo due prove:

- > .IT, la più facile, con un profilo fortemente disallineato a destra
- > CELI, con un profilo compatibile con una distribuzione normale.

La relativa facilità della prova .IT è evidente se si considera la percentuale degli studenti che hanno raggiunto l'eccellenza, e cioè che hanno risposto correttamente a 5 item su 5: 45,3%, cifra ben superiore rispetto a quelle relative a quanti hanno raggiunto l'eccellenza nelle altre prove (tab. 3).

La relativa difficoltà della prova CELI, come dicevamo, è immediatamente inferibile considerato il profilo della curva che presenta una distribuzione pressoché normale (con un leggerissimo disallineamento a destra). Si tratta di un profilo che non ci si attende con un campione rappresentativo come il nostro, cioè con studenti che si stanno preparando sistematicamente ad affrontare una prova di questo livello.

Il profilo della prova CILS è simile a quello della prova .IT, salvo presentare una percentuale molto bassa in riferimento all'eccellenza (solo 13 candidati hanno risposto a tutti gli item: 8,1%). Ciò si spiega con la forte sovracalibrazione di un item (#3), come avremo modo di vedere nel paragrafo successivo.

Nel complesso pare che alla prova PLIDA corrisponda il profilo più armonico e atteso.

## 2.2. La calibrazione degli item

Nella tab. 3 rappresentiamo le percentuali di studenti che hanno risposto correttamente ai singoli item (*Facility Value*, d'ora in poi *FV*).

A quali valori attenerci, sotto i quali o oltre i quali si danno, rispettivamente, casi di sotto- e di sovracalibrazione? Nella letteratura i pareri sono discordanti, con oscillazioni significative (per un prospetto: Green, 2013). A nostro giudizio, un ente certificatore, nella più flessibile delle ipo-

Tab. 3. Facility Value  
n=161

tesi, si può orientare attorno all'intervallo 35-90, e cioè ritenere che con un  $FV \leq 35$  vi sia *evidente* sovracalibrazione (vale a dire che se solo un candidato su 3 indovina la risposta, l'item è *evidentemente* difficile per il livello *target*), così come, al contrario, con un  $FV \geq 90$  vi è *evidente* sottocalibrazione (se oltre 9 studenti su 10 indovinan la chiave, l'item è *evidentemente* facile in riferimento al livello *target*). In termini di ricerca, sempre nell'ottica di un contesto certificatorio, suggeriamo valori più ristretti: 40-85; e cioè con item con

- >  $FV \leq 40$  abbiamo una sovracalibrazione che merita di essere indagata
- >  $FV \geq 85$  abbiamo una sottocalibrazione che, al pari, costringe a un'analisi accurata.

Alla domanda, dunque, se si diano *item* sotto- o sovracalibrati, la risposta è affermativa. Abbiamo, in effetti:

- > due *item* sovracalibrati nella prova CELI
- > due *item* orientati in maniera opposta nella prova CILS, uno sotto- e uno sovracalibrato
- > un *item* sottocalibrato nella prova PLIDA
- > tre *item* sottocalibrati nella prova .IT

In alcuni casi, come avremo modo di appurare, i casi di sotto- e sovradimensionamento si spiegano per via di problemi legati alla confezione degli item stessi.

### 2.3. La qualità degli item

La tab. 4 riprende la tab. 3: riportiamo ancora una volta gli FV dei singoli item. In aggiunta, rappresentiamo con un doppio asterisco ed evidenziati in grassetto i casi di cattiva fattura degli item. Come anticipato, in alcuni frangenti, l'errato dimensionamento è attribuibile a una scrittura poco sorvegliata dei quesiti.

In sintesi, più casi richiederebbero una riformulazione dei quesiti e un conseguente riadattamento delle prove:

- > CELI, tre casi: *item* ## 1,2,4
- > CILS, cinque casi: *item* ## 1,2,3,6,7
- > PLIDA, un caso: *item* # 2
- > .IT, tre casi: *item* ## 1,2,3

	1	2	3	4	5	6	7
<b>CELI</b>	55,9	27,3	72,7	37,9	82,6		
<b>CILS</b>	75,8	76,4	37,9	71,4	83,9	93,2	72
<b>PLIDA</b>	88,2	54,7	60,9	73,9	70,2		
<b>.IT</b>	90,1	88,8	71,4	67,2	87,6		

Tab. 4. Item di cattiva fattura  
n=161

	1	2	3	4	5	6	7
<b>CELI</b>	<b>55,9**</b>	<b>27,3**</b>	72,7	<b>37,9**</b>	82,6		
<b>CILS</b>	<b>75,8**</b>	<b>76,4**</b>	<b>37,9**</b>	71,4	83,9	<b>93,2**</b>	<b>72**</b>
<b>PLIDA</b>	88,2	<b>54,7**</b>	60,9	73,9	70,2		
<b>.IT</b>	<b>90,1**</b>	<b>88,8**</b>	<b>71,4**</b>	67,2	87,6		

Tab. 5. Mappa della criticità.

	CELI	CILS	PLIDA	.IT
Distrattore parzialmente vero	<b>1,2,4</b>			
Assenza di una chiave		<b>3</b>		<b>3</b>
Ambiguità della chiave			<b>2</b>	
Ambiguità del testo di riferimento		<b>1</b>		
Risalto dalla chiave		<b>6</b>		
Dipendenza tra gli item		<b>2</b>		<b>2</b>
Inadeguata distribuzione tra gli item		<b>7</b>		<b>1</b>

Ancora una volta, la prova PLIDA pare essere la meno approssimativa, mentre la prova CILS comporta il numero maggiore di criticità.

Nella tab. 5 illustriamo i tipi di criticità. I numeri riportati fanno riferimento ai singoli item delle prove (es. il numero 3 riferito alla prova CILS rappresenta il terzo item dell'esercizio).

Illustriamo a seguire la natura delle criticità:

- > *Distrattore parzialmente vero*: si tratta di un distrattore che, per via inferenziale, può essere colto come possibile – non è cioè totalmente smentito dal testo.
- > *Assenza di una chiave*: nessuna tra le opzioni presentate dal certificatore è

perfettamente sovrapponibile alle informazioni veicolate dal testo; di conseguenza non vi è una risposta corretta.

> *Ambiguità della chiave*: la chiave è formulata in maniera tale da ammettere margini di interpretazione.

> *Ambiguità del testo di riferimento*: il testo presenta aspetti di incoerenza logica (contraddizioni) o tematica (annuncia tratterà un certo argomento, ma passa a trattare tutt'altro).

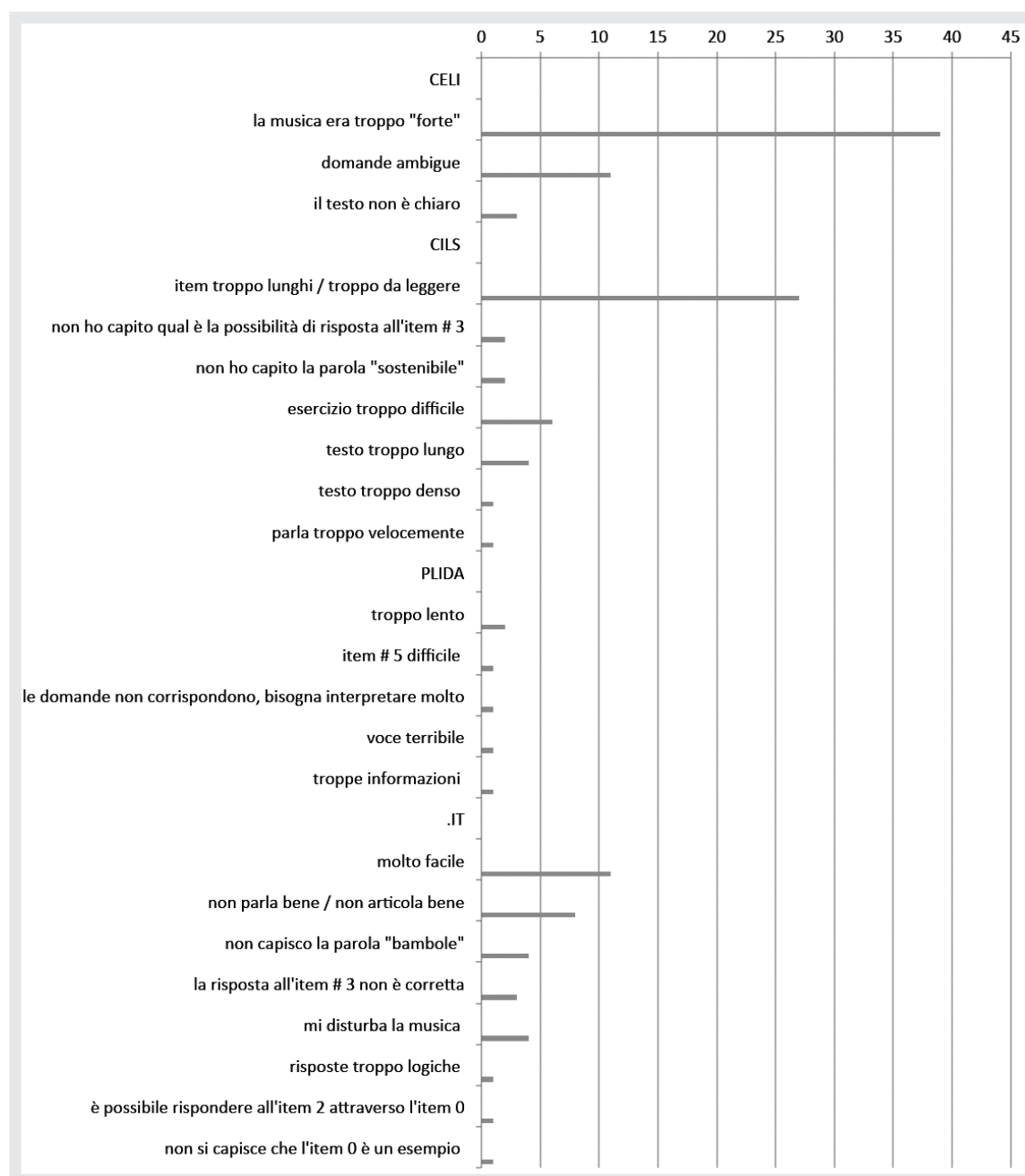
> *Risalto della chiave*: la chiave si distingue graficamente dai distrattori perché più lunga e/o più elaborata; è quindi

facile la sua individuazione da parte del candidato.

> *Dipendenza tra gli item*: si può rispondere a un item attraverso la presa visione di un item precedente o di uno successivo.

> *Inadeguata distribuzione degli item*: gli item rimandano a informazioni troppo ravvicinate tra loro, se non addirittura 'incastrate' l'una nell'altra; oppure l'ordine degli item non rispetta la progressione delle informazioni presenti nel testo, o infine gli item si riferiscono a informazioni trasmesse nei primi se-

Fig. 2. Osservazioni critiche di una parte del campione



condi del brano, sicché l'ascoltatore fa fatica a individuarle.

Per il lettore interessato a una disamina attenta, caso per caso, rimandiamo alla versione estesa online di questo nostro articolo, sul sito della rivista.

#### 2.4. Il *feedback* del campione

Mentre i dati ottenuti tramite la scala Likert circa il grado di preferenza non risultano significativi, data la dispersione delle risposte, i commenti critici alle prove (fig. 2) ci illuminano su aspetti sfuggiti all'indagine, oltre a ribadire questioni già trattate durante l'analisi degli item. In particolare, quali punti finora non trattati, rileviamo:

##### Per la prova CELI

- › Il disagio provocato dalla musica di sottofondo, che, oltre ad essere presente negli stacchi si sovrappone spesso al parlato; per alcuni, la musica catalizza l'attenzione ("se sento la musica non mi concentro più sulle parole")

##### Per la prova CILS

- › La generale complessità del testo (denso, lungo, intricato, con velocità di eloquio eccessiva)
- › La lunghezza degli item e la complessità, in particolare, di alcuni. Ciò costituisce un fattore di invalidità, data la presenza di una variabile esterna rispetto al costrutto (*construct-irrelevant variance*): la lettura, cioè, rispetto all'ascolto
- › L'incomprensione della parola-chiave "sostenibile"

##### Per la prova PLIDA:

- › La lentezza del monologo
- › La qualità della voce
- › La densità delle informazioni

##### Per la prova .IT:

- › La difficoltà a capire il parlato di una delle donne intervistate
- › L'incomprensione della parola-chiave "bambole"
- › La predittività delle risposte ("risposte troppo logiche"), anch'esso fattore di invalidità, in quanto a variabile esterna
- › Difetti di *layout* (non è chiaro che l'item 0 costituisca un esempio).

### 3. Conclusioni

Nel nostro studio abbiamo analizzato prove di ascolto di livello B2 (nella forma del quesito a scelta multipla), tratte dalle certificazioni di italiano LS. Sono emerse difficoltà di vario tipo, legate in particolare alla confezione degli item. Le differenze in merito sono abbastanza rilevanti; ciò significa che gli esiti che un candidato ipotetico avrebbe potuto raggiungere – nell'ambito della valutazione dell'abilità di ascolto, a un livello B2 – sarebbero potuti variare se costui si fosse sottoposto alla prova di una certificazione piuttosto che a quella di un'altra.

Lo scopo di questo contributo non è, comunque, quelli di auspicare prove 'perfette', dall'oggettività assoluta, assodato del resto che ogni misurazione delle abilità linguistiche equivale pur sempre a un'istantanea sfocata e parziale (Douglas, 2011); piuttosto è quello di stimolare a un maggiore impegno, affinché i margini di approssimazione si possano ridurre. Va precisato, in ogni caso, che poiché l'analisi è limitata a una singola prova e dato il numero ristretto di item, lo studio non è rappresentativo, non consente cioè di formulare un giudizio in generale sul modo con cui le singole certificazioni valutano la comprensione orale a livello B2. Ciò che emerge ha, altrimenti, un valore circoscrivibile ai soli dati a disposizione. L'impianto metodologico dell'indagine può valere comunque da riferimento per analisi più estese.

### Bibliografia

Cardillo, G., Menzinger, C. & Vecchio, P. (ed.) (2015). *Livello B1 PLIDA. Quaderno delle specifiche*. Roma: Società Dante Alighieri. <http://plida.it/certificazione-plida/documenti.html> (consultazione: 18.01.2016)

Douglas, D. (2011). *Understanding Language Teaching*. Londra: Hodden.

Green, R. (2013). *Statistical analyses for language teachers*. Basingstoke: Palgrave MacMillan.

Novello, A. (2014). *La valutazione delle lingue straniere e seconde nella scuola. Dalla teoria alla pratica*. Venezia: Edizioni Ca' Foscari.

### Sitografia

Prova ascolto B2 CELI  
<http://www.cvcl.it/categorie/categoria-104>  
(consultazione: 18.01.2016)

Prova ascolto B2 CILS  
[http://cils.unistrasi.it/89/198/Prove\\_Liv\\_B2.htm](http://cils.unistrasi.it/89/198/Prove_Liv_B2.htm) (consultazione: 18.01.2016)

Prova ascolto B2 PLIDA (vecchio formato)  
<http://plida.it/certificazione-plida/che-cos-e-la-certificazione-plida/esempi-di-prove-d-esame/esempi-sessioni-esame-plida.html>  
(consultazione: 18.01.2016)

Prova ascolto B2 .IT  
<http://www.certificazioneitaliano.uniroma3.it/B2-intlT.aspx> (consultazione: 18.01.2016)