

ACCESSIBILITÀ E INCLUSIVITÀ NELLA CERTIFICAZIONE LINGUISTICA: UNO STUDIO DI CASO NELL'ITALIANO L2

Marco Mezzadri¹, Paola Vecchio²

1. INTRODUZIONE

Questo contributo intende presentare il progetto di adattamento, in risposta a bisogni educativi speciali, delle procedure di gestione e somministrazione del test di certificazione di italiano come lingua straniera, Progetto di Lingua Italiana Dante Alighieri (PLIDA), rilasciato dalla Società Dante Alighieri. Le ipotesi di adattamento sono state testate tramite una sperimentazione svolta in un contesto coeso e relativamente uniforme in quanto a età, provenienza e ruolo sociale dei soggetti in esame e dei componenti del gruppo di controllo. Di tale percorso vengono forniti alcuni dati iniziali che sono stati raccolti e analizzati e che riteniamo possano essere utili a promuovere una riflessione, tutt'ora in divenire, sull'applicazione di adattamenti volti a migliorare l'accessibilità del test di certificazione e, di conseguenza, a rendere più inclusiva la prova, con attenzione posta sui candidati con BES, in particolare con DSA e con disturbi del linguaggio.

Un ulteriore confronto viene sviluppato tra i gruppi sperimentale e di controllo da un lato e, dall'altro l'insieme complessivo a livello mondiale dei candidati della sessione in esame.

Il progetto di adattamento nasce da una considerazione generale sulla necessità di esplorare nuove soluzioni per rendere maggiormente efficaci, in termini di inclusività, le misure da tempo adottate per favorire l'accesso ai test da parte di candidati con disabilità o DSA.

La certificazione PLIDA, infatti, prevede da regolamento l'applicazione di misure compensative o dispensative, in base alle linee guida del MIUR allegate al DM 12 luglio 2011, per i candidati che ne facciano richiesta e che producano la documentazione relativa al loro disturbo o disabilità. Le misure compensative adottate sono: tempo maggiorato, ascolti ripetuti, fascicoli delle prove stampati con font dedicati per candidati con dislessia, fascicoli con dimensioni e risoluzione aumentati per candidati ipovedenti, istruzioni lette dagli assistenti agli esami per consentire la lettura da parte di candidati con dislessia o non vedenti; uso delle cuffie per le prove d'ascolto per candidati ipoudenti, computer, sintetizzatori vocali. Le misure dispensative comprendono l'esonero da alcune prove, l'esclusione dalla valutazione dei parametri che possono essere influenzati da disturbi o disabilità.

A fronte di un impianto già strutturato, la raccolta dei dati relativa al periodo 2017-2022 ha evidenziato che la libera segnalazione di bisogni speciali da parte dei candidati è riconducibile a pochi casi a livello mondiale a fronte di circa 90.000 test sostenuti. Inoltre, si è trattato, per la maggior parte dei casi, di richieste motivate da disabilità e non da disturbi.

¹ Università di Parma.

² Società Dante Alighieri.

Il contributo è stato pensato e strutturato assieme dai due autori. L'attribuzione della scrittura dei singoli paragrafi è così distribuita: M. Mezzadri 1, 2, 3, 4.3, 6.1, 7.1; P. Vecchio 4.1, 4.2, 5, 6.2, 7.2.

La ricerca ha dunque inteso cercare di determinare parametri di accessibilità e inclusività del test PLIDA in grado di rispondere in maniera più adeguata ai bisogni educativi speciali dei candidati. La prima domanda di ricerca formulata ha riguardato l'individuazione di parametri e criteri di analisi che andassero oltre le indicazioni normative nazionali e che posassero sulla letteratura scientifica a disposizione³.

A seguire, ci si è chiesti quali elementi fossero necessari per arrivare al risultato operativo di fornire una versione del test PLIDA maggiormente accessibile e inclusiva, giungendo alla divisione della ricerca in due filoni: il primo è volto ad analizzare gli elementi che determinano il livello di accessibilità testuale sulla base di criteri che prendono in esame aspetti linguistici, cognitivi e relativi alle conoscenze culturali. Il secondo filone di ricerca è finalizzato alla sperimentazione di adeguamenti e ristrutturazioni applicabili al test PLIDA. In questa sede, presentiamo alcune riflessioni riguardanti la seconda pista, tramite l'analisi dei dati acquisiti attraverso la sperimentazione svolta, durante l'anno scolastico 2022-23, nel contesto della provincia di Bolzano. Tutti i candidati, tanto del gruppo sperimentale quanto di quello di controllo, erano iscritti al penultimo e ultimo anno delle scuole secondarie di secondo grado di questa zona. La scelta di concentrarsi su questo contesto, avendo a disposizione un'ampia gamma di soluzioni alternative in un grande numero di paesi in tutti i continenti, deriva dal fatto che la condizione di partenza conferisce ai dati raccolti un livello di omogeneità notevole. Inoltre, la presenza nell'area della provincia di Bolzano di un sistema efficiente di educazione pubblica a più livelli ha consentito di aggregare e coordinare le candidature.

A ciò si aggiunge un dato particolarmente significativo: nella serie storica di sessioni di test PLIDA presa in esame, dal 2017 al primo semestre 2023, soltanto 12 candidati provenienti da queste scuole dichiaravano di avere bisogni educativi speciali, a fronte di una media tra le più elevate in Italia⁴ di alunni con DSA iscritti.

Per concludere, il concetto di inclusività a cui si ispira il nostro gruppo di ricerca induce a considerare la necessità di riflettere attentamente sulle caratteristiche del test PLIDA, sui metodi di creazione, somministrazione e amministrazione delle prove, nonché sui risultati conseguiti dai candidati, in modo da poter eventualmente migliorare il livello di accessibilità del test per tutti, non soltanto per i candidati con BES e DSA. È questo un ulteriore filone di ricerca che ci stiamo apprestando ad affrontare, ma di cui non trattiamo in questo contributo.

Da questa rapida disamina del contesto della PLIDA emerge, su tutte, la convinzione che sia possibile intraprendere un cammino che colga le opportunità offerte dalla certificazione per esplorare anche nuovi ambiti dell'educazione linguistica. Si evince, infatti, che la via spontanea alla partecipazione ai percorsi di certificazione linguistica da parte di studenti con DSA o disabilità non sta dando i risultati che ci si potrebbero attendere data la percentuale di casi presenti nelle realtà educative e quindi nella società in genere, ovviamente con le numerose rilevanti differenze che i differenti contesti culturali presentano. Non è un caso,

³ Le prime riflessioni erano già state oggetto dell'intervento di Francesca Bariviera e Giammarco Cardillo, "Equità e accesso ai test di italiano LS/L2: l'esperienza della certificazione PLIDA" al Convegno scientifico "Verso un'educazione linguistica inclusiva: problemi di oggi, prospettive di domani", Università degli Studi di Parma (11 novembre 2019).

⁴ Secondo il documento dell'allora Ministero dell'Istruzione intitolato *I principali dati relativi agli alunni con DSA* (p. 9), gli alunni con DSA in Trentino-Alto Adige di cui fa parte la Provincia di Bolzano erano il 5,5% della popolazione scolastica totale nell'anno scolastico 2020-21, a fronte di una media nazionale del 5,3%. Il documento è consultabile all'indirizzo:

www.miur.gov.it/%2Fdocuments/%2F20182/%2F6891182/%2FFocus/%2B Sugli/%2B alunni/%2B con/%2B Disturbi/%2B Specifici/%2B dell'/%2527Apprendimento_aa.ss.201920_202021.pdf.

infatti, che le candidature di studenti con bisogni speciali provengano da paesi e contesti educativi in cui è presente una maggiore e più matura sensibilità, supportata da norme giuridiche, verso i temi dell'inclusione. La possibilità di realizzare un progetto di ricerca innovativo facendovi partecipare un gruppo coeso di studenti candidati con bisogni speciali è stata fornita grazie alla collaborazione con l'Ispettorato all'inclusione e con la Ripartizione pedagogica della Direzione Istruzione e Formazione tedesca della Provincia Autonoma di Bolzano.

2. VALIDITÀ E AFFIDABILITÀ DI UN TEST

Il concetto di validità è al centro della ricerca e delle relative applicazioni nell'ambito del testing linguistico. Negli anni, i ricercatori hanno dato corpo a diverse prospettive riguardo non solo alla validità ma anche alla validazione. In uno dei suoi contributi, Chapelle (2012: 21-33) fornisce un excursus storico del concetto di validità e del processo di validazione, assieme a una riflessione su temi di questo ambito considerati oggi critici. L'autrice sostiene che approcci diversi hanno portato ad attribuire significati diversi al concetto di validità e che le domande focali non sono cambiate: come si interpretano i risultati di un test e per cosa possono essere utilizzati?

La ricerca internazionale condotta fin qui, sostenuta da un bisogno crescente di testing linguistico e di certificazioni, è così ampia che lungo i decenni sono state prodotte estese rassegne bibliografiche sullo stato dell'arte in materia. Basti citare alcuni dei più prolifici ricercatori in questo campo e alcune delle loro rassegne più significative: Alderson (1991, 1995), Alderson, Banerjee (2001, 2002); Bachman, Palmer (2010); Chapelle (1999), Chapelle, Lee (2021); Hamp-Lyons, Lynch (1998); Kane (2001); Kunnan (1998, 1999). Alcune di queste rassegne hanno preso in esame i diversi aspetti della validità con metodi quantitativi: Cumming (1996); Bachman, Eignor (1997); Kunnan (1998). Chapelle (2012) dedica attenzione anche a recenti sviluppi, come quelli descritti da Kane (2006) con enfasi posta su riflessioni interpretative, affermando che «*validity in mainstream language assessment may be moving forward in harmony with educational measurement*»; anche se, e concordiamo: «*[...] within educational measurement, the conception of validity is not a settled matter, but rather a source of continuing discussion [...]*» (Chapelle, 2012: 26).

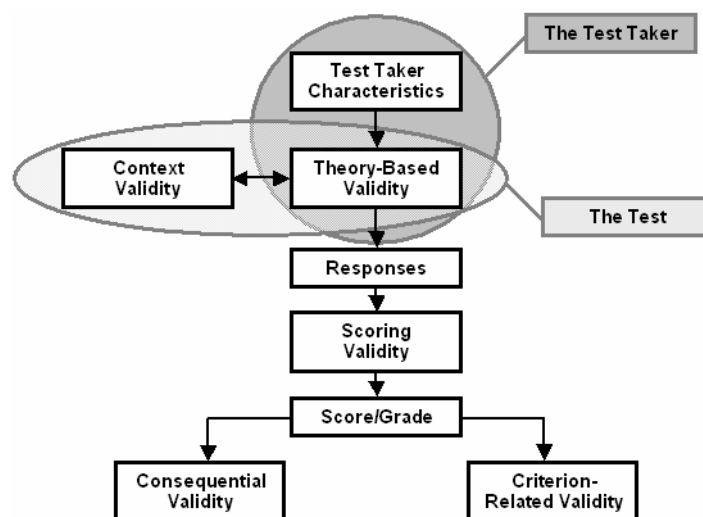
Due altri influenti studiosi, McNamara (1996) e Douglas (2000), avevano sottolineato in precedenza una particolarità dei test riguardanti la lingua per fini specifici; test che non possono fare a meno di prendere in esame la complessità della competenza comunicativa, includendo le strategie, la *background knowledge* e il contesto in cui la lingua si realizza. Nel caso particolare oggetto della nostra ricerca, relativo a studenti con DSA nelle scuole secondarie di secondo grado della provincia di Bolzano, crediamo occorra applicare un approccio in tutto simile, nonostante il test PLIDA somministrato rientri nella tipologia di certificazioni finalizzate alla valutazione delle competenze linguistico-comunicative generali. Ciò perché, tra l'altro, appare sempre più evidente che in primis le strategie risultano essere cruciali nella gestione di un test linguistico indipendentemente dalla natura della lingua testata, sia essa per scopi comunicativi generali o specifici e per fini di studio. Le nostre scelte circa le misure di adeguamento messe in atto e le attività riguardanti la ricerca su aspetti cognitivi e culturali ancora in sviluppo si ispirano e testimoniano questo approccio.

Secondo Chapelle (2012: 28), «[t]he construct definition has to include the domain of language use as well as the ability to make appropriate linguistic choices and interpretations in order to make meaning».

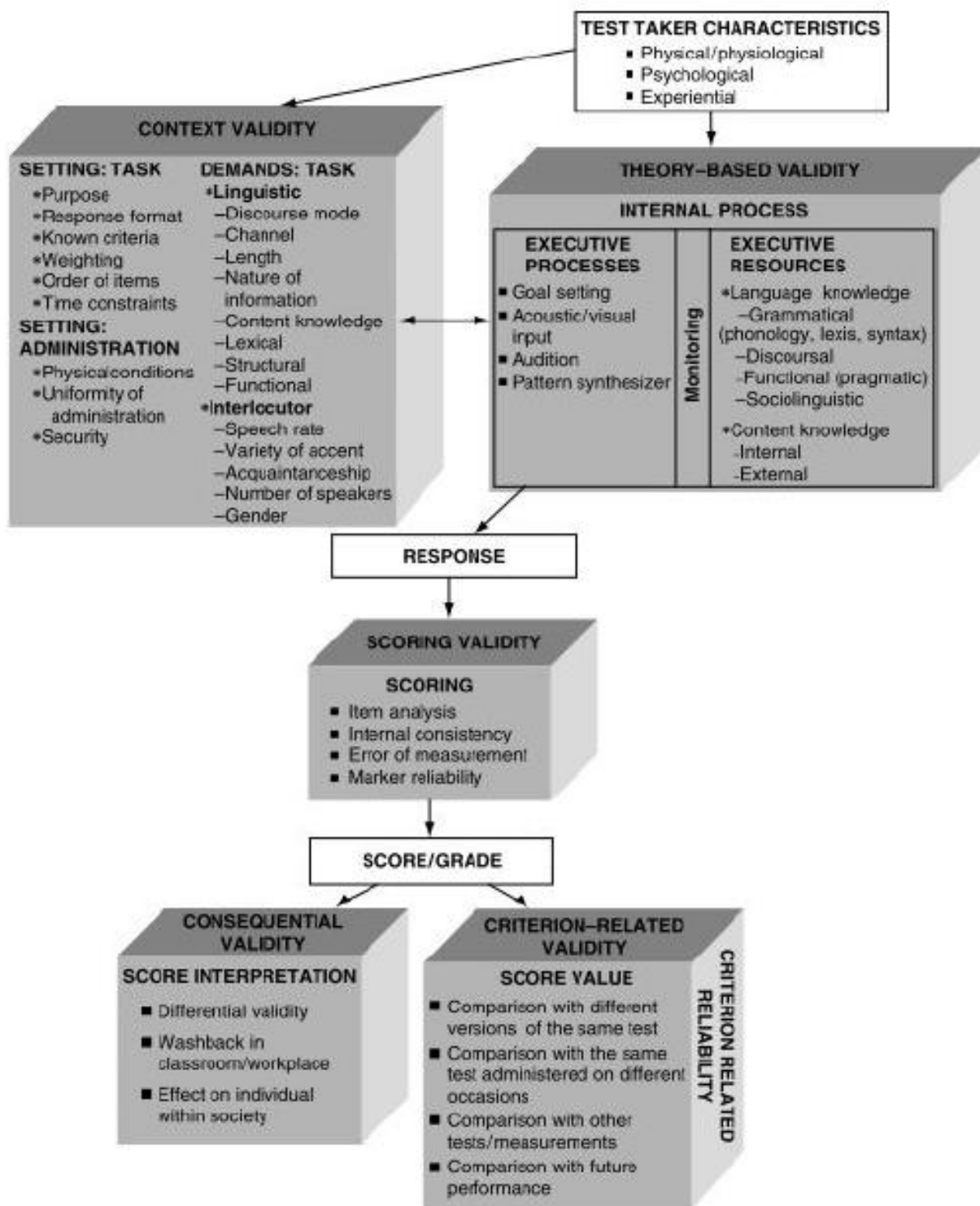
Questo modo di accostarsi ai temi in oggetto esige una revisione delle interpretazioni tradizionali del concetto di validità, riassumibili con Lado (1961: 312) nel modo seguente: «Does a test measure what it is supposed to measure? If it does, it is valid». Weir suggerisce (2005: 13) una modifica affermando che «validity is multifaceted and different types of evidence are needed to support any claims for the validity of scores on a test» e che (2005: 12) «[v]alidity is perhaps better defined as the extent to which a test can be shown to produce data, i.e., test scores, which are an accurate representation of a candidate's level of language knowledge or skills. In this revision, validity resides in the scores on a particular administration of a test rather than in the test per se».

L'affidabilità è l'altra caratteristica principale di un test (cfr. Jones, 2012). Quando il termine “affidabilità” è riferito a un test, il significato è relativamente diverso dall'accezione usuale. Nel significato corrente qualcosa di affidabile è “degno di fiducia”, mentre quando si tratta di un test il significato più prossimo a “affidabile” è “coerente”. Jones (2012: 352) afferma che «[a] reliable test is consistent in that it produces the same or similar result on repeated use; that is, it would rank-order a group of test takers in nearly the same way». Ciononostante, la sola affidabilità non è sufficiente, in quanto se un test è affidabile, ciò non vuole necessariamente dire che è anche un buon test, cioè, come sostiene Weir (2005: 12), che è accurato nel trasmettere una corretta rappresentazione del livello di conoscenze linguistiche o di abilità di un candidato al test. Così come anche la validità da sola non è sufficiente. Per un approfondimento sui rapporti che possono instaurarsi tra “affidabilità” e “validità”, si veda il contributo di M. Barni (2000: 161-163).

Alla luce di queste considerazioni, la creazione di un test linguistico, che nel caso della PLIDA è di tipo certificatorio, comporta un'ampia riflessione su numerosi fattori che concorrono a costruire l'architettura degli ineludibili processi di validazione del test. Tale analisi può essere eseguita utilizzando una serie composita di strumenti. Alcuni studiosi hanno sviluppato dei *framework* per rendere maggiormente esplicita e applicabile la loro visione. Di uno di questi è autore Weir (2005): lo adottiamo in questo testo in quanto lo riteniamo esaustivo e coerente con i fini del nostro studio. Ne forniamo a seguire una rappresentazione sintetica (ALTE, 2005: 26):



Si evince con facilità quali sono gli ambiti da tenere in considerazione nei processi di validazione. Questi ambiti possono essere ulteriormente analizzati in maggior dettaglio così come suggerito dal seguente diagramma, riportato a titolo esemplificativo e riferito all'abilità di ascolto (Weir, 2005: 45, ripreso da Mezzadri, 2017: 39).



Lo stesso modo di rappresentare la struttura viene applicato alle altre macroabilità con il mantenimento di molte parti e la modifica di quelle in cui si registrano differenze tra le abilità, cioè nelle sezioni dedicate alla *theory-based validity* e alla *context validity*.

La rappresentazione proposta evidenzia in maniera immediata le sei aree su cui concentrarsi per un'analisi della validità di un test: *test taker characteristics*, *context validity*, *theory-based validity*, *scoring validity*, *consequential validity*, *criterion-related validity/reliability*. Weir (2005: 49) sottolinea che per essere certi della validità e della correttezza – etica, aggiungiamo noi – gli sviluppatori di un test debbono porsi, in modo rigoroso, una domanda per ognuno di questi ambiti. Le riproponiamo in ordine, nella nostra traduzione:

- Quali sono le caratteristiche fisiche/fisiologiche, psicologiche ed esperienziali dei candidati a cui si rivolge il test?
- Le caratteristiche del contesto dei compiti presentati nel test e i processi di amministrazione del test risultano corretti per i candidati da un punto di vista delle situazioni proposte?
- I processi cognitivi richiesti per completare i compiti sono autentici sul piano dell'interazione?
- Quanto sono affidabili i risultati?
- Che impatto ha il test sui diversi *stakeholder*?
- Che prove esterne ci sono che il test funzioni?

3. LA VALIDAZIONE DEL TEST SPERIMENTALE PLIDA PER STUDENTI CON DSA

Le considerazioni esposte nel paragrafo precedente permettono di avvicinarsi meglio attrezzati alla trattazione del tema delle modifiche apportate, in chiave sperimentale, alla certificazione PLIDA per studenti con BES e in particolare con DSA.

Nella letteratura anglofona si ritrovano due ordini di modifiche, *accommodations* e *modifications*, termini difficili da rendere in italiano, in quanto in Italia molto spesso il focus viene posto principalmente sulla distinzione tra misure compensative e dispensative. Adottiamo “adeguamento” e “ristrutturazione” quali traduzioni rispettivamente di *accommodations* e *modifications*, in linea con Daloiso (2017).

L'analisi dei fattori e le conseguenti ipotesi di modifica, sotto forma di adeguamenti e ristrutturazioni, comportano un processo lungo e complesso che è ancora in atto. Al momento, la ricerca è in corso e siamo in grado di proporre alcuni risultati limitatamente ad alcuni elementi che vengono descritti nei paragrafi successivi.

In generale, ma a maggior ragione nel caso specifico in esame, a nostro avviso, occorre enfatizzare una variabile, comunque inclusa nel *framework* di Weir sopra descritto: l'equità di un test (*fairness* nella letteratura anglofona). L'etica – una sorta di iperonimo che accoglie il concetto di equità – rappresenta un tema cruciale che non ha sempre avuto fortuna in ambito linguistico e linguistico-educativo. Secondo Hamp-Lyons (1997b: 324), come ricorda Barni (2023: 67), le ragioni risiedevano inizialmente in un approccio positivisticò alla lingua, vista come esistente a sé, al di fuori e oltre l'individuo. Una maggior attenzione verso l'apprendente si cominciò a sviluppare a partire dagli anni '90 del secolo scorso, comportando l'inaugurazione di filoni di indagine promettenti anche sui temi dell'etica nel testing linguistico. Per un dettagliato excursus storico e la relativa rassegna bibliografica, rimandiamo a Hamp-Lyons (1997b: 323-328) e a Barni (2023: 63-71).

Sul tema dell'etica, Daloiso (2017: 159) ricorda che «validity and fairness are not intrinsic features of a test, as they both depend on the intended purpose of the test and the context in which it is used», e aggiunge, intrecciando i due elementi, che la validità e l'equità sono determinati dal livello di limitazioni di cui è portatore il candidato al test, dal tipo di supporto fornito nelle aree linguistiche che il test prende in esame e dall'interpretazione che sarà data ai risultati.

Gli interventi da approntare per rendere accettabile e valido un test linguistico per candidati con DSA comportano, quindi, la riflessione sulla dimensione etica, che, tuttavia, non si esaurisce nella identificazione di misure definibili quali adeguamenti e ristrutturazioni, ma tocca aspetti cognitivi e culturali relativi, in particolare, alla interazione tra testo e *background knowledge* in riferimento ai processi di comprensione (cfr. Mezzadri, 2017: 20-35). In questa sede ci limitiamo a rendere conto di interventi relativi agli adeguamenti scelti, in quanto i dati a oggi disponibili sulle ristrutturazioni non sono ancora stabili, nonostante sia in corso un ampliamento della ricerca in questo senso.

Negli *Standards for Educational and Psychological Testing* (2014: 215 e 221), così viene definito il concetto di *accommodation*: «[a]djustments that do not alter the assessed construct that are applied to test presentation, environment, content, format (including response format), or administration conditions for particular test takers, and that are embedded within assessments or applied after the assessment is designed». Mentre la definizione di *modification* recita: «[a] change in test content, format (including response formats), and/or administration conditions that is made to increase accessibility for some individuals but that also affects the construct measured and, consequently, results in scores that differ in meaning from scores from the unmodified assessment».

Daloiso (2017: 160) offre esempi di entrambe: cambiare il *layout* e dare tempo aggiuntivo per l'esecuzione del test, per le prime; per le seconde, esentare dallo svolgimento di compiti basati sull'ortografia.

Acosta (2008: 4-6) propone un'ulteriore distinzione tra adeguamenti diretti e indiretti. La riflessione risulta interessante in quanto coglie una tendenza iniziale nello stabilire l'efficacia di forme di intervento diretto in ambito linguistico, come l'utilizzo di glossari e dizionari o di lettori, oppure come la possibilità di rispondere oralmente ai quesiti scritti, ecc. Modalità indirette quale l'aumento del tempo a disposizione o la divisione in sottogruppi dei candidati si rendono necessarie in conseguenza anche delle scelte di compensazione linguistica: l'uso di un glossario, ad esempio, porta inevitabilmente a estendere i tempi.

Con una prospettiva differente che prende spunto dagli adeguamenti proposti per il testing di studenti con disabilità estendendo la riflessione agli apprendenti di inglese L2, in un contributo a cui rimandiamo Abedi (2012: 62-76) presenta una puntuale disamina dell'efficacia delle tipologie più diffuse di adeguamenti, oltre a un'utile panoramica storica.

A completamento di quest'analisi, aggiungiamo che non sarebbe possibile comprendere la ratio con cui sono stati adottati adeguamenti, e in futuro ristrutturazioni, nella certificazione PLIDA per studenti con BES senza tener conto delle riflessioni svolte circa il tema dell'ansia nel testing linguistico. Si tratta di un ambito che ha suscitato grande interesse fin da tempi lontani (cfr. Blau, 1955; Scovel, 1978); per Horwitz, Horwitz e Cope (1986: 128), l'ansia nell'apprendimento linguistico in classe è un insieme di percezioni, credenze, sentimenti e comportamenti che insorge a causa dell'unicità dei processi di apprendimento linguistico. In quest'ultimo studio è contenuta la Foreign Language Classroom Anxiety Scale (FLCAS), che nei decenni produrrà una serie di declinazioni sotto forma di scale specifiche, così come ricordano Teimouri *et al.* (2019: 14).

Gettando uno sguardo agli studenti con DSA, secondo Daloiso (2017: 159), le persone con dislessia soffrono di ansia da lingua straniera e hanno performance non buone quando si trovano sotto pressione per quanto concerne la gestione dei tempi. Questa condizione si realizza in particolare durante i test. Ciò comporta, da parte degli insegnanti o dei somministratori, la necessità di ricordare che un test come quello di certificazione «potrebbe non risultare affidabile quanto si potrebbe essere portati a pensare quando si è in presenza di studenti dislessici» (2017: 160, nostra traduzione). Anche altri disturbi del linguaggio e dell'apprendimento conducono alla creazione di stati d'ansia (cfr. Im-Bolter, Cohen, 2007).

Le conseguenze dell'ansia, secondo Cheng e Zheng (2021: 178), colpiscono i meccanismi cognitivi, la capacità di affrontare le difficoltà e il controllo dell'attenzione, ingenerando uno stato di disagio e di nervosismo.

Nella situazione specifica di un test linguistico, l'ansia può essere scatenata da diversi fattori quali la già ricordata pressione dovuta al tempo o la posta in gioco del test.

4. LA CERTIFICAZIONE PLIDA DALL'INTERNO

4.1. *Uno sguardo d'insieme*

La certificazione PLIDA è una delle quattro certificazioni della competenza in italiano come lingua straniera riconosciute dallo Stato italiano. In quanto ente certificatore, la Società Dante Alighieri è socio fondatore dell'associazione CLIQ e, dal 2016, membro associato dell'ALTE (Association of Language Testers in Europe).

Il progetto PLIDA ha preso l'avvio a metà degli anni '90 del secolo scorso ed era originariamente inteso a supportare – oltre che con la certificazione linguistica, anche con la produzione di sillabi e materiali didattici per l'insegnamento e con iniziative di formazione glottodidattica – i comitati esteri della Società Dante Alighieri nelle attività di promozione della lingua italiana nei territori di competenza. Negli anni successivi la certificazione PLIDA si è estesa al di fuori della rete dei comitati e sono stati aperti centri d'esame presso università straniere, Istituti italiani di cultura, CPIA, scuole e associazioni private in Italia e all'estero. Oggi il 30% degli esami vengono sostenuti in centri autorizzati presso istituzioni al di fuori della rete dei comitati.

La certificazione PLIDA prevede due programmi distinti: uno destinato a candidati adulti (PLIDA ordinario); uno destinato a candidati in età scolare, tra i 13-17 anni (PLIDA Juniores).

I test PLIDA sono *criterion-referenced* (Bachmann, 1990: 8), allineati al *Quadro comune europeo di riferimento per le lingue* (d'ora in avanti QCER) per la definizione delle competenze attese ai vari livelli compresi tra l'A1 e il C2.

Il costrutto degli esami, descritto nel *Sillabo della certificazione PLIDA* e nei *Quaderni delle specifiche* dei singoli livelli d'esame consultabili all'indirizzo <https://plida.dante.global>, riflette il principio dell'orientamento all'azione adottato e promosso dal QCER, secondo il quale «[l']uso della lingua, incluso il suo apprendimento, comprende le azioni compiute da persone che, in quanto individui e attori sociali, sviluppano una gamma di **competenze**, sia **generali** sia, nello specifico, **linguistico-comunicative**. Gli individui utilizzano le proprie competenze in **contesti** e **condizioni** differenti e con **vincoli** diversi per realizzare delle **attività linguistiche**. Queste implicano i **processi** linguistici di produrre e/o ricevere **testi** su determinati **temi** in **domini** specifici, con l'attivazione delle **strategie** che sembrano essere più adatte a portare a buon fine i **compiti** previsti» (QCER, 2001: 12; grassetto nell'originale).

Il programma di certificazione per adulti prevede esami per i livelli dall'A1 al C2, mentre il programma denominato Juniores offre esami per i livelli dall'A1 al B2. Gli esami di ciascun livello si compongono di 4 prove dedicate ciascuna a una delle 4 abilità fondamentali, Ascoltare, Leggere, Scrivere e Parlare (per quest'ultime vengono proposti due task distinti che verificano sia la competenza di produzione, sia la competenza di interazione).

La creazione delle prove d'esame avviene all'interno dell'unità del PLIDA della Società Dante Alighieri con il contributo di *item writer* formati appositamente per l'elaborazione di *item* che rispecchino le specifiche d'esame, secondo il seguente schema (*Manual for Language Test Development and Examining* dell'ALTE, 2011: 26):

Test specifications → *producing materials* → *quality control* → *constructing tests* → *live test materials*

Il controllo di qualità avviene con il contributo di *item writer* esperti e prevede, alla fine del processo di correzione e riscrittura, sessioni di pilotaggio interne all'unità del PLIDA. Fino al 2019, prima della pandemia, venivano regolarmente organizzate sessioni di pre-testing delle prove di comprensione scritta e orale con un campione di studenti di italiano LS (tra i 50 e i 150) rappresentativo della popolazione complessiva che si sottopone agli esami PLIDA.

La correzione degli esami è centralizzata per quanto riguarda le prove a risposta chiusa (Ascoltare e Leggere) e la prova di produzione e interazione scritta. La valutazione della prova di produzione e interazione orale viene invece realizzata in due fasi: una prima valutazione viene espressa dagli esaminatori che compongono la commissione che somministra gli esami; una seconda valutazione viene espressa centralmente durante la fase di controllo e riascolto delle registrazioni della prova Parlare. In caso di discrepanze tra le due valutazioni, viene considerata valida quella espressa dai valutatori della sede centrale.

I valutatori delle prove di produzione e interazione scritta e orale vengono selezionati sulla base del possesso di specifici requisiti accademici e professionali e formati all'utilizzo delle griglie di valutazione del PLIDA attraverso un programma che prevede dapprima una formazione individuale e successivamente sessioni periodiche di *benchmarking*, con correzioni collettive che hanno lo scopo di ridurre la variabilità nel giudizio dei valutatori (Shohamy *et al.*, 1992).

Al termine di ogni sessione i risultati delle prove di ricezione (prove chiuse) vengono analizzati con la tecnica dell'*item analysis* classica e vengono raccolti dati sulla tenuta in sessione sia delle prove, sia dei singoli *item*.

Per le prove Scrivere e Parlare si raccolgono dati sulla produttività dei task proposti, sia nei termini dell'effettiva elicitazione delle competenze che si intendevano osservare, sia nei termini delle preferenze dei candidati laddove il formato d'esame prevede la scelta tra task equivalenti.

Viene inoltre prodotta una reportistica sulla popolazione d'esame per età, genere e provenienza dei candidati che, insieme ai dati risultanti dall'*item analysis*, riorienta il processo di creazione delle prove d'esame per le successive sessioni.

Per rispondere ai principi di equità ed eticità della certificazione (Messick, 1989; Kunnan, 2004, 2014; AERA, APA, NCME, 2014), nonché per allinearsi al *Codice professionale dell'ALTE* e agli *Standard minimi per stabilire profili di qualità negli esami ALTE* (cfr. <https://www.alte.org/>) e per ottemperare a quanto disposto dalla normativa italiana per la valutazione delle competenze di candidati con DSA o portatori di handicap (L. 8/10/2010 n. 170, D.L. n. 5669 del 12/07/2011, L. 104 del 15/02/1992, art. 8, 16, 20), il Regolamento del PLIDA prevede

l'applicazione di misure dispensative o compensative in sede d'esame per tutti i candidati che ne facciano richiesta all'atto dell'iscrizione.

La richiesta dell'applicazione di misure speciali non è tuttavia molto frequente e, pur facendo registrare un incremento dalle poche unità del 2017 alle diverse decine del 2022 e del primo semestre del 2023, continua ad essere estremamente esigua a fronte delle circa 20.000 iscrizioni totali ogni anno.

4.2. La certificazione PLIDA nella Provincia Autonoma di Bolzano

In virtù di un accordo esclusivo stipulato la prima volta nel 2011 tra la Società Dante Alighieri, il Comitato di Bolzano e la Ripartizione pedagogica presso la Direzione Istruzione e Formazione tedesca della Provincia Autonoma di Bolzano, ogni anno centinaia di studenti di madrelingua tedesca che frequentano il penultimo e l'ultimo anno degli istituti superiori di secondo grado partecipano, a condizioni economiche agevolate, agli esami di certificazione PLIDA di livello B2 e C1; i certificati così ottenuti hanno il valore di titoli sostitutivi per ottenere il patentino di bilinguismo nella regione Trentino-Alto Adige (Decreto Presidenziale 86 del 14 maggio 2010).

Il Secondo Statuto di Autonomia della Provincia, risalente al 1972 e tuttora vigente (art. 19), prevede infatti il mantenimento della separazione degli insegnamenti nelle scuole materne, elementari e secondarie, per gli alunni appartenenti ai diversi gruppi etnici della Regione, come nello Statuto del 1948, e introduce, a partire dalle scuole elementari, l'obbligo dell'insegnamento della L2 (italiana o tedesca), che deve essere «impartito da docenti per i quali tale lingua sia ugualmente quella materna» (Statuto di Autonomia 1972, art. 19). Il livello target in uscita dalla scuola secondaria di secondo grado è il B2.

Sul bilinguismo dell'Alto Adige si veda Dal Negro (2017). Per un'analisi sociolinguistica sul bilinguismo degli studenti delle scuole della Provincia di Bolzano e sui livelli di competenza in L2 raggiunti si vedano Vettori e Abel (2017), Vettori e Abel (2021), Masillo *et al.* (2021).

Anche in questo contesto la richiesta di applicazione di misure speciali non è frequente. Nella tabella seguente sono riportati i dati degli anni compresi tra il 2017 e il primo semestre 2023 relativi agli esami svolti nelle Scuole della Provincia Autonoma di Bolzano e alle richieste dell'applicazione di misure speciali.

Tabella 1. *Richieste di applicazione di misure speciali negli anni 2017-2023*

Anno	Iscrizioni totali	Richieste di applicazione di misure speciali
2017	882	1
2018	811	2
2019	814	5
2020	375	1
2021	685	1
2022	1020	4
2023 (primo semestre)	427	2

4.3. Adegamenti nella versione sperimentale per studenti con BES della PLIDA

Come già indicato, la Certificazione PLIDA può contare su un impianto strutturato di misure compensative e dispensative ispirate alla vigente normativa italiana in materia di bisogni educativi speciali. A seguito delle riflessioni sviluppate in seno al gruppo di ricerca e avendo tenuto presente considerazioni in merito al contesto in cui la ricerca si realizzava, sono state proposte misure utili a promuovere condizioni sperimentali rispondenti ai bisogni di studenti con BES. Le misure messe in atto avevano lo scopo primario di aiutare i candidati che presentavano bisogni educativi speciali nella gestione della prova di esame.

Il numero esiguo di soggetti con BES, tutti con DSA, coinvolti ha permesso di massimizzare la personalizzazione dell'intervento, nell'intento di ridurre quanto possibile elementi di disturbo al contesto sperimentale.

La prima fase ha riguardato l'assunzione di elementi utili a comprendere il tipo di assistenza che poteva essere offerta. Come strumento di riferimento è stato preso in esame il PDP/PEP redatto dalla scuola di provenienza dei candidati e si è agito in contatto diretto con il referente per i bisogni educativi speciali.

Dopo aver definito le misure di adeguamento del test, si è proceduto a organizzare la prova in base a quanto concordato. Non si è rilevata la necessità di applicare misure di ristrutturazione della prova.

La valutazione della prova d'esame è avvenuta applicando i criteri utilizzati per i candidati del gruppo di controllo.

Alla base delle scelte attuate stanno le misure di adeguamento e ristrutturazione ipotizzate nella sperimentazione della PLIDA che sono illustrate nella tabella seguente. Di queste ne sono state applicate soltanto alcune, avendo preso in esame le esigenze dei soggetti partecipanti alla sperimentazione.

Tabella 2. *Le misure di adeguamento e ristrutturazione della PLIDA*

Tipologia di disturbo	Tempo	Strumenti	Altro
Disturbi della vista		Stampa in caratteri leggibili per ipovedenti o utilizzo di computer con software per la lettura dello schermo. Lenti di ingrandimento o altri supporti analoghi	Le stampe sono fornite ai candidati. Il computer con software, le lenti e altri strumenti di supporto sono invece in dotazione allo studente con BES.
Disturbi specifici dell'apprendimento	30% in più per ogni singola prova Gestione alternativa della prova di ascolto Pause con supervisore	<u>Dislessia</u> : uso del computer con software di lettura dello schermo. Il computer non può avere accesso alla rete e alle funzioni di controllo ortografico, grammaticale, dizionario e vocabolario dei sinonimi e dei contrari.	È prevista l'esecuzione della prova in una sessione unica con studenti senza BES. Il personale per la somministrazione è coadiuvato da un'unità dedicata. Viene inoltre messo a disposizione un locale attrezzato in più. I tempi di svolgimento della prova si dilatano del 30%.

		<u>Disgrafia</u>: uso del computer non connesso alla rete e senza accesso alle funzioni di controllo ortografico, grammaticale, dizionario e vocabolario dei sinonimi e dei contrari. <u>Disortografia</u>: dispensa dalla valutazione della correttezza ortografica.	I dispositivi sono in dotazione allo studente con BES.
Disturbi dell'udito	30% in più per le prove di comprensione e di produzione orale Gestione alternativa della prova d'ascolto Pause con supervisore	A seconda del livello di disabilità, si può giungere alla dispensa dalla prova d'ascolto e a misure compensative per la prova di produzione orale.	La prova si svolge in sessione unica con studenti senza BES. Esonero dalla prova d'ascolto. Se il candidato è portatore di un impianto cocleare o di un apparecchio acustico, affronta la prova senza misure di adeguamento o ristrutturazione.
Altri disturbi o disabilità certificate	30% in più gestione alternativa della prova di ascolto; pause con supervisore	A seconda delle necessità, uno o più strumenti tra quelli previsti e indicati	La prova si svolge in sessione unica insieme a studenti senza BES. A seconda delle necessità si possono prevedere misure anche di ristrutturazione, ad esempio dalla prova di produzione orale in casi di documentate difficoltà che potrebbero rendere inattuabile la prova di verifica delle competenze.

Per agevolare i candidati con DSA o disabilità certificate che possono richiedere accorgimenti integrativi o alternativi a quelli in tabella, è possibile applicare due ulteriori misure:

- gestione speciale della prova di ascolto, in alternativa al tempo aggiuntivo nella misura del 30%. Chi somministra la prova
 - interrompe la riproduzione della registrazione prima di ogni parte per dare il tempo sufficiente per leggere le domande;
 - interrompe la riproduzione della registrazione in certi punti per dare il tempo sufficiente per scrivere le risposte;
- oltre al tempo aggiuntivo, o in alternativa, è possibile richiedere delle pause con supervisione. Questo significa che è possibile interrompere la prova e fare una pausa in un'altra stanza con supervisione. Questo potrebbe essere d'aiuto a chi ha difficoltà a concentrarsi per lunghi periodi. Il momento delle pause e la durata sono a discrezione dell'organizzazione della prova;
- se la scrittura risulta illeggibile, il candidato può richiedere l'assistenza di un copiatore che riprodurrà fedelmente (errori inclusi) il testo scritto che il candidato detterà (incluso la punteggiatura).

Per l'osservazione durante la prova, è stata sviluppata e conseguentemente utilizzata la seguente griglia.

Tabella 3. *Griglia di osservazione per le prove PLIDA*

Fase iniziale di organizzazione

	Studente 1	Studente 2
Ha difficoltà a gestire il materiale ricevuto		
Ha difficoltà nell'organizzazione del lavoro prima delle spiegazioni dei somministratori		
Ha bisogno che le istruzioni vengano ripetute più volte		
Altro:		
Altro:		

Rispondere: sempre, spesso, raramente, mai. Aggiungere eventualmente commenti.

<i>Inserire eventuali commenti liberi su qualsiasi elemento di osservazione</i>

Fase di svolgimento del test

	Studente 1	Studente 2
Mostra difficoltà a restare seduto/a		
Mostra difficoltà a restare concentrato a lungo		
Mostra variabilità nei tempi di esecuzione delle attività: a volte risulta frettoloso/a, a volte lento/a		
Mostra segnali di insofferenza e frustrazione (indicare in quale/i parte/i del test)		
Mostra di apprezzare la presenza dei somministratori come supporto psico-affettivo		
Mostra ansia		
Chiede aiuto per la comprensione dei task		
Riesce a gestire i tempi delle varie parti (indicare in quale/i parte/i del test incontra eventuali difficoltà)		
Utilizza agevolmente il pc fornito		
Utilizza il lettore digitale agevolmente (per studenti dislessici)		
Mostra di essere maggiormente a proprio agio con alcune parti del test e tecniche (indicare quali)		
Altro:		
Altro:		

Rispondere: Sempre, spesso, raramente, mai. Aggiungere eventualmente commenti.

5. LA SESSIONE DI SPERIMENTAZIONE

5.1. *I partecipanti*

Alla sperimentazione degli esami di certificazione del livello B2 hanno partecipato in totale 6 studenti e 5 studentesse iscritti/e al IV e V anno di corso e nati/e tra il 2003 e il 2005:

- 4 con disturbo specifico della lettura e dell'ortografia (studenti 2, 5, 6, 9);
- 2 con disturbo specifico della compitazione (studenti 3, 10);
- 1 con disturbo evolutivo specifico delle abilità scolastiche (studente 8);
- 1 con disturbo misto delle abilità scolastiche (studente 4);
- 1 con disturbo dell'attenzione, iperattività (studente 11);
- 2 con disturbo specifico delle abilità aritmetiche (studenti 1, 7).

Alla sperimentazione degli esami di certificazione del livello C1 hanno partecipato in totale 5 studenti (3 studenti e 2 studentesse), anch'essi iscritti al IV e V anno e nati/e tra il 2003 e il 2005:

- 1 con disturbo specifico della lettura e dell'ortografia (studente 4);
- 2 con disturbo specifico della compitazione (studenti 1, 3);
- 2 con disturbo dell'attenzione, iperattività (studenti 2, 5).

Gli esami si sono svolti in due giorni distinti presso la sede del *Liceo Vogelweide* di Bolzano⁴ e sono stati somministrati da due collaboratori esperti della PLIDA, che hanno svolto anche il ruolo di intervistatori delle prove orali. Le prove scritte si sono svolte nella mattinata, con una pausa di 15 minuti tra l'una e l'altra. Le prove orali si sono svolte nel pomeriggio.

Le misure adottate sono state le seguenti:

1. 30% in più del tempo per svolgere le prove (tutti);
2. fascicoli d'esame stampati in Open sans (font indicato per i dislessici);
3. PC a disposizione, con strumento di lettura ad alta voce (tutti tranne gli studenti 1 e 7 della sessione B2). Uno studente che ha partecipato a entrambe le sessioni (studente 11 della sessione B2, studente 5 della sessione C1) aveva fatto esplicita richiesta del PC in fase di iscrizione ma in sede d'esame ha rinunciato a utilizzarlo;
4. nella prova Ascoltare 30% del tempo in più per leggere gli *item* prima della riproduzione degli audio; aumento dell'intervallo tra la prima e la seconda riproduzione dei testi orali; aumento dell'intervallo tra una parte e l'altra della prova di ascolto;
5. pausa di 5 minuti a metà della prova Ascoltare;
6. nella prova Parlare 30% del tempo in più per leggere i testi input.

5.2. *I risultati dell'osservazione dei comportamenti*

Ai somministratori della sperimentazione è stata consegnata la griglia di osservazione, presentata al paragrafo precedente, con l'istruzione di compilarla indicando la frequenza (sempre, spesso, raramente o mai) dei comportamenti indicati.

Venendo a quanto osservato durante la somministrazione delle prove, riguardo alla fase iniziale di organizzazione i due somministratori sono stati concordi nell'indicare che:

- Nessuno dei candidati ha dimostrato difficoltà nel gestire il materiale ricevuto. Tuttavia, nella sessione B2, 6 studenti su 11 (1, 3, 5, 6, 9, 10) hanno separato i fogli del fascicolo *Leggere* che erano stati consegnati spillati, così da poter lavorare contemporaneamente sui testi e sui quesiti stampati su pagine diverse. Analogamente, nella sessione C1, 4 studenti su 5 hanno separato i fogli dei fascicoli ricevuti: lo studente 3 i fogli del fascicolo *Leggere*; lo studente 1 tutti i fascicoli (*Ascoltare*, *Leggere*, *Scrivere*); lo studente 4 i fascicoli *Leggere* e *Scrivere*; lo studente 2 il fascicolo *Scrivere*.
- Nessuno ha dimostrato difficoltà nell'organizzazione del lavoro prima delle spiegazioni dei somministratori.
- Quasi nessuno ha avuto bisogno che le istruzioni venissero ripetute più volte. Solo lo studente 2 del livello B2 e lo studente 2 del livello C1 hanno avuto bisogno che venissero chiarite le istruzioni per lo svolgimento del primo esercizio della prova *Ascoltare* (prima prova dell'esame).

Nella fase di svolgimento del test:

- Nessuno ha mostrato difficoltà a restare seduto/a, né a restare concentrato a lungo.
- Pochi studenti hanno mostrato variabilità nei tempi di esecuzione delle attività: lo studente 6 del livello B2 è stato più lento nella prova *Leggere*; lo studente 4 del livello C1 ha ripetuto due volte gli esercizi della prova *Leggere*; lo studente 1 dello stesso livello non ha sfruttato il secondo ascolto dei brani audio della prova *Ascoltare* per risolvere i quesiti ma ha usato il tempo per trascrivere le sue risposte sul foglio a lettura ottica (la cui compilazione è prevista alla fine della prova).
- Lo studente 9 del livello B2 ha mostrato segnali di insofferenza e frustrazione durante la prova *Scrivere*. In un primo momento sembrava voler rinunciare a svolgere la prova.
- Lo studente che ha sostenuto le prove di entrambi i livelli (identificato come studente 11 del livello B2, 5 del livello C1) ha mostrato di apprezzare la presenza dei somministratori come supporto psico-affettivo, cercandone costantemente il contatto visivo.
- Lo studente 2 del livello B2 ha mostrato ansia all'inizio della prova. Lo studente 3 del livello B2 ha mostrato ansia nella prova *Parlare*.
- Gli studenti 1 e 7 del B2 hanno chiesto aiuto per la comprensione dei task, con domande sul lessico della traccia della prova *Scrivere*.
- Tutti sono riusciti a gestire i tempi delle varie parti.
- Tutti gli studenti del B2 hanno utilizzato agevolmente il PC fornito (solo lo studente 9 ha chiesto aiuto all'inizio della prova per selezionare la lingua del lettore digitale). Ma degli 8 studenti che avevano a disposizione il PC, 3 non lo hanno mai usato (studenti 2, 6, 8) e gli studenti 3, 4, 5 e 9 raramente. Lo studente 4 lo ha usato solo per leggere i testi della prova *Leggere*. Nell'esame C1 solo lo studente 4 ha usato il PC, e sempre agevolmente.
- Gli studenti 5, 6 e 9 del B2 hanno utilizzato il lettore digitale agevolmente: 5 e 6 lo hanno usato solo nella prova *Ascoltare*; 9 lo ha usato solo nella prova *Leggere*. Lo studente 4 del C1 ha usato agevolmente il lettore digitale in tutte le parti della prova.
- Lo studente 9 del B2 ha mostrato di essere maggiormente a proprio agio con le prove *Ascoltare* e *Leggere*.
- Nella sessione B2 il 30% di tempo in più per lo svolgimento delle prove è stato sfruttato da tutti ad eccezione degli studenti 4, 8 e 11 che hanno finito 30 minuti prima la prova

Leggere. Nella sessione C1, invece, il tempo aggiuntivo non è servito a nessuno, tranne che allo studente 4. Va notato che lo studente 4 ha svolto entrambe le prove e che ha utilizzato il tempo aggiuntivo solo nella sessione C1, riferendo ai somministratori di aver avuto necessità di più tempo in questa prova per inferire dal contesto il significato di molte parole non conosciute.

- Nella sessione B2 la pausa tra la seconda e terza parte della prova *Parlare* è stata utilizzata per riposarsi solo dagli studenti 10 e 11, mentre gli altri sono rimasti seduti per compilare il foglio risposte. Nella sessione C1 tutti gli studenti hanno usato la pausa per riposarsi, gli studenti 1 e 5 alzandosi dal banco.

6. ANALISI E DISCUSSIONE DEI DATI

6.1. Riflessioni sui risultati

I dati raccolti forniscono la possibilità di sviluppare diverse riflessioni. In questo contributo, li interrogheremo attraverso due ordini di strumenti statistici: il primo riguarda la media dei risultati ottenuti a livello mondiale dall'insieme dei candidati alla sessione di marzo 2023 della PLIDA di livello B2 e C1. A questo insieme, costituito inizialmente da 1088 candidati per il livello B2 e 372 per il C1, sono stati tolti gli studenti delle scuole di Bolzano, i candidati che ripetevano una o più parti non superate in precedenza e i soggetti di età inferiore ai 20 anni. Il criterio dell'età è stato adottato in funzione di riflessioni sull'impatto del fattore culturale e cognitivo nei processi di comprensione legati ai testi e agli *item* della sezione dedicata alla lettura e all'ascolto della prova somministrata; il criterio dell'esclusione dei candidati ripetenti, per creare uguali condizioni di partenza con il gruppo di studenti con DSA, che sostenevano tutti le prove per la prima volta. In seguito, si è proceduto a un raffronto con gli esiti del gruppo sperimentale.

Una fase successiva dello studio si è basata su alcuni elementi dell'*item analysis* svolta sull'intero corpus di prove a disposizione.

Nella tabella seguente sono evidenziati i risultati ottenuti dai tre gruppi a confronto. Anche per il gruppo degli studenti delle scuole di Bolzano sono stati presi in considerazione solo i soggetti che sostenevano l'esame per la prima volta, su tutt'e quattro le abilità.

Tabella 4. *Risultati a confronto*

Percentuali di successo esami B2

		Numero partecipanti	Ascoltare	Leggere	Scrivere	Parlare
A	Sessione	737	82%	78%	72%	79%
B	Scuole	121	74%	76%	87%	78%
C	DSA	11	64%	45%	73%	82%

Percentuali di successo esami C1

		Numero partecipanti	Ascoltare	Leggere	Scrivere	Parlare
A	Sessione	204	74%	78%	72%	79%
B	Scuole	121	78%	87%	88%	91%
C	DSA	5	60%	60%	100%	80%

I dati a disposizione rimandano a una composizione asimmetrica del campione. Da un lato ci si trova di fronte a un gruppo, quello degli studenti delle scuole di Bolzano, coeso: i candidati hanno un'età e un background socio-culturale e linguistico affini. Dall'altro lato, i candidati al test a livello globale presentano condizioni maggiormente diversificate. La significatività della comparazione crediamo stia nella rappresentazione globale che se ne può evincere. Non sono stati effettuati confronti con altri gruppi coesi per età e background socio-culturale e linguistico perché ritenuti non pertinenti per lo studio in atto.

Dall'osservazione dei dati emerge che la performance nel livello B2 sia del gruppo globale che del gruppo delle scuole è sostanzialmente allineata tanto nelle abilità ricettive, quanto in quelle produttive, con una sostanziale sovrapposizione nel parlato e nella lettura. Tale analogia si evidenzia anche nel livello C1 con una sostanziale sovrapposizione anche nell'ascolto e nel parlato.

Un confronto tra il gruppo sperimentale composto da studenti con DSA e gli altri due gruppi sembrerebbe sottolineare una differenza particolarmente marcata nella performance nelle abilità ricettive, con un livello di insuccesso del 55% nella lettura da parte del gruppo sperimentale B2 e del 36% nell'ascolto, a fronte rispettivamente del 24% e del 26% del gruppo di controllo. Crediamo che questo dato possa far supporre una tendenza che vale la pena approfondire.

La consistenza numerica del campione è però troppo esigua per potersi avventurare in un'analisi comparativa e per poter considerare come generalizzabili i dati restituiti dalle percentuali di successo. Tuttavia, la novità dello studio che porta per la prima volta nell'ambito delle certificazioni di lingua italiana a isolare e analizzare le performance di candidati con DSA, confrontandole con gruppi di controllo, giustifica il tentativo di individuare le ragioni di quella che ora è poco più di una ipotesi di indagine, ma che in futuro potrebbe rivelarsi una tendenza significativa. L'approfondimento potrebbe avvenire attraverso l'analisi delle risposte degli studenti con DSA ai singoli quesiti e prendendo in considerazione alcuni elementi di tipo testuale che stimolano riflessioni sul ruolo della domanda cognitiva e della *background knowledge*.

Un primo tentativo di analisi in questo senso viene presentato nel paragrafo seguente.

6.2. Spunti dall'item analysis

Nella sessione sperimentale con studenti DSA le prove Ascoltare sono state superate da 7 studenti su 11 nel livello B2 e da 3 studenti su 5 nel livello C1.

Nella prova B2 gli *item* che registrano un numero di risposte corrette inferiore alla metà dei partecipanti (5 o meno di cinque) sono i seguenti: 5, 6, 15, 16, 17 e 18.

Gli *item* 5 e 6 sono 2 quesiti a scelta multipla che fanno parte della stessa attività e che risultano relativamente difficili anche per gli studenti del gruppo di controllo rappresentato dagli studenti delle Scuole (percentuali di facilità di 0,56% e 0,55%). Significativamente, ad essi rispondono in maniera corretta solo gli studenti con DSA che ottengono un punteggio complessivo sufficiente a superare la prova Ascoltare nel suo complesso.

Gli *item* 15, 16, 17 e 18 costituiscono insieme la quarta parte della prova, in cui i candidati ascoltano due testi espositivi monologici, di argomento scientifico divulgativo, di 2 minuti circa e devono completare 2 frasi (rispettivamente 15 e 16 per il primo brano, 17 e 18 per il secondo) individuando la soluzione tra 6 possibilità. Si tratta della parte più complessa della prova, in cui ciascun *item* corretto assegna un punteggio di 2,25 punti invece di 1,5 punti come gli altri *item*. La complessità è data dal tipo di testo di cui si vuole verificare la comprensione, ma anche dalla tecnica utilizzata che propone al candidato 6 opzioni possibili per ogni *item*.

Mentre gli *item* 17 e 18 risultano difficili solo per gli studenti con DSA che non superano la prova, per gli *item* 15 e 16 si registrano risposte errate anche da parte degli studenti con performance migliori. Gli stessi risultano difficili anche per il gruppo di controllo degli studenti delle Scuole, con percentuali di facilità relativamente basse, inferiori al 0,60%; l'*item* 15 risulta difficile anche per i candidati dell'intera sessione (facilità pari a 0,42%).

Il testo dell'esercizio in questione era di argomento storico ed era contestualizzato dal titolo "Lo storico Alessandro Barbero racconta le origini del Regno di Prussia".

L'*item* 15 è il seguente:

Gli imperatori _____

- A) catturarono un orso in mezzo alle paludi.
- B) provenivano dal sud della Germania.
- C) erano parenti di Federico il Grande.
- D) erano leali nei confronti degli Hohenzollern.
- E) conquistarono in battaglia la città di Berlino.
- F) promossero la conquista dell'Europa orientale.

La porzione di testo su cui insiste l'*item* è:

All'inizio c'era una famiglia di principi tedeschi, gli Hohenzollern. È una famiglia che compare sulla scena nel Medioevo, non particolarmente antica e neanche particolarmente illustre, che erano originari della Germania meridionale, ma nel corso del Medioevo qualcuno di loro cominciò a cercare fortuna da un'altra parte, lontano, verso Est, come molti altri tedeschi. D'altra parte, bisogna sapere che nel Medioevo per i tedeschi l'Est era un po' come il vecchio West per gli americani: una frontiera, un paese immenso da colonizzare. All'inizio questa spinta – questa immigrazione, anzi – verso l'Europa orientale, fu diretta dagli imperatori del Sacro Romano Impero: sotto la loro guida venne creato un paese nuovo, la Marca di Brandeburgo, e fu creata anche una nuova città che si chiamò Berlino.

L'informazione rilevante per risolvere correttamente l'*item* (opzione F: "promossero la conquista dell'Europa orientale") è contenuta nella stringa di testo "questa spinta – questa immigrazione, anzi – verso l'Europa orientale, fu diretta dagli imperatori del Sacro Romano Impero". Pesa, sul basso numero di risposte corrette, il fatto che il distrattore E ("conquistarono in battaglia la città Berlino") risulti particolarmente attraente per sia gli studenti DSA sia per il gruppo a loro coetaneo delle Scuole, mentre il distrattore B ("provenivano dal sud della Germania") risulta attraente per i candidati della sessione.

Avanziamo l'ipotesi che la scelta dell'opzione E sia determinata da una strategia di decodifica del testo che, una volta individuato il referente di cui si parla nell'*item* ("gli imperatori del Sacro Romano Impero"), spinge a cercare la risposta nella porzione di testo che viene immediatamente dopo ("fu creata anche una nuova città che si chiamò Berlino") piuttosto che recuperare l'informazione precedente "questa spinta - questa immigrazione, anzi - verso l'Europa orientale, fu diretta dagli imperatori del Sacro Romano Impero". Inoltre, sulla scelta dell'opzione E ha sicuramente influito la presenza della parola "Berlino", sia nel distrattore sia nel testo orale, mentre per scegliere l'opzione corretta era necessario operare un'inferenza: "gli imperatori del Sacro Romano Impero danno inizio all'immigrazione verso Est con l'intento di colonizzare, ovvero conquistare quelle terre". Un processo inferenziale di questo tipo comporta un aumento della richiesta cognitiva.

Per la scelta dell'opzione B ("provenivano dal sud della Germania"), invece, ipotizziamo che ci sia stato un problema di corretta identificazione del referente "gli imperatori del Sacro Romano Impero". Questi non sono stati riconosciuti come soggetti diversi dagli Hohenzollern, per cui agli "imperatori" è stata attribuita la predicazione "provenivano dalla Germania meridionale" di inizio testo. La confusione dei due referenti può essere stata determinata da un problema di *background knowledge*.

Nella prova Ascoltare del livello C1 gli *item* che pongono difficoltà agli studenti con DSA sono l'*item* 10 e l'*item* 17. Mentre l'*item* 17, compreso nella quarta parte della prova (che ha la stessa struttura della parte corrispondente del livello B2 analizzata sopra), risulta difficile anche per gli studenti delle scuole, l'*item* 10 viene invece risolto senza problemi dagli altri due gruppi di controllo. Si tratta di un quesito a scelta multipla che testa la comprensione di una intervista televisiva di 2 minuti circa sul "lato pop di Galileo Galilei" (questa è la contestualizzazione data al candidato nel titolo dell'esercizio).

L'*item* è:

Nei suoi scritti Galileo aveva uno stile

- A) informale.
- B) elegante.
- C) essenziale.

La porzione di testo su cui insiste l'*item* è:

Perché poi Galileo scriveva e polemizzava e usava delle immagini molto così... ehm... forti e divertenti, e il suo linguaggio era... era... era molto... molto colloquiale, molto spesso. E tutta questa cosa appunto, di averlo ingabbiato invece in una figura un po' così... un po' più seriosa, forse ha un po' impedito che diventasse un'icona pop, un... un personaggio della scienza che uno potrebbe immaginarsi su delle magliette ecco. Questa cosa non è successa.

Per rispondere in maniera corretta era necessario riconoscere in questo contesto la sinonimia tra "linguaggio informale" e "linguaggio colloquiale", come fa solo uno studente, mentre gli altri quattro si orientano equamente verso i due distrattori. In questo caso si può forse ipotizzare una mancanza di *background knowledge* relativa alla dimensione sociolinguistica del linguaggio.

Nella prova Leggere B2 (15 *item*) gli *item* problematici per gli studenti con DSA sono l'1 e il 9. Mentre l'*item* 9 registra valori di facilità nella norma (range tra 0,60% e 0,80%) nelle prove dei due gruppi di controllo, l'*item* 1 risulta difficile anche per gli studenti delle scuole.

L'*item* 1 è un quesito a scelta multipla con 4 opzioni, riferito a un brano tratto da un'opera di lettura contemporanea, contestualizzato in questo caso dal titolo "Una notte difficile". La prima porzione, quella su cui insiste l'*item* 1 è la seguente:

Il vicequestore aggiunto Giovanna Guarrasi, detta Vanina, aprì gli occhi di colpo balzando sul materasso, il cuore in gola per il boato che aveva appena scosso i vetri. Un'esplosione, aveva temuto il suo inconscio di sbirra, prima che il cielo fuori dalla finestra si illuminasse a giorno, producendo un tuono ancora più roboante di quello che l'aveva appena svegliata e mostrando la cortina d'acqua che si stava abbattendo sul paese di Santo Stefano.

Il quesito proposto è:

Giovanna si sveglia a causa di

- A) un temporale.
- B) un incubo.
- C) un incendio.
- D) un terremoto.

All'opzione A, che è quella corretta e a cui rimandano gli elementi del boato che scuote i vetri, del cielo che si illumina a giorno, del tuono e della cortina d'acqua, il 19% degli studenti del gruppo B e 5 degli 11 studenti con DSA preferiscono l'opzione D "un terremoto". È possibile che gli studenti siano stati tratti in inganno dalla presenza, più avanti nel testo, della parola Etna, e che abbiano interpretato così i fenomeni metereologici estremi citati a inizio testo come le conseguenze di un'eruzione del vulcano. La *background knowledge* potrebbe in questo caso aver avuto un effetto negativo, innescando un'inferenza errata.

Nella prova Leggere C1 risultano difficili solo per gli studenti con DSA gli *item* 10 e 11 che, insieme all'*item* 9, costituiscono i quesiti della terza parte della prova, nella quale i candidati devono ripristinare la coerenza e la coesione di un testo tramite un'attività di incastro di paragrafi. Dal punto di vista dei meccanismi cognitivi è probabilmente una delle parti più impegnative della prova, richiedendo una lettura prima globale, poi analitica del testo, la conoscenza di meccanismi di coesione testuale e la capacità di riconoscere la progressione logica del discorso.

I valori di facilità degli stessi *item* registrati dalle prove dei gruppi di controllo sono medio-alti; tuttavia, ogni *item* risulta, stando alle percentuali, più impegnativo per gli studenti delle Scuole che per i candidati dell'intera sessione.

7. CONCLUSIONI

7.1. Riflessioni sulle modalità di svolgimento della prova sperimentale

Nel paragrafo 5.2 abbiamo ritenuto opportuno evidenziare i risultati dell'osservazione, dai quali, tuttavia, a causa del numero esiguo dei candidati, così come del comportamento privo di tratti inusuali, non è possibile cogliere alcuna tendenza di distinzione dalle situazioni standard. Crediamo, però, che la possibilità di confronto nel tempo con test sostenuti da

candidati con DSA in altre sessioni e l'abbinamento dei comportamenti con la tipologia di disturbo potrebbero risultare di una certa utilità per la ricerca.

Un elemento merita, da ultimo, di essere messo in evidenza: le misure compensative e le scelte riguardo alle modalità di svolgimento delle prove (tempo aggiuntivo, pausa nell'ascolto ecc.) hanno permesso di personalizzare maggiormente il percorso, pur senza derogare alle regole previste, mantenendo sotto controllo il livello d'ansia generato da fattori esogeni al confronto diretto con il testo e consentendo a ciascuno di gestire il tempo a disposizione. La gestione del tempo rappresenta solitamente uno dei maggiori fattori ansiogeni, non rivelatosi tale nella sperimentazione descritta, in quanto il tempo è risultato sufficiente e ben gestito lungo tutto il corso della somministrazione.

Da un punto di vista della riflessione che sottende al lavoro di ricerca e degli obiettivi di aggiornamento della certificazione PLIDA è, in conclusione, possibile confermare la validità delle scelte effettuate al fine di realizzare condizioni di svolgimento del test maggiormente inclusive. Si riscontra, tuttavia, la necessità di proseguire l'applicazione delle scelte descritte ad altri contesti in nuove sessioni di test con l'impegno a monitorarne l'andamento in maniera costante e con il fine di realizzare una banca dati specifica in grado di permettere studi longitudinali.

7.2. Riflessioni a partire dai risultati nel test

Per quanto concerne i risultati del test, l'*item analysis*, di cui si è data un'esemplificazione nel paragrafo 6.2, ha messo in luce risposte simili tra gli studenti delle Scuole, con DSA e senza DSA, ai problemi posti dalle attività più complesse delle prove di ricezione. Nella fattispecie si sono evidenziati meccanismi di semplificazione nelle strategie di *problem solving*, difficoltà nell'attivazione di processi inferenziali e scarsa competenza testuale; tutte manifestazioni che rimandano probabilmente a un fattore condizionante collegato all'età, il cui impatto, insieme a quello della *background knowledge*, andrebbe maggiormente indagato in un'ottica di confronto con i comportamenti della popolazione globale dei candidati alla certificazione PLIDA e che sarà oggetto di ulteriori studi.

RIFERIMENTI BIBLIOGRAFICI

- Abedi J. (2012), "Validity issues in designing accommodations for English language learners", in Fulcher G., Davidson F. (eds.), *The Routledge handbook of language testing*, Routledge, Abingdon, pp. 62-76.
- Acosta B. D., Rivera C., Shafer Willner L. (2008), *Best Practices in State Assessment Policies for Accommodating English Language Learners: A Delphi Study*, The George Washington University Center for Equity and Excellence in Education, Arlington (VA), pp. 1-29.
- Acosta B. D., Rivera C., Shafer Willner L. (2009), "Ensuring Accommodations Used in Content Assessments Are Responsive to English-Language Learners", in *The Reading Teacher*, 62, 8, pp. 696-698.
- AERA, APA, NCME (2014), *The Standards for Educational and Psychological Testing*, AERA, American Psychological Association (APA), Washington, DC.

- Alderson J. C. (1991), *Language Testing in the 1990s*, Modern English Publication in association with the British Council, London.
- Alderson J. C. (1995), *Language Test Construction and Evaluation*, Cambridge University Press, Cambridge.
- Alderson J. C. (2000), *Assessing Reading*, Cambridge University Press, Cambridge.
- Alderson J. D., Banerjee J. (2001), "Language testing and assessment" (Part 1), in *Language Teaching*, 34, pp. 213-236.
- Alderson J. D., Banerjee J. (2002), "Language testing and assessment" (Part 2), in *Language Teaching*, 35, pp. 79-113.
- ALTE (2001), *Manual for Language Testing Development and Examining*, Council of Europe, Strasbourg.
- ALTE (2005), *ALTE materials for the guidance of test item writers*:
www.alte.org/attachments/files/item_writer_guidelines.pdf.
- Aydin S. (2009), "Test anxiety among foreign language learners: A review of literature", in *Journal of language and linguistic studies*, 5, 1, pp. 127-137.
- Aydin S. et al. (2021), "The Foreign Language Test Anxiety Scale: Tests of Reliability and Validity", in *Eurasian Journal of Language Teaching and Linguistic Studies*, 1, 2, pp. 96-109.
- Bachman L. F., Eignor D. R. (1997), "Recent advances in quantitative test analysis", in Beckham S. (2015), *Effects of Linguistic Modification Accommodation on High School English Language Learners' Academic Performance*, Doctoral dissertation, Nova Southeastern University, retrieved from NSUWorks, Abraham S. Fischler College of Education, http://nsuworks.nova.edu/fse_etd/3.
- Bachman L. F., Palmer A. S. (2010), *Language Assessment in Practice*, Oxford University Press, Oxford.
- Barni M. (2000), "La verifica e la valutazione", in De Marco A. (ed.), *Manuale di glottodidattica*, Carocci, Roma, pp.155-174.
- Barni M. (2023), *Valutare le competenze nella L2*, Carocci, Roma.
- Blau A. (1955), "A unitary hypothesis of emotion: Anxiety, Emotions of Displeasure, and Affective Disorders", in *Psychoanalytic Quarterly*, 24, pp. 75-103.
- Chapelle C. A. (1999), "Validity in language assessment", in *Annual Review of Applied Linguistics*, 19, 1, pp. 254-272.
- Chapelle C. A. (2012), "Conceptions of validity", in Fulcher G., Davidson F. (eds.), *The Routledge Handbook of Language Testing*, Routledge, Abingdon, pp. 21-33.
- Chapelle C. A., Lee H. W. (2021), "Conceptions of validity", in Fulcher G., Harding L. (eds.), *The Routledge handbook of language testing*, Routledge, Abingdon, pp. 17-31.
- Cheng L., Zheng, Y. (2021), "Measuring anxiety", in Winke P., Brunfaut T. (eds.), *The Routledge Handbook of Second Language Acquisition and Language Testing*, Routledge, Abingdon, pp. 177-186.
- Clapham C., Corson D. (eds.), *Encyclopedia of language and education*, vol. 7: *Language testing and assessment*, Kluwer Academic Publishers, Dordrecht, pp. 227-242.
- Cumming A. (1996), "Introduction: The concept of validation in language testing", in Cumming A., Berwick R. (eds.), *Validation in language testing*, Multilingual Matters Clevedon, pp. 1-14.
- Dal Negro S. (2017), "Bilinguismo asimmetrico in Alto Adige: lo spazio sociolinguistico dell'italiano", in Bombi R. (a cura di), *Nuovi spazi comunicativi per l'italiano nel mondo*, Forum Editrice Universitaria Udinese, Udine, pp. 59-67.

- Daloiso M. (2017), *Supporting learners with dyslexia in the ELT classroom*, Oxford University Press, Oxford.
- Douglas D. (2000), *Assessing Languages for Specific Purposes*, Cambridge University Press, Cambridge.
- Hamp-Lyons L., Lynch B. K. (1998), "Perspectives on validity: A historical analysis of language testing conference abstracts", in Kunnan A. J. (ed.), *Validation in language assessment: selected papers from the 17th Language Testing Research Colloquium, Long Beach*, Erlbaum, Mahwah (NJ), pp. 253-276.
- Hamp-Lyons L. (2001), "Ethics, fairness(es) and developments in language testing", in *Experimenting with Uncertainty: Essays in Honour of Alan Davies. Studies in Language Testing*, 11, pp. 222-227.
- Hamp-Lyons L. (1997a), "Washback, impact and validity: Ethical concerns", in *Language testing*, 14, 3, pp. 295-303.
- Hamp-Lyons L. (1997b), "Ethics in Language Testing", in Clapham C., Corson D. (eds.), *The Encyclopedia of Language and Education*, Volume 7: *Language Testing and Assessment*, Kluwer, Dordrecht, pp. 323-333.
- Horwitz E. K., Horwitz M. B., Cope, J. (1986), "Foreign language classroom anxiety", in *Modern Language Journal*, 70, 2, pp. 125-132.
- Im-Bolter N., Cohen N. J. (2007), "Language impairment and psychiatric comorbidities", in *Pediatric Clinics of North America*, 54, 3, pp. 525-542.
- Jones N. (2012), "Reliability and dependability", in Fulcher G., Davidson F. (eds.), *The Routledge Handbook of Language Testing*, Routledge, Abingdon, pp. 350-363.
- Kane M. T. (2001), "Current concerns in validity theory", in *Journal of Educational Measurement*, 38, pp. 319-342.
- Kunnan A. J. (ed.) (1998), *Validation in language assessment: Selected papers from the 17th Language Testing Research Colloquium, Long Beach*, Erlbaum, Mahwah (NJ).
- Kunnan A. J. (1999), "Recent developments in language testing", in *Annual Review of Applied Linguistics*, 19, 1, pp. 235-253.
- Kunnan A. J. (2004), "Test fairness", in Milanovic M., Weir C. J. (eds.), *European language testing in a global context: Proceedings of the ALTE Barcelona Conference*, Cambridge University Press, Cambridge, pp. 27-48.
- Kunnan A. J. (2014), "Fairness and justice in language assessment", in Kunnan A. J. (ed.), *The Companion to Language Assessment*, Wiley-Blackwell, Boston (MA), pp. 1098-1114.
- Lado R. (1961), *Language testing: The construction and use of foreign language tests*, McGraw Hill, New York.
- Masillo P., Gelo A., Nasimi A., Peri G. (2021), "L'italiano nelle scuole di lingua tedesca: un progetto per la rilevazione delle competenze linguistiche degli studenti nella Provincia Autonoma di Bolzano", in Favilla M. E., Machetti S. (a cura di), *Lingue in contatto e linguistica applicata: individui e società*, Studi AItLA, 13, Officinaventuno, Milano, pp. 215-238: http://www.aitla.it/images/pdf/StudiAItLA13/014Masillo_et_al.pdf.
- McNamara T. (1996), *Measuring second language performance*, Longman, London.
- Messick S. (1989), "Validity", in Linn R. L. (ed.), *Educational measurement*, American Council on education and Macmillan, New York, pp. 13-104.
- Mezzadri M. (2017), *Testing academic language proficiency*, Cambridge Scholars Publishing, Newcastle upon Tyne.
- Scovel T. (1978), "The effect of affect on foreign language learning: A review of the anxiety research", in *Language Learning*, 28, pp. 129-142.

- Shohamy E., Gordon C. M., Kraemer R. (1992), “The effects of rater’s background and training on the reliability of direct writing tests”, in *Modern Language Journal*, 76, 1, pp. 27-33.
- Teimouri Y., Goetze J., Plonsky L. (2019), “Second language anxiety and achievement: A meta-analysis”, in *Studies in Second Language Acquisition*, 41, pp. 363-387.
- Vettori C., Abel A. (2021), “La scuola come unico contesto di apprendimento della seconda lingua? Il caso dell’Alto Adige”, in Favilla M. E., Machetti S. (a cura di), *Lingue in contatto e linguistica applicata: individui e società*, Studi AItLA, 13, Officinaventuno, Milano, pp. 239-254: http://www.aitla.it/images/pdf/StudiAItLA13/015Vettori_Abel.pdf.
- Vettori C., Abel A. (a cura di) (2017), *KOLIPSI II. Gli studenti altoatesini e la seconda lingua: indagine linguistica e psicosociale*, Accademia Europea di Bolzano, Bolzano.
- Weir C. J. (2005), *Language Testing and Validation*, Palgrave Macmillan, Basingstoke.
- Yan J. X., Liang J. (2022), “Foreign language anxiety and dependency distance in English–Chinese interpretation classrooms”, in *Frontiers in Psychology*, 13: <https://doi.org/10.3389/fpsyg.2022.952664>.

