

Dalla valutazione in DaD alle scale di adeguatezza funzionale

FEDERICA DEL BONO

Università degli Studi Roma Tre/Sapienza Università di Roma

ELISABETTA BONVINO

Università degli Studi Roma Tre

Received 2 May 2021; accepted 7 August 2021

ABSTRACT

IT Il presente contributo analizza la valutazione delle competenze linguistiche nella Didattica a Distanza (DaD) durante il periodo di emergenza sanitaria (Covid-19). Dall'analisi delle risposte a un questionario somministrato a docenti di lingue, emerge che la valutazione delle lingue straniere ha risentito notevolmente di questa nuova modalità di insegnamento e che alcune delle difficoltà valutative riscontrate in DaD possano essere in parte risolte assumendo una prospettiva di valutazione globale, che integri le diverse competenze e abilità degli apprendenti, attraverso attività task-based. Dopo aver illustrato le diverse soluzioni adottate dai docenti durante la DaD, vengono presentati i presupposti della valutazione task-based e si evidenziano i potenziali vantaggi dell'applicazione didattica delle scale di adeguatezza funzionale.

Parole Chiave: EDUCAZIONE LINGUISTICA, VALUTAZIONE LINGUISTICA TASK-BASED, ADEGUATEZZA FUNZIONALE, DIDATTICA A DISTANZA

EN This article analyzes the evaluation of linguistic competence in distance learning during the crisis caused by the Covid-19 pandemic. Based on the analysis of responses to a questionnaire administered to language teachers, language teachers perceive that assessment of foreign language has regressed notably due to this new mode of teaching, but that some of the difficulties in assessing distance learning can be resolved, in part, by adopting an approach based on global assessment, which integrates various competencies and abilities of learners through task-based activities. After having illustrated the varying solutions adopted by teachers during this period of distance learning, suppositions about task-based learning are presented, followed by the potential advantages of the pedagogical application of the scale of functional adequacy.

Key words: LANGUAGE EDUCATION, TASK-BASED ASSESSMENT, FUNCTIONAL ADEQUACY, DISTANCE LEARNING

ES Esta contribución analiza la evaluación de las competencias lingüísticas en educación a distancia durante la crisis sanitaria provocada por la pandemia Covid-19. El análisis de las respuestas a un cuestionario administrado al profesorado de idiomas extranjeros muestra un retroceso notable en la evaluación de la lengua extranjera debido a esta nueva forma de enseñanza, aunque algunas de las dificultades en la evaluación de la educación a distancia pueden ser resueltas, en parte, mediante la adopción de un enfoque de evaluación global que integre las diversas competencias y habilidades del alumnado a través de actividades basadas en tareas. Después de haber ilustrado las distintas soluciones adoptadas por el profesorado durante este período de educación a distancia, se presentan las conjeturas sobre el aprendizaje basado en tareas así como las potenciales ventajas de la aplicación pedagógica de la escala de adecuación funcional.

Palabras clave: EDUCACIÓN LINGÜÍSTICA, EVALUACIÓN POR TAREAS, ADECUACIÓN FUNCIONAL, ENSEÑANZA A DISTANCIA

1. Introduzione

La situazione di emergenza sanitaria causata dalla malattia da Coronavirus (Covid-19) iniziata in Italia nel marzo 2020 ha costretto tutte le persone coinvolte nei processi di formazione, dalla scuola all'università, a importanti e repentini cambiamenti: i corsi in presenza sono stati infatti riconvertiti in corsi on-line, con un conseguente stravolgimento delle abitudini di insegnamento, apprendimento e valutazione. Nell'ambito del progetto di rilevamento delle pratiche nella Didattica a Distanza (DaD) presentato in questo numero, abbiamo ideato, con il gruppo GRAAL di Roma Tre, un questionario (cfr. Conti, in stampa, in questo numero), somministrato a docenti di lingue, per cercare di comprendere quale sia stato l'impatto del passaggio a una didattica prevalentemente a distanza nell'ambito dell'insegnamento delle lingue seconde e straniera.

La valutazione è uno degli ambiti che ha maggiormente risentito di questa situazione improvvisa e che ha trovato tutti impreparati. Insegnanti e apprendenti hanno infatti dovuto affrontare notevoli problemi legati principalmente all'inadeguatezza degli strumenti tecnologici disponibili e alla difficoltà di dover adattare le consuete modalità valutative ad una realtà diversa. I mezzi di comunicazione italiani hanno riportato numerosi episodi che vedono i docenti protagonisti di vani tentativi di assicurare sorveglianza e rigore durante lo svolgimento delle prove valutative, contrapposti ad apprendenti abilissimi nell'uso delle nuove tecnologie e quindi in grado di eludere facilmente il controllo. Questi episodi evidenziano la difficoltà della valutazione a distanza, testimoniata anche dal fatto che in molti casi si è preferito rimandare il momento valutativo al ritorno in presenza.

Il termine valutazione è qui utilizzato in senso molto ampio come un complesso di operazioni che consente di raccogliere informazioni volte a favorire il processo decisionale (Bachman & Palmer, 2010). In questo senso, la nozione di valutazione si estende a comprendere anche altri termini, quali: esame, prova di verifica, test, valutazione alternativa, e altri, e può essere utilizzata anche per indicare le singole attività, ad esempio colloqui, interrogazioni orali o risposte ad una *check-list* di autovalutazione (cfr. Lenz & Berthele, 2010).

La valutazione costituisce un momento fondamentale del processo formativo che rientra nell'ambito dei procedimenti di comunicazione, controllo e *feedback* esercitati, in misura diversa, sia dall'insegnante sia dall'apprendente. Vi è infatti un legame strettissimo tra verifica, valutazione e azione didattica: si verificano le competenze linguistiche per poterle classificare in una scala di valori, tale scala permette di comprendere se gli obiettivi proposti sono stati adeguatamente raggiunti e/o se non sia il caso di aggiustare l'intervento formativo (Novello, 2014). Il processo valutativo ha, pertanto, sempre importanti ricadute sul processo didattico, e può portare a riflessioni, ripensamenti e approfondimenti sia da parte dell'insegnante che da parte dell'apprendente. In particolare, le informazioni raccolte nella valutazione linguistica permettono di determinare la qualità e la quantità dell'acquisizione di una determinata lingua da parte di un apprendente, attribuendo a questi dati un valore, un punteggio, o un livello. Per queste ragioni, all'interno del progetto di indagine della DaD, presentato in questo numero della rivista, ci è parso importante capire quali siano state quindi le conseguenze per la valutazione e pertanto una sezione è stata dedicata a questa importante fase del processo formativo. Ci siamo infatti interrogati se i criteri di qualità della valutazione siano stati sempre soddisfatti in DaD.

La qualità della valutazione è solitamente individuata attraverso una serie di requisiti. La validità costituisce il più importante requisito qualitativo e riguarda principalmente l'ambito dell'interpretazione dei risultati (Bachman, 1990). Una procedura di valutazione è valida nella misura in cui si può dimostrare che ciò che si sta valutando è ciò che, in quel contesto, si vuole effettivamente valutare, e che l'informazione ottenuta rappresenta con precisione la competenza del candidato in questione (Consiglio d'Europa, 2002, p.217). In altri termini la validità deriva dalla coerenza tra il contenuto della prova (le richieste fatte al candidato, il formato di queste richieste, le abilità chiamate in gioco, il campione di lingua utilizzato e/o richiesto ecc.) e ciò che il test si prefigge di valutare. Il termine può comunque essere inteso nella sua accezione più ampia, in cui si intende anche la validità di quanto viene richiesto al candidato, validità dei criteri di selezione, validità delle scelte fatte ecc. (Bonvino, 2008). Un secondo fattore, collegato alla validità, e da cui la validità di una prova dipende, riguarda invece la rappresentatività e l'appropriatezza dell'aspetto valutato, sia per quanto riguarda la dimensione e la durata sia per quanto riguarda il contenuto. Le prove, infatti, dovrebbero essere indicative di un'abilità soggiacente nel candidato e rappresentative della lingua e della cultura oggetto della valutazione. La qualità della valutazione è considerata anche rispetto al criterio della comparabilità, con il quale si intende il fatto che i dati ottenuti possono essere confrontati con prestazioni di altri candidati o dello stesso candidato anche in momenti diversi dell'apprendimento. Infine, due aspetti imprescindibili di una prova sono la sua accettabilità, criterio che indica che la prova proposta deve essere considerata accettabile sia da parte del

docente che dell'apprendente. Quest'ultimo criterio è spesso collegato a quello relativo alla sua fattibilità ed economicità, cioè da fattori determinati dal costo in termini di materiali, tempo e sforzi che l'organizzazione della prova richiede. La fattibilità di un test dipende pertanto soprattutto dalle condizioni e dalla struttura in cui lo si svolge. Inoltre, un test per essere considerato fattibile dovrebbe anche poter essere applicabile agli ordinamenti didattici in vigore ed essere con questi coerente.

Le risposte di questa sezione del questionario (cfr. § 2) hanno confermato le grandi difficoltà incontrate dagli insegnanti nel processo valutativo in questo particolare momento. Oltre alle difficoltà, è però stato possibile rilevare anche aspetti positivi e innovativi, che sono emersi proprio per far fronte alle esigenze di una valutazione a distanza. Alcuni docenti hanno infatti abbandonato una valutazione tradizionale, basata su verifiche puntuali della competenza grammaticale, o di singole abilità, e hanno invece optato per la creazione di prove globali basate su compiti in cui si richiede agli apprendenti un'elaborazione personale e il ricorso ad abilità linguistiche integrate. Le prove utilizzate dagli insegnanti in DaD possono essere ricondotte alla definizione di task, inteso come un compito che richiede agli apprendenti di utilizzare le proprie risorse linguistiche per raggiungere un obiettivo non linguistico (Cortés Velásquez & Nuzzo, 2018). I task richiedono una valutazione adeguata, che, come mostreremo, può essere effettuata attraverso strumenti di verifica idonei a un utilizzo in DaD.

Il presente lavoro ha pertanto un duplice obiettivo che si manifesta in una struttura bipartita: la prima sezione (§2), che costituisce il punto di partenza e il presupposto della seconda, analizza i risultati del questionario sulla valutazione in DaD con l'intento di tracciare un bilancio; la seconda sezione (§3) discute i presupposti della valutazione task-based con l'obiettivo di evidenziare i vantaggi di una valutazione basata sul concetto di adeguatezza funzionale (d'ora in poi AF, dall'inglese *Functional Adequacy*, Kuiken e Vedder, 2017). Le scale di AF, anche se oggi sono utilizzate prevalentemente in contesti di ricerca, costituiscono a nostro avviso un utile strumento valutativo, in grado di tener conto sia del raggiungimento degli obiettivi del task che dell'adeguatezza della performance.

2. Il questionario

2.1. Contesto e campione

Il questionario somministrato dal gruppo GRAAL ai docenti di lingue straniere ha previsto una sezione interamente dedicata alla valutazione, vista l'importanza che essa riveste nell'ambito della didattica delle lingue seconde/straniere. Questa sezione era suddivisa in due parti, una prima riguardante le difficoltà incontrate nel valutare in DaD e una seconda relativa al tipo di pratica valutativa applicata, in termini di strumenti usati, test creati e aspetti valutati. Nel questionario si alternavano domande con risposta a scelta multipla di diverso tipo, in cui i docenti erano chiamati a selezionare una o più opzioni, oppure in cui erano invitati a rispondere indicando un valore scalare numerico e non. In alcuni casi, inoltre, era anche richiesto di argomentare le proprie risposte tramite domande aperte. In questo contributo non saranno presentate nel dettaglio le domande, ma soltanto i risultati più salienti su cui si basa la proposta didattica. Per una descrizione dettagliata del questionario e per la sua versione originale si veda Conti (in stampa, in questo numero).

Ha risposto all'indagine un totale di 134 insegnanti¹, afferenti a strutture educative diverse (scuola primaria, scuola secondaria di I e II grado, università, CPIA, scuole di lingua private e associazioni). La maggior parte degli insegnanti che ha partecipato all'indagine insegna in scuole secondarie di I e II grado (40,3%) o università (39,6%), mentre gli altri in scuole primarie (2,2%) o altri tipi di istituti (17,9%)². Il 40,3% degli informanti ha dichiarato di aver ricevuto indicazioni specifiche da parte dell'istituto di appartenenza riguardo le modalità di svolgimento della valutazione, senza specificare però in merito a quali aspetti.

2.2. I risultati

Nei paragrafi seguenti saranno presentati i risultati dell'indagine sulla valutazione, partendo da un'analisi delle difficoltà avute nella valutazione in DaD e passando poi a descrivere nel dettaglio come questa è stata svolta.

¹ Due dei 136 insegnanti che hanno compilato il questionario generale hanno dichiarato di non aver valutato e pertanto sono stati esclusi dall'analisi.

² Nella presentazione dei risultati non teniamo conto del parametro relativo alla tipologia di istituto perché dall'analisi dei dati non sono emerse differenze sostanziali nelle preferenze espresse dai docenti.

2.2.1. Valutazione a distanza: difficoltà e problemi

La prima parte del questionario indagava, come si è detto, i problemi incontrati dai docenti nel valutare in DaD, osservando innanzitutto il cambiamento della valutazione in termini organizzativi. Innanzitutto, è stato chiesto agli insegnanti di indicare il loro grado di accordo – da 0 (per niente) a 5 (del tutto) – con affermazioni riguardanti l'organizzazione della valutazione in DaD (Tabella 1).

Tabella 1

Organizzazione della valutazione in DaD

Scala	1. Ho dovuto modificare la mia modalità di valutazione rispetto alla didattica in presenza	2. Il tempo che dedico alla valutazione è aumentato rispetto a quando lavoravo in presenza	3. Ho incontrato dei problemi nell'adattare la mia valutazione alla modalità a distanza
0 (per niente)	4,5%	11,2%	10,4%
1	11,9%	4,5%	17,2%
2	8,2%	20,1%	17,2%
3	17,9%	9,7%	20,1%
4	31,3%	27,6%	22,4%
5 (molto)	26,1%	26,9%	12,7%

I risultati nella Tabella 1 mostrano che la metà dei docenti ha segnalato un alto livello di accordo sia con l'affermazione "1) *Ho dovuto modificare la mia modalità di valutazione rispetto alla didattica in presenza*" (31,3% per il punteggio 4; 26,1% per il punteggio 5); che con l'affermazione "2) *Il tempo che dedico alla valutazione è aumentato rispetto a quando lavoravo in presenza*" (27,6% per il punteggio 4; 26,9% per il punteggio 5). È risultato essere minore, invece, il grado di accordo segnalato per la terza affermazione "3) *Ho incontrato problemi nell'adattare la mia valutazione alla modalità a distanza*". Trattandosi di una domanda a risposta chiusa, non è stato possibile approfondire le ragioni di queste risposte nel breve spazio del questionario, né il motivo per cui ci fosse una differenza nel grado di accordo espresso tra l'affermazione 3 e le affermazioni 1 e 2. Ciò nonostante, le risposte fornite dai docenti ci permettono di fare delle deduzioni generali. A nostro avviso, infatti, l'aumento del tempo dedicato alla valutazione e i problemi incontrati, potrebbero essere una conseguenza della necessità di modificare la modalità di valutazione, probabilmente data dall'inadeguatezza dei tradizionali strumenti e metodi di valutazione utilizzati in presenza.

Un secondo aspetto osservato è stato il tipo di difficoltà riscontrate nella valutazione in DaD, focalizzando l'attenzione da un lato su problemi nel raggiungimento della qualità della valutazione (Figura 1), dall'altro su difficoltà riguardanti la gestione delle prove e la valutazione delle abilità degli apprendenti (Figura 2).

Per quanto riguarda il primo aspetto indagato, è stato chiesto ai docenti di indicare il grado di difficoltà (per niente, poca, abbastanza, molta) nel raggiungere la qualità dei seguenti criteri di valutazione: la validità, la rappresentatività e l'appropriatezza, l'accettabilità e la fattibilità ed economicità³ (Figura 1).

³ Di ognuno di questi criteri il questionario forniva una breve definizione.

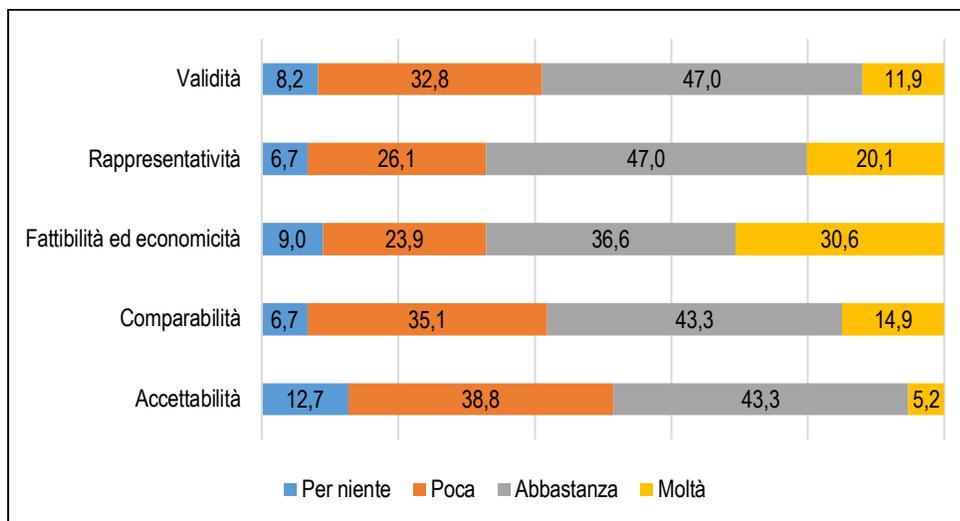


Figura 1. Difficoltà nel raggiungere la qualità della valutazione (valori percentuali)

Dalle risposte al questionario riportate nella Figura 1, emerge abbastanza chiaramente che il criterio più difficile da soddisfare è stato quello relativo alla fattibilità ed economicità della prova, in quanto la maggior parte dei docenti ha affermato di aver avuto “abbastanza” (36,6%) o “molte” (30,6%) difficoltà. Questo dato è in linea con le dichiarazioni dei docenti presentate nella Tabella 1, dalle quali emerge un aumento del tempo dedicato alla valutazione in DaD rispetto alla modalità in presenza. Per quanto riguarda gli altri aspetti osservati, invece, la maggior parte delle risposte fornite sono state per le opzioni “abbastanza” o “poche”, suggerendo quindi che i docenti hanno avuto meno problemi nel confrontarsi con questi criteri di qualità della valutazione.

In merito al tipo di difficoltà, invece, è stata proposta ai docenti una domanda in cui era richiesto di selezionare una o più aree della valutazione in cui avevano riscontrato problemi, e che potrebbero essere raggruppate in due macrocategorie: problemi legati alla pianificazione e somministrazione delle prove (Tabella 2); problemi legati alla valutazione degli aspetti della competenza linguistico-comunicativa (Tabella 3).

Tabella 2

Difficoltà nella pianificazione e somministrazione delle prove

Tipo di difficoltà	Risposte
Creare test adatti	54,5%
Utilizzare strumenti di valutazione adatti	62,7%
Verificare l'autenticità e l'integrità della prova	76,9%

Tabella 3

Difficoltà nella valutazione di aspetti della competenza linguistico-comunicativa

Tipo di difficoltà	Risposte
Valutare le competenze metalinguistiche	29,9%
Valutare le conoscenze linguistiche	20,1%
Valutare l'interazione orale	43,3%
Valutare l'interazione scritta	31,3%
Valutare le produzioni orali	35,1%
Valutare le produzioni scritte	32,8%

I risultati nelle Tabelle 2 e 3 mostrano che i docenti hanno avuto difficoltà, in maniera maggiore o minore, in tutte le opzioni proposte.

Per quanto riguarda i problemi relativi alla valutazione di aspetti della competenza linguistico-comunicativa, le difficoltà sono state basse per la valutazione delle conoscenze linguistiche e metalinguistiche

(20,1% per le prime, 29,9% per le seconde), moderate per l'abilità di interazione scritta (31,3%) e per quelle di produzione scritta (32,8%) e orale (35,1), mentre più alte quelle per l'interazione orale (43,3%).

Dai risultati si evince, inoltre, che la maggior parte dei docenti ha incontrato difficoltà elevate nella pianificazione e somministrazione delle prove. L'aspetto più critico è stato quello relativo alla verifica "dell'autenticità e dell'integrità della prova" (76,9%), che nel questionario abbiamo inteso e definito come la possibilità di accertarsi che i test fossero svolti correttamente dagli apprendenti senza supporto di strumenti non autorizzati o aiuti esterni. Inoltre, più della metà degli intervistati ha dichiarato di aver avuto problemi sia nell'utilizzare degli strumenti adatti alla valutazione (62,7%), che nella creazione di prove adeguate (54,5%).

Come per i risultati presentati nella Tabella 1, trattandosi anche in questo caso di risposte chiuse, non è stato possibile indagare in maniera approfondita nei confronti di quale specifica attività messa in pratica avessero riscontrato difficoltà i docenti. Confrontando, però, questi risultati con quelli presentati precedentemente, si potrebbe pensare che le tradizionali modalità di valutazione solitamente utilizzate in classe fossero inadeguate per la valutazione di specifici aspetti della competenza linguistico-comunicativa in DaD, poiché i docenti non avevano né gli strumenti adatti, né la possibilità di creare delle prove che fossero fattibili in questa specifica modalità didattica.

Un dato emerso dalle risposte presentate nella Tabella 2, e che a nostro avviso è particolarmente importante, è quello riguardante i problemi relativi alla sorveglianza delle modalità di svolgimento delle prove. Per i docenti è fondamentale, infatti, controllare che le prove siano condotte in maniera corretta, ma gli strumenti tecnologici non permettono di avere la stessa garanzia che si ha quando la prova è svolta in presenza. Al fine di comprendere le soluzioni messe in atto dagli insegnanti rispetto a tale problema, è stato chiesto loro di selezionare una o più modalità di sorveglianza applicate durante lo svolgimento delle prove tra condivisione dello schermo, sorveglianza tramite webcam, uso di software appositi, oppure di indicare "altro" nel caso in cui avessero scelto di utilizzare altri mezzi. Quest'ultima opzione poteva essere selezionata anche se venivano utilizzate le altre modalità incluse nella domanda.

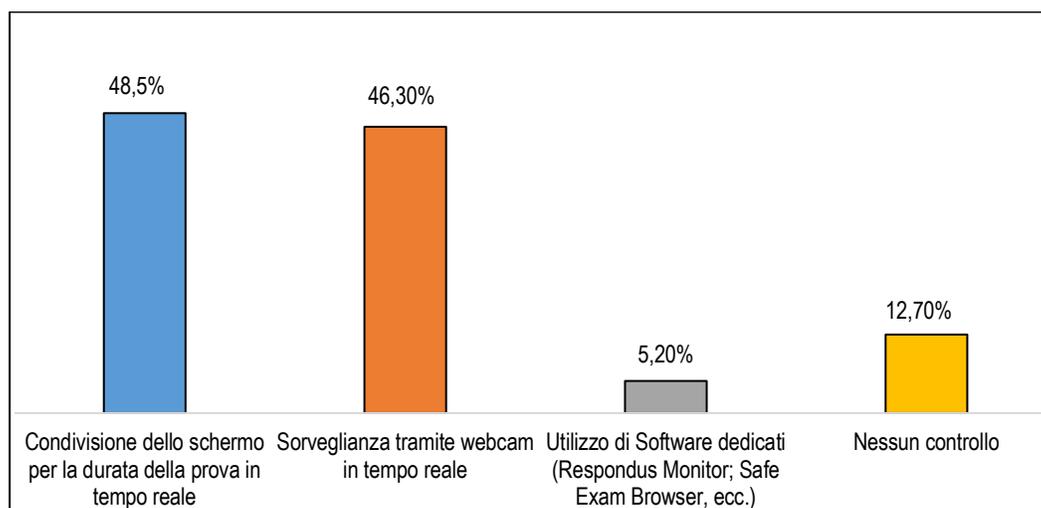


Figura 2. Modalità di controllo dello svolgimento della prova

Dai risultati presentati nella Figura 2 emerge che la maggior parte dei docenti ha deciso di controllare lo svolgimento della prova attraverso la condivisione dello schermo (48,5%) o la sorveglianza tramite webcam (46,3%). Al contrario, soltanto il 5,2% ha optato per l'utilizzo di software appositamente ideati per la sorveglianza del computer. Un aspetto interessante è che spesso i docenti hanno deciso di combinare queste tre opzioni, mettendo in pratica modalità di controllo diverse contemporaneamente. Inoltre, il 29,9% dei docenti ha selezionato l'opzione "altro", in quanto le loro modalità di sorveglianza non rientravano in quelle previste dalla domanda. Innanzitutto, un dato importante riguarda coloro che nel selezionare "altro" hanno dichiarato di non aver controllato in alcun modo lo svolgimento della prova (12,7%). Le motivazioni di quest'ultima scelta sono state molteplici. Alcuni docenti hanno fatto leva sul senso di responsabilità degli alunni e sulla fiducia riposta in loro, mentre altri hanno seguito le indicazioni dell'istituto o non hanno considerato necessario verificare. Altri docenti che hanno selezionato l'opzione "altro", invece, hanno dichiarato di aver messo in atto tipi di sorveglianza diversi e alternativi. Alcuni di loro, ad esempio, hanno utilizzato test con limite

di tempo, che rendevano quindi difficile consultare materiali o utilizzare strumenti non autorizzati. Importante è sottolineare che alcuni docenti, invece, hanno optato per la creazione di prove personalizzate e non giudicate in base a risposte oggettivamente corrette o sbagliate, ma in cui gli apprendenti dovevano mettere in gioco tutte le loro abilità linguistiche (cfr. par. 2.2.2.).

2.2.2. Valutazione a distanza: modalità, strumenti e test

La seconda parte del questionario è stata dedicata ad indagare le pratiche valutative utilizzate in DaD, in termini sia di tipo di valutazione che di attività proposte.

Una prima osservazione ha riguardato il cambiamento del tipo di valutazione utilizzato in presenza (Figura. 3) e in DaD (Figura. 4). A tal proposito, è stato richiesto ai docenti di indicare su una scala da “mai” a “spesso” il livello di utilizzo in presenza e in DaD di sei tipi di valutazione: valutazione separata; valutazione globale; verifica finale; valutazione continua; auto-valutazione; valutazione tra pari. Questi diversi tipi di valutazione sono alcuni di quelli presentati nel Quadro Comune Europeo di Riferimento (Consiglio d’Europa, 2002), pertanto sono concetti con cui si presume che i docenti abbiano familiarità. La valutazione separata è intesa come verifica dei diversi aspetti della competenza linguistico-comunicativa, come ad esempio il lessico e la grammatica, in maniera isolata. Al contrario, la valutazione globale valuta complessivamente le diverse abilità e conoscenze degli apprendenti. La verifica finale indica la valutazione del raggiungimento di specifici obiettivi in maniera puntuale, quindi tramite un test. La valutazione continua, invece, si occupa di verificare in itinere le competenze degli apprendenti tramite diverse attività individuali o in gruppo. Infine, l’auto-valutazione fa riferimento alla verifica da parte dell’apprendente stesso delle proprie abilità, mentre la valutazione tra pari indica il controllo da parte di un apprendente del livello in cui si trova un altro (cfr. Consiglio d’Europa, 2002).

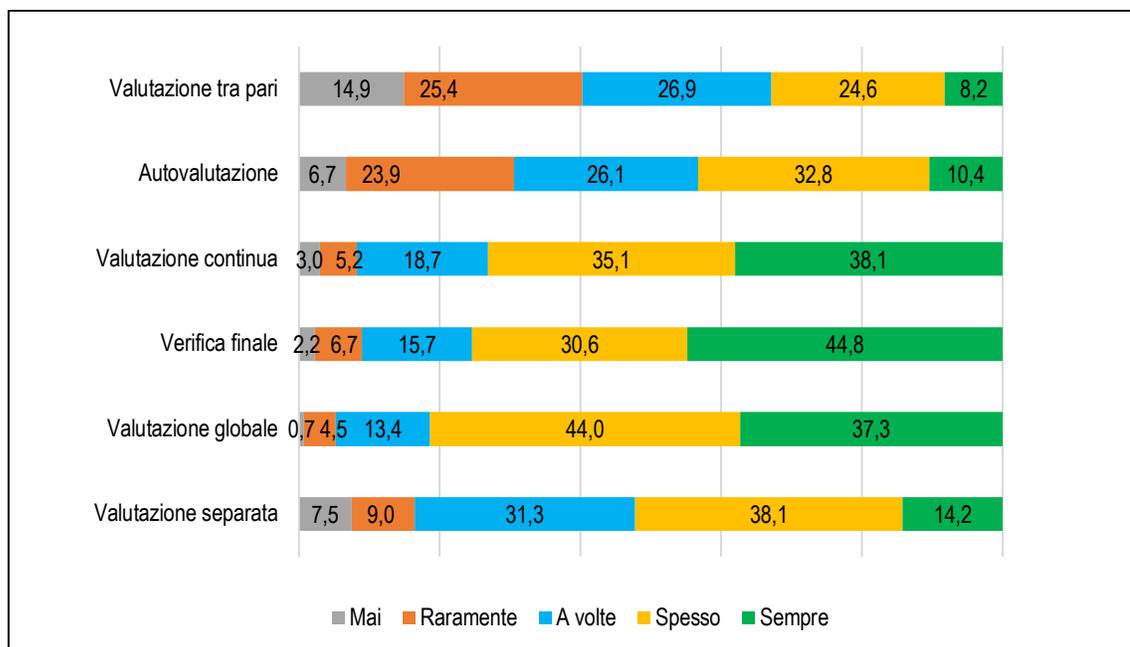


Figura 3. Tipo di valutazione in presenza

I risultati relativi al tipo di valutazione utilizzato in presenza (Figura 3), mettono in luce una preferenza dei docenti per la verifica finale. Quasi la metà degli insegnanti, infatti, ha affermato di utilizzare sempre (44,8%) questa modalità valutativa. Risulta essere abbastanza frequente, ma in maniera leggermente inferiore rispetto alla verifica finale, l’utilizzo della valutazione continua (38,1% per “sempre”) e di quella globale (37,3 per “sempre”). Sono state meno utilizzate, invece, la valutazione separata, la valutazione tra pari e l’autovalutazione. I dati qui presentati, quindi, suggeriscono che nella modalità in presenza i docenti tendono a preferire in modo assoluto un tipo specifico di valutazione.

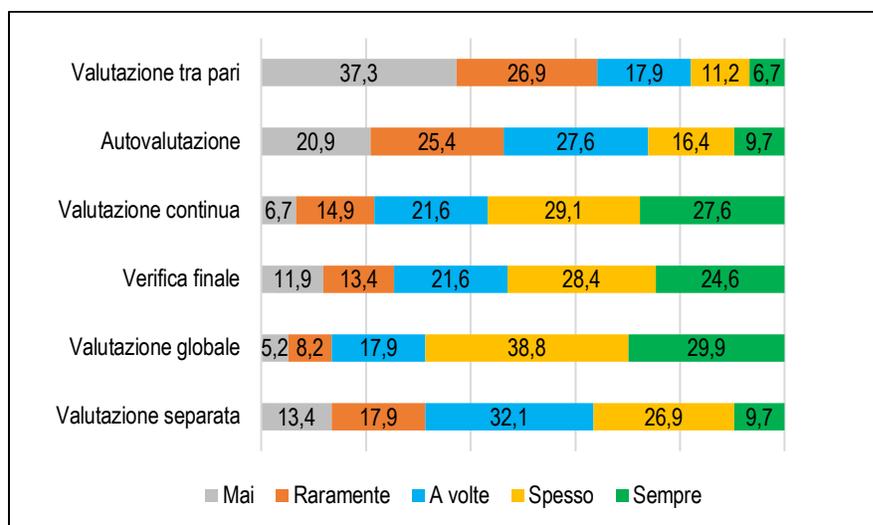


Figura 4. Tipo di valutazione in DaD (valori percentuali)

Per quanto riguarda, invece, il tipo di valutazione utilizzato in DaD, i risultati presentati nella Figura 4 mostrano una situazione diversa rispetto alla didattica in presenza. Nella modalità valutativa a distanza, infatti, non c'è la tendenza ad avere una preferenza assoluta per un tipo di valutazione rispetto ad un altro. Al contrario, dai dati emerge un uso combinato di diversi tipi di valutazione, suggerito dal numero maggiore di docenti che ha risposto di utilizzare "spesso" o "a volte" uno specifico tipo di valutazione. Inoltre, a differenza della didattica in presenza, in cui la scelta preferita dei docenti era la verifica finale, i dati mostrano che in DaD c'è un uso più diffuso della valutazione globale. È confermata, invece, la tendenza al mancato ricorso alla valutazione tra pari e all'autovalutazione, il cui utilizzo risulta ulteriormente diminuito rispetto alla didattica in presenza.

Il confronto tra il tipo di valutazione attuato in presenza e quello a distanza suggerisce quindi un cambiamento nelle scelte valutative dei docenti, i quali passano da una preferenza quasi assoluta per un singolo tipo di valutazione ad una tendenza ad attuare in maniera complementare diversi tipi di valutazione. Questo permette di osservare la competenza degli apprendenti da diversi punti di vista e soprattutto in maniera più completa.

Al fine di indagare come questi diversi tipi di valutazione fossero stati declinati nella pratica, è stato chiesto ai docenti di segnalare il tipo di attività utilizzate per valutare i diversi aspetti della competenza linguistico-comunicativa, dando loro la possibilità di selezionare per ogni aspetto una o più opzioni (Tabella 4).

Tabella 4
Tipo di attività (valori in percentuale)

	Conoscenze		Comprensione		Produzione		Interazione	
	linguistiche	metalinguistiche	scritta	orale	scritta	orale	scritta	orale
Nessuna	3,0	11,9	7,5	6,0	7,5	7,5	29,9	10,4
Role-play	26,9	20,9	6,0	26,9	4,5	40,3	5,2	43,3
Monologhi	23,9	15,7	3,0	16,4	6,0	45,5	6,0	9,0
Chat/forum	20,9	20,9	14,2	11,9	26,1	13,4	38,8	18,7
Cloze	38,1	11,9	29,9	16,4	13,4	2,2	10,4	6,0
Domande aperte	58,2	38,8	47,8	46,3	43,3	30,6	20,1	29,1
Domande a scelta multipla	51,1	26,1	54,5	43,3	16,4	6,7	17,2	4,5
Presentazioni orali	22,4	11,2	3,7	15,7	2,2	35,8	5,2	20,9
Lavori di gruppo	14,2	11,9	6,0	10,4	11,2	22,4	10,4	15,7
Testi scritti	30,6	9,0	34,3	2,2	41,0	2,2	17,9	1,5

Dai risultati riportati nella Tabella 4 emerge innanzitutto che tutti gli aspetti sono stati considerati in fase di valutazione. Infatti, per ogni tipo di conoscenza e abilità presa in considerazione, solo una piccola fascia di docenti ha dichiarato di non aver utilizzato alcun tipo di attività per valutarla. Fa eccezione soltanto l'interazione scritta, che il 29,9% dei docenti ha affermato di non valutare. Questo potrebbe essere dato dalla difficoltà nel giudicare questa abilità, come è emerso dalle dichiarazioni degli stessi docenti (cfr. Tabella 3, par. 2.2.1.). Inoltre, alcuni insegnanti hanno selezionato l'opzione "Altro" dichiarando di aver utilizzato alcune di queste attività anche per valutare le conoscenze relative alla cultura e civiltà del paese della lingua studiata.

Per quanto riguarda le attività più frequenti, dai dati si evince che c'è stato un uso diffuso di domande aperte e domande a scelta multipla. Le seconde sono state per lo più utilizzate nel modo più tradizionale, ossia per valutare le conoscenze e le abilità di comprensione, mentre le prime sono state usate molto anche per valutare le abilità di produzione. A queste si affiancano anche molti compiti comunicativi, come ad esempio *role-play* o monologhi, nonché altre attività, come ad esempio presentazioni orali e lavori o progetti di gruppo. In merito a questi tipi di attività, i dati suggeriscono anche che c'è una preferenza ad usarne alcune rispetto ad altre in base all'aspetto che si valuta. Ad esempio, c'è un frequente utilizzo dei *role-play* per verificare le abilità di produzione (40,3%) e interazione orale (43,3%). Infine, gli altri tipi di attività vengono utilizzati complessivamente in maniera inferiore rispetto a quelli presentati precedentemente e soprattutto in maniera spesso concentrata per valutare uno o due aspetti specifici, come ad esempio l'uso di testi scritti per valutare la produzione scritta (41%) o le conoscenze linguistiche (30,6%), oppure i cloze per le attività di comprensione scritta (29,9%) o per valutare le conoscenze linguistiche (38,1%).

I risultati presentati in questa seconda parte, quindi, suggeriscono che in DaD i docenti abbiano preferito mettere in atto una valutazione che tenesse conto nel complesso delle diverse abilità e conoscenze degli alunni, integrando diverse pratiche e diverse attività. Questa modalità valutativa permette di avere un'idea completa della competenza linguistico-comunicativa degli apprendenti, in quanto la stessa viene verificata tenendo conto di diversi elementi e avendo una visione d'insieme. Infatti, dai dati è emerso che insieme alle classiche prove oggettive a cui solitamente corrisponde un'unica risposta corretta, come ad esempio i *cloze* e le domande a scelta multipla, gli apprendenti sono stati chiamati a svolgere delle attività in cui era richiesto loro di utilizzare la lingua in diverse situazioni e in cui era necessario mettere in gioco diverse abilità e competenze.

Trattandosi di risposte chiuse, anche in questo caso, non è stato possibile osservare nel dettaglio il tipo di prove proposte, né le ragioni esistenti dietro le scelte dei docenti, ma i dati presentati fin qui ci permettono di fare delle considerazioni generali. A nostro avviso, infatti, le scelte valutative dei docenti sono state dettate dalla necessità di superare i problemi di adattamento alla DaD, ma soprattutto dalle preoccupazioni relative all'impossibilità di controllare lo svolgimento delle prove. Nonostante queste soluzioni derivino da problemi legati all'inadeguatezza delle tradizionali pratiche valutative per la DaD, riteniamo che rappresentino anche uno dei punti forti della modalità a distanza e mettano in luce un importante elemento da tener conto. La necessità dei docenti di riadattare i propri strumenti e le proprie pratiche didattiche alla nuova situazione, infatti, li ha portati ad orientarsi verso una nuova modalità di valutazione, in cui le prove mirano a verificare le effettive capacità di uso della lingua da parte degli apprendenti, piuttosto che misurare le conoscenze grammaticali o singole abilità. Questo tipo di attività, a nostro avviso, sono molto simili a quelli in letteratura sono conosciuti come *task* (cfr. par. 3).

3. La valutazione task-based e l'adeguatezza funzionale

Nella sezione precedente, i dati emersi dalle risposte al questionario hanno mostrato come alcuni docenti abbiano optato per la creazione di prove globali basate su compiti in cui si richiede agli apprendenti un'elaborazione personale e il ricorso ad abilità linguistiche integrate. Per valutare questo tipo di compiti è pertanto necessaria una valutazione volta a misurare la capacità degli apprendenti di utilizzare le loro abilità linguistiche per raggiungere gli obiettivi di un *task*, in linea con i principi del *task-based performance assessment* (cfr. Brown, Hudson, Norris & Bonk, 2002, p.9). Dopo aver introdotto e discusso brevemente i principi della valutazione task-based, illustreremo una proposta di valutazione basata sul concetto di AF.

3.1. Valutare con i task

Nel caso dell'apprendimento linguistico, l'obiettivo della valutazione è legato alla concezione che si ha di lingua, di apprendimento linguistico e di didattica delle lingue. La conoscenza dichiarativa, il contenuto di una particolare sequenza di un corso di lingua, le abilità linguistiche, la capacità comunicativa, il grado di

acquisizione, la competenza metalinguistica o strategica e altre caratteristiche che possono costituire l'obiettivo di una valutazione sono descritte in termini di "costrutto". Quest'ultimo è un concetto teorico astratto che deve essere esplicitamente descritto e specificato nella progettazione dell'intervento valutativo, al fine di definire una base non solo per lo sviluppo di una data valutazione ma anche per l'interpretazione e l'uso dei suoi risultati (Bachman & Palmer, 2010).

In designing, developing, and using language assessments, we can define the construct from a number of perspectives, including everything from the content of a particular part of a language course to a needs analysis of the components of language ability that may be required to perform language use tasks in a target language use domain, to a theoretical model of language ability. (Bachman & Palmer, 2010, p.43)

A partire dagli anni Sessanta del secolo scorso, gli approcci didattici sono incentrati sul concetto di competenza comunicativa e culturale, e, di conseguenza, il costrutto della valutazione si è basato sulla capacità del candidato nel saper usare operativamente la lingua obiettivo. In questa stessa direzione, nel QCER la competenza è descritta in termini di "saper fare" e gli strumenti di verifica sono elaborati per elicitarne un dato comportamento dal quale poter comprendere il grado di competenza raggiunta in una L2 da un individuo. Partendo dall'osservazione di una determinata prestazione richiesta in base a specifici compiti o consegne si può rilevare un determinato comportamento linguistico e interattivo, misurarlo in base a parametri precedentemente stabiliti e valutare la competenza sottostante che lo ha determinato (Ambroso, 1993, p.3). Quindi possiamo affermare che, nell'ambito della valutazione linguistica, il ricorso a task, più o meno collegati con l'uso reale della lingua, è abbastanza diffuso e consolidato e deriva direttamente dall'orientamento didattico comunicativo⁴.

In letteratura sono presenti numerose definizioni di task (cfr. Ellis, 2003; Long, 2015; Nunan, 2004), che, pur mettendo l'accento su aspetti diversi, concordano nel descrivere il task come un compito che richiede agli apprendenti di raggiungere un obiettivo extra-linguistico tramite l'utilizzo delle loro risorse linguistiche, focalizzando l'attenzione sul significato e non sulla forma (Cortés Velásquez & Nuzzo, 2018). Possiamo adottare una definizione abbastanza ampia di task:

a pedagogical task is a piece of classroom work that involves learners in comprehending, manipulating, producing or interacting in the target language while their attention is focused on mobilizing their grammatical knowledge in order to express meaning, and in which the intention is to convey meaning rather than to manipulate form. (Nunan, 2004, p. 4)

Quando si parla di valutazione linguistica task-based, si intende un tipo di pratica valutativa in cui il costrutto è interamente basato sulla realizzazione di task legati a specifiche situazioni comunicative (Brindley, 1994). Il costrutto, in questo approccio non si basa sulla lingua come oggetto, intesa in termini di conoscenze linguistiche dichiarative, ma sul concetto di task come unità di analisi e sulla prestazione come mezzo per verificare la capacità degli apprendenti di portare a termine il compito (Brown et al., 2002; Norris, Brown, Hudson & Bonk, 2002). I task replicano situazioni, contenuti, interlocutori e altri aspetti al fine di ottenere indicazioni affidabili della misura in cui gli apprendenti sanno gestire situazioni e compiti della vita reale, e presentano numerosi vantaggi sia in ambito di valutazione formativa che sommativa. L'introduzione della valutazione linguistica task-based risponde alla necessità di allineare insegnamento, curriculum e valutazione, al fine di promuovere un apprendimento effettivo della lingua (Norris, 2016, p.237). In questa linea sono stati sviluppati sistemi valutativi interamente basati sul task come quello della certificazione brasiliana Celpe-Bras (cfr., fra gli altri, Scaramucci et al. 2019) o l'approccio basato sullo scenario (cfr., fra gli altri, Bonvino & Fiorenza 2020; Fiorenza & Diego Hernández, 2020; Purpura, 2016).

Benché l'impegno di task nella didattica sia abbastanza diffuso in ambito valutativo, come osservano Cortés Velásquez e Nuzzo (2018), nella maggior parte dei casi si tratta di un uso 'strumentale' del task, utilizzato per elicitarne una determinata prestazione linguistica, che viene poi valutata in base a criteri di efficacia comunicativa e anche, o talvolta esclusivamente, di accuratezza morfosintattica e lessicale.

Questo disallineamento tra insegnamento e valutazione è riconducibile ad una delle questioni ancora irrisolte nella letteratura sulla valutazione linguistica task-based, ossia il modo in cui bisogna valutare la prestazione elicitata dal task. Possiamo infatti distinguere due posizioni opposte: una concezione forte e una concezione più debole. Nella prima, la valutazione linguistica è focalizzata sulla capacità di portare a

⁴ Cfr. Bachman (2002); Ellis (2003); Nunan (2004).

compimento il task e la *performance* viene vista come finalizzata a verificare il raggiungimento di questo obiettivo. Nella concezione più debole, invece, il task è visto come uno strumento per elicitare una *performance* valutata secondo alcuni criteri (Timpe-Laughlin, 2018). Quali siano questi criteri è il nodo centrale ampiamente dibattuto, soprattutto per quanto riguarda la valutazione dell'accuratezza della prestazione linguistica (cfr., tra gli altri, Bachman, 2000, Norris et al., 2002). Come sostiene Long (2015), infatti, non solo è difficile stabilire fino a che punto sia necessario tener conto degli aspetti formali della produzione, ma anche riuscire a trovare degli strumenti per poterli valutare in maniera oggettiva. Inoltre, un aspetto fondamentale di cui tener conto è che la scelta del tipo di valutazione da mettere in atto dipende principalmente dall'obiettivo della valutazione e dall'uso se ne vuole fare (Norris, 2016; Timple-Laughlin, 2018).

In tal senso, una possibile soluzione al problema della valutazione della performance del task è stata data da Kuiken e Vedder (2017) con l'introduzione dell'AF intesa come valutazione di una performance in termini di riuscita nell'esecuzione di un task. Come vedremo nel prossimo paragrafo, infatti, valutare la performance durante un task in termini di AF, permette di giudicare il raggiungimento degli obiettivi del task, senza però trascurare gli aspetti linguistici della produzione.

3.2. L'adeguatezza funzionale: uno strumento per valutare i task

L'AF è definita come un costrutto interpersonale, legato al *task* e alla sua "felice" esecuzione (Kuiken & Vedder 2017, p. 323). Il messaggio prodotto, infatti, è considerato dal ricevente funzionalmente adeguato a seconda del tipo di task che viene svolto ed è giudicato come tale in relazione alle massime conversazionali di Grice (1975). In quest'ambito assumono particolare importanza sia il contesto sociale, sia il task stesso (Kuiken & Vedder, 2018, p. 267). In questo senso, la valutazione dell'AF è un tipo di valutazione task-based che permette di tener conto sia del raggiungimento degli obiettivi del task, sia della valutazione dell'adeguatezza della performance da un punto di vista più formale. L'AF, infatti, è caratterizzata da "a strong relationship between successful task fulfillment and the appropriateness and effectiveness of linguistic output" (Nuzzo & Bove, 2020, p. 10).

Kuiken e Vedder (2017) hanno ideato la scala di valutazione dell'AF, riadattando la scala olistica di efficacia comunicativa utilizzata per lo studio *CALC* (Kuiken, Vedder, & Gilabert, 2010). Le caratteristiche di questa scala di valutazione sono le seguenti⁵:

- a) la suddivisione dell'AF in quattro dimensioni valutabili (contenuto, requisiti del *task*, comprensibilità, coerenza e coesione);
- b) l'elaborazione di descrittori oggettivi e quantificabili, che possono essere utilizzati sia da parte di valutatori esperti che da valutatori non esperti;
- c) la possibilità di applicazione sia per la valutazione di produzioni in L1 che in L2;
- d) la dimensione funzionale della performance del task considerata in maniera indipendente rispetto agli indici di complessità, accuratezza e fluenza (CAF, cfr. Pallotti, 2009).

La scala di valutazione dell'AF di Kuiken e Vedder (2017) per la produzione scritta prevede quattro sotto-scale corrispondenti alle quattro dimensioni che compongono il costrutto: contenuto, requisiti del *task*, comprensibilità, coerenza e coesione. Il contenuto valuta l'adeguatezza numerica e la pertinenza delle idee e dei concetti presenti nel testo. I requisiti del task giudicano il rispetto delle istruzioni del task in riferimento a genere testuale, registro e atti linguistici. La comprensibilità valuta lo sforzo che il lettore deve compiere per capire i contenuti e le idee del testo. Infine, la dimensione di coerenza e coesione, verifica la presenza o assenza di strumenti coesivi e salti logici nel testo. Ogni dimensione è rappresentata da una scala Likert a sei punti, con descrittori specifici per ogni punto⁶.

Molti studi hanno dimostrato la possibilità di applicare le scale della AF per valutare diversi tipi di testi scritti (cfr. per il testo argomentativo Kuiken & Vedder, 2017; per il testo narrativo Pagliara, 2017; per il testo regolativo, Faone & Pagliara; per tutti e tre i generi testuali Del Bono, 2019; Orrù, 2019) utilizzando test di tipo statistico per verificarne l'affidabilità in termini di *inter-rater reliability* (cfr. Le Breton & Senter, 2008). Altri studi hanno adattato le scale per valutare monologhi argomentativi (Kuiken & Vedder, 2018; Nuzzo & Cortes-Velasquez, 2017;), confermando la possibilità di utilizzarle anche per questo tipo di produzioni orali. Inoltre, De Meo, Maffia e Vitale (2019) hanno messo in luce i benefici che questo strumento di valutazione potrebbe

⁵ Adattati da Kuiken e Vedder (2017, p. 326).

⁶ Si veda Kuiken e Vedder (2017) per la versione originale delle scale.

avere nell'ambito della certificazione per migranti, soprattutto in termini di valorizzazione delle abilità di apprendenti con un livello basso di competenza.

Nonostante i risultati di queste ricerche confermino la validità della scala di valutazione dell'AF, lo strumento rimane appannaggio della ricerca teorica, senza trovare applicazioni didattiche. A nostro avviso, invece, sono numerose le implicazioni pedagogiche e i vantaggi che la scala potrebbe apportare in ambito didattico, sia in termini di valutazione sommativa che formativa.

Innanzitutto, come suggerito dagli autori stessi delle scale, la presenza di descrittori specifici per ogni livello scalare permette ai docenti di dare un *feedback* preciso agli apprendenti per ognuna delle dimensioni (Kuiken e Vedder, 2017, p.332).

Un altro elemento molto importante sottolineato da Kuiken e Vedder (2017) è la possibilità di utilizzare le scale da parte di valutatori non esperti, in quanto questo ha notevoli implicazioni in ambito didattico. Nel loro studio Kuiken e Vedder (2017) hanno definito valutatori non esperti coloro che non hanno alcun tipo di esperienza in ambito valutativo, ma che tramite un training preparatorio possono acquisire familiarità con l'utilizzo delle scale e quindi giudicare i testi in maniera affidabile. I risultati dei vari studi sull'applicabilità della scala (cfr. Del Bono, 2019; Kuiken & Vedder, 2017; Nuzzo & Cortes-Velasquez, 2017; Pagliara, 2017) hanno confermato la possibilità da parte di valutatori non esperti di poterla utilizzare per valutare produzioni scritte e orali. Questo dato suggerisce un'importante implicazione per l'applicazione delle scale in ambito didattico, in quanto la possibilità di acquisire familiarità con lo strumento anche da parte di coloro che non sono soliti valutare, permette ai docenti di utilizzarlo per svolgere attività di autovalutazione o valutazione tra pari nell'ambito dei corsi in classe. I valutatori non esperti che hanno preso parte ai diversi studi di applicabilità della scala, infatti, erano spesso studenti universitari di lingue, con una discreta familiarità con i concetti linguistici affrontati. Sebbene questo possa risultare difficile per apprendenti con scarsa competenza in L2, si potrebbero suggerire degli adattamenti della scala al contesto, come ad esempio l'uso di una versione nella L1 degli apprendenti, oppure un aumento delle sessioni di *training* per essere sicuri che gli apprendenti abbiano capito come utilizzare lo strumento.

L'applicazione della scala per attività di autovalutazione e valutazione tra pari, inoltre, potrebbe favorire la riflessione sulle abilità di scrittura e avere un ulteriore vantaggio in termini di aumento della consapevolezza nella scrittura in L1 e in L2 (cfr. Nuzzo & Bove 2020, p. 23).

I vantaggi delineati fin qui sono validi sia per il tradizionale contesto di classe, ma anche e soprattutto per la DaD. Le difficoltà avute dai docenti nel valutare in DaD, infatti, potrebbero essere state superate mettendo in pratica un tipo di valutazione della performance del task in termini di AF, in quanto risulta essere un tipo di valutazione flessibile e facilmente adattabile a contesti e situazioni diverse. Nel paragrafo seguente ci occuperemo di presentare un esempio di applicazione didattica della scala di AF.

3.3. Scala di valutazione della AF: un esempio di applicazione

Come si è affermato nel paragrafo precedente, la scala di AF è uno strumento flessibile che si presta a valutare un testo in merito sia al raggiungimento degli obiettivi del *task*, sia all'adeguatezza della produzione scritta e che potrebbe essere adatto sia alla valutazione in presenza sia a quella in DaD. Tale affermazione può essere esemplificata riportando alcuni esempi di applicazione della scala di AF⁷ utilizzati per valutare diversi generi testuali⁸.

Partiamo analizzando innanzitutto lo strumento utilizzato per elicitare la performance del parlante, ossia il *task*. Come è stato confermato dagli studi sull'applicabilità delle scale (si veda §3.2, possono essere utilizzati diversi tipi di *task* a seconda del tipo di testo che si vuole valutare. Di seguito viene riportato ad esempio un *instruction task*⁹, nel quale viene richiesto all'apprendente di scrivere un testo regolativo, in questo caso un bigliettino con delle istruzioni relative ad una casa vacanze (Figura 5). Si tratta di un tipo di attività con la quale solitamente gli apprendenti hanno una certa familiarità, in quanto è molto comune sia in ambito didattico che certificatorio.

⁷ I dati qui riportati sono stati raccolti nell'ambito dello studio di Del Bono (2019) sulla verifica dell'applicabilità della scala su diversi generi testuali.

⁸ È di particolare interesse per questo contributo sottolineare che le attività riportate nel §3.3 sono state svolte completamente online, tramite risposte fornite su Moduli Google e in modalità asincrona, senza creare problemi in termini di adattabilità dello strumento o di autenticità della valutazione

⁹ Il task utilizzato è un adattamento in italiano di Del Bono (2019) a partire da un task in inglese ideato da Sofia Martín Laguna (Universitat Jaume I).

TASK D'ISTRUZIONI

Tramite un sito di prenotazioni online hai affittato la tua casa ad una giovane coppia, Maria e Max, che sta visitando la tua città. Dato che non sarai lì quando arriveranno, decidi di lasciare un biglietto con le istruzioni relative alla casa.

Il tuo biglietto deve includere le seguenti informazioni:

- Inizia salutando i tuoi ospiti.
- Spiega dove possono trovare ciò di cui possono aver bisogno durante la loro permanenza.
- Spiega brevemente come usare alcuni apparecchi elettronici.
- Spiega come dovrebbero lasciare la casa.

Figura 5. Task d'istruzioni

Come si evince dal testo riportato nella Figura 5, il *task* utilizzato ha un input che indica all'apprendente il tipo di testo da produrre, gli argomenti da trattare e in che ordine, guidandolo in fase di stesura. Questa organizzazione del *task*, inoltre, fornisce al docente una serie di informazioni utili per valutare l'AF della produzione scritta. Il *task*, infatti, permette facilmente di individuare quali siano gli aspetti da valutare in riferimento alle quattro dimensioni di AF. Ad esempio, per valutare il contenuto, sarà sufficiente verificare che siano presenti nel testo le informazioni richieste nell'elenco puntato e che le idee espresse siano efficaci e pertinenti all'argomento trattato. In questo senso, quindi, è possibile comprendere con più chiarezza perché valutare l'AF della performance dipende dal particolare *task* svolto e quindi è interpretabile in termini di riuscita esecuzione del *task*, ma allo stesso tempo è strettamente legata ad aspetti prettamente legati alla produzione valutata.

Osserviamo, adesso, uno dei testi regolativi (Figura 6) elicitati a partire dall'*instruction task* appena presentato. Si tratta di una produzione scritta svolta da una studentessa universitaria, iscritta al primo anno di un corso di laurea in studi italiani dell'Università di Amsterdam, con un livello intermedio di italiano LS¹⁰.

Ciao Maria e Max,

Benvenuti alla mia casa. Come vedete il mio appartamento è molto piccolo, dunque non sarà difficile trovare tutto. Nella cucina potete trovare tutto quello che avete bisogno per il cuocere e mangiare. Aprite proprio tutto e troveremo le cose. Nello spazio dove entrate, trovate un'armadio in cui sune gli asciugamani. Alla sinistra c'è il bagno dove trovate un'altra porta dove c'è la lavatrice, su cui ho attaccato una carta con le istruzioni. Lì è anche un'armadio in cui c'è la roba con cui potete lavare i vestiti. Poi al camera da letto c'è un'armadio per i vestiti. Al balcone potete asciugare i vestiti. La tv è nel soggiorno e dovete usare senza il remote. Non c'è una macchina per fare caffè. L'appartamento è molto semplice ma comodo. State attenti di non disturbare i vicini di casa, perché sono abbastanza vecchi e non piaciono di essere disturbati. Quando andate via, per favore chiudete tutte le finestre e la poi la porta con la chiave.

Figura 6. Produzione elicitata dal *task* d'istruzioni

L'AF del testo riportato nella Figura 6, di cui riportiamo di seguito i giudizi (in Tabella 5), è stata valutata da 5 parlanti nativi di italiano, iscritti ad un corso di laurea magistrale o triennale in lingue presso l'Università degli Studi Roma Tre, considerati valutatori non esperti e sottoposti a due *training* prima dell'utilizzo delle scale.

¹⁰ Il testo qui riportato è una trascrizione fedele della produzione scritta dell'informante. Errori ortografici e morfosintattici, nonché suddivisione in paragrafi del testo, non sono stati alterati.

Tabella 5
Giudizi di AF del testo regolativo

Dimensione	VALUTATORI					MEDIA	SD	CV
	V1	V2	V3	V4	V5			
Contenuto	5	5	6	5	6	5,4	0,548	10,1%
Requisiti del task	5	4	5	4	5	4,6	0,548	11,9%
Comprensibilità	5	5	4	4	5	4,6	0,548	11,9%
Coerenza e coesione	4	4	4	4	3	3,8	0,447	11,8%
Globale	4,75	4,5	4,75	4,25	4,75	4,6	0,224	4,9%

I giudizi presenti nella Tabella 5¹¹ permettono di collocare l'apprendente ad uno specifico livello di AF non solo in termini globali, ma anche in termini di singole dimensioni. Da un lato, infatti, ci danno un'indicazione circa il livello generale di AF dell'informante, ottenuto facendo una media dei singoli punteggi ottenuti nelle quattro dimensioni. Secondo questi risultati si può affermare che l'apprendente abbia un livello intermedio di AF, con una media di 4,6. Questo valore risulta essere anche abbastanza rappresentativo in quanto la deviazione standard (0,224) e il coefficiente di variazione (4,9%) indicano una bassa variabilità tra le medie dei giudizi dei singoli valutatori. Dall'altro, i giudizi dati alle singole dimensioni forniscono informazioni più precise sul livello di AF, permettendo al docente di individuare quali siano gli aspetti nei quali l'apprendente deve migliorare e poter pianificare in maniera più precisa la pratica didattica. Inoltre, in linea con quanto affermato da De Meo, Maffia e Vitale (2019), questi giudizi permettono sin da subito di mettere in luce le potenzialità dell'apprendente. Dai dati, infatti, emerge che l'informante ha ricevuto un ottimo giudizio nel contenuto, con una media di 5,4, ma uno più basso per la coerenza e coesione, con una media di 3,8. Di conseguenza, nonostante l'apprendente nel vedere i voti ottenuti sia consapevole di dover migliorare nell'organizzazione testuale, si rende anche conto di riuscire già ad avere una buona capacità nel presentare alcune idee o concetti (contenuto, comprensibilità) e nel rispettare i requisiti del task.

La possibilità di fornire delle valutazioni separate per le singole dimensioni si ricollega ad un altro dei vantaggi principali della scala. L'insegnante, infatti, può utilizzare i singoli descrittori per dare agli studenti un *feedback* mirato sugli elementi da migliorare, non limitandosi a dare un voto, bensì fornendo una valutazione analitica che illustri punti di forza e debolezza della performance (Kuiken & Vedder, 2017). Nel dare questo specifico tipo di feedback, il docente, oltre a riportare i descrittori di AF relativi ai singoli livelli, può anche aggiungere dei commenti in cui spiega il motivo per cui ha attribuito un dato punteggio ad ogni singola dimensione (Tabella 6)¹².

Tabella 6
Giudizi di AF di un testo argomentativo

Dimensione	Voto	Descrittore livello	Motivazione fornita dal valutatore
Contenuto	4	Il numero di idee è piuttosto adeguato e le idee sono sufficientemente efficaci.	Ho assegnato il punteggio 4 perché secondo me il numero di idee è piuttosto adeguato anche se l'ultima argomentazione (...) risulta a mio avviso un po' debole e poco efficace.
Requisiti del task	4	La maggior parte dei requisiti e dei criteri specifici del task è stata soddisfatta.	Ho assegnato il punteggio di 4 perché ritengo che il registro linguistico sia in alcuni punti troppo colloquiale, come si vede per esempio nell'uso dei punti esclamativi.

¹¹ I giudizi sono espressi su una scala Likert a sei punti, dove 1 è il punteggio più basso e 6 il più alto.

¹² I descrittori riportati nella tabella 6 appartengono alla versione italiana delle scale di AF (cfr. Vedder, 2016), riadattata da Nuzzo & Cortes-Velasquez (2017).

Comprensibilità	4	Il testo è abbastanza comprensibile: in alcune parti non è formulato chiaramente, ma si può capire con qualche sforzo.	Il punteggio che ho dato è 4 perché il testo è in generale abbastanza comprensibile, ma in alcuni punti ho avuto bisogno di ragionare sulla frase e rileggerla più volte per capire il significato che l'autore voleva comunicare, come per esempio in (...)
Coerenza e coesione	3	Il testo è parzialmente coerente: alcuni salti logici provocano interruzioni della coerenza. Il testo è parzialmente coeso: vengono usati alcuni connettivi, per lo più delle congiunzioni. Vengono usati alcuni riferimenti anaforici, ma non sempre in modo adeguato (per es. più di due frasi di seguito contengono lo stesso soggetto esplicito, che risulta ridondante oppure non lo esplicitano quando è necessario). In alcuni casi l'autore fa ricorso a ripetizioni non necessarie.	Ho assegnato 3 perché, benché il testo sia abbastanza coerente e non presenti salti logici, a livello coesivo ci sono moltissime ripetizioni (...) ed è spesso ridondante a livello anaforico.

Nella Tabella 6 sono riportati i giudizi di AF attribuiti da una valutatrice e le motivazioni corrispondenti. A differenza dell'esempio precedente, le valutazioni presentate nella Tabella 6 riguardano un testo argomentativo utilizzato durante uno dei due *training*. In questa fase iniziale, per prendere familiarità con la scala, le valutatrici erano chiamate a giustificare i singoli punteggi dati ad ogni dimensione e questo ha permesso quindi di raccogliere una serie di commenti e riflessioni interessanti. Come si può leggere dalle motivazioni riportate nella Tabella 6, infatti, la valutatrice fornisce delle indicazioni molto dettagliate su quali siano gli aspetti carenti del testo. Ad esempio, in riferimento alla dimensione del contenuto, la stessa sostiene che il punteggio "4" non è relativo al numero di idee, che a suo avviso risultano piuttosto adeguate, ma riguarda l'efficacia di una delle argomentazioni, da lei considerata "un po' debole e poco efficace". Altrettanto dettagliato è il commento fornito per le altre dimensioni, in cui viene specificato quale aspetto particolare ha determinato l'attribuzione dei singoli punteggi, riportando anche delle parti di testo come esempio.

Inoltre, come affermato precedentemente (cfr. §3.2, la struttura della scala di AF permette di utilizzarla anche per attività di autovalutazione o valutazione tra pari che potrebbero avere numerosi benefici in termini di aumento della consapevolezza nella scrittura in L1 e in L2 (cfr. Nuzzo & Bove, 2020). La chiarezza e la semplicità con cui sono formulati i descrittori della scala permettono di poter far acquisire familiarità con il loro utilizzo dopo qualche sessione di *training* iniziale. Come si evince, infatti, dai commenti presentati nella Tabella 6, anche i valutatori non esperti riescono ad utilizzare le scale, giustificando in maniera dettagliata le proprie valutazioni.

I possibili benefici che l'utilizzo della scala come strumento di autovalutazione e valutazione tra pari potrebbe avere in ambito didattico sono ben espressi da alcuni commenti e riflessioni di alcune valutatrici che riportiamo di seguito.¹³

Innanzitutto, dai commenti delle valutatrici sono emerse in generale una serie di vantaggi portati dall'utilizzo della scala come strumento di autovalutazione e valutazione tra pari (1-2).

- 1) "L'utilizzo delle scale mi ha aiutato ad individuare i punti che non andavano bene, perché quando ho scritto il testo per la prima volta non avevo notato certi aspetti."
- 2) "Nel momento in cui mi sono trovata a leggere e valutare i testi delle altre informanti mi sono resa conto di come avessi commesso degli errori in fase di produzione che avrei potuto evitare. Per questo ho deciso di procedere alla correzione anche grazie alle riflessioni che sono scaturite in seguito alle valutazioni delle produzioni delle altre ragazze"

Il commento riportato in (1), ad esempio, conferma la possibilità di utilizzare la scala come strumento di autovalutazione, in quanto i parlanti riescono facilmente ad individuare errori che altrimenti non avrebbero

¹³ I commenti qui riportati provengono da uno studio pilota (Del Bono, 2018) sull'utilizzo delle scale dell'AF per migliorare la scrittura in L1.

notato. La riflessione in (2), invece, mette in luce i benefici che la valutazione tra pari può avere in termini di aumento della consapevolezza nella scrittura. La partecipante, infatti, afferma che non avrebbe notato molti degli aspetti carenti del suo testo se non fosse stata chiamata anche a valutare i testi delle sue colleghe.

Inoltre, la suddivisione in quattro dimensioni e la presenza di descrittori specifici per ognuna permette di individuare in maniera più dettagliata gli errori commessi (3-6).

- 3) "Il numero di idee è adeguato ma non sono pienamente efficaci. Avrei potuto dare delle motivazioni migliori, o comunque esporle meglio"
- 4) "ho commesso un errore a mio parere evitabile, ossia quello di non inserire delle formule di apertura e chiusura che invece sarebbero state necessarie per completare il testo e renderlo più adeguato"
- 5) "Ho notato che il mio testo ha delle carenze dal punto di vista della coesione, infatti penso che forse avrei potuto utilizzare più connettivi e riferimenti anaforici."
- 6) (in riferimento al numero di idee e alla loro efficacia) "introdurrei il punto del contenuto che avevo tralasciato e ampliarei l'altro che a mio parere era troppo sintetico" (seguono esempi di frasi che aggiungerebbe).

Nel commento riportato in (3)), ad esempio, la partecipante afferma di essersi resa conto di aver utilizzato delle idee poco efficaci a sostegno della sua argomentazione, specificando cosa avrebbe potuto fare per non commettere quell'errore. In (4), invece, notiamo come la persona che ha prodotto il testo si sia resa conto di non aver rispettato un requisito del task, in quanto non erano presenti formule di apertura e chiusura necessarie a rendere il testo conforme al genere testuale richiesto. L'utilizzo della scala di AF per l'attività di valutazione, però, non solo permette di riconoscere i propri errori, ma anche di riflettere sugli stessi trovando possibili alternative che renderebbero il testo più adeguato. In (5), ad esempio, la partecipante si rende conto che il suo testo è carente in termini di coesione, presumendo che un maggiore utilizzo di connettivi e riferimenti anaforici l'avrebbe reso più coeso. Nel commento in (6), invece, la valutatrice non solo è riuscita ad individuare il tipo di errore commesso e un'eventuale soluzione, ma fornisce anche una serie di frasi esemplificative che aggiungerebbe al testo per rendere le idee più efficaci.

Infine, da una delle riflessioni delle valutatrici (7) è emerso che le scale potrebbero essere anche utilizzate come guida alla produzione di un testo scritto, in quanto la presenza di descrittori dettagliati per ogni dimensione permette agli studenti di verificare passo per passo se hanno inserito nel testo tutte le informazioni necessarie e se queste sono presentate in maniera adeguata e comprensibile, rispettando i criteri specifici del tipo di task.

- 7) "l'autovalutazione, grazie al riferimento alle scale di adeguatezza funzionale, è un grande strumento di aiuto (...) per produrre testi in una lingua con cui si ha poca dimestichezza, perché consente di potersi attenere ad uno schema ben preciso (le quattro dimensioni delle scale) da utilizzare come mezzo guida per la produzione scritta."

Come abbiamo cercato di illustrare, la scala di AF costituisce quindi, a nostro avviso, uno strumento flessibile, adeguato alla valutazione sia in modalità in presenza che a distanza. I task utilizzati per elicitarle le produzioni degli studenti possono essere somministrati indipendentemente dal contesto in cui avviene la prova e pertanto non richiedono un ulteriore adattamento nel passaggio da una modalità all'altra. Infine, spostare l'oggetto di valutazione dalla lingua del candidato e le sue componenti (grammaticali, lessicali, ecc.) all'adeguatezza della realizzazione del task, che richiede una valutazione globale dell'effettivo apporto del candidato, rende più difficile, anche se naturalmente non impossibile, il ricorso ad ausili esterni.

4. Conclusioni

I risultati dell'indagine sulla valutazione in DaD hanno permesso di descrivere il modo in cui i docenti si sono adattati a questa particolare situazione e di suggerire possibili implicazioni per la didattica delle lingue in generale. Le principali difficoltà incontrate dai docenti sono legate alle mutate esigenze di organizzazione e pianificazione della valutazione, in quanto la DaD ha reso inadeguati gli strumenti tradizionalmente utilizzati per valutare in presenza e ha fatto aumentare notevolmente il tempo dedicato alla valutazione. Per quanto riguarda la qualità della valutazione, sono dunque venuti meno due criteri fondamentali: l'economicità e la fattibilità di una prova.

La maggiore preoccupazione dei docenti riguarda comunque la verifica del corretto svolgimento della prova e in particolare che i candidati possano copiare. Quando si somministra un test in classe, infatti, il docente può facilmente controllare che gli studenti rispettino le regole di svolgimento e non copino; mentre tale controllo risulta estremamente complesso in DaD, a causa degli strumenti tecnologici che non permettono talora nemmeno di individuare con certezza l'identità di chi svolge la prova. Tali difficoltà sono acute qualora la prova sia svolta in modalità differita. Per superare questi problemi, gli insegnanti hanno optato per soluzioni diverse e a volte combinate tra loro, e soltanto una piccola parte ha affermato di non aver attuato nessun tipo di controllo sullo svolgimento delle prove. La maggior parte dei docenti, infatti, si è affidata a strumenti tecnologici per sorvegliare lo svolgimento delle prove, come ad esempio software dedicati o condivisione dello schermo.

Alla luce di tali preoccupazioni, vale la pena chiedersi quale sia l'obiettivo di un percorso linguistico e, di conseguenza, quale sia l'oggetto della valutazione. Se l'obiettivo di un percorso linguistico è quello di insegnare ad usare una determinata L2/LS nel mondo reale, allora non ci si dovrebbe preoccupare del fatto che gli apprendenti possano aver accesso ad ausili di varia natura (controllo ortografico, traduttori automatici, articoli, scampoli di lingua autentica ecc.). Ci si dovrebbe preoccupare invece che l'ausilio di questi strumenti sia sufficiente a superare la prova, cioè che la prova non comporti un apporto originale del candidato. La valutazione dovrebbe, infatti, mettere in rilievo la reale capacità di portare a termine un task, in un'elaborazione personale, mobilitando tutte le conoscenze dell'individuo e del suo repertorio linguistico, in modo che l'uso eventuale di determinati ausili non sia di per sé sufficiente. Si può al limite sostenere che persino l'uso adeguato di determinati strumenti potrebbe far parte delle competenze valutabili, in quanto si avvicina all'uso della lingua nel mondo reale.

In questa direzione ci pare siano andati effettivamente alcuni docenti che per accertare le competenze dei candidati hanno creato dei compiti personalizzati, in cui era richiesto agli studenti di mettere in gioco le loro abilità linguistiche nel complesso.

Abbiamo illustrato come una valutazione task-based, e in particolare l'uso del criterio di AF possa essere fruttuosamente utilizzata per favorire questo cambiamento, permettendo inoltre di superare alcuni dei problemi riscontrati in DaD. La scala di valutazione dell'AF è infatti uno strumento flessibile, facilmente adattabile sia alla didattica in presenza che a quella a distanza e pertanto particolarmente utile in questo momento, in cui l'incertezza della situazione sanitaria costringe i docenti a continui riadattamenti della modalità didattica. I task, infatti, possono essere somministrati sia in presenza che a distanza, senza richiedere una pianificazione diversa a seconda del tipo di modalità utilizzato. Inoltre, indipendentemente dalla modalità di svolgimento, le prove possono essere corrette utilizzando lo stesso strumento di valutazione, ossia la scala dell'AF.

In conclusione, quindi, a nostro avviso l'utilizzo della scala di valutazione dell'AF e più in generale un tipo di valutazione basata sui task può rappresentare un passo verso una valutazione più in linea con i principi di un insegnamento linguistico basato sul "saper fare" in L2/LS, da tanto tempo considerato l'obiettivo dell'educazione linguistica. Inoltre, le scale di AF sembrerebbero una valida risposta alla discussione sul modo in cui bisogna improntare la valutazione basata sui task, in quanto tengono conto sia della capacità degli apprendenti di portare a termine il task, sia degli aspetti linguistici della performance e quindi rappresentano un punto d'incontro tra la posizione "forte" e "debole" presente in letteratura (cfr. §3.1).

Si può quindi pensare che in una certa misura la DaD abbia permesso un ripensamento della valutazione riportando al centro dell'azione valutativa la competenza linguistico-comunicativa intesa come la capacità degli apprendenti di utilizzare la lingua per raggiungere obiettivi, e non semplicemente come conoscenza di nozioni lessicali e grammaticali.

Riferimenti bibliografici

- Ambroso, Serena (1993). *Prove per la verifica dell'apprendimento*, Roma: Istituto dell'Enciclopedia Italiana fondata da G. Treccani.
- Bachman, Lyle F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.

- Bachman, Lyle F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453-476.
- Bachman, Lyle F. (1990). *Fundamental considerations in language testing*. Oxford University Press,
- Bachman, Lyle F., & Palmer, Adrian (2010). *Language assessment in practice*. Oxford University Press.
- Bonvino, Elisabetta (2008), *Valutazione e misurazione dell'apprendimento linguistico*. ICoN - Italian Culture on the Net.
- Bonvino Elisabetta, & Fiorenza Elisa (2020), Valutare per includere, *EL.LE*, 9(2), 197-218. Consultato all'url:<https://edizionicafoscarini.unive.it/media/pdf/article/elle/2020/2/art-10.14277-ELLE-2280-6792-2020-02-003.pdf>.
- Brindley, Geoff (1994). Task-centred assessment in language learning: The promise and the challenge. In Norman Bird, Peter Falvey, Amy Tsui, Allison Desmond, & Arthur McNeill (a cura di), *Language and learning* (pp. 73-94). Institute of Language in Education.
- Brown, James Dean, Hudson, Thom D., Norris, John M., & Bonk, William (2002). *An investigation of second language task-based performance assessments*. University of Hawaii Press.
- Consiglio d'Europa (2002). *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*. La Nuova Italia-Oxford.
- Conti, Sergio (in stampa). Didattica delle lingue a distanza durante l'emergenza Covid19: il quadro generale. *E-JournALL - EuroAmerican Journal of Applied Linguistics and Languages*, 8(2). 9-52. <http://doi.org/10.21283/2376905X.14.245>
- Cortés Velásquez, Diego, & Nuzzo, Elena (19-21 aprile, 2017). *Assessing L1 functional adequacy: Can we use the same scale as for L2?*. Contributo presentato alla "7th Task Based Language Teaching Conference - Tasks in Context". Barcellona: Spagna.
- Cortés Velásquez, Diego, & Nuzzo, Elena (2018). La didattica con i task: principi, applicazioni, prospettive. In Cortés Velásquez Diego & Nuzzo Elena (a cura di), *Il task nell'insegnamento delle lingue. Percorsi tra ricerca e didattica al CLA di Roma Tre* (pp. 13-39) (CLARo, Vol. 2). Roma TrE-Press.
- De Meo, Anna, Maffia, Marta, & Vitale, Giuseppina (2019). La competenza scritta in italiano L2 di apprendenti vulnerabili. Due scale di valutazione a confronto. *EL.LE*, 8(3), 637-654.
- Del Bono, Federica (6-8 giugno 2018). L'autovalutazione per migliorare la produzione scritta in italiano L1: uno studio pilota sull'adeguatezza funzionale in testi argomentativi. Contributo presentato al "X Convegno-Seminario CIS - La scrittura per l'insegnamento dell'italiano L2". CIS - Centro Italiano per Stranieri, Università di Bergamo.
- Del Bono, Federica (2019). Aspetti pragmatici nella valutazione di testi scritti: uno studio sull'adeguatezza funzionale in italiano L2. In Elena Nuzzo & Ineke Vedder (a cura di). *Lingua in contesto: la prospettiva pragmatica* (pp. 231-244). Officinaventuno-AItLA.
- Ellis, Rod (2003). *Task-based language learning and teaching*. Oxford University Press.
- Faone, Serena, & Pagliara, Francesca (2017, 19-21 aprile). *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: Spagna.
- Fiorenza Elisa, & Diego-Hernández, Elena (2020). The challenge of assessing plurilingual repertoires: The EVAL-IC project. In *What does plurilingualism mean for language assessment?*, Research Notes, Cambridge Assessment, 78, 43-50.
- Grice, Herbert P. (1975). Logic and conversation. In Peter Cole & Jerry L. Morgan, *Syntax and semantics: Speech acts* (pp. 41-58). Academic Press.
- 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 Naoko Taguchi & Youjin Kim (a cura di), *Task-based approaches to teaching and assessing pragmatics* (pp. 266-285). John Benjamins.
- Kuiken, Folkert, Vedder, Ineke, & Gilabert, Roger (2010). Communicative adequacy and linguistic complexity in L2 writing. In Inge Bartning, Maisa Martin, & Ineke Vedder, *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 81-100), Eurosla Monograph Series 1.
- LeBreton, James M., & Senter, Jenell L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational research methods*, 11(4), 815-852.
- Lenz, Peter, & Berthele, Raphael (2010). *Prise en compte des compétences plurilingue et interculturelle dans l'évaluation*. Conseil de l'Europe.
- Long, Michael H. (2015). *Second language acquisition and task-based language teaching*. John Wiley & Sons.
- Norris, John M. (2016). Current uses for task-based language assessment. *Annual Review of Applied Linguistics*, 36, 230-244.
- Norris, John M., Brown, James Dean, Hudson, Thom D., & Bonk, William (2002). Examinee abilities and task difficulty in task-based L2 performance assessment. *Language Testing*, 19(4), 395-418.
- Novello, Alberta (2014). *La valutazione delle lingue straniere e seconde nella scuola. Dalla teoria alla pratica*. SAIL Edizioni Ca' Foscari.
- Nunan, David (2004). *Task-based language teaching*. Cambridge University Press.
- Nuzzo, Elena, & Bove, Giuseppe (2018). Assessing functional adequacy across tasks: A comparison of learners and native speakers' written texts. *E-JournALL - Euro-American Journal of Applied Linguistics and Languages*, 7(2), 9-27. <http://doi.org/10.21283/2376905X.12.175>
- Orrù, Paolo (2019). Misurare l'adeguatezza funzionale in testi scritti di apprendenti di italiano L2. *Italiano LinguaDue*, 11(1), 45-58.
- Pagliara, Francesca (2017, 6-7 ottobre). *Valutare l'adeguatezza funzionale in produzioni scritte di studenti Marco Polo*. [Presentazione in convegno]. Dieci anni di didattica dell'italiano a studenti cinesi. Risultati, esperimenti, proposte. Siena, Italia.
- Pallotti, Gabriele (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30(4), 590-601.
- Purpura, James E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(S1), 190-208.
- Timpe-Laughlin, Veronika (2018). Pragmatics in task-based language assessment. In Naoko Taguchi & Youjin Kim (a cura di), *Task-based approaches to teaching and assessing pragmatics* (pp. 288-304). John Benjamins.
- Scaramucci, Matilde, Santos Junior, Elysio, & Rodrigues Alves Diniz, Leandro (2019). Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras): histórico, questões controversas e perspectivas. *Em aberto*, 32(104).
- Vedder, Ineke (2016). Il ruolo dell'adeguatezza funzionale nelle produzioni scritte in lingua seconda: proposta per una scala di valutazione. In Elisabetta Santoro e Ineke Vedder (a cura di), *Pragmatica e interculturalità in italiano lingua seconda* (pp. 79 - 92). Franco Cesati Editore.

Federica Del Bono, Università degli Studi Roma Tre/ Sapienza Università di Roma
federica.delbono@uniroma3.it

- IT** **Federica Del Bono** è dottoranda di ricerca in Linguistica presso l'Università degli Studi Roma Tre, in consorzio con la Sapienza Università di Roma. Il suo progetto di dottorato riguarda l'apprendimento incidentale di strategie di mitigazione degli atti linguistici in italiano come lingua straniera. Da gennaio 2020 collabora con il "Gruppo di Ricerca e Azione sull'Apprendimento delle Lingue" (GRAAL).
- EN** **Federica Del Bono** obtained her Ph.D. in Linguistics from the University Roma Tre, in collaboration with the Sapienza University of Rome. Her doctoral thesis project is on the incidental learning of mitigation strategies of speech acts in learning Italian as a foreign language. Since January 2020, she has been collaborating with the "Grupo di Ricerca e Azione sull'Apprendimento delle Lingue (GRAAL).
- ES** **Federica Del Bono** se doctoró en Lingüística de la Universidad Roma Tre, en colaboración con la Sapienza Universidad de Roma. Su proyecto de tesis doctoral trata sobre el aprendizaje incidental de estrategias de mitigación de los actos del habla en el aprendizaje del italiano como lengua extranjera. Desde enero de 2020 colabora con el "Grupo di Ricerca e Azione sull'Apprendimento delle Lingue (GRAAL).

Elisabetta Bonvino, Università degli Studi Roma Tre
elisabetta.bonvino@uniroma3.it

- IT** **Elisabetta Bonvino** è professoressa ordinaria in Didattica delle lingue moderne presso l'Università di Roma Tre. Ha conseguito il Dottorato di Ricerca in Linguistica presso École Pratique des Hautes Études (Parigi). Attualmente ricopre il ruolo di Direttrice del Centro Linguistico di Ateneo che gestisce l'Ufficio della Certificazione della lingua italiana come L2 di Roma Tre (Cert.IT). Dal 2017 è Presidente della Società italiana di didattica delle lingue e linguistica educativa" (DILLE).
- EN** **Elisabetta Bonvino** is a full professor of Modern Language Didactics at the University of Roma Tre. She obtained her Ph.D. in Linguistics from the École Pratique des Hautes Études (Paris). She is currently serving as the Director of the Center for Linguistics Research, which manages the Office of Certification of Italian as a Second Language at Roma Tre (Cert.IT). Since 2017 she has been the President of the Italian Society for Language Didactics and Linguistic Education (DILLE).
- ES** **Elisabetta Bonvino** es catedrática de Didáctica de Lenguas Modernas en la Universidad de Roma Tre. Obtuvo su doctorado en Lingüística en la École Pratique des Hautes Études (París). Actualmente realiza la labor de directora del Centro de Investigación Lingüística que gestiona la Oficina de Certificación de Italiano como Segunda Lengua en Roma Tre (Cert.IT). Desde 2017 es Presidenta de la Sociedad Italiana de Didáctica de la Lengua y Educación Lingüística (DILLE).