



UNIVERSITÀ DI NAPOLI L'ORIENTALE

ANNALI

*del Dipartimento di Studi
Letterari, Linguistici e Comparati
Sezione linguistica*

AION-L

N.S. 11

2022



UniorPress

ANNALI

*del Dipartimento di Studi
Letterari, Linguistici e Comparati
Sezione Linguistica*

AION-L

N.S. 11
2022



Direttore/Editor-in-chief: *Alberto Manco*

Comitato scientifico/Scientific committee: *Ignasi-Xavier Adiego Lajara, Françoise Bader, Annalisa Baicchi, Philip Baldi, Giuliano Bernini, Carlo Consani, Pierluigi Cuzzolin, Paolo Di Giovine, Norbert Dittmar, Annarita Felici, Laura Gavioli, Nicola Grandi, Marco Mancini, Andrea Moro, Vincenzo Orioles, Paolo Poccetti, Diego Poli, Ignazio Putzu, Giovanna Rocca, Velizar Sadovski, Domenico Silvestri, Francisco Villar*

Comitato di redazione/Editorial board: *Anna Cardinaletti, Mario Cardona, Valeria Caruso, Marina Castagneto, Francesca Chiusaroli, Anna De Meo, Lucia di Pace, Francesca Dragotto, Elena Favilla, Gloria Gagliardi, Franco Lorenzi, Sabrina Machetti, Alberto Manco, Antonietta Mara, Johanna Monti, Andrea Nuti, Rossella Pannain, Lorenzo Spreafico, Massimo Vai, Iride Valenti*

Segreteria di redazione/Editorial assistant: *Valeria Caruso*
e-mail: segreteriaion@unior.it

Annali-Sezione Linguistica, c/o *Alberto Manco, Università di Napoli L'Orientale, Dipartimento di Studi Letterari, Linguistici e Comparati, Palazzo Santa Maria Porta Coeli, Via Duomo 219, 80138 Napoli - albertomanco@unior.it*

ISSN 2281-6585

Registrazione presso il Tribunale di Napoli n. 2901 del 9-1-1980

web: <http://www.serena.unina.it/index.php/aionlin/index>



This work is licensed under a Creative Commons Attribution 4.0 International License

UniorPress, Via Nuova Marina 59 - 80133 Napoli

Per la redazione delle proposte i collaboratori sono invitati ad attenersi con cura alle "norme" disponibili nel sito della rivista.

Le proposte di pubblicazione inviate alla rivista vengono valutate da revisori anonimi. A tal fine una loro copia dev'essere priva di qualunque riferimento all'autore.



UNIVERSITÀ DI NAPOLI L'ORIENTALE

ANNALI

*del Dipartimento di Studi
Letterari, Linguistici e Comparati
Sezione Linguistica*

AION-L

N.S. 11

2022



UniorPress

PROPRIETÀ RISERVATA

INDICE

LETTERE APERTE, DISCUSSIONI, PROPOSTE

Problemi e prospettive di ricerca, convegni e tavole rotonde, notizie e suggerimenti

- A. M. THORNTON, *Genere e igiene verbale: l'uso di forme con ə in italiano* 11

ARTICOLI, NOTE, SAGGI

Analisi linguistiche di testi arcaici, riflessioni su aspetti e problemi linguistici del mondo antico, linee e momenti di preistoria e protostoria linguistica

- R. BATISTI, *L'arroganza e l'ingiuria: su alcune forme greche in στεμφ- e στε(μ)β- e sul problema della 'deaspirazione postnasale'* 57
- M. COSENTINO, *Sprachmischung e civilisation mixte nella stratificazione toponomastica di un abitato calabrese: Rota Greca* 79
- F. COTUGNO, *-d / -t alternation in the Vindolanda corpus: insights from Octavius' letter* 105
- G. SAMO, F. A. URSINI, G. CARACCILO, *Quantifying formulaic syntax: a quantitative and computational study on temporal complements in Latin subject relatives extracted from the Epigraphic collection of the Catacombs in Chiusi* 129

Ricerche e problemi linguistici di ambito teorico e applicato

- V. COLONNA, A. ROMANO, *La prosodia di Seamus Heaney: un approccio fonetico alle sue letture* 155
- M. FRONTERA, *Tracce di variazione nei dialetti ereditari di parlanti italo-argentini* 181
- M. MAFFIA, V. BOCCIA, A. DE MEO, *Apprendenti vulnerabili di italiano L2 e valutatori (in)esperti: uno studio sull'Adeguatezza Funzionale* 229
- F. MORLEO, *A Portuguese interjection and its process of pragmaticalization: Pá* 253

BIBLIOGRAFIE, RECENSIONI, RASSEGNE

- IGNAZIO MAURO MIRTO, *Grammatica, didattica linguistica, tecniche di scoperta*, Pisa, ETS, 2021, 93 pp. 281
- NUNZIO LA FAUCI, *Cinema e parole*, Pisa, ETS, 2022, 138 pp. 287
- SILVIA PIERONI, *Lezioni di sintassi*, Ospedaletto-Pisa, Pacini editore, 2022, 102 pp. 293
- ANDREA MARINI, *Che cos'è la psicolinguistica*, Roma, Carocci, 2021, 146 pp. 297

MARTA MAFFIA – VITTORIA BOCCIA – ANNA DE MEO*

Apprendenti vulnerabili di italiano L2 e valutatori (in)esperti: uno studio sull’Adeguatezza Funzionale

*Vulnerable learners of L2 Italian and (non)expert raters:
A study on Functional Adequacy*

Abstract

Lo studio intende verificare l’applicabilità della scala sviluppata per l’Adeguatezza Funzionale nella valutazione di testi scritti in italiano L2 di apprendenti vulnerabili, richiedenti asilo debolmente scolarizzati, mettendo a confronto i giudizi di valutatori con diversi livelli di esperienza. I punteggi ottenuti sono stati comparati con quelli attribuiti dagli stessi valutatori attraverso l’uso di una seconda scala, maggiormente centrata sull’accuratezza formale.

Parole chiave: Adeguatezza Funzionale, apprendenti vulnerabili, italiano L2, valutazione, scrittura

The study aims to verify the applicability of a Functional Adequacy-based rating scale in the evaluation of written texts produced in L2 Italian by vulnerable learners, all of them being asylum seekers and having a low level of education. For this purpose, raters with different degrees of teaching experience were involved and their judgments were analyzed and compared. The results thus obtained were then compared with those achieved by the same raters when applying a second rating scale focusing on formal accuracy.

Keywords: Functional Adequacy, vulnerable learners, L2 Italian, assessment, writing task

1. Introduzione

La formazione linguistica di immigrati vulnerabili, spesso debolmente alfabetizzati, è un compito divenuto centrale negli ultimi anni, data la tipologia di ospiti della rete territoriale italiana che realizza progetti di accoglienza, protezione, integrazione in favore di richiedenti e titolari di protezione internazionale.

* Marta Maffia, Vittoria Boccia, Anna De Meo, Università di Napoli L’Orientale, mmaffia@unior.it; vboccia@unior.it; ademeo@unior.it.

Nel 2018, ad esempio, il 12% dei 41.113 beneficiari dei centri del circuito SPRAR (Sistema di Protezione per Richiedenti Asilo e Rifugiati) in Italia era composto da analfabeti e il 63% da individui con una breve storia di scolarizzazione nel paese d'origine (Cittalia, Fondazione Anci, Ministero dell'Interno 2019). Nel 2020, dei circa 15mila beneficiari iscritti *ex novo* nella rete SAI (Sistema Accoglienza Integrazione) che hanno frequentato con continuità almeno un corso d'italiano, il 20,3% era inserito in un corso di pre-alfabetizzazione (Cittalia, Fondazione Anci, Ministero dell'Interno 2021).

La conoscenza della lingua italiana, in questo contesto, è presupposto per l'inclusione nelle comunità locali, per l'acquisizione di un sentimento di appartenenza comunitaria e per l'esercizio della cittadinanza attiva (Ibidem 2021). Tuttavia, la gestione dei percorsi formativi è spesso affidata a volontari o a operatori privi di una competenza glottodidattica adeguata (Nitti 2018).

In considerazione della vulnerabilità degli apprendenti, della frequente limitata esperienza professionale dei docenti volontari e della necessità di prevedere momenti di verifica e valutazione delle competenze acquisite, emerge la necessità di avere a disposizione strumenti gestibili anche da docenti non esperti.

Particolarmente significativa in questo senso appare la proposta di Ineke Vedder e colleghi (Kuiken & Vedder 2017; 2018; Vedder 2016) di una scala di valutazione sviluppata per l'Adeguatezza Funzionale (AF) che mira a superare i limiti dei consueti strumenti valutativi basati sugli indici di complessità, accuratezza e fluenza (CAF – Housen, Kuiken & Vedder 2012), integrando la dimensione linguistica con quella comunicativo/funzionale. Il costrutto multidimensionale dell'Adeguatezza Funzionale trova originaria ispirazione nelle massime conversazionali di Grice (1975) e si basa sull'assunto che la valutazione dell'adeguatezza di qualsiasi messaggio linguistico non possa prescindere dal tipo di compito (o *task*) che il parlante/scrivente sta eseguendo. La scala di valutazione dell'AF si propone quindi come uno strumento in grado di "fotografare" la (in)felice esecuzione di un testo scritto o orale, attraverso descrittori che siano indipendenti dal grado di complessità sintattica e lessicale (C), di accuratezza formale (A) e dalla fluenza (F) dei testi, che possano essere oggettivi e quantificabili e che siano utilizzabili sia per le lingue straniere/secondo sia per le lingue materne da parte di valutatori esperti e non esperti.

L'uso della scala anche da parte di non esperti sarebbe, secondo Kuiken e Vedder (2014), non solo possibile ma, anzi, auspicabile, al fine di evitare *bias* e giudizi troppo severi o troppo clementi, condizionati da pregresse esperienze di valutazione o didattica. La possibilità di coinvolgimento di parlanti nativi *naive*, inoltre, confermerebbe la capacità di questo strumento valutativo di rendere conto dell'efficacia comunicativa di un testo e della sua funzionalità, così come percepita e recepita fuori dall'ambiente scolastico/accademico e in contesti di comunicazione simmetrica.

L'applicabilità del costrutto e della scala dell'AF è stata testata e confermata da Kuiken e Vedder (2017, 2018) nella valutazione di testi argomentativi in italiano e olandese L2. I valutatori coinvolti nella ricerca erano non esperti, studenti universitari e parlanti nativi delle due lingue, adeguatamente formati al compito di valutazione e all'uso della scala.

In seguito, diversi studi hanno applicato la scala dell'AF nella valutazione di testi scritti e orali in italiano L2 di apprendenti con differenti lingue materne e su testi elicitati attraverso diversi *task*¹. Nella ricerca di Faone, Pagliara e Vitale (2017) la scala è stata applicata nella valutazione di testi regolativi e narrativi prodotti da apprendenti sinofoni di italiano L2, ottenendo risultati eccellenti di affidabilità, sebbene in assenza di accordo alto tra i giudizi dei valutatori (esperti e non esperti). Orrù (2019) e Del Bono (2019) hanno invece raccolto e sottoposto a valutazione testi narrativi, regolativi e argomentativi prodotti, rispettivamente, da apprendenti ungheresi e olandesi di italiano L2. In questi studi la valutazione è stata affidata a non esperti (cinque nel primo caso e quattro nel secondo), dai quali è stata riportata una certa difficoltà nell'uso di una delle dimensioni della scala, in particolare di quella denominata *Coerenza e coesione* (vedi § 2.2). Tale criticità è attestata anche nel lavoro di Nuzzo e Bove (2020), in cui si testa l'uso della scala dell'AF nella valutazione di testi in italiano L1 e L2 da parte di sette valutatori inesperti. I risultati di questa ricerca sembrano indicare una minore applicabilità dei descrittori dell'AF su produzioni di parlanti nativi. Tuttavia, in un lavoro successivo (Nuzzo & Bove 2022), i due autori esplorano le potenzialità della

¹ Per una rassegna completa degli studi condotti sull'Adeguatezza Funzionale e per una riflessione anche sulle implicazioni didattiche nell'uso di tale scala si veda Kuiken e Vedder (2022).

(auto)valutazione globale come strumento educativo, per lo sviluppo di abilità di scrittura accademica di studenti universitari.

Del Bono e Bonvino (2021) sono partite dalla constatazione delle difficoltà incontrate dai docenti di lingua nelle pratiche valutative nella didattica a distanza. Le autrici hanno proposto l'uso della scala dell'AF, testando con successo la possibilità di valutazione asincrona di testi regolativi in italiano L2, elicitati attraverso *task* d'istruzioni, da parte di cinque valutatori non esperti.

La scala dell'AF è stata utilizzata per la prima volta anche nella valutazione di testi scritti in italiano L2 prodotti da rifugiati e richiedenti protezione internazionale con debole livello di alfabetizzazione nella lingua materna da De Meo, Maffia e Vitale (2019). I punteggi attribuiti da due valutatori esperti sono stati comparati a quelli ottenuti attraverso l'uso di una scala più focalizzata sull'accuratezza formale dei testi². Tali dati hanno confermato le criticità e la lentezza nello sviluppo dell'abilità di scrittura per questo particolare target di apprendenti vulnerabili, con scarsi progressi nell'arco dei sei mesi della ricerca. È emerso, però, come la valutazione basata sul costrutto dell'AF fosse maggiormente in grado di valorizzare le competenze, seppur basiche, degli apprendenti, favorendo giudizi in media più alti. L'uso di tale strumento valutativo si è rivelato, quindi, potenzialmente più incoraggiante nei confronti di apprendenti privi di una solida alfabetizzazione in L1 e facilmente esposti a fenomeni di frustrazione e perdita di motivazione all'apprendimento linguistico. Allo stesso tempo si è considerato come, per tali apprendenti, l'ottenimento di una certificazione di italiano L2 fosse una necessità impellente e fondamentale alla buona riuscita del processo di richiesta e ottenimento della protezione internazionale.

2. Lo studio

2.1 Obiettivi dello studio

Il presente contributo intende proseguire la riflessione già avviata dalle ricerche sopra citate e in particolare da De Meo, Maffia e Vitale (2019) sulla

² Per una descrizione della scala si rimanda al § 2.2.

valutazione di testi scritti da apprendenti vulnerabili di italiano L2 e si pone quattro domande di ricerca:

1. L'uso della scala di Adeguatezza Funzionale da parte di valutatori con diversi livelli di esperienza e diversi gradi di familiarità con la scrittura in italiano L2 (inesperti e in formazione) può condurre a risultati comparabili a quelli ottenuti con valutatori esperti?
2. Cosa accade nel confronto con i giudizi ottenuti attraverso l'applicazione di una seconda scala di valutazione focalizzata sull'accuratezza formale da parte degli stessi valutatori e sugli stessi testi?
3. Il tipo di compito svolto costituisce una variabile significativa nella valutazione con le due scale da parte dei diversi valutatori?
4. Qual è il grado di *interrater agreement* e *interrater reliability* tra i giudizi dei valutatori coinvolti?

2.2 Le due scale di valutazione

Per favorire la comparazione dei risultati ottenuti attraverso l'applicazione delle due scale, si è deciso di attribuire a entrambe un punteggio massimo complessivo di 6 punti.

La scala dell'Adeguatezza Funzionale (cui si farà riferimento in seguito con la sigla V-AF) include quattro dimensioni, già proposte da Kuiken e Vedder (2017, 2018), a ciascuna delle quali corrisponde un punteggio massimo attribuibile pari a 1,5 punti³:

- la dimensione del *Contenuto* (CO) fa riferimento all'adeguatezza del numero e del tipo di unità informative espresse nel testo, alla loro pertinenza e rilevanza, considerate indipendentemente dai requisiti specifici richiesti dal *task*;

³ Nella scala originale proposta da Kuiken e Vedder a ciascuna delle dimensioni è possibile attribuire fino a 6 punti, ridotti a 1,5 in questo studio per agevolare la comparabilità della valutazione finale complessiva con quella ottenuta con la seconda scala utilizzata. La scala così modificata è stata testata con i due valutatori esperti, prima di essere utilizzata con gli altri gruppi di valutatori coinvolti nella ricerca.

- la dimensione riguardante i *Requisiti del task* (RT) considera la misura in cui nel testo si è tenuto conto dei requisiti specifici del compito comunicativo proposto allo scrivente, relativi al genere testuale, all'atto linguistico, alla macrostruttura e al registro del testo;
- la *Comprensibilità* (COMPR) tiene conto dello sforzo richiesto dal ricevente per capire lo scopo e le idee espresse nel testo;
- infine, la dimensione della *Coerenza e Coesione* (CC) si misura in termini di presenza/assenza di anafore, connettivi, salti logici e argomenti non collegati, ripetizioni, e altri fenomeni di ridondanza.

La seconda scala, utilizzata come termine di paragone e che da ora in poi verrà definita della Valutazione Standard (V-ST), è stata elaborata sulla base delle scale comunemente utilizzate nelle prove degli Enti certificatori per l'italiano L2, che presentano criteri ai quali vengono attribuiti pesi diversi per rapporto al livello di competenza atteso (Vedovelli 2005: 275-278).

Si riportano di seguito i parametri di tale scala, con i punteggi massimi attribuibili⁴:

- *Efficacia Comunicativa* (EC - 2,5 punti) che include:
 - contenuto, aderenza della performance ricevuta dal soggetto rispetto all'input fornito (1 punto);
 - coerenza e coesione, adeguatezza dei concetti sviluppati e dei meccanismi coesivi usati (1 punto);
 - uso della lingua in contesto, adeguatezza dei concetti della lingua utilizzata rispetto al contesto pragmatico di uso della stessa (0,5 punti);
- *Correttezza Morfosintattica* (CM - 1,5 punti), che prende in considerazione le regole che riguardano le forme e gli usi degli elementi di una lingua: accordi, concordanze, modi e tempi verbali, posizione dei costituenti frasali, presenza/assenza delle preposizioni, ecc.;
- *Appropriatezza Lessicale* (AL - 1 punto), che si riferisce all'uso di un repertorio lessicale che permette all'apprendente di gestire la propria

⁴In una prima fase della ricerca il parametro attualmente denominato *Efficacia Comunicativa* è stato etichettato come *Adeguatezza Testuale* (De Meo, Maffia & Vitale 2019). La modifica è motivata dalla necessità di una maggiore chiarezza per rapporto ai contenuti dei descrittori.

performance adeguandosi al destinatario del testo, all'argomento e al registro richiesto dalla situazione in questione, esprimendosi con efficacia e precisione terminologica;

- *Ortografia e Punteggiatura* (OP - 1 punto), si riferisce al rispetto delle convenzioni ortografiche dell'italiano standard, compreso il criterio di impaginazione come uso funzionale dello spazio pagina in relazione al tipo di testo.

Poiché in entrambe le scale utilizzate il punteggio massimo raggiungibile è 6, nella presentazione dei dati si considererà 4 come un risultato equivalente alla sufficienza per entrambi gli strumenti di valutazione.

2.3 Tre gruppi di valutatori

Al fine di verificare l'incidenza della variabile legata alla tipologia di valutatore sull'applicazione delle due scale nella valutazione di testi di apprendenti vulnerabili in italiano L2, alla coppia di esperti già coinvolti nello studio precedente sono stati aggiunti due nuovi gruppi di valutatori. Poiché le ricerche presenti in letteratura riportano dati relativi a gruppi di valutatori numericamente contenuti, si è scelto di verificare la gestibilità delle due scale anche da parte di gruppi più numerosi e di diversa ampiezza, pur consci dell'impatto che tali variabili avrebbero potuto avere sui risultati finali.

In sintesi, i dati che di seguito discuteremo sono il risultato delle valutazioni condotte da:

- due valutatori esperti, cui da ora in poi si farà riferimento con la sigla V1;
- 50 docenti/valutatori in formazione, corsisti del Master in Didattica dell'italiano L2 all'Università di Napoli L'Orientale, la cui sigla di riferimento sarà da ora in poi V2;
- 280 valutatori inesperti, studenti e studentesse di diversi corsi di studio triennali nello stesso Ateneo, rappresentati dalla sigla V3.

Ai due valutatori esperti (V1), i primi a essere stati coinvolti in ordine cronologico, è stato chiesto di applicare entrambe le scale di valutazione su tutti i testi in italiano L2 (per una descrizione dei testi si veda il § 2.4). Essi hanno svolto, in presenza, sessioni di standardizzazione dei risultati in fase iniziale, intermedia e finale del processo di valutazione, allo scopo di assicurarne la

validità e l'affidabilità e di limitare al minimo l'incidenza di variabili soggettive⁵. I valutatori hanno quindi effettuato dei controlli incrociati sui punteggi attribuiti a un campione di testi, riflettendo sui casi dubbi e cercando un accordo di valutazione nei casi di discrepanza di punteggio.

Sia i docenti in formazione (V2) sia i valutatori inesperti (V3) sono stati formati e addestrati all'uso delle due scale, prima di procedere alla effettiva fase di valutazione: i primi nel corso di una lezione del Master della durata di quattro ore; i secondi in occasione di due lezioni del corso di Apprendimento Linguistico, ciascuna della durata di due ore. Sono stati illustrati nel dettaglio i descrittori di entrambi gli strumenti valutativi e si è proceduto subito dopo a una attività laboratoriale, nella quale tutti i valutatori coinvolti hanno avuto la possibilità di provare ad applicare le due scale su un campione di testi in italiano L2 rappresentativi di diversi generi testuali. A conclusione del laboratorio, si è discusso dei risultati ottenuti e delle criticità riscontrate, chiarendo i dubbi e risolvendo alcune incertezze.

La fase di valutazione vera e propria per i gruppi V2 e V3 è stata condotta in maniera individuale e online, per motivi legati all'emergenza sanitaria da Covid-19 e anche per facilitare gli aspetti organizzativi, dato l'alto numero di persone coinvolte. La procedura utilizzata per i due gruppi è stata, però, diversa, data la necessità di adeguarsi ai tempi a disposizione nei due differenti casi.

A ciascuno dei 50 docenti in formazione è stato chiesto di valutare con entrambe le scale 12 dei testi dal corpus selezionato, proposti in ordine casuale, affinché fosse ridotto qualsiasi effetto dovuto alla distribuzione dei testi. Ogni testo del corpus è stato quindi valutato da almeno due valutatori con ciascuna scala. Solo per questo gruppo, inoltre, sono state raccolte anche osservazioni in forma scritta sull'esperienza valutativa e sulle eventuali difficoltà incontrate.

Nel caso del gruppo V3, invece, i 280 valutatori inesperti sono stati suddivisi in due gruppi di uguale ampiezza, ciascuno dei quali ha utilizzato una sola scala di valutazione su tutti i testi del corpus, proposti in tre diversi blocchi e in ordine casuale. In questo caso, quindi, ciascun testo è stato valutato da 140 valutatori con una scala e 140 con l'altra scala.

⁵Data la procedura di standardizzazione dei risultati, le valutazioni dei due docenti esperti, raccolte nella prima fase della ricerca, non sono state sottoposte ad analisi statistica.

2.4 Gli apprendenti e i testi

Per questo studio è stata selezionata una porzione del corpus *ScrItAV* (*Scritture in Italiano L2 di Apprendenti Vulnerabili*) consistente in testi prodotti da apprendenti di italiano L2 inseriti in percorsi di formazione linguistica di livello A2 offerti da CAS (Centri di Accoglienza Straordinaria) e centri del circuito SPRAR (Sistema di Protezione per Richiedenti Asilo e Rifugiati) nella città di Napoli. Si tratta di un gruppo di apprendenti eterogeneo, composto da 50 soggetti (di cui 5 donne, età media 26 anni) di diversa nazionalità, provenienti in particolare da paesi dell’Africa occidentale e centrale (Costa D’Avorio, Gambia, Guinea Bissau, Mali, Nigeria, Senegal) ma anche dall’Asia meridionale (Afghanistan e Pakistan) e con 20 lingue materne differenti. La vulnerabilità che caratterizza tali soggetti è legata a diversi aspetti: un passato difficile, spesso costellato di eventi traumatici, e un presente nel quale ci si trova, contro il proprio volere, “bloccati” in una situazione socio-politica precaria e disagiata, in attesa del completamento di complicate e lunghe trafale burocratiche; dal punto di vista linguistico, la diffusa debole alfabetizzazione nella lingua materna o nella lingua di scolarizzazione del paese d’origine provoca uno *shock* anche di carattere semiotico, oltre che sociale, nel paese d’arrivo (“*document shock*”, Adami 2008) e fa dei rifugiati e dei richiedenti protezione internazionale degli apprendenti “fragili”. Tutti questi fattori incidono inevitabilmente in maniera negativa sulla motivazione all’apprendimento linguistico, provocando spesso un blocco nel processo di sviluppo dell’interlingua, già di per sé lento e non sempre lineare, come riportato, tra gli altri, anche da Gordon 2011, Galos et al. 2017, Nitti 2018.

Nel corso delle regolari lezioni di italiano L2, nell’arco temporale di sei mesi e con una cadenza tendenzialmente bisettimanale, sono stati proposti in classe agli apprendenti coinvolti nove *task*, di cui tre semistrutturati di interazione scritta e sei aperti monologici, suddivisi in tre di tipo narrativo e tre di tipo descrittivo. Ai fini del presente studio, dal corpus *ScrItAV* sono stati selezionati 55 testi per ciascuna tipologia di *task*, per un totale di 165 produzioni scritte. La lunghezza media di questi testi è di circa 50 parole⁶.

⁶ La lunghezza dei testi prodotti dagli apprendenti coinvolti rispecchia il numero di parole atteso nelle prove di scrittura proposte dagli Enti Certificatori per il livello di competenza A2 (40-60 parole per i *task* aperti; 15-30 parole per i compiti di interazione scritta).

Si è scelto di utilizzare queste tre tipologie di *task* poiché risultano quelle abitualmente proposte anche nei test di certificazione della lingua italiana per i livelli elementari di competenza linguistico-comunicativa. La selezione dei contenuti, inoltre, è stata effettuata tenendo conto dell'esigenza di rappresentatività della ricchezza strutturale e funzionale dei compiti nella comunicazione reale extra-didattica (Vedovelli 2005)⁷.

I testi manoscritti degli apprendenti raccolti nelle classi di italiano L2 sono stati scansionati e successivamente digitalizzati, per poter essere oggetto di valutazione. Riportiamo di seguito un esempio di produzione scritta per ciascun tipo di *task*. Nell'operazione di digitalizzazione si è cercato di preservare il più possibile le caratteristiche formali dei testi (alternanza di grafie, autocorrezioni, ecc.). Nonostante questo, alcune informazioni relative, ad esempio, al grado di incertezza del tratto della scrittura o alla gestione dello spazio del foglio e delle righe non sono state registrate, per cui non hanno contribuito alla valutazione complessiva e non saranno oggetto di attenzione in questo lavoro.

Il primo (1) è un testo narrativo prodotto da DS, nigeriano di 27 anni, che ha dichiarato di parlare yoruba, egba, inglese e italiano. La traccia fornita dall'insegnante riguardava il racconto di una festa tradizionale del proprio Paese d'origine.

(1)

*UNA fESTiviTA tradizionale del mio paese
 Ci sono tante feste in mio paese molto importanti:
 *Nuovo anno: Questo dura per una settimana
 del primo Gennaio al 7 Gennaio. Noi siamo
 andate in Chiesa con mia famiglia l mattina
 dalle 8:00 alle 13:00 5 ore dopo la Chiesa
 Noi andiamo a casa, Noi abbiamo mangiato
 insieme, Noi pregiato insieme, dopo sono
 andato a festa con mia amici. Noi siamo
 andati fuori. Abbaimo bevuto dopo sono*

⁷ I compiti di tipo narrativo richiedevano di raccontare tradizioni e feste del proprio paese d'origine o dell'Italia e di parlare della propria giornata tipo. I *task* descrittivi riguardavano la descrizione fisica di persone, luoghi della città o di hobby e interessi. Infine, nei compiti di interazione scritta era richiesto di scrivere una e-mail a un amico o di formulare semplici annunci per la ricerca di una casa o di un lavoro.

*andata a casa per festa con la mia
famiglia.*

Il secondo testo (2) è una descrizione prodotta da EN, apprendente nigeriana di 22 anni, che ha dichiarato l'edo come lingua materna. La traccia richiedeva di descrivere l'aspetto fisico e il carattere di una persona conosciuta.

(2)

*NOME di compagno - Emmanuel
Aspetto fisico - lui e magro, più alto e
ha capelli corti, occhio neri.
Abbigliamento - lui ti piace completo
vestiti, lui ti piace Adidas scarpe.
Lui ti piace orologio nero.
Carattere - lui gentile, stmpatico, lui GiocaRE,
lui palare, lui cantare e intelligente.*

Si propone, infine, un esempio di interazione scritta (3), prodotta da RD, cubano di 28 anni, ispanofono con una competenza elementare anche nella lingua inglese. In questo caso la traccia prevedeva l'elaborazione di un messaggio personale rivolto a un amico con l'obiettivo di disdire un appuntamento.

(3)

*CARO MiCHELe, oggi è UN gioRNo iMPORtANte
PeR Noi, Pero mi deve SCUSARe NON Mi SEN-
to BeNe, HO MALe Di goLA è UN Po' Di
FeBBRe, LA VeRiTA' sto MORENDO. COMPAGNO
MiO Mi DeVe CAPiRe iO non ho voluto
STARe MALe, Pero ti Posso SCRiVeRe UN E-MAiL
CON iL iNDiRizzo DOVE DeVe ANDARe
A FARE iL SeRViZio, VA BeNe PeR te,
OK Mi Fai SAPeRe DOPO, CIAO*

2.5 Analisi statistiche

I dati raccolti dalle valutazioni condotte dai due gruppi V2 e V3 sono stati sottoposti ad analisi statistica tramite il software R⁸.

Per la misurazione dell'*interrater agreement*, indice che determina il consenso

⁸ R versione 4.0.5 (2021-03-31).

in termini assoluti tra i giudizi assegnati dai valutatori a ciascun testo, sono stati calcolati i coefficienti di correlazione intraclasse (ICC – *Intraclass Correlation Coefficient*) per ogni parametro delle due scale. Secondo quanto riportato in Koo & Li (2016), valori tra 0,5 e 0,75 corrispondono a un accordo moderato, tra 0,75 e 0,9 a un buon livello di accordo, oltre 0,9 a un eccellente livello di accordo.

Si è inteso, inoltre, testare l'*interrater reliability*, ossia l'affidabilità dei giudizi attribuiti nell'uso delle scale da parte dei valutatori nei due gruppi, tramite il calcolo del coefficiente di Cronbach (1951), un importante indice generale per la verifica della coerenza interna in un insieme di dati. Questo indicatore determina la proporzione di varianza condivisa dagli item che è attribuibile al costrutto indagato e assume anch'esso valori compresi tra 0 e 1. De Vellis (1991) specifica per questo coefficiente sei classi di valori: sotto lo 0,60 valori inaccettabili, tra 0,60 e 0,65 valori indesiderabili, tra 0,65 e 0,70 valori accettabili, tra 0,70 e 0,80 valori rispettabili, tra 0,80 e 0,90 valori molto buoni, e nel caso di valori superiori allo 0,90 la scala risulta molto affidabile.

Al fine di determinare la significatività delle differenze tra i giudizi ottenuti con le due scale di valutazione nei diversi gruppi di valutatori, inoltre, sono stati applicati test T per campioni appaiati. La soglia di significatività è stata fissata a 0,05.

3. Risultati

3.1 Le valutazioni

In Tabella 1 sono riportati i valori medi delle valutazioni attribuite dai tre gruppi di valutatori con le due diverse scale a tutte le produzioni scritte degli apprendenti vulnerabili di italiano L2. Si ricorda che il gruppo V1 è costituito da due valutatori esperti, il gruppo V2 da 50 valutatori in formazione, il gruppo V3 da 280 studenti, valutatori inesperti, ulteriormente suddivisi in due sottogruppi di eguale grandezza (140 hanno applicato la scala V-AF e 140 la scala V-ST).

	V-AF	V-ST	<i>p-value*</i>
V1	4,58	4,04	$p < 0,001$
V2	3,6	3,2	$p < 0,01$
V3	2,77	2,67	$p < 0,05$

Tabella 1. Valori medi delle valutazioni attribuite con le due scale a tutte le produzioni dai tre gruppi di valutatori (*test T).

Si nota immediatamente che i tre gruppi si comportano in maniera diversa, con punteggi che in media superano la soglia della sufficienza (4 pt) solo nel gruppo dei valutatori esperti (V1). Inoltre, benché in tutti i gruppi il punteggio medio raggiunto con la V-AF sia maggiore di quello della V-ST, solo nel caso dei valutatori esperti si riscontra una marcata significatività nel confronto tra i giudizi ottenuti con le due scale di valutazione ($p < 0,001$).

Nelle Tabelle 2 e 3 sono riportati i punteggi medi attribuiti ai diversi parametri di valutazione delle due scale su tutti i testi e dai tre gruppi. Si osserva come nel caso della V-AF i dati intragruppo siano abbastanza omogenei, con un andamento decrescente da V1 a V3. Il parametro CC (*Coerenza e Coesione*) riceve in media i punteggi più bassi da parte di tutti i valutatori. Anche per i dati relativi alla V-ST, è evidente come lo scarto tra il massimo attribuibile a ciascun parametro e il punteggio medio ottenuto aumenti in maniera progressiva dal gruppo V1 al gruppo V3. Fa eccezione il parametro CM (*Correttezza Morfosintattica*), per il quale sembra esserci un maggiore accordo di giudizio tra i tre gruppi.

V-AF	V1	V2	V3	p. m.
CO	1,22	0,96	0,74	1,5
RT	1,15	0,95	0,72	1,5
COMPR	1,19	0,97	0,75	1,5
CC	1,04	0,78	0,56	1,5

Tabella 2. Valori medi delle valutazioni attribuite ai diversi parametri della scala V-AF in tutte le produzioni e dai tre gruppi di valutatori (p. m.: punteggio massimo attribuibile).

V-ST	V1	V2	V3	p. m.
EC	1,75	1,5	1,09	2,5
CM	0,73	0,7	0,63	1,5
AL	0,82	0,7	0,60	1
OP	0,76	0,4	0,36	1

Tabella 3. Valori medi delle valutazioni attribuite ai diversi parametri della scala V-ST in tutte le produzioni e dai tre gruppi di valutatori (p. m.: punteggio massimo attribuibile).

Al fine di visualizzare in maniera più chiara l'andamento delle valutazioni, si è deciso di osservare le differenze tra le valutazioni medie attribuite da ciascun gruppo per ciascun parametro. Tale misura, indicata con il simbolo Δ , è stata applicata al rapporto tra V1-V2, V2-V3 e V1-V3. In Figura 1 si riportano i valori differenziali per le due scale di valutazione: il grafico riporta anche il Δ_{\min} attestato, pari a 0,03, e il Δ_{\max} , pari a 0,66.

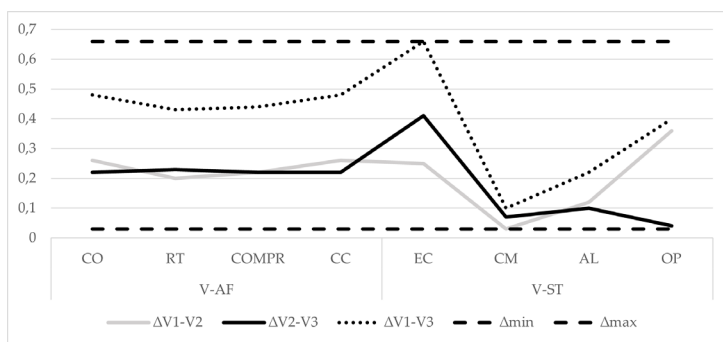


Figura 1. Differenze tra valutazioni medie di V1, V2 e V3 per parametro (scale V-AF e V-ST).

Nei dati relativi all'applicazione della scala V-AF è evidente come i valori di Δ_{V1-V2} e Δ_{V2-V3} siano molto vicini, attestandosi intorno a una differenza di 0,20 pt tra ciascun gruppo e in maniera omogenea per ciascun parametro, con un andamento delle due linee perfettamente sovrapponibile. Questo dato è confermato dai valori della linea relativa a Δ_{V1-V3} , che presentano il raddoppiamento dello scarto.

Nel caso della scala V-ST, i dati per parametro appaiono meno omogenei. Si evidenzia la vicinanza tra i tre gruppi per il parametro CM (*Correttezza Morfosintattica*) che vede uno scarto ridottissimo, pari o minore a 0,1 pt. Il parametro OP (*Ortografia e Punteggiatura*) avvicina i valutatori in formazione e gli inesperti, separandoli marcatamente dai valutatori esperti. Il parametro che appare più complesso da gestire per i valutatori inesperti è EC (*Efficacia Comunicativa*), per il quale si attestano il Δ_{\max} con il gruppo di valutatori esperti e uno scarto importante con il gruppo dei valutatori in formazione.

Con riferimento alla variabile tipo di compito svolto (narrazione, descrizione, interazione scritta), i dati dello studio non permettono di identificare una significativa correlazione con il tipo di scala e con il livello di esperienza

dei valutatori. Invece, mettendo a confronto le valutazioni ottenute con le due scale per i tre gruppi di valutatori, è emersa per V1 una differenza significativa per la narrazione ($p < 0,001$) e per l'interazione scritta ($p < 0,005$); per il gruppo V2 una differenza significativa solo nel caso delle valutazioni attribuite alle descrizioni ($p < 0,01$); nel gruppo V3 la variabile tipologia di task non è risultata mai statisticamente significativa.

3.2 L'accordo e l'affidabilità dei giudizi

Il calcolo dei coefficienti di correlazione intraclassa ha prodotto risultati di *interrater agreement* diversi in V2 e V3, probabilmente a causa delle differenze numeriche tra i due gruppi di valutatori. Nel gruppo dei docenti in formazione (Tabella 4) l'accordo nei giudizi attribuiti ai parametri della scala V-AF risulta moderato per ciascuno dei parametri di valutazione; nel caso della scala V-ST, l'accordo risulta omogeneamente buono.

V-AF	V2	V-ST	V2
CO	,75	EC	,84
RT	,72	CM	,88
COMPR	,70	AL	,83
CC	,73	OP	,84

Tabella 4. Accordo tra i giudizi nel gruppo V2 dei valutatori in formazione (ICC).

Nel gruppo V3 dei valutatori inesperti, invece, i dati mostrano un eccellente livello di accordo, con valori che superano sempre lo 0,9, per entrambe le scale di valutazione (Tabella 5).

V-AF	V3	V-ST	V3
CO	,98	EC	,97
RT	,97	CM	,97
COMPR	,98	AL	,96
CC	,98	OP	,97

Tabella 5. Accordo tra i giudizi nel gruppo V3 dei valutatori inesperti (ICC).

I risultati del calcolo dell'alpha di Cronbach applicato alle diverse dimensioni delle due scale nel gruppo V2, riportati in Tabella 6, dimostrano che i valutatori usano le due scale in modo coerente. I dati per l'*interrater reliability* sono globalmente significativi ($p < 0,01$). Utilizzando le classi di valori di De Vellis (1991), i risultati ottenuti con la scala V-AF possono essere interpretati come accettabili o rispettabili; quelli relativi alla scala V-ST appaiono tutti molto buoni.

V-AF	V2	V-ST	V2
CO	,75	EC	,85
RT	,66	CM	,88
COMPR	,69	AL	,83
CC	,69	OP	,84

Tabella 6. Affidabilità dei giudizi nel gruppo V2 dei valutatori in formazione (alpha di Cronbach).

Per quanto riguarda l'*interrater reliability* nelle valutazioni del gruppo V3, anche in questo caso le correlazioni risultano significative ($p < 0,01$). Per entrambe le scale i risultati confermano la marcata affidabilità dei giudizi dei valutatori sui diversi parametri, con dati molto omogenei (Tabella 7).

V-AF	V3	V-ST	V3
CO	,99	EC	,99
RT	,99	CM	,98
COMPR	,99	AL	,97
CC	,99	OP	,98

Tabella 7. Affidabilità dei giudizi nel gruppo V3 dei valutatori inesperti (alpha di Cronbach).

4. Discussione e conclusioni

Nel presente contributo si è proposto uno studio sulla applicabilità della scala di valutazione basata sull'Adeguatezza Funzionale a testi in italiano L2 di apprendenti vulnerabili, con un livello di competenza in italiano L2 pari all'A2 del QCER e un generale basso livello di scolarizzazione nel paese d'ori-

gine. Sono stati coinvolti nella ricerca valutatori con diversi percorsi di formazione e un diverso grado di familiarità con le produzioni in interlingua e le procedure valutative. Sono stati utilizzati tre tipi di *task* scritti (narrazione, descrizione e interazione) e i dati ottenuti con l'applicazione della scala V-AF sono stati confrontati con quelli scaturiti dall'uso di una scala di valutazione maggiormente focalizzata sull'accuratezza formale (V-ST).

I risultati dello studio permettono di rispondere alle domande di ricerca obiettivo di questo lavoro.

1. L'uso della scala di Adeguatezza Funzionale da parte di valutatori con diversi livelli di esperienza e diversi gradi di familiarità con la scrittura in italiano L2 (inesperti e in formazione) può condurre a risultati comparabili a quelli ottenuti con valutatori esperti?

Il confronto tra i giudizi assegnati ai testi in italiano L2 di apprendenti vulnerabili con la scala V-AF dai tre gruppi di valutatori ha evidenziato un rapporto inverso tra il grado di esperienza dei soggetti coinvolti e la severità dei giudizi⁹. I valutatori esperti risultano infatti gli unici ad attribuire punteggi medi che superano la soglia della sufficienza; i docenti in formazione si collocano in una fascia intermedia e i valutatori inesperti attribuiscono ai testi i punteggi più bassi.

L'omogeneità globale delle valutazioni di questo ultimo gruppo sembra testimoniare una generale difficoltà nel calibrare il proprio giudizio su una produzione poco familiare, come quella prodotta da un parlante non nativo, per giunta con competenza elementare in italiano L2 e debole scolarizzazione. Vi è, probabilmente, un eccessivo scollamento tra le attese del valutatore nativo *naïf* e la resa linguistica degli apprendenti.

Tali risultati sembrano indicare che la tipologia di valutatore coinvolto può determinare notevoli differenze nei risultati ottenuti, in particolare se la valutazione è applicata a testi peculiari e particolarmente “devianti” dal punto di vista formale, come quelli prodotti da apprendenti debolmente alfabetizzati nella lingua materna.

⁹Per una riflessione sul coinvolgimento di valutatori esperti e non esperti e sui possibili *bias* nelle procedure valutative si vedano Kuiken, Vedder e Gilbert (2010) e Vedder (2019).

I dati di questa ricerca, inoltre, confermano la difficoltà già riscontrata in studi precedenti nell'applicazione del parametro *Coerenza e Coesione*, che riceve i punteggi più bassi da parte di tutti i valutatori. La natura complessa di tale parametro e la difficoltà di valutare con un solo punteggio due aspetti diversi di un testo (la sua coerenza dal punto di vista dei contenuti e l'adeguatezza dei mezzi utilizzati per esprimere la coesione) sono evidenziate, infatti, anche nei commenti di diversi valutatori del gruppo V2, uno dei quali è riportato di seguito:

[...] a volte è stato problematico definire il livello di coerenza e coesione in un unico punteggio poiché quasi tutti i testi presi in esame si dimostrano coerenti, talvolta a mancare è soprattutto la coesione del testo ed è evidente la scarsità di connettivi e la loro ripetizione. (V2-7)

2. Cosa accade nel confronto con i giudizi ottenuti attraverso l'applicazione di una seconda scala di valutazione focalizzata sull'accuratezza formale da parte degli stessi valutatori e sugli stessi testi?

Nonostante le differenze tra le valutazioni dei tre gruppi ottenute con la scala V-AF, i risultati appaiono globalmente migliori rispetto a quelli ottenuti con la scala V-ST. A influire negativamente sulla V-ST è in particolare il “peso” che possiede il parametro della *Correttezza Morfosintattica*, che restituisce la scarsa accuratezza formale dei testi degli apprendenti coinvolti nella ricerca. Al contrario, la scala V-AF, puntando sull'efficacia comunicativa e sulla felice esecuzione del compito, riesce a valorizzare maggiormente le pur minime abilità di scrittura di tali soggetti e i parametri che la costituiscono ricevono valutazioni equilibrate.

3. Il tipo di compito svolto costituisce una variabile significativa nella valutazione con le due scale da parte dei diversi valutatori?

Il gruppo dei valutatori inesperti (V3) non mostra alcuna differenza significativa tra il tipo di *task* e la scala di valutazione utilizzata. Per i docenti in formazione (V2) appare una significatività solo nel caso della descrizione, mentre per i docenti esperti (V1) la significatività si manifesta nei compiti di narrazione e interazione scritta.

Emerge una progressione graduale: il gruppo V3 sembra rimanere maggiormente ancorato a una valutazione standard, in cui la variabile *task* non

ha alcun effetto poiché tutti i testi sono ugualmente devianti dalla lingua standard; il gruppo V2 attribuisce punteggi migliori con la scala V-AF esclusivamente ai testi formalmente più semplici, quelli descrittivi¹⁰; il gruppo V3, infine, riesce a prescindere maggiormente dalla complessità strutturale dei testi e a riconoscere l'Adeguatezza Funzionale anche in produzioni fortemente marcate dalla scorrettezza dell'ortografia e della morfosintassi.

I dati confermano che il livello di esperienza del valutatore incide sulla capacità di valorizzare gli aspetti funzionali di un testo scritto in maniera indipendente dalla sua accuratezza formale, in particolare nel caso di testi caratterizzati da una forte devianza dalla norma (e dall'uso) dei nativi, prodotti da apprendenti con un livello elementare di competenza linguistico-comunicativa nella seconda lingua nonché una debole alfabetizzazione nella lingua materna.

4. Qual è il grado di interrater agreement e interrater reliability tra i giudizi dei valutatori coinvolti?

È stato osservato come il livello di accordo in termini assoluti, ottenuto attraverso il calcolo dell'ICC, differisca nei due gruppi di valutatori V2 e V3: i punteggi del gruppo V2 raggiungono un livello di accordo moderato con la scala V-AF e buono nel caso della V-ST; il gruppo V3 raggiunge, invece, livelli eccellenti di accordo con entrambi gli strumenti di valutazione, con giudizi che quindi si presentano significativamente omogenei e spesso coincidenti. Tali dati testimoniano un maggiore grado di incertezza tra i docenti in formazione, in particolare nell'uso della scala basata sull'Adeguatezza Funzionale. La *Comprensibilità* riporta gli indici più bassi di accordo nel gruppo V2, probabilmente a causa della natura fortemente soggettiva di tale parametro, che tiene conto dello sforzo richiesto dal lettore nella comprensione dello scopo e delle idee espresse in un testo.

I giudizi dei valutatori inesperti, invece, appaiono globalmente severi, indipendentemente dalla scala utilizzata e in maniera equilibrata nei vari parametri.

¹⁰ La capacità di descrivere qualcosa o qualcuno molto semplicemente è prevista già al livello A1 dei descrittori del Volume complementare del QCER relativi alla Scrittura creativa (Consiglio d'Europa 2020: 73). Le abilità narrative e di interazione scritta testate attraverso i *task* proposti in questa ricerca sono invece previste al livello A2 (ibidem: 73 e 92).

L'analisi statistica ha dimostrato, inoltre, che *l'interrater reliability*, impiegata come misura dell'affidabilità dello strumento valutativo, raggiunge livelli di accettabilità/rispettabilità nel gruppo V2 in relazione all'uso della scala V-AF, mentre risulta molto buona nel caso dei punteggi attribuiti da tale gruppo con la scala V-ST. I giudizi dei valutatori inesperti anche in questo caso presentano dati più omogenei e una marcata affidabilità nell'applicazione di entrambe le scale.

I risultati, la cui interpretazione non può prescindere dal considerare la diversa composizione numerica dei gruppi di valutatori e le diverse procedure valutative (cfr. § 2.3), appaiono tuttavia globalmente soddisfacenti e sono indice di una formulazione precisa dei singoli parametri di ciascuna scala.

In conclusione, i dati di questo studio hanno confermato che l'uso di una scala di valutazione basata sull'Adeguatezza Funzionale è utile a far emergere competenze comunicative che non trovano riscontro in una scala più "tradizionale", in cui si consideri anche l'accuratezza formale. Questo appare particolarmente importante nel caso di apprendenti vulnerabili di italiano L2, spesso debolmente scolarizzati nel paese d'origine, il cui percorso di formazione linguistica è lento e faticoso, ostacolato da una motivazione non sempre stabile, da un contesto di vita precario e un passato a volte traumatico.

La scala dell'Adeguatezza Funzionale, inoltre, si è dimostrata uno strumento utilizzabile anche da parte di valutatori in formazione e inesperti, sebbene il livello di esperienza abbia inciso notevolmente sui valori medi dei punteggi attribuiti: da V1 a V3 si è assistito a uno scarto graduale in negativo (fino a circa 2 punti) e una sempre minore variabilità tra i giudizi attribuiti ai diversi *task*, con i punteggi del gruppo V3 dei valutatori inesperti che risultano, infatti, in assoluto i più severi e i più costanti.

È difficile stabilire se sia l'esperienza nella didattica delle lingue a rendere più "docili" o se la scarsa familiarità con interlingue di livello basico sia causa di maggiore rigidità. Al di là, però, dei valori assoluti, sicuramente condizionati anche dalla diversa composizione numerica dei gruppi di valutatori coinvolti, è importante constatare che la scala dell'Adeguatezza Funzionale sembra rappresentare uno strumento prezioso, perché potenzialmente applicabile anche in contesti in cui la formazione glottodidattica e la valutazione

linguistica sono affidati a operatori e volontari non sempre esperti o con diversi livelli di esperienza, come accade in Italia nelle strutture del sistema di accoglienza per richiedenti e titolari di protezione internazionale.

In tale prospettiva, i risultati della presente ricerca andranno validati da ulteriori indagini sul caso peculiare della valutazione delle competenze linguistico-comunicative di apprendenti vulnerabili. L'applicabilità della scala fondata sull'Adeguatezza Funzionale potrebbe essere testata, ad esempio, su testi scritti di soggetti con diversi livelli di competenza nella seconda lingua nonché sulle produzioni orali, tenendo sotto attento controllo le variabili legate alla scolarizzazione nel Paese d'origine (livello di alfabetizzazione, tipologia di percorso formativo).

Riferimenti bibliografici

- Adami, Hervé, 2008, "The role of literacy in the acculturation process of migrants", Strasbourg, Council of Europe, <https://www.coe.int/lang-migrants>.
- Cittalia; Fondazione Anci; Ministero dell'Interno, 2019, *Atlante SPRAR/SIPROIMI 2018. Rapporto annuale SPRAR/SIPROIMI 2018. Sistema di protezione per titolari di protezione internazionale e per minori stranieri non accompagnati*, <https://www.retesai.it/wp-content/uploads/2019/11/Atlante-Sprar-Siproimi-2018-leggero.pdf>.
- Cittalia; Fondazione Anci; Ministero dell'Interno, 2020, *Atlante SPRAR/SIPROIMI 2019. Rapporto annuale SPRAR/SIPROIMI 2019. Sistema di protezione per titolari di protezione internazionale e per minori stranieri non accompagnati*, <https://www.retesai.it/wp-content/uploads/2020/12/ATLANTE-SIPROIMI-2019.pdf>.
- Consiglio d'Europa, 2020, "Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione. Volume complementare" (traduzione italiana a cura di Barni, M.; Lugarini, E.; Cardinaletti, A.), *Italiano LinguaDue* 12 (2).
- Cronbach, Lee J., 1951, "Coefficient alpha and the internal structure of tests", *Psychometrika* 16, 297-334.
- De Vellis, Robert F., 1991, *Scale development: Theory and applications*, Newbury Park CA, Sage Publications.
- Del Bono, Federica, 2019, "Aspetti pragmatici nella valutazione di testi scritti: uno studio sull'adeguatezza funzionale in italiano L2", in Nuzzo E.; Vedder I. (a cura di), *Lingua in contesto: la prospettiva pragmatica*, Milano-Bologna, Officinaventuno- AITLA, 231-244.
- Del Bono, Federica; Bonvino, Elisabetta, 2021, "Dalla valutazione in DaD alle scale di adeguatezza funzionale", *EuroAmerican Journal of Applied Linguistics and Languages* 8 (2), 105-124.

- Faone, Serena; Pagliara, Francesca; Vitale, Giuseppina, 2017, “How to assess L2 information-gap tasks through functional adequacy rating scales”, contributo presentato alla *7th Task Based Language Teaching Conference - Tasks in Context*. Barcellona, 19-21 aprile.
- Galos, Eliza; Bartolini, Laura; Cook, Harry; Grant, Naomi, 2017, *Migrant Vulnerability to Human Trafficking and Exploitation: Evidence from the Central and Eastern Mediterranean Migration Routes*. International Organization for Migration, Ginevra, IOM.
- Gordon, Daryl, 2011, “Trauma and Second Language Learning Among Laotian Refugees”, *Journal of Southeast Asian American Education & Advancement* 6 (1), <https://docs.lib.purdue.edu/jsaea/vol6/iss1/13>.
- Grice, Herbert P., 1975, “Logic and conversation”, in Cole P.; Morgan J. L. (eds.), *Syntax and semantics: Speech acts*, Academic Press, 41-58.
- Housen, Alex; Kuiken, Folkert; Vedder, Ineke (eds.), 2012, *Dimensions of L2 performance and proficiency. Complexity, accuracy and fluency in SLA*, Amsterdam-Philadelphia, Benjamins.
- Koo, Terry K.; Li, Mae Y., 2016, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research”, *Journal of Chiropractic Medicine* 15, 155-163.
- Krippendorff, Klaus, 2004, “Reliability in content analysis: Some common misconceptions and recommendations”, *Human Communication Research* 30 (3), 411-433.
- Kuiken, Folkert; Vedder, Ineke, (eds.), 2022, “The Assessment of Functional Adequacy in Language Performance”, Special issue of *Journal on Task-Based Language Teaching and Learning* 2 (1).
- Kuiken, Folkert; Vedder, Ineke, 2014, “Raters’ decisions, rating procedures and rating scales”, *Language Testing* 31 (3), 279-284.
- Kuiken, Folkert; Vedder, Ineke, 2017, “Functional adequacy in L2 writing. Towards a new rating scale”, *Language Testing* 34 (3), 321-336.
- Kuiken, Folkert; Vedder, Ineke, 2018, “Assessing functional adequacy of L2 performance in a task-based approach”, in Taguchi N.; You-Jin K. (eds.), *Task-based approaches to teaching and assessing pragmatics*, Amsterdam-Philadelphia, Benjamins, 266-285.
- Kuiken, Folkert; Vedder, Ineke; Gilabert, Roger, 2010, “Communicative adequacy and linguistic complexity in L2 writing”, in Bartning I.; Martin M.; Vedder I. (eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, Eurosla Monographs Series 1, 81-100.
- Nitti, Paolo, 2018, “I bisogni linguistici nei corsi di italiano L2 rivolti ad utenti vulnerabili. Un’indagine sui corsi di lingua seconda erogati dai centri di accoglienza in Piemonte”, *EL.LE* 7 (3), 413-428.

- Nuzzo, Elena; Bove, Giuseppe, 2020, "Assessing functional adequacy across tasks: A comparison of learners' and native speakers' written texts" *E-JournALL, EuroAmerican Journal of Applied Linguistics and Languages* 7 (2), 9-27.
- Nuzzo, Elena; Bove, Giuseppe, 2022, "Exploring the pedagogical use of the rating scale for functional adequacy in L1 writing instruction", in Kuiken F.; Vedder I. (eds.), "The Assessment of Functional Adequacy in Language Performance", Special issue of *Journal on Task-Based Language Teaching and Learning* 2 (1), 115-136.
- Orrù, Paolo, 2019, "Misurare l'adeguatezza funzionale in testi scritti di apprendenti di italiano L2", *Italiano LinguaDue* 11 (1), 54-68.
- Vedder, Ineke, 2016, "Il ruolo dell'adeguatezza funzionale nelle produzioni scritte in lingua seconda: proposta per una scala di valutazione", in Vedder, I.; Santoro, E. (a cura di), *Pragmatica e interculturalità in italiano lingua seconda*, Firenze, Cesati, 79-92.
- Vedder, Ineke, 2019, "La valutazione dell'adeguatezza funzionale di produzioni orali e scritte in italiano L2 in tipologie di task differenti", *Testi e linguaggi* 13, 165-177.
- Vedovelli, Massimo (a cura di), 2005, *Manuale della certificazione dell'italiano L2*, Roma, Carocci.