

VALUTARE TESTI REGOLATIVI ATTRAVERSO LA SCALA DELL'ADEGUATEZZA FUNZIONALE

Francesca Pagliara¹

1. INTRODUZIONE

Nell'ambito di una didattica basata sull'impiego di compiti comunicativi di realtà (*task*), la valutazione della competenza comunicativa assume un ruolo centrale, sia in termini di apprendimento che di verifica. Sebbene numerosi studi abbiano approfondito il concetto di complessità linguistica, la valutazione dell'efficacia comunicativa all'interno di un task rimane un ambito ancora oggetto di indagine (Pallotti, 2009; Kuiken *et al.*, 2010; Vedder, 2012; Kuiken, Vedder, 2014). Un task si considera completato, quando l'apprendente ne raggiunge l'obiettivo in modo efficace e adeguato. Tale aspetto della comunicazione è concettualizzato dalla nozione di Adeguatezza Funzionale (d'ora in poi AF), intesa come capacità di trasmettere efficacemente un messaggio e di portare a termine il compito assegnato (Kuiken *et al.*, 2010; Kuiken, Vedder, 2014; Vedder, 2016). Al fine di misurare e valutare empiricamente questi aspetti, il costrutto teorico dell'AF è stato operazionalizzato in una scala di valutazione multidimensionale (Vedder, 2012; Kuiken, Vedder, 2014; Vedder, 2016). Il presente contributo, a partire da uno studio condotto presso il Centro Linguistico di Ateneo dell'Università degli Studi di Roma Tre (Faone, Pagliara, 2017), presenta i risultati di una sperimentazione volta a valutare gli aspetti funzionali delle produzioni scritte di studenti Marco Polo al termine di un corso di italiano L2 caratterizzato da un approccio *task-supported* (Ellis, 2003).

2. L'ADEGUATEZZA FUNZIONALE: DAL COSTRUTTO TEORICO ALLA SCALA DI VALUTAZIONE

Una questione cruciale della linguistica applicata è cercare di definire che cosa rende un parlante non nativo un utilizzatore competente della L2 e come può essere adeguatamente misurata la competenza comunicativa dal momento che essa non è direttamente osservabile. Da qui la necessità di identificare degli indici di misurazione in base ai quali poter fare inferenze sulla competenza di un parlante non nativo.

Negli ultimi anni, è stato dimostrato che la competenza in L2 non è un costrutto unitario bensì multidimensionale e che i suoi principali componenti sono la complessità, l'accuratezza e la fluenza (CAF). Questo modello triadico della competenza in L2 si è rivelato un valido complemento dei precedenti modelli di misurazione: il tradizionale modello basato sulle quattro abilità linguistiche (Bachman, 1990), quello sociolinguistico (Canale, Swain, 1980) e quello cognitivo (Bialystock, 1994).

Più recentemente, tuttavia, è stato messo in rilievo che gli indici CAF non riescono a valutare un aspetto fondamentale della comunicazione ovvero la felicità pragmatica, intesa come successo comunicativo. Uno studio di De Jong *et al.* (2012a), che intende misurare

¹ Università degli Studi Roma Tre.

come e quanto la complessità di un task incida sul prodotto del task stesso, ha dimostrato che l'aumento della complessità del compito ha un effetto sulla fluenza, sull'appropriatezza lessicale (considerato un aspetto della complessità in tale studio), ma anche sul successo comunicativo delle performance degli apprendenti. Il loro studio sugli effetti della complessità linguistica sulle produzioni in L2, oltre a misurare quanto e come questo avviene, porta dunque alla luce anche un altro dato e cioè che il costrutto della competenza deve essere ampliato con un'altra dimensione, ovvero quella dell'AF. L'importanza dell'AF- come componente essenziale della competenza in L2 è stata rilevata da diversi autori (Pallotti, 2009; De Jong *et al.*, 2012a, 2012b; Kuiken *et al.*, 2010), i quali mettono in evidenza che la AF è una dimensione a sé stante della competenza linguistico-comunicativa. Kuiken, Vedder e Gilabert (2010), tuttavia, rilevano che pochi studi si occupano della dimensione comunicativa e degli aspetti pragmatici della padronanza linguistica – a dispetto dell'impostazione socio-pragmatica del *Quadro di Riferimento Europeo per le lingue* (QCER) – e che dunque valutare produzioni linguistiche, tenendo conto solo delle caratteristiche strettamente linguistiche quali gli indici CAF, significhi abbracciare una visione ristretta e riduttiva di ciò che costituisce la competenza d'uso di una lingua. Una possibile ragione del basso numero di studi sugli aspetti funzionali delle produzioni scritte e orali in L2 è probabilmente da ricercarsi nella mancanza di una definizione precisa del costrutto dell'AF. McNamara e Roever (2007) definiscono l'AF in termini di appropriatezza socio-pragmatica, rispetto alla scelta del registro, al genere testuale e alle convenzioni pragmatico-retoriche. Knoch (2009) si focalizza, invece, sui meccanismi e sulle strategie di coerenza e coesione impiegate nel testo, mentre Hismanoglu (2011) interpreta l'AF innanzitutto come competenza comunicativa interculturale. Altri studi (Upshur, Turner 1995, Alanen *et al.* 2010; De Jong *et al.* 2012a, 2012b) mettono in rilievo soprattutto l'efficacia comunicativa della trasmissione del messaggio o l'esecuzione riuscita del compito. Kuiken e Vedder (2010; 2014; 2016) definiscono invece l'AF come costrutto dinamico e interazionale, inteso in termini di felicità di esecuzione di un atto comunicativo (Searle, 1969) e ispirato alle massime conversazionali di quantità, rilevanza, modo e qualità di Grice (1975). Vista in questa ottica, l'AF è determinata dal successo della trasmissione del messaggio del parlante A recepito dall'interlocutore B. Proprio sulla base delle massime di Grice, gli stessi autori hanno decostruito l'AF in quattro dimensioni costitutive: contenuto, comprensibilità, requisiti del compito, coerenza e coesione. Secondo questa definizione, l'AF è dunque intesa come un costrutto interattivo, dipendente dal contesto e dipendente dalle attività, ovvero dai compiti comunicativi (task). L'AF è un costrutto interattivo perché lo scambio comunicativo tra A e B determina ciò che è funzionalmente adeguato in tale scambio; inoltre, dipende anche dal contesto perché un messaggio è funzionalmente adeguato in un dato contesto e non in ogni contesto; infine, l'AF è dipendente dalle attività, perché un messaggio è funzionalmente adeguato rispetto al compito linguistico e agli atti linguistici.

Per poter misurare e valutare questi aspetti della comunicazione, il costrutto teorico è stato operazionalizzato in una scala di valutazione multidimensionale (Vedder, 2012; Kuiken, Vedder, 2014; Vedder, 2016). I criteri della scala che scaturiscono da questi presupposti sono pertanto che, in primo luogo, il costrutto dell'AF è stata scomposto in quattro componenti distinte; in secondo luogo, i descrittori sono indipendenti dai descrittori linguistici CAF; infine, i descrittori sono oggettivi e quantificabili.

Le quattro dimensioni, in cui è stata scomposta l'AF e costruita la scala di valutazione (Vedder, 2016), sono:

1. il contenuto (CONT), dimensione che prende in considerazione l'adeguatezza del numero e del tipo di unità informative espresse nel testo dello scrivente, la loro

- pertinenza e rilevanza, indipendentemente dai requisiti specifici riguardo all'esecuzione del task (Grice 1975; massime di quantità, di relazione e di qualità);
2. i requisiti del task (RT), che considera la misura in cui nel testo si è tenuto conto dei requisiti specifici del task che lo scrivente deve svolgere, relativi al genere testuale, all'atto linguistico, alla macro-struttura e al registro del testo (Grice 1975; massime di relazione e di modo);
 3. la comprensibilità (COMP), che si riferisce al grado di comprensibilità del testo e allo sforzo richiesto al destinatario per capire lo scopo e le idee espresse nel testo (Grice 1975; massima di modo);
 4. coerenza e coesione (CO&CO), dimensione che si riferisce alla coerenza e alla coesione del testo dello scrivente, in termini di presenza o assenza di strategie anaforiche, connettivi, salti logici e argomenti non collegati, ripetizioni, ecc. (Grice 1975, massima di modo).

Ogni dimensione della scala è valutabile attraverso una scala Likert che prevede un punteggio da 1 a 6. Ad ogni punto della scala Likert corrisponde un descrittore (cfr. Appendice 1).

Nei vari studi condotti da Kuiken e Vedder finora citati, la scala di valutazione dell'AF è stata validata da valutatori esperti e non esperti, su produzioni sia in L1 (olandese) sia in LS (spagnolo e italiano), su task argomentativi. Gli apprendenti coinvolti nelle loro sperimentazioni sono studenti di italiano o di spagnolo LS presso l'Università di Amsterdam. I risultati dell'articolo del 2016 hanno dimostrato che la scala di valutazione dell'AF è uno strumento valido e affidabile per valutare le produzioni elicitate a partire da task argomentativi sia in L1 sia in LS e facilmente utilizzabile da valutatori sia esperti sia non esperti.

3. LO STUDIO

Tenendo conto del quadro teorico dell'AF, il presente studio vuole indagare alcuni punti lasciati aperti da Kuiken e Vedder (2016), in particolare la necessità di testare la scala dell'AF su ulteriori produzioni scritte e orali; su differenti tipi di task, su diversi profili di apprendenti e in vari contesti di apprendimento. Pertanto, per quanto riguarda il presente studio, il primo obiettivo è testare la scala di valutazione su un tipo di task diverso da quello argomentativo, ovvero l'*information gap task*²; il secondo obiettivo è quello di valutare l'AF attraverso le produzioni scritte di studenti sinofoni, cioè su un profilo di studenti diversi dallo studio modello; infine, il terzo obiettivo è quello di verificare l'uso della scala da parte degli insegnanti all'interno di un corso di italiano L2 condotto in modalità *task-supported* (cfr. Tabella 1). Per modalità *task-supported* si intende un corso in cui l'insegnante utilizza i task per creare dei momenti di produzione libera e di pratica di strutture linguistiche affrontate già nel corso delle lezioni, quest'ultime tipicamente strutturate secondo la sequenza propria del metodo comunicativo della "Presentazione, Pratica e Produzione – PPP" (Ellis 2003, 28-30). Tale scelta didattica è stata motivata dalla necessità di coniugare le esigenze didattiche legate al syllabo del centro che ha erogato il corso e dalla necessità di finalizzare il percorso didattico al momento dell'esame di certificazione con la volontà di utilizzare i task nella pratica didattica. Inoltre, ma non secondariamente,

² Un *information gap task* è definito in senso ristretto come un compito in cui «one student or group of students has one set of information, and another student or group has a complementary set of information. They must negotiate and find out what the other party's information is in order to complete an activity» (Nunan, 2004: 59). In un senso più ampio rientrano, tuttavia, sotto questa definizione tutti i task che «involve a transfer of given information from one person to another» (Prabhu, 1987).

la modalità *task-supported* è stata selezionata, perché considerata dalle insegnanti del corso³ più idonea ad andare incontro alle differenti abitudini di apprendimento degli studenti di madrelingua cinese, giacché «tendenzialmente più abituati a strategie di tipo analitico, mentre si trovano disorientati di fronte a ragionamenti di tipo globale e olistico» (Ingrassia, 2014, 396-397).

Tabella 1. *Confronto tra il presente studio e lo studio modello: variabili del disegno sperimentale*

	Kuiken, Vedder (2012, 2014, 2016)	Presente studio
Tipo di task	Task argomentativo con testo argomentativo come prodotto atteso	<i>Information gap task</i> con testo regolativo come prodotto atteso
Profilo apprendenti	Studenti olandesi dell'Università di Amsterdam	Studenti sinofoni del progetto MP
Contesto di apprendimento	Corsi di italiano LS	Corso di italiano L2

Lo studio è stato condotto presso il Centro linguistico dell'Università di Roma Tre; si inserisce in una serie di casi di studio e sperimentazioni sulla didattica e sulla valutazione basata sull'impiego di task⁴, condotti all'interno dei corsi Marco Polo e Turandot dal Gruppo di ricerca sul Task dell'Università degli Studi di Roma Tre, coordinato da Elena Nuzzo e Diego Cortés-Velasquez.

Nello studio sono stati coinvolti 15 studenti di italiano L2 provenienti dalla Cina, facenti tutti parte del programma Marco Polo. Gli studenti avevano iniziato il corso di italiano con un livello di competenza tra il livello A1 e il livello A2 del QCER. Il livello obiettivo era il B1. Gli studenti avevano tra i 20 e i 25 anni ed erano in attesa di iscriversi a corsi universitari o presso istituti AFAM⁵. Alla fine del corso ai partecipanti sono stati somministrati due task scritti con la collaborazione delle insegnanti del corso. Il primo era un task di tipo narrativo⁶ usato come studio preliminare, importante per testare e fissare il protocollo sperimentale; il secondo era un *information gap task* che aveva come prodotto atteso un testo regolativo (Appendice 2). Nella fattispecie il task finale consisteva nell'invviare ad un amico un'e-mail in cui fornire istruzioni dettagliate per l'accesso a un alloggio affittato, le regole della casa e cosa fare in caso di problemi. La scelta per testare la scala è ricaduta proprio sul testo regolativo perché permette di osservare al meglio la dimensione interattiva, che è alla base del costrutto teorico della FA, come visto prima (§ 2). In effetti, in un *information gap task* c'è una notevole dimensione performativa che coinvolge lo scrittore e il suo destinatario, anche se fittizio.

Per testare l'applicabilità della scala sul testo regolativo l'*information gap task* è stato valutato da tre valutatori esperti e tre valutatori semi-esperti, tutti di madrelingua italiana.

³ Un resoconto di questa esperienza didattica si trova in Lania, Mastrocesare, Rossi, (2017).

⁴ *Task-based Language Teaching* (TBLT); *Task-based Language Assessment* (TBLA).

⁵ L'Alta Formazione Artistica Musicale e Coreutica (AFAM) raggruppa tutte le istituzioni della formazione superiore italiana il cui scopo è la formazione nei settori dell'arte della musica, della danza e del teatro. Comprende essenzialmente le Accademie di Belle Arti, le Accademie Nazionali di Arte Drammatica e di Danza, gli Istituti Superiori per le Industrie Artistiche (ISIA), i Conservatori di Musica e gli Istituti superiori di studi Musicali (<http://www.miur.gov.it>).

⁶ Un task narrativo è un task in cui «subjects had to construct a story» (Foster, Skehan, 1996).

I valutatori esperti sono realizzatori o correttori di prove di certificazione dell'italiano come lingua seconda. Per "semi esperti" si intendono insegnanti di italiano L2 che hanno senz'altro esperienza nella valutazione formativa finalizzata al monitoraggio del processo di apprendimento e dell'efficacia del metodo di insegnamento, ma non hanno familiarità con l'uso di scale di valutazione per valutare il prodotto di un task. Gli insegnanti stessi coinvolti come valutatori, infatti, si consideravano non esperti, soprattutto per quanto riguarda la valutazione di task. Affinché i valutatori non esperti familiarizzassero con la scala di valutazione, sono state organizzate due sessioni di formazione. Durante le sessioni di *training*, sono stati spiegati gli obiettivi e l'uso della scala di valutazione e i valutatori sono stati coinvolti in sessioni di correzione.

Lo scopo principale del presente studio è pertanto indagare se la scala di valutazione dell'AF sia uno strumento di valutazione valido e affidabile anche per *l'information gap task* somministrato all'interno di un corso di italiano L2 per sinofoni. Per rendere conto di questo macro-obiettivo, è stato necessario approfondire alcuni aspetti riguardanti i valutatori per verificare l'affidabilità dei punteggi dati dai valutatori ad ogni produzione scritta; per quantificare il grado di accordo tra i giudizi dei valutatori; e, infine, per indagare la correlazione statistica tra le quattro sottodimensioni che costituiscono il costrutto dell'AF. Considerato ciò, sono state formulate le seguenti domande di ricerca:

- c'è uniformità tra i giudizi dei valutatori? I giudizi espressi dai valutatori sono affidabili?
- C'è accordo tra i giudizi dei valutatori?
- Come si correlano i giudizi dei valutatori sulle quattro dimensioni?

I dati sono stati analizzati tramite il programma di elaborazione statistica SPSS.24; il software è stato impiegato principalmente per calcolare statistiche descrittive sulla affidabilità delle scale. A tal proposito gli indici statistici presi in considerazione sono: l'*interrater reliability* (IRR o Cronbach's α) per misurare il grado di uniformità (o affidabilità) dei giudizi; l'*Inter Class Correlation* (ICC-*single measures*) per misurare l'accordo tra i giudizi dei valutatori (*interrater agreement* o IRA) e, infine, il coefficiente di correlazione lineare (*Pearson's r*) per verificare la correlazione tra le quattro dimensioni della scala della FA.

4. RISULTATI E ANALISI

In questa sezione si riportano i risultati ottenuti. La prima domanda di ricerca intende verificare che ci sia uniformità tra i giudizi dei valutatori. Per misurare il grado di uniformità di valutazione tra i valutatori è stata calcolata l'alpha di Cronbach (IRR). I risultati evidenziano che i giudizi dei valutatori sulle quattro dimensioni sono piuttosto uniformi, come si può vedere nella Tabella 2:

Tabella 2. IRR

<i>Interrater reliability - Cronbach's Alpha (IRR)</i>	
CONTENUTO	.928
REQUISITI DEL TASK	.915
REQUISITI DEL TASK	.915
COERENZA E COESIONE	.888

Dalla tabella è possibile vedere che l'affidabilità dei giudizi riporta valori elevati, giacché compresi tra .888 e .928⁷. Inoltre, valori che si approssimano all'1 parlano a favore di un grado di uniformità pressoché assoluto nei giudizi assegnati dai valutatori. Quando l'alpha di Cronbach è superiore allo 0,7 è infatti indice del fatto che ci sia costanza nei punteggi assegnati dai singoli valutatori. Tuttavia, l'IRR da sola non dice se i punteggi assegnati dai vari valutatori sono tra loro concordi: per capire se le scale sono interpretate in maniera omogenea da tutti, è necessario analizzare specificamente l'indice di accordo (*agreement*) tra i valutatori.

La seconda domanda di ricerca che lo studio si pone, infatti, è verificare il grado di accordo tra i valutatori. I dati in questo caso indicano che l'accordo tra i valutatori è tendenzialmente basso, come visibile nella Tabella 3:

Tabella 3. ICC: IRA

ICC: Interrater Agreement (IRA)	
CONTENUTO	,447
REQUISITI DEL TASK	,402
COMPRESIBILITÀ	,563
COERENZA E COESIONE	,440

Tutti i valori si trovano al di sotto di 0.7: tra i valutatori, pertanto, l'accordo nell'assegnare lo stesso punteggio per le quattro dimensioni rispetto alle singole produzioni è debole.

Con la terza domanda di ricerca si vogliono infine rilevare le correlazioni tra i giudizi dei valutatori per le quattro dimensioni dell'AF. A tale scopo è stato calcolato il coefficiente di correlazione di *Pearson's r*. I risultati contenuti nella Tabella 4 dimostrano che vi è correlazione tra le quattro dimensioni della scala:

Tabella 4. Coefficiente di correlazione di *Pearson's r*

	REQUISITI DEL TASK	COMPRESIBILITÀ	COERENZA E COESIONE
CONTENUTO	,954	,502	,739
REQUISITI DEL TASK		,567	,845
COMPRESIBILITÀ			,812

La relazione tra contenuto e requisiti del task appare particolarmente elevata (.954). Risultano solide anche le correlazioni della coerenza e della coesione con le altre dimensioni dell'AF (.739 con il contenuto; .845 con i requisiti del task; .812 con la comprensibilità). Sono più bassi, invece, i valori ottenuti per le correlazioni tra la comprensibilità e le altre dimensioni dell'AF (.502 con il contenuto; .576 con i requisiti del task).

⁷ In letteratura, correlazioni superiori a 0.7 sono considerate indicative di una relazione elevata/forte tra le variabili (cfr. Cohen, 1988 che per primo fornisce linee guida per interpretare le correlazioni).

5. CONCLUSIONI

Il presente lavoro, sebbene con il limite di basarsi su un esiguo campione di informanti, ha permesso di fornire nuovi dati riguardo l'utilizzo della scala di valutazione dell'AF, variando, rispetto allo studio modello, il tipo di task e il genere testuale atteso come prodotto, il contesto di apprendimento e il profilo degli apprendenti. Dall'analisi dei dati è emerso che:

- in riferimento alla IRR, i punteggi dei valutatori sono uniformi e la scala di valutazione dell'AF risulta affidabile per classificare un gruppo di produzioni scritte e per stabilire se il prodotto del task valutato è adeguato funzionalmente oppure no (domanda di ricerca 1);
- l'accordo assoluto (IRA) tra i valutatori non è sufficientemente alto (domanda di ricerca 2);
- le correlazioni sono abbastanza forti, ma le due correlazioni più basse - contenuto/comprendibilità e requisiti del task/comprendibilità – indicano che c'è una difficoltà nell'interpretazione del descrittore della comprendibilità (domanda di ricerca 3).

Rispetto allo studio modello (Kuiken, Vedder, 2016), il presente lavoro, dunque, conferma un'alta uniformità di giudizi tra i valutatori e una buona correlazione tra le quattro dimensioni, seppur con un nuovo dato: tra le quattro dimensioni, la comprendibilità sembra essere quella che correla meno con le altre. Sebbene già nello studio sul testo argomentativo emergesse una correlazione più bassa tra comprendibilità e requisiti del task (Kuiken, Vedder, 2016: 9), nel presente caso anche la correlazione con il contenuto risulta bassa. Ne consegue che l'alta correlazione del contenuto con tutte le altre dimensioni non viene qui confermata (per una *summa* dei risultati ottenuti, si cfr. la Tabella 5).

Tabella 5. *Confronto con lo studio modello*

RISULTATI	AF – TASK ARGOMENTATIVO (KUIKEN, VEDDER, 2016)	AF – INFORMATION GAP TASK
Alta uniformità dei giudizi (Cronbach's <i>Alpha</i>)	✓	✓
Bassa correlazione tra RT e COMPR	✓	✓
CONT altamente correlato con le altre quattro dimensioni dell'AF	✓	✓
Alta correlazione tra CONT e COMP	✓	✗

La bassa correlazione tra contenuto e comprendibilità potrebbe spiegarsi con il fatto che non necessariamente un testo che presenti molte unità di significato – come indica il descrittore dell'AF – sia sufficientemente comprensibile. Parimenti, la correlazione minore tra comprendibilità e requisiti del task può essere interpretata come indice del fatto che un testo, a fronte della presenza di unità informative adeguate per numero e pertinenza, non sia necessariamente adeguato per scelte di genere e di registro.

In linea generale, tuttavia, le dimensioni risultano correlate in maniera elevata o moderatamente elevata e questo dimostra che l'AF può essere valutata attraverso le quattro dimensioni postulate come soggiacenti al costrutto. Anche le correlazioni meno intense si avvicinano a un valore di 0.7, suggerendo una relazione non trascurabile, e la loro minore forza potrebbe essere dovuta a limitazioni dello studio (cfr. *infra*).

Per quanto riguarda i valutatori, lo studio ha messo in evidenza che la scala può essere adoperata tanto da valutatori esperti quanto da valutatori inesperti e che, perciò, si conferma affidabile per essere utilizzata come strumento di valutazione dell'AF anche all'interno di corsi di italiano L2. I valori più bassi relativi all'accordo assoluto, in attesa di altri studi, per il momento possono essere letti come segnale del fatto che la sessione di formazione è una condizione necessaria, ma non sufficiente, per il corretto impiego delle scale e che probabilmente, per arrivare a una normalizzazione dei giudizi soddisfacente, sia necessario incrementare il numero delle sessioni minime di *training* da prevedere.

È senz'altro da menzionare qui un limite dello studio a riguardo: la mancanza di retrospettive sistematiche con i valutatori; queste, infatti, sono avvenute in maniera poco organizzata e strutturata, attraverso scambi di e-mail, piuttosto che con sedute plenarie.

In base, dunque, ai risultati ottenuti ed ai limiti dello studio messi in evidenza nel corso della discussione, ci si propone di ampliare il presente lavoro, aumentando il campione di testi da sottoporre a valutazione, aumentando e sistematizzando le sessioni di formazione ai valutatori e prevedendo un gruppo di controllo di apprendenti non MP, in un corso non supportato dalla didattica per *task*.

RIFERIMENTI BIBLIOGRAFICI

- Alanen R., Huhta A., Tarnanen M. (2010), "Designing and assessing L2 writing tasks across CEFR proficiency levels", in Bartning I., Martin M., Vedder I. (eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*, Eurosla Monographs Series, 1, pp. 21-56.
- Bachman L. (1990), *Fundamental Considerations in Language Testing*, Oxford University Press, Oxford.
- Byalistok E. (1994), "Analysis and control in the development of second language proficiency", in *Studies in Second Language Acquisition*, 16, 2, pp. 157-168.
- Canale M., Swain M. (1980), "Theoretical bases of communicative approaches to second language teaching and testing", in *Applied Linguistics*, 1, pp. 1-7.
- Cohen J. (1988), *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum Associates.
- De Jong N., Steinel M. P., Florijn A. F., Schoonen R., Hulstijn J. H. (2012a), "The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers", in Housen A., Kuiken F., Vedder I. (eds.), *Dimensions of L2 performance and proficiency. Investigating complexity, accuracy and fluency in SLA*, John Benjamins, Amsterdam, pp. 121-142.
- De Jong N. H., Steinel M. P., Florijn A. F., Schoonen R., Hulstijn J. H. (2012b), "Facets of speaking proficiency", in *Studies in Second Language Acquisition*, 34, 1, pp. 5-34.
- Ellis R. (2003), *Task-based Language Learning and Teaching*, Oxford University Press, Oxford.
- Faone S., Pagliara F. (2017), *How to assess L2 information-gap tasks through Functional Adequacy rating scales*, paper presented at the Seventh International Conference on Task – Based Language Teaching, Barcellona April 19-21, 2017.
- Foster P., Skehan P. (1996), "The influence of planning and task type on second language performance", in *Studies in Second Language Acquisition*, 18, pp. 299-323.

- Grice H. P. (1975), "Logic and conversation", in Cole P., Morgan J. L. (eds.) *Speech acts*, Academic Press, New York, pp. 41-58.
- Hismanoglu M. (2011), "An investigation of ELT students' intercultural communicative competence in relation to linguistic proficiency, overseas experience and formal instruction", in *International Journal of Intercultural Relations*, 35, 6, pp. 805-817.
- Housen A., Kuiken F. (2009), "Complexity, accuracy and fluency in second language acquisition", in *Applied linguistics*, 30, 4, pp. 461-473.
- Knoch U. (2009), "Diagnostic assessment of writing: A comparison of two rating scales", in *Language Testing*, 26, 2, pp. 275-304.
- Kuiken F., Vedder I. (2011), "Task complexity and linguistic performance in L2 writing and speaking: The effect of mode", in Robinson P. (ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance*, John Benjamins, Amsterdam, pp. 91-104.
- Kuiken F., Vedder I. (2014), "Raters' decisions, rating procedures and rating scales", in *Language testing*, 31, 3, pp. 279-284.
- Kuiken F., Vedder I. (2016), "Functional adequacy: Towards a new rating scale", in *Language Testing*, 34, 3, pp. 321-336.
- Kuiken F., Vedder I., Gilbert R. (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, pp. 81-100.
- Ingrassia M. (2014), "La didattica del Task in un corso di italiano L2 per studenti sinofoni. in *ItalianoLinguaDue*, 6, 2, pp. 379-399:
<https://riviste.unimi.it/index.php/promoitals/article/view/4706>.
- Lania E., Mastrocesare D., Rossi V. (2017), "I Task con studenti Marco Polo Turandot al Cla di Roma Tre: pratiche didattiche contro vecchi stereotipi e implicazioni culturali", in Bagna C., Chiapedi N., Salvati L., Scibetta A., Visigalli M. (a cura di), *La lingua italiana per studenti internazionali: aspetti linguistici e didattici*, OL3 srl, Perugia, pp. 91-97.
- McNamara T., Roever C. (2007), *Testing: The social dimension*, Blackwell Publishing, Malden (MA).
- Nunan D. (1989), *Designing tasks for the communicative classroom*, Cambridge University Press, Cambridge.
- Nunan D. (2004), *Task-Based Language Teaching*, Cambridge University Press, Cambridge.
- Oxford R. L. (1990), *Language learning strategies: What every teacher should know*, Heinle & Heinle, Boston.
- Pallotti G. (2009), "CAF: Defining, Refining and Differentiating constructs", in *Applied Linguistics*, 30, 4, pp. 590-601.
- Prabhu N. S. (1987), *Second Language Pedagogy*, Oxford University Press, Oxford.
- Searle J. R. (1969), "Indirect speech acts", in *Syntax and Semantics*, 3, pp. 59-82.
- Upshur J. A., Turner C. E. (1995), "Constructing rating scales for second language tests", in *ELT Journal*, 49, 1, pp. 3-12.
- Vedder I. (2012), "Adeguatezza comunicativa e complessità linguistica in italiano L2 e italiano L1: la valutazione delle produzioni scritte di nativi e non nativi", in Bernini G., Lavinio C., Valentini A., Voghera M. (a cura di), *Competenze e formazione linguistiche. In memoria di Monica Berretta*. Atti dell'XI Congresso dell'Associazione Italiana di Linguistica Applicata, Guerra Edizioni, Perugia, pp. 9-11.
- Vedder I. (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*, Franco Cesati Editore, Firenze, pp. 68-79.

APPENDICE 1

Scala di valutazione dell'adeguatezza funzionale (Vedder, 2016)

Contenuto: il numero delle unità informative espresse nel testo è adeguato e rilevante?

1. *Per niente*: il numero di idee è insufficiente, non è affatto adeguato, e non sono coerenti le une con le altre.
2. *Appena*: il numero di idee non è sufficientemente adeguato, le idee sono poco consistenti.
3. *Parzialmente*: il numero di idee è abbastanza adeguato anche se non sono molto consistenti.
4. *Sufficientemente*: il numero di idee è adeguato e sono sufficientemente consistenti.
5. *Largamente*: Il numero di idee è molto adeguato e appaiono molto consistenti e coerenti le une con le altre.
6. *Assolutamente*: il numero di idee è assolutamente adeguato e appaiono molto consistenti e coerenti le une con le altre.

Sono stati soddisfatti i requisiti e i criteri specifici del task (ad es. genere, atto linguistico, registro)?

1. *Per niente*: nessuno dei requisiti e dei criteri specifici del task è stato soddisfatto.
2. *Appena*: pochi requisiti e criteri specifici del task (meno della metà) sono stati soddisfatti.
3. *Parzialmente*: circa la metà dei requisiti e dei criteri specifici del task sono stati soddisfatti.
4. *Sufficientemente*: più della metà dei requisiti e delle condizioni del task sono stati soddisfatti.
5. *Largamente*: quasi tutti i requisiti e i criteri specifici del task sono stati soddisfatti.
6. *Assolutamente*: tutti i requisiti e i criteri specifici del task sono stati soddisfatti.

Comprensibilità: quanto sforzo è richiesto per capire lo scopo del testo e le idee?

1. *Per niente*: il testo non è affatto comprensibile. Le idee e lo scopo sono espressi in modo oscuro e il lettore, anche sforzandosi, non riesce a capire.
2. *Appena*: il testo non si comprende facilmente, i suoi scopi non sono chiari e il lettore deve sforzarsi molto per capire le idee dell'autore. Il lettore deve cercare di indovinare la maggior parte delle idee e degli scopi del testo.
3. *Parzialmente*: il testo è parzialmente comprensibile, ma alcune frasi non si capiscono bene a una prima lettura. Un'ulteriore rilettura è utile per chiarire gli scopi del testo e le idee espresse, ma rimangono alcuni dubbi.
4. *Sufficientemente*: il testo è sufficientemente comprensibile, solo certe frasi sono poco chiare ma si possono capire senza grandi sforzi con una rilettura.
5. *Largamente comprensibile*: il testo è facile da comprendere e si legge agevolmente, non ci sono problemi di comprensibilità.
6. *Assolutamente*: il testo è molto facile da comprendere e si legge molto agevolmente, le idee e gli scopi sono espressi con chiarezza.

Coerenza e coesione: il testo è coerente e coeso (ad es. ci sono connettivi e strategie di coesione)?

1. *Per niente*: ci sono frequenti salti logici e argomenti non collegati. L'autore non usa riferimenti anaforici (pronomi, frasi con soggetto sottinteso chiaramente interpretabile). Il testo non è affatto coeso. I connettivi sono praticamente assenti e le idee non sono collegate tra loro.
2. *Appena*: l'autore spesso non collega gli argomenti tra loro; se c'è coerenza, questa è espressa per lo più da ripetizioni. Pochi riferimenti anaforici. Alcuni salti logici. Il testo è poco coeso. Vengono usati pochi connettivi, che non collegano bene le idee.

3. *Parzialmente*: ci sono frequenti salti di argomento e/o ripetizioni. Più di due frasi di seguito esprimono esplicitamente lo stesso soggetto, anche quando questo sarebbe chiaro. Vengono usati alcuni riferimenti anaforici. Possono esservi alcune interruzioni della coerenza. Il testo è abbastanza coeso. Vengono usati alcuni connettivi, ma sono per lo più delle congiunzioni.
4. *Sufficientemente*: i salti di argomento sono abbastanza rari, ma l'autore a volte riesce a essere coerente solo mediante ripetizioni non necessarie. Si trovano sufficienti riferimenti anaforici. Possono esservi interruzioni della coerenza. L'autore fa un buon uso dei connettivi, che a volte vanno oltre le semplici congiunzioni.
5. *Largamente*: quando si introduce un nuovo argomento, di solito ciò avviene mediante l'uso di connettivi o espressioni di collegamento esplicite. Le ripetizioni sono molto rare. Si trovano numerosi riferimenti anaforici. Nessuna interruzione della coerenza. Il testo è molto coeso e le idee sono ben collegate tra loro mediante connettivi nominali o avverbiali.
6. *Assolutamente*: l'autore produce un'ottima coerenza integrando le nuove idee nel testo con connettivi o espressioni di collegamento esplicite. I connettivi anaforici sono usati regolarmente. Qualche raro caso di argomenti non collegati e nessuna interruzione della coerenza. La struttura del testo è estremamente coesa, grazie a un abile uso dei connettivi (in particolare nominali, avverbiali e formule di collegamento), usati spesso per descrivere le relazioni tra le idee.

APPENDICE 2

Hai affittato un alloggio per un amico che viene a trovarti nella tua città. Purtroppo, non potrai essere presente al suo arrivo. Scrivi un'e-mail per fornirgli tutte le informazioni necessarie:

- come accedere all'alloggio (es. codice, dove trovare le chiavi);
- quali sono le regole della casa;
- cosa fare in caso di problemi.

