

LA VALUTAZIONE DEL PARLATO DEGLI STUDENTI: IMPLICAZIONI SUL TESTING LINGUISTICO

*Cecilia Varcasia*¹

1. VALUTARE IL PARLATO

L'abilità del parlato costituisce un elemento importante nell'insegnamento linguistico, di conseguenza la sua valutazione nell'ambito della valutazione delle competenze linguistiche è altrettanto importante. La valutazione del parlato è però al tempo stesso un compito difficile, non solo perché la valutazione di quanto una persona sappia parlar bene una lingua è influenzata da numerosi fattori, ma anche perché ci si aspetta che la valutazione dia dei giudizi accurati, obiettivi e appropriati rispetto all'obiettivo iniziale che la valutazione stessa si è posto (Luoma, 2004: 1).

Testare e valutare il parlato pone però innanzitutto un quesito teorico di definizione: cosa significa testare il parlato? Quali sono gli elementi caratteristici del parlato che possano essere operazionalizzati e misurati in una valutazione? Già Lado (1961: 239) sosteneva che l'abilità di parlare in una L2 fosse senza dubbio quella più importante, ma nonostante ciò la ricerca sul *testing* di questa abilità è forse quella meno sviluppata e meno praticata a causa di una lacuna nella comprensione di quali siano le componenti dell'abilità del parlato. Se questo è vero in termini generali per quanto riguarda il *testing* sul parlato di lingue quali l'inglese, nell'ambito del quale la ricerca si è mossa con particolare interesse nel corso degli ultimi quindici anni (cfr. Coombe *et al.* 2012; Davis, 2009; Ducasse & Brown, 2009; Fulcher, 2003; Lazaraton, 2008; Luoma, 2004; May, 2009; McNamara & Lumley, 1997; Nakatsuhara, 2011; O'Sullivan, 2002, 2008; Powers *et al.*, 1999), lo è soprattutto per quanto riguarda la ricerca in Italia. Il presente lavoro vuole contribuire alla riflessione sul *testing* della competenza orale in Italia, proponendo una riflessione innanzitutto per quanto riguarda lo sviluppo di un test del parlato in generale, prendendo poi in considerazione le attuali certificazioni di italiano per scopi generali e i loro sillabi e tipologie di *task* usati per la valutazione di questa abilità e concludendo con la presentazione di un test del parlato sviluppato dalla nostra Università, discutendo di alcune delle questioni metodologiche che si pongono nelle fasi iniziali dello sviluppo di un test.

¹ Centro di Competenza Lingue, Libera Università di Bolzano.

2. LO SVILUPPO DI UN TEST DEL PARLATO: UN PROCESSO CICLICO

Lo sviluppo di un test richiede diverse fasi di elaborazione ed è caratterizzato da una struttura ciclica. Quando ne viene richiesta l'elaborazione, molti docenti cominciano direttamente con la scrittura degli *item* perché ad uno sguardo ingenuo questo sembra essere il punto di partenza. Le agenzie di *testing*, però, investono molto tempo ed energie nel processo di sviluppo di test standardizzati per cercare di garantirne l'obiettività e l'imparzialità (*fairness*). Il primo passo nello sviluppo di un test è quindi rappresentato dalla definizione dello scopo per il quale viene creato, piuttosto che direttamente dalla scrittura degli *item*. In questa fase è dunque importante porsi una serie di domande che aiutino a definire lo scopo del test che si va a costruire, come ad esempio:

- Per quale motivo è richiesto/necessario sviluppare un test?
- Perché le persone dovrebbero sostenere il test? Che obiettivo spererebbero di raggiungere?
- Quale può essere l'impatto della valutazione sui candidati?
- Come possono essere identificate le questioni importanti?
- Chi deve essere coinvolto?
- Come possono essere raccolte e analizzate le informazioni?
- Chi è più indicato per raccogliere e analizzare le informazioni?
- Quali sono le fonti di informazione più utilizzabili?
- Quali sono le risorse disponibili (di tempo, potere politico, forze lavoro, ecc.)? (cfr. Harris & Bell, 1994).

Nella definizione dello scopo di un test è necessario includere anche la risposta alla domanda "che cosa si sta testando? Qual è l'informazione necessaria e per quale motivo è necessaria?". La risposta a queste domande costituisce la base di partenza per poter poi prendere le decisioni successive.

Fulcher (2006) e Fulcher e Davidson (2009) usano una metafora presa dal mondo dell'architettura per descrivere il processo di elaborazione di un test: quando un architetto comincia a progettare la costruzione di un edificio deve avere ben chiaro il motivo per il quale viene costruito, se si tratta di un supermercato piuttosto che di una villa, o di un grattacielo piuttosto che di un centro commerciale. Da questo dipende il tipo di struttura da progettare e costruire. Non avrà senso progettare di costruire una bella villetta con il giardino intorno se si dovrà usare l'edificio come supermercato, e così via. Lo sviluppo di un test linguistico è molto simile alla progettazione di un edificio. Ciascun test deve essere elaborato per uno scopo specifico poiché, come già sosteneva Ingram, «if a test is made up without a clear idea of what it is for, is no good» (1968: 70, cit. in Fulcher, 2010: 95).

Se lo scopo è quello di valutare il raggiungimento delle competenze proposte nel syllabo del corso nel quale si è insegnato negli ultimi due mesi si dovrà elaborare un test che sia legato al contenuto del corso. Per fare questo si potrà partire dal mettere insieme i contenuti del syllabo e sviluppare degli *item* per la valutazione che riflettano i processi e le abilità oggetto dell'apprendimento. Alternativamente si potrà partire dall'osservazione degli obiettivi di apprendimento e basare la valutazione su questi. Se invece il test è usato per una

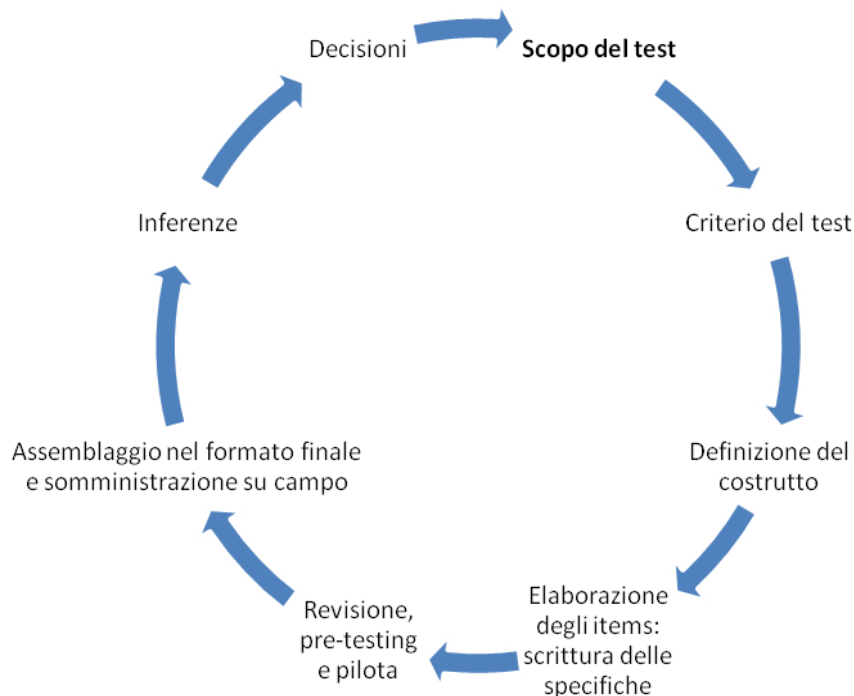
certificazione, la definizione del suo scopo è più complessa e di maggiore importanza, data dall'impatto della certificazione sui candidati e il loro futuro percorso. In altri casi lo scopo del test sarà molto preciso e puntuale, come ad esempio nel caso in cui si debbano certificare le abilità di scrittura e lettura di un gruppo di ingegneri spaziali. Nel processo di valutazione si deve tenere conto di questa specificità e non pretendere di valutare competenze più ampie.

Bygate (1987: 3) paragona il parlato alla guida di una macchina e si domanda quale tipo di conoscenze debba avere un autista per poter guidare. Certamente dovrà sapere come si chiamano i comandi, dove si trovano e come funzionano, ma avrà anche bisogno di essere capace di usarli per poter essere in grado di guidare la macchina lungo una strada senza sbattere contro vari ostacoli che si possono trovare sul percorso. L'autista dovrà anche essere capace di camminare a una andatura normale, di gestire le variazioni del manto stradale, ecc. Nel parlare, afferma Bygate, eseguiamo delle operazioni simili. Fulcher (2003: 47) sostiene che l'analogia con il saper guidare un'automobile sia estendibile anche al *testing* linguistico quando rileva che lo scopo di testare il parlato in una seconda lingua è simile a un esame di guida.

Lo scopo di un test del parlato è quello di raccogliere in maniera sistematica, attraverso dei *task* o altre tecniche di elicitazione, evidenze che supportino le inferenze fatte sulla base del costruito. Ci interesserà quindi valutare la potenziale capacità dell'apprendente di parlare in una varietà di situazioni comunicative molto più ampia rispetto a quelle che possono essere osservate durante un test, così come non è possibile testare tutti i tipi di strada e manovre durante l'esame di guida. Dal campione elicitato si dovranno quindi fare delle inferenze riguardo alla probabilità di successo o fallimento in future situazioni comunicative che non siano oggetto di valutazione diretta del nostro apprendente. Questa è una delle principali sfide nel testare il parlato: è necessario sviluppare dei *task* che elicitino un parlato del tipo e della quantità che permetta di formulare le inferenze significative riguardo il costruito per il quale il test è stato sviluppato (cfr. Fulcher, 2003: 47).

La descrizione dello scopo per il quale il test viene costruito è strettamente legata alla definizione del criterio in base al quale il test viene costruito. Ciò costituisce la seconda fase nel processo ciclico di sviluppo di un test, come mostrato dalla Figura 1. Il criterio di un test è rappresentato da un'esemplificazione del tipo di competenze richieste per il raggiungimento dell'obiettivo. Questo viene elaborato nel collezionare campioni rappresentativi del tipo di competenza desiderata. Se per esempio si intende valutare la capacità di comprendere le istruzioni per il corretto funzionamento di un macchinario, si potrà risalire a diversi campioni di manuali di istruzioni già esistenti che forniranno il materiale per sviluppare il test.

Figura 1. *Le fasi dello sviluppo di un test* (cfr. Fulcher, 2010: 94).



Lo scopo e il criterio del test vengono esplicitati formalmente e vanno a far parte delle *specifiche* del test, delle quali si parlerà nella sezione 4:

a statement of test purpose is likely to include information on the target population and its ability range. Test developers normally state target domains of language use, and the range of knowledge, skills or abilities that underpin the test. This statement justifies the selection of constructs and content by articulating a direct link between the intended score meaning and the use to which the scores will be put in decision making (Fulcher & Davidson, 2009: 123-124).

Nella prossima sezione si tratterà invece del passo successivo alla definizione del criterio di un test: la definizione del costrutto.

3. LA DEFINIZIONE DEL COSTRUTTO

La prima domanda da porsi al momento di sviluppare un test del parlato è quella di chiedersi ‘che cosa è il parlato?, cosa costituisce la mia valutazione dell’abilità orale?’. La risposta a questo genere di domande contribuisce alla definizione del costrutto del test.

Il costrutto è costituito dall'insieme di abilità che si presuppone stiano alla base della performance che uno studente esegue in una L2, non osservabili in maniera diretta. Queste abilità possono essere inizialmente definite come concetti, identificabili perché vengono associate a dei sostantivi astratti. Tra i costrutti più utilizzati nell'insegnamento delle lingue troviamo il costrutto 'intelligenza', 'attitudine', 'fluenza'. Non è possibile indicare concretamente e con precisione un esempio di questi costrutti, ma il nostro giudizio sull'esistenza di tali concetti ci è dato dall'estrazione di vari elementi dal comportamento degli individui.

La definizione di un costrutto è un aspetto molto importante nell'elaborazione di un test, proprio perché si presuppone una valutazione. È importante sapere che cosa chiedere di fare ad un apprendente in modo da poter osservare il suo comportamento e decidere se e in quale misura questa abilità astratta sia presente. Fulcher e Davidson (2009: 126-127) descrivono i costrutti come dei "*design patterns*" in architettura, che possono essere usati nella progettazione di un edificio specifico. I costrutti sono generalmente descritti in modelli. Questi operano a un livello più alto rispetto alla struttura del test. Sono generali e astratti per natura, mentre la struttura del test è una selezione dei costrutti rilevanti per uno scopo specifico.

Questo tipo di concettualizzazione riconosce che esiste un gran numero di costrutti del linguaggio e infinite situazioni linguistiche, ma che solo alcuni costrutti o abilità possono essere realizzati in ogni situazione. Inoltre, è necessario considerare che persino il parlante più competente non sarà competente in egual misura in tutti i contesti. Già Lado sosteneva che «le situazioni nelle quali il linguaggio è il mezzo di comunicazione sono potenzialmente infinite. Nessuno, nemmeno il parlante più istruito, può parlare e comprendere la propria L1 in tutte le situazioni in cui questa può essere utilizzata» (1961: 26). Il costrutto sarà elaborato come contenente una serie di competenze ritenute utili in relazione allo scopo del test stesso e nella misura in cui i risultati della valutazione saranno generalizzabili anche in altri contesti.

Si propone nella Tabella 1 una sintesi di ciò che potrebbe contribuire alla definizione operativa di un costrutto per un test del parlato in L2 (cfr. Fulcher, 2003: 48). Il modello è adattato da Bachman & Palmer (1996) allo specifico della valutazione dell'abilità orale.

La decisione riguardo a quali elementi includere nel costrutto dipenderà quindi da quanto questi contribuiranno alla formulazione di un "*validity argument*" che rinforzi la relazione che deve esistere tra i risultati del test e il significato che viene attribuito a questi risultati. Fulcher (2003: 20) sostiene quindi che ciò che è necessario fare per la definizione di un costrutto è selezionare e mettere insieme le varie competenze rispetto all'uso e l'obiettivo che ci si pone, prendendo in considerazione anche le ambizioni e i desideri di coloro che sosterranno il test, e le necessità di chi utilizzerà i risultati del test stesso.

Tabella 1. *Un modello per descrivere il costrutto del parlato* (cfr. Fulcher, 2003: 48).

COMPETENZA LINGUISTICA	
<u>Fonologia</u> <ul style="list-style-type: none"> - Pronuncia - Accenti - Intonazione 	<u>Accuratezza</u> <ul style="list-style-type: none"> - Sintassi - Lessico - Coesione
<u>Fluente</u> <ul style="list-style-type: none"> - Esitazioni - Ripetizioni - Ri-selezione di parole inappropriate - Ri-strutturazione delle frasi/enunciati - Coesione 	<u>Competenze strategiche</u> <i>Strategie di conseguimento</i> <ul style="list-style-type: none"> - Iper-generalizzazione - Parafrasi - Conio di parole - Ristrutturazione - Strategie cooperative - Code-switching - Strategie non linguistiche <i>Strategie di evitamento</i> <ul style="list-style-type: none"> - Evitamenti formali - Evitamenti funzionali
<u>Competenza testuale</u> <i>La struttura della conversazione</i> <ul style="list-style-type: none"> - Presa dei turni - Coppie adiacenti - Aperture e chiusure 	
<u>Competenze pragmatiche</u> <ul style="list-style-type: none"> - Appropriatazza - Implicature - Espressività 	<u>Competenze sociolinguistiche</u> <ul style="list-style-type: none"> - Situazionali - Argomentative - Culturali

4. LA SCRITTURA DELLE SPECIFICHE

Una volta definito il costrutto del test da sviluppare si può procedere alla scrittura degli *item* e conseguentemente delle specifiche del test. Le specifiche sono il documento di riferimento nella produzione di diverse forme dello stesso test. Vengono definite come il livello più alto in termini di dettagli riguardo la struttura portante del test. Sono chiamate anche *blueprints* per l'analogia con il mondo dell'architettura e dell'ingegneria (Alderson, Clapham & Wall, 1995: 9). Le specifiche sono quei documenti che mostrano come costruire un edificio, una macchina o appunto un test. Sono diverse da un sillabo, come sostengono Alderson *et al.* (1995): il sillabo è infatti normalmente orientato verso i bisogni degli studenti

che dovranno sostenere il test e degli insegnanti, mentre le specifiche sono usate da chi scrive e sviluppa il test.

Le *specifiche* possono essere scritte secondo diversi formati. Possono essere contenute in un unico documento in cui viene descritto lo scopo, il costrutto, alcuni esempi degli *item* e informazioni riguardo il numero di *item* che il test deve avere nel suo complesso. In alcuni casi può essere necessario avere dei documenti separati per ciascun tipo di *item*, se il test è composto di *item* diversi. Le specifiche saranno più complesse nella misura in cui il test stesso è complesso. Modelli e schemi per la scrittura delle *specifiche* sono proposti dal Consiglio d'Europa (2009) nelle appendici del manuale *Relating Language Examinations to the CEFR*, da Fulcher (2010: 127-158) e da O'Sullivan (2012a: 47-58). Riportiamo in appendice un adattamento delle griglie proposte da O'Sullivan (2012a).

5. LE CERTIFICAZIONI DELLA LINGUA ITALIANA E LE PROVE ORALI

In questa sezione vengono presentati i formati della componente “produzione orale” di alcuni test esistenti per la certificazione della lingua italiana: il CELI, il CILS e il PLIDA. Ci si soffermerà in particolare sui livelli alti di queste certificazioni, il B2 e il C1.

Il CELI, *Certificato della Lingua Italiana*, è prodotto e sviluppato dal *Centro per la Valutazione e le Certificazioni Linguistiche* (CVCL) dell'Università per stranieri di Perugia (cfr. Grego Bolli & Spiti, 2000) e mira a certificare le competenze e capacità d'uso della lingua italiana generale spendibili per scopi professionali e di studio.

Il CILS, la *Certificazione di Italiano come Lingua Straniera*, è sviluppato e rilasciato dall'Università per Stranieri di Siena. È un test per la valutazione delle abilità comunicative che verifica la capacità di uso della lingua in situazioni comunicative e in rapporto alle loro caratteristiche socio-culturali e socio-linguistiche (cfr. Barni *et al.*, 2009). Questa certificazione basa la propria elaborazione su un modello di competenza legato ai contesti d'uso della lingua, sull'analisi dei testi usati come *input* linguistico e su una loro scelta accurata.

Il PLIDA, *Progetto Lingua Italiana Dante Alighieri*, è il diploma di certificazione che attesta la competenza in italiano come lingua straniera rilasciato dalla Società Dante Alighieri. «Le prove di certificazione sono costruite sui principi dell'approccio comunicativo e si basano su attività pratiche collegate alla vita reale che richiedono l'impiego delle quattro abilità linguistiche di base: ascoltare, leggere, parlare, scrivere. Le situazioni comunicative presentate nelle prove seguono una precisa scala di difficoltà: dai contesti vicini alla vita familiare e quotidiana si passa a quelli legati ad ambiti professionali di alta qualificazione» (Patota & Pizzoli, 2004: 2).

Le tre istituzioni offrono certificazioni autonome e complete per ciascun livello di competenza, ciascuno allineato ai livelli del Quadro Comune di riferimento per le lingue (2002). La Tabella 2 sintetizza le principali caratteristiche delle tre certificazioni nei due livelli presi in esame.

Tabella 2. *Le certificazioni di italiano per i livelli B2-C1.*

Test		Struttura	No. task	Durata	Tempo di preparazione
B2	CELI 3	Commento e descrizione di una foto Riassunto Role-play	3	15 min.	15 min. prima dell'inizio della prova
	CILS Due	Intervista con esaminatore Monologo su un tema dato	2	5 min. ca.	nd
	PLIDA B2	Presentazione Interazione guidata (situazione imprevista) Monologo a partire da tema + immagini	3	10 min. ca.	nd
C1	CELI 4	Commento analogie/differenze di due foto Riassunto Illustrazione grafico/tabella	3	20 min.	15 min. prima dell'inizio della prova
	CILS Tre	Role-play Monologo	2	7 min. ca.	3 min. prima di ciascuna prova
	PLIDA C1	Presentazione Interazione guidata Monologo a partire da tema + immagini	3	13 min. ca.	nd

Le tre certificazioni richiedono ai candidati sia una produzione di tipo monologico, sia di tipo dialogico. Il parlato monologico viene elicitato con la richiesta di una produzione di un riassunto per il CELI, di un discorso elicitato a partire da un breve testo come nel caso del CILS, di un discorso prodotto a partire dalla descrizione di immagini su un tema specifico come nel caso del PLIDA e del CELI. Per quanto riguarda invece il parlato dialogico, le tipologie di elicitazione variano dall'intervista classica con l'esaminatore nel CILS due, al *role-play* o "interazione guidata", come viene definito nel PLIDA.

Una differenza che emerge dal confronto delle tre certificazioni riguarda i tempi per l'esecuzione delle produzioni. Si può notare come per il livello più avanzato venga dato progressivamente più tempo rispetto al livello precedente in tutte e tre le certificazioni, ad indicare che per la valutazione di un livello avanzato, caratterizzato dalla complessità e astrattezza della produzione, sia necessario consentire più tempo per una valutazione obiettiva. Quello che varia è il tempo per l'esecuzione della prova per lo stesso livello, per cui se nel CELI 3 viene richiesta una produzione di circa 15 minuti, nel PLIDA viene richiesta una produzione di circa 10 minuti e nel CILS della metà del tempo, 5 minuti. La quantità di parlato che viene elicitata pone quindi un quesito riguardo quanto estesa debba essere la produzione di un candidato per poter offrire un campione di parlato sufficiente per una valutazione obiettiva e che renda conto della effettiva competenza di ciascuno.

Nelle prossime sezioni (5 e 6) verranno illustrate alcune delle tipologie di *task* usate in queste certificazioni, distinte per i livelli di competenza da valutare, il B2 e il C1.

5. LE PROVE DI PRODUZIONE ORALE A LIVELLO B2

Nel CELI 3 il candidato deve essere in grado di prendere parte in modo sufficientemente spedito a una conversazione faccia a faccia con i suoi esaminatori in lingua non formale. Deve saper descrivere e raccontare, esprimendo in modo semplice ma logico e corretto, la propria opinione e confrontarla con quella degli esaminatori.

Dopo una breve presentazione, il candidato deve descrivere e commentare una foto o una illustrazione che rappresenta persone, oggetti o situazioni riferiti a realtà quotidiane. In questa fase è quindi richiesto al candidato di produrre un discorso autonomo a partire dall'input proposto e poi di proseguire la conversazione sulla base di alcune domande poste dall'esaminatore per guidare e stimolare la produzione.

In una seconda fase, il candidato deve riassumere un breve testo come nell'esempio 1. L'esaminatore ha poi a disposizione alcune domande guida per stimolare la conversazione.

Esempio 1.

C'è un appendino, nel bagno, di quelli che si attaccano a pressione, che ogni tanto si scolla. Una notte è caduto con un fracasso agghiacciante. Poi ci sono i mobili, che di notte si lamentano come fossero persone, come se nel silenzio ritrovassero una vita segreta.

Forse, mi son detta, è un rumore del legno, uno di quei suoni naturali che di giorno non riusciamo a distinguere, sopraffatti da altri rumori.

Le mie dita erano ancora serrate al bordo del lenzuolo, quando il rumore di un passo, questa volta netto e preciso, mi ha fatto gelare il sangue. Sentivo che avrei dovuto balzare dal letto e correre a chiudere a chiave la porta della camera, ma non riuscivo a farlo. E nemmeno a lanciare un grido, nel timore di richiamare su di me quei passi.

(C. Castellaneta, *Questa primavera*)

Domande guida

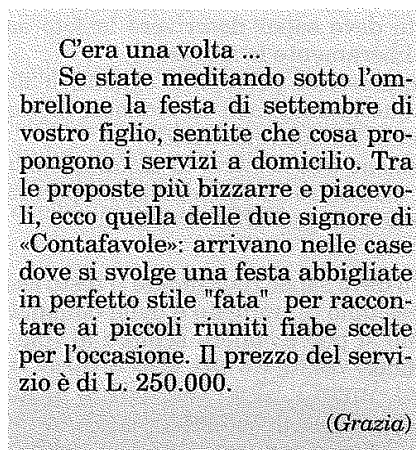
- Riassuma il testo.
- Come immagina che finirà il racconto?
- Come reagisce Lei se di notte Le sembra di sentire un rumore in casa?

(Grego-Bolli & Spiti, 2000: 213)

Infine, nella terza parte della prova, viene chiesto al candidato di misurarsi all'interno di un *role-play*, avendo come interlocutore uno degli esaminatori della commissione. L'esempio 2 propone una situazione di *role-play*.

Esempio 2.

Un'amica italiana Le chiede consiglio per organizzare una festa per il figlio che compie cinque anni. Lei ricorda di aver letto questo annuncio



Come propone quest'idea all'amica?

(Grego-Bolli & Spiti, 2000: 214)

Nel CILS le prove orali devono essere interamente registrate (su audiocassette o su CD) e vengono inviate all'Università per Stranieri per la valutazione centrale. A questo livello

viene chiesto al candidato – le prove sono individuali – di sostenere innanzitutto una conversazione con l'esaminatore per i primi due-tre minuti su un tema a scelta, ad esempio tra quelli proposti nell'esempio 3.

Esempio 3.

- Se dovesse parlare del suo paese ad un amico che vuole trasferirsi lì, quali aspetti positivi e negativi descriverebbe?
- Pensa che sia importante conoscere una o più lingue straniere? Perché?
- Come immagina una società ideale dove vivere?
- Che cosa farebbe per migliorare la vita nella sua città?

(materiali CILS, giugno 2009²).

Nella seconda parte della prova viene invece chiesto al candidato di produrre un monologo di circa un minuto e mezzo nel quale descrivere una situazione come, ad esempio, *la scoperta scientifica che ritiene fondamentale per il miglioramento delle condizioni di vita dell'uomo* o *l'epoca del passato che l'affascina di più* (cfr. materiali CILS, giugno 2009) o un'immagine tra due proposte.

Nel PLIDA la prova orale può essere sostenuta dal candidato sia singolarmente sia in coppia con un altro, a discrezione dei candidati stessi. La prova comincia con una prima fase di presentazione, della durata di un minuto, che ha lo scopo di allentare la tensione. In questa fase l'intervistatore, dopo essersi presentato, chiede al candidato di spiegare il motivo dello studio dell'italiano. Segue poi una "interazione guidata", della durata di 5-7 minuti, in cui è prevista una partecipazione attiva del candidato: gli viene presentata una situazione imprevista in contesti non quotidiani (un incidente, un reclamo). La partecipazione del candidato in questa fase consiste nella formulazione di domande legate all'argomento, di commenti e proposte alternative per la risoluzione della situazione rivolte al proprio interlocutore, all'intervistatore o ad un altro candidato, così come nell'esempio 4.

Esempio 4.

Dopo diversi mesi trascorsi in Italia per lavoro, avete finalmente una settimana libera e volete visitare alcune città d'arte italiane. Vorreste invitare un vostro amico/una vostra amica italiano/a a venire con voi ma non sapete se è interessato/a all'itinerario che avete in mente. Lo/la chiamate al telefono:

- gli/le proponete il vostro programma;
- gli/le chiedete che cosa ne pensa;
- cercate di convincerlo/convincerla a fare il viaggio.

(Prove d'esame PLIDA, interazione guidata, novembre 2010³)

² Materiali CILS: <http://www.gedi.it/cils/file/89/198/file/b2-po.pdf>

³ Materiali PLIDA: <http://www.plida.it/plida/certificazione/prove-desame.html>

Infine, nell'ultima fase viene chiesto al candidato di produrre un monologo scelto a partire da una lista di tre temi accompagnati da immagini. Viene lasciato qualche minuto per scegliere l'argomento e raccogliere le idee, anche prendendo qualche appunto su cosa dire. L'intervistatore «è tenuto a sostenere il colloquio con osservazioni, suggerimenti e brevi commenti» (Patota e Pizzoli, 2004: 6).

6. LE PROVE DI PRODUZIONE ORALE A LIVELLO C1

Nel CELI 4 dopo la presentazione vengono fatte esaminare al candidato, che deve metterne in risalto analogie e differenze, due foto dal soggetto simile. Poi viene richiesto al candidato di riassumere un testo ed esprimere su di esso il proprio punto di vista. Nell'ultima parte del test gli viene chiesto di illustrare una tabella o un grafico o uno schema esponendone i dati rappresentati. Un esempio è fornito nell'esempio 5, in cui viene mostrata una tabella riguardo ad alcune organizzazioni di volontariato. La prova dura complessivamente 20 minuti.

Esempio 5.

INTERVENTI IN OSPEDALE, APOGGIO AGLI HANDICAPPATI			
CONTO CORRENTE	SCOPO	LAVORO DEL VOLONTARIO	NOTE
c.c. postale 88885005	Centro di accoglienza per anziani e persone con handicap fisici e mentali.	Sia gli anziani sia gli handicappati hanno bisogno di compagnia, di qualcuno che li ascolti e parli.	E' un centro di accoglienza, con unica sede a Roma, fondato da un gruppo di volontari nel 1981.
c.c.postale 74885009S	Sostegno psicologico e pratico agli ammalati e alle loro famiglie. Tutela dei diritti e diffusione di informazioni.	Assiste i pazienti down e le famiglie e organizza la segreteria e le raccolte di fondi.	E' presente in 16 città ed è stata fondata nel 1979. Ogni volontario lavora come minimo mezza giornata alla settimana.
c.c. postale 912302	Sostegno pratico e psicologico alle persone spastiche e alle loro famiglie.	Il volontario deve creare un buon rapporto con il malato, trascorrere con lui parte del suo tempo libero, come fossero amici.	Nata nel 1990 ha 171 sezioni autonome. L'impegno minimo è di mezza giornata alla settimana.
c.c.postale 13613203	Realizzazioni di programmi di formazione e ricerca in favore dei disabili verbali.	Il volontario gestisce direttamente l'associazione e impara a parlare con i disabili.	E' nata nel 1985, ha sede a Milano ma agisce in tutta Italia. Per i volontari è previsto un corso.
c.c. postale 46716007	Raccolte di fondi per la ricerca scientifica, collaborazione con i reparti di ematologia e attività di sostegno ai pazienti.	Può essere utilizzato nel lavoro amministrativo e di segreteria ma anche nei vari reparti di ematologia con i pazienti.	Nata nel '69, è presente in 14 città italiane. (L'Espresso)

(Grego-Bolli, Spiti, 2000: 234)

Nel CILS l'esame a livello C1 consiste nella prima parte in una presa di ruoli a scelta da un elenco come nell'esempio 6 seguente.

Esempio 6.

- Il candidato è membro di un'associazione di volontari per la tutela degli animali. Vuole organizzare una campagna di sensibilizzazione. Si reca presso una banca per chiedere di sponsorizzare l'iniziativa e spiega il progetto ad un funzionario.
- Il candidato è appena tornato da un viaggio organizzato. La gita è stata un vero disastro e quindi si reca all'agenzia che l'ha organizzata per protestare e per chiedere un risarcimento.
- Il candidato si presenta ad un colloquio di lavoro presso una casa editrice. Presenta il suo curriculum e si informa sulle condizioni di lavoro.
- Il candidato è stato scippato per strada. Si reca al posto di polizia per sporgere denuncia e spiega come si sono svolti i fatti.

(Materiali d'esame CILS, giugno 2009)

Nella seconda parte dell'esame CILS, come per il livello B2, viene chiesto al candidato di fare una breve esposizione nella quale esprimere la propria opinione personale su un tema a scelta. Una volta scelto il tema i candidati hanno tre minuti per preparare il loro discorso e devono parlare per altri tre minuti circa.

Nel PLIDA anche al livello C1 la prova è costituita da una prima fase di presentazione in cui l'intervistatore chiede al candidato di descrivere gli usi che, in futuro, intende fare della lingua italiana. Nella fase successiva, come per il livello B2, viene presentata un'attività di *role-play*, l'"interazione guidata", la cui durata deve essere tra i 5 e i 7 minuti. In questo caso viene presentata al candidato, anche in modo ironico e scherzoso, una situazione imprevista nella quale gli viene chiesto di risolvere un compito. L'esempio 7 mostra una delle situazioni presentate in un esame per il livello C1.

Esempio 7.

Lavorate da molti anni in un negozio che vende articoli di lusso (orologi, gioielli, ecc.). Avete sempre dimostrato grande professionalità, affidabilità e onestà nel vostro lavoro e avete conquistato nel tempo la fiducia del proprietario del negozio. Un lunedì mattina nel negozio dove lavorate mancano alcuni oggetti di valore, ma non ci sono segni di effrazione. Il proprietario del negozio vi accusa ingiustamente di avere commesso il furto poiché siete l'unica persona ad avere le chiavi del negozio oltre a lui:

- dimostrate l'impossibilità del vostro coinvolgimento nel furto;
- esprimete la vostra indignazione per l'accusa che vi è stata mossa e ricordate al proprietario la vostra lunga esperienza insieme;

- tentate di ricostruire insieme al proprietario quello che può essere accaduto e ipotizzate il coinvolgimento di altre persone nel fatto.

(Indicazioni per l'intervistatore: siete il proprietario del negozio, siete fuori di voi dalla rabbia e dalla preoccupazione per il furto subito; ve la prendete con la persona che lavora con voi, ma poi vi rendete conto che la state accusando ingiustamente).

(Prove d'esame PLIDA, novembre 2010)

Nella terza e ultima fase dell'esame PLIDA viene chiesto al candidato di produrre un monologo a scelta tra tre temi complessi accompagnati da delle immagini. Anche in questo caso l'intervistatore interviene nel colloquio con osservazioni, suggerimenti o brevi commenti.

Quello che emerge dalla struttura delle diverse fasi delle prove nelle tre certificazioni è il ruolo molto attivo rappresentato dall'esaminatore. Nelle parti dialogiche l'interazione si svolge sempre tra candidato e esaminatore, dando luogo ad un formato di interazione di tipo asimmetrico. Le tre certificazioni hanno dunque optato per questo formato di interazione piuttosto che per uno nel quale i candidati si confrontino tra loro in questa fase in maniera simmetrica, come avviene, ad esempio, in altre certificazioni per l'inglese (cfr. *First Certificate of English, FCE*). Un'eccezione è rappresentata da alcuni casi dell'esame PLIDA, in cui i candidati possono optare per un formato di esame a coppie e in cui, quindi, l'interazione prevista dai *task* dialogici è da loro direttamente gestita.

Il formato scelto prevalentemente per le certificazioni di italiano è dunque quello che garantisce maggiore controllo della produzione da parte dell'esaminatore. Se, da un lato, in questo tipo di formato si può controllare la produzione dell'esaminatore attraverso la sua formazione e guidando i suoi interventi con delle istruzioni precise (cfr. Lazaraton, 1996; O'Loughlin, 2001), dall'altro i suoi interventi hanno un effetto marcato sulla performance dei candidati (cfr. Brooks, 2009; Brown 2003, 2005; Lumley & Brown, 1996; O'Sullivan, 2000; Ross, 1992), tanto che il contributo dell'esaminatore nel test deve essere considerato comunque come una "*wild card*" (Brooks 2009: 344) per l'insieme di variabili che entrano in gioco.

7. UNA PROPOSTA DI TEST PER LA VALUTAZIONE DELLE COMPETENZE ORALI

La Libera Università di Bolzano (LUB) si inserisce nel contesto socio-linguistico locale offrendo una formazione accademica plurilingue, in cui non solo vengono offerti insegnamenti in italiano e tedesco, le due lingue parlate dalle comunità linguistiche locali, ma a questi si aggiunge anche l'inglese che è una delle lingue ufficiali dell'università. Il carattere plurilingue dell'ateneo è uno dei motivi per i quali gli studenti da altre parti d'Italia e dall'estero decidono di studiare in questa città. Ne risulta un ateneo a vocazione internazionale in cui il 15% della popolazione studentesca proviene dall'estero e sceglie di studiare alla LUB proprio per il suo profilo linguistico (cfr. Franceschini, in corso di stampa). Il nuovo modello linguistico in vigore a partire dal 2011-12 prevede che, per

garantire il carattere plurilingue dell'istituzione, gli studenti debbano certificare le loro competenze linguistiche sia in ingresso sia in uscita. Così, se in ingresso è richiesta una competenza pari al livello B2 del QCER in almeno due delle tre lingue ufficiali dell'università, alla conclusione degli studi gli studenti devono dimostrare di avere raggiunto il livello C1 nella prima lingua, il livello B2 potenziato nella seconda e il livello B2 nella terza lingua. La tabella 3 riassume i livelli di competenza richiesti per entrambi i momenti.

Tabella 3. *Le competenze linguistiche richieste dal nuovo modello linguistico.*

Lingua ufficiale LUB	Livello minimo di ingresso	Livello minimo di uscita
1° lingua	B2	C1
2° lingua	B2	B2+
3° lingua	---	B2

Si potrebbe pensare che la prima lingua corrisponda alla L1 degli studenti, questo corrisponde solo in parte alla situazione reale perché non sempre la prima lingua scelta per lo studio corrisponde alla L1 dello studente. Per questo motivo, nei casi in cui non ci sia una corrispondenza con la L1, si è reso necessario pensare a uno strumento di valutazione anche per la prima lingua dichiarata.

Il nuovo modello linguistico ha quindi tra i suoi obiettivi quello di dare maggiore risalto e attenzione alla formazione linguistica degli studenti che si aggiunge e integra con quella delle discipline caratterizzanti di ciascun corso di laurea.

Per quanto riguarda la valutazione delle competenze orali degli studenti viene qui presentato un test sviluppato in particolare per la valutazione alla fine dei corsi di laurea da un gruppo di ricerca coordinato da chi scrive⁴.

Il test del parlato sviluppato mira ad offrire a tutti gli studenti una valutazione ed una certificazione esplicita, interna all'università, delle competenze linguistiche in ambito accademico generale. Il test è stato sviluppato, con lo stesso impianto, in tre lingue: italiano, inglese e tedesco. Le competenze vengono valutate in maniera progressiva per due ragioni. Prima di tutto i livelli soglia da valutare sono quelli che attestano la competenza avanzata dei parlanti tra il B2 e il C1. In secondo luogo il numero di studenti che potrebbero sostenere il test per ciascun livello non è elevato, in quanto una parte della popolazione studentesca decide di investire nelle competenze linguistiche optando per una certificazione internazionale. A differenza quindi dei test sopra esaminati che attestano le competenze ad un livello specifico, B2 o C1, ecc., il test sviluppato dal nostro gruppo di ricerca dell'università misura le competenze lungo il *continuum* tra il B2 e il C1.

⁴ Ringrazio Felicità Bettoni, Renata Cavosi Silbernagl, Rosmarie De Monte Frick, Sharon Hartle, Federica Ledda, Sandra Montali, Marcella Perisutti, Claudia Provenzano, Cristina Tomei, Brigitte Widmann per la loro attiva collaborazione al progetto nell'Anno Accademico 2011-2012.

Il parlato che si intende valutare rientra nel linguaggio accademico generale, di persone istruite, e mira a testare le competenze pragmatiche e sociolinguistiche degli studenti sia in situazioni dialogiche, sia monologiche. L'attenzione è rivolta in particolare, oltre alla capacità di sintesi, alle capacità dei candidati di argomentare e di interagire con un certo grado di formalità.

Con questo test si vogliono quindi esaminare da un lato la capacità di sostenere un monologo su argomenti astratti, così come avviene all'interno dei corsi e seminari in brevi presentazioni richieste dal docente, e poi durante gli esami di laurea. Dall'altro lato si intende valutare la capacità degli studenti di interagire e di esprimere la propria opinione, di fare confronti, di argomentare i pro e contra in diverse situazioni. Infine si intende valutare la capacità dei candidati di saper gestire situazioni sia informali sia formali.

Il formato del test prevede la presenza di due esaminatori, un intervistatore e un valutatore, e due candidati. Il test è composto di tre parti e dura circa 15 minuti:

- nella prima parte, dopo una breve presentazione, ciascuno studente risponde a delle domande poste dall'interlocutore riguardanti la propria carriera universitaria (facoltà, esami, tirocini, aspettative per il futuro, club studenteschi, ecc.). Questa fase dura circa 4 minuti;
- nella seconda parte viene chiesto a ciascun candidato di produrre un monologo di circa 2 minuti a partire da un breve testo seguito da alcune domande guida, utili per la strutturazione del discorso. Ciascun candidato ha a disposizione un minuto per raccogliere le idee. Alla fine del suo discorso il compagno gli pone una domanda su ciò che ha esposto, alla quale il candidato risponde. Questa sequenza dura al massimo 1 minuto, al termine del quale l'intervistatore procede con il secondo candidato;
- nella terza parte infine viene chiesto ai candidati di interagire tra loro, commentare e confrontare tre immagini contestualizzate dall'intervistatore in una situazione concreta. Alla fine della conversazione è chiesto a ciascun candidato di scegliere un'immagine e di convincere il compagno dei vantaggi della soluzione scelta. Anche in questo caso i candidati hanno a disposizione 30 secondi per raccogliere le idee, e 3-4 minuti per completare il compito.

Si è cercato anche di implementare il carattere accademico del test nel contestualizzare le diverse fasi in contesti perlopiù accademici o legati alla vita universitaria e questo contribuisce ad aumentare l'autenticità del compito stesso.

8. ALCUNI PROBLEMI METODOLOGICI NELLO SVILUPPO DI UN TEST

Sviluppare un test linguistico è un processo complesso in sé, qualsiasi sia l'abilità da dover valutare, per le implicazioni di tipo sia teorico nella definizione del costrutto sia metodologico e pratico. In particolare lo sviluppo di un test per la valutazione delle abilità di produzione orale per gli studenti della nostra università ha posto diversi quesiti dovuti

anche alla specificità del contesto nel quale si opera, soprattutto per quanto riguarda la scelta dei materiali da usare.

Una questione che si è posta nella fase iniziale di definizione degli obiettivi, del costrutto e del formato da usare è quella della diversità di approcci e tradizioni didattiche, così come di valutazione e *testing* nelle tre lingue, italiano, inglese e tedesco. Se infatti per la lingua inglese la tradizione relativa al *testing* è ben consolidata e lunga è la lista di ricerche in questo campo, non si può dire lo stesso per l'italiano e per il tedesco. La difficoltà principale in questa prima fase di elaborazione è stata quindi quella del rapporto tra rigidità del formato del test che si andava a costruire e il tipo di istruzioni e di comportamento che avrebbero dovuto assumere gli esaminatori, con un conseguente impatto sullo stile comunicativo di ciascuno. La decisione presa in questa prima fase di elaborazione è stata quella di cercare di fare in modo che tutti gli esaminatori fossero in grado di seguire le istruzioni, ma che allo stesso tempo potessero mantenere il loro stile comunicativo. Un esempio riguarda l'uso dei materiali per gli esaminatori e la loro formazione. Se nella tradizione anglosassone l'uso dei materiali per gli esaminatori e delle istruzioni (*rubrics*) è assodato e gli esaminatori richiedono indicazioni dettagliate sul loro modo di comportarsi durante la sessione di esame, lo stesso non avviene da parte degli esaminatori di tedesco e italiano. Il seguire pedissequamente le istruzioni è percepito negativamente dagli esaminatori di tedesco e italiano, i quali tendono ad avere una visione più libera del materiale da usare. Nel caso degli esaminatori di italiano poi, si aggiunge una ulteriore questione: l'atteggiamento negativo da parte degli esaminatori nei riguardi delle istruzioni proposte da un collega, quasi a voler significare che l'esaminatore in questione non sia capace di gestire un'interazione durante un esame. Il contesto locale ha spesso visto vicini in questo tipo di atteggiamento negativo nei confronti della stesura delle istruzioni per gli esaminatori sia gli insegnanti di tedesco che quelli di italiano. Il nostro gruppo di ricerca si è dunque chiesto come fare per garantire un comportamento comune e affidabile degli esaminatori durante l'esame e che allo stesso tempo non risultasse impositivo nei confronti degli esaminatori. Si è cercato quindi da un lato di scrivere le istruzioni per gli esaminatori nel modo più semplice possibile per ciascun *task*, limitando la verbosità dei testi che sarebbero risultati problematici non solo perché poco comunicativi, ma anche poco leggibili durante l'esame orale, soprattutto per chi non è abituato a seguire in maniera sistematica le istruzioni. Dall'altro lato è chiaramente emersa la necessità prioritaria di formare gli esaminatori in maniera puntuale, costruttiva e costante, alla luce anche del comportamento degli esaminatori durante le sessioni di *pre-trialling* del test.

Data la grande variabilità del comportamento degli esaminatori e la diversità di formazione sul *testing* si è deciso quindi di optare per il formato di esame con coppie di candidati. Questa scelta è stata fatta anche per garantire maggiore autenticità nella performance degli studenti e per offrire loro maggiore possibilità di esprimersi. Gli studi sull'inglese che confrontano i diversi tipi di formato di test hanno infatti mostrato come il test a coppie elicitò un tipo di interazione più complessa tra i partecipanti, quali il sollecitare, l'elaborare, il completare gli enunciati, il fare riferimento alle idee del partner, la parafrasi (Brooks, 2009; Galaczi, 2008), e bilanciò la simmetria dei rapporti nell'interazione. Se nel formato classico il rapporto tra esaminatore e candidato è sempre asimmetrico, nel formato

di esame con la coppia di candidati si cerca di bilanciarlo con alcune parti dell'esame gestite in maniera simmetrica tra i candidati e altre caratterizzate da un'interazione di tipo asimmetrico tra candidati e esaminatore (Galaczi & French, 2011). Questo tipo di formato non è chiaramente esente da altre variabili che entrano in gioco e che possono risultare problematiche, quali il carattere introverso o estroverso dei candidati (cfr. Van Moere, 2006), il diverso livello di competenza dei candidati nella coppia (Iwashita, 1996), il grado di familiarità con l'interlocutore (O'Sullivan, 2002). Nonostante ciò, si è comunque deciso per questo formato. Gli studi esistenti infatti, sembrano provare che siano più numerosi i vantaggi che gli effetti negativi delle variabili che entrano in gioco nell'uso di questo formato. Tra i possibili aspetti negativi non vi è, per altro, nessun accordo tra gli studiosi. Infine, un ultimo aspetto da prendere in considerazione riguarda l'obiettivo del test di valutare progressivamente la competenza dei candidati per i livelli B2-C1, caratterizzati l'uno da maggiore concretezza e familiarità nelle situazioni quotidiane, l'altro da una capacità di saper parlare anche di argomenti più astratti e complessi. Il nostro tentativo è stato quindi quello di produrre *item* su argomenti più concreti e *item* su argomenti più astratti, per operare poi un'associazione fra *task* più semplici e *task* più complessi che potessero essere somministrati all'interno di ciascuna prova e dare così la possibilità ai candidati di mostrare la loro abilità.

9. CONCLUSIONI

Si è cercato in questo contributo di dare maggiore luce agli aspetti metodologici primari nell'elaborazione di un test per l'abilità del parlato, ponendo l'accento sulle prime fasi di questo processo. Lo sviluppo di test per la valutazione dell'abilità orale, in definitiva, non è tanto più difficile rispetto a quello per la valutazione delle altre abilità, come afferma O'Sullivan (2012b). Come si è infatti più sopra sostenuto, le fasi iniziali nello sviluppo di un test per il parlato sono le stesse di quelle relative allo sviluppo di test per le altre abilità. Senza un obiettivo chiaro di quale sia lo scopo e senza una definizione puntuale del costrutto nessun test può essere sviluppato solidamente. È proprio la definizione del costrutto, cioè del tipo di parlato che si vuole valutare, coerentemente con lo scopo che il test si prefigge, l'elemento che costituisce la difficoltà maggiore nelle fasi iniziali dello sviluppo di un test, in quanto da questo dipendono poi le decisioni successive.

Diverse possono essere le attività che caratterizzano i test per la valutazione del parlato, come mostrano le differenti attività scelte per la valutazione di questa abilità nelle tre certificazioni prese in esame, così come diversi possono essere i formati dei test, con prove individuali o a coppie di candidati. Quale sia il formato più opportuno e di successo per valutare le competenze sul parlato dell'italiano degli stranieri richiede una maggiore ricerca scientifica in questo ambito. Molto nella ricerca sul *testing* in Italia deve ancora essere fatto, soprattutto nel campo della validazione delle prove orali dei test esistenti, sicuramente oggetto di maggiore difficoltà proprio per il carattere sfuggente del parlato. Il presente contributo, anche con lo sviluppo di un test per la certificazione interna del nostro ateneo, intende essere una prima riflessione in questo ambito.

RIFERIMENTI BIBLIOGRAFICI

- Alderson J. C., Clapham C. & Wall D. (1995), *Language test construction and evaluation*, Cambridge University Press, Cambridge.
- Bachman L.F. & Palmer A.S. (1996), *Language testing in practice*, Oxford University Press, Oxford.
- Barni M., Bandini A., Sprugnoli L., Lucarelli S., Scaglioso A.M., Strambi B., Fusi C., Arruffoli A.M. (2009), *Linee guida CILS. Certificazione di Italiano come Lingua Straniera*, Guerra, Perugia.
- Brooks L. (2009), "Interacting in pairs in a test of oral proficiency: co-constructing better performance", in *Language Testing*, 26/3, pp. 341-366.
- Brown A. (2003), "Interviewer variation and the co-construction of speaking proficiency", in *Language Testing*, 20, pp. 1-25.
- Brown A. (2005), *Interviewer variability in oral proficiency interviews*, Peter Lang, Frankfurt.
- Bygate M. (1987), *Speaking*, Oxford University Press, Oxford.
- Consiglio d'Europa (2002), *Quadro Comune Europeo di Riferimento per le Lingue: Apprendimento, Insegnamento, Valutazione*, trad. it. di Bertocchi D. e Quartapelle F., Milano-Firenze, RCS Scuola/La Nuova Italia-Oxford University Press (tit. or. *Common European Framework for Languages: Learning, Teaching, Assessment*, Council of Europe, Strasbourg, 2001).
- Coombe C., Davidson P., O'Sullivan B., Stoyhoff S. (eds.) (2012), *Second Language Assessment*. Cambridge University Press, Cambridge.
- Council of Europe (2009), *Relating language examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual*. Council of Europe Language Policy Division, Strasbourg:
http://www.coe.int/t/dg4/linguistic/Manuel1_EN.asp#TopOfPage.
- Davis L. (2009), "The influence of interlocutor proficiency in a paired oral assessment", in *Language Testing*, 26/3, pp. 367-396.
- Ducasse A.M. & Brown A. (2009), "Assessing paired orals: rater's orientation to interaction", in *Language Testing*, 26/3, pp. 423-443.
- Franceschini, R. (in corso di stampa), "CLIL universitario: una risposta europea per l'inclusione della diversità linguistica", in M. Gotti e C. Nickenig (a cura di), *Multilinguismo, CLIL e innovazione didattica*, Bozen-Bolzano: Bozen-Bolzano University Press.
- Fulcher G. & Davidson F. (2009), "Test architecture, test retrofit", in *Language Testing*, 26/1, pp. 123-144.
- Fulcher G. (2003), *Testing Second Language Speaking*, Pearson, London.
- Fulcher G. (2006), "Test architecture", in *Foreign language education research*, 9, pp. 1-22.
- Fulcher G. (2010), *Practical Language Testing*, Hodder Education, London.
- Galaczi E. D. (2008), "Peer-peer interaction in a speaking test: the case of the First certificate in English examination", in *Language Assessment Quarterly*, 5, pp. 89-119.

- Galaczi E.D. & French A. (2011), "Context validity", in L. Taylor. (ed.), *Examining Speaking. Research and practice in second language speaking*, Cambridge University Press, Cambridge, pp. 112-170.
- Grego Bolli G. & Spiti M.G. (2000), *La verifica delle competenze linguistiche: Misurare e valutare nella certificazione CELI. Linee guida alla certificazione*, Guerra, Perugia.
- Harris D. & Bell C. (1994), *Evaluating and assessing for learning*, Kogan Page, London.
- Ingram E. (1968), "Attainment and diagnostic testing", in A. Davies (ed.), *Language Test in Symposium: a psycholinguistic approach*, Oxford University Press, Oxford.
- Iwashita N. (1996), "The validity of the paired interview format in oral performance assessment", in *Melbourne papers in language testing*, 5/2, pp. 51-66.
- Lado R. (1961), *Language Testing*, Longman, London.
- Lazaraton A. (1996), "Interlocutor support in oral proficiency interviews. The case of CASE", in *Language Testing*, 13/2, pp. 151-172.
- Lazaraton A. (2008), *A qualitative approach to the validation of oral language tests*, Cambridge University Press, Cambridge.
- Lumley T. & Brown A. (1996), "Specific purpose language performance tests: task and interaction", in G. Wigglesworth & C. Elder (eds.), *The Language Testing Cycle: from inception to washback*, *Australian Review of Applied Linguistics*, Series S/13, pp. 105-136.
- Luoma S. (2004), *Assessing speaking*, Cambridge University Press, Cambridge.
- May L. (2009), "Co-constructed interaction in a paired speaking test: the rater's perspective", in *Language Testing*, 26/3, pp. 397-421.
- McNamara T. & Lumley T. (1997), "The effect of the interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings", in *Language Testing*, 14/2, pp. 140-156.
- Nakatsuhara F. (2011), "Effects of test-taker characteristics and the number of participants in group oral tests", in *Language Testing*, 28/4, pp. 483-508.
- O'Loughlin K. (2001), *The equivalence of direct and semi-direct speaking tests*, Cambridge University Press, Cambridge.
- O'Sullivan B. (2000), "Exploring gender and oral proficiency interview performance", in *System*, 28, pp. 373-386.
- O'Sullivan B. (2002), "Learner acquaintanceship and oral proficiency test pair-task performance", in *Language Testing*, 19, pp. 277-295.
- O'Sullivan B. (2008), *Modelling performance in oral language tests*, Peter Lang, Frankfurt.
- O'Sullivan B. (2012a), "The assessment development process", in C. Coombie, P. Davidson, B. O'Sullivan, S. Stoyloff (eds.), *Second Language Assessment*, Cambridge: Cambridge University Press, Cambridge, pp. 47-58.
- O'Sullivan B. (2012b), "Assessing Speaking", in C. Coombie, P. Davidson, B. O'Sullivan, S. Stoyloff (eds.), *Second Language Assessment*, Cambridge University Press, Cambridge, pp. 234-246.
- Patota G. e Pizzoli L. (2004), *Il sillabo della certificazione Società Dante Alighieri*, Società Dante Alighieri, Roma: <http://www.plida.it/plida/images/stories/documenti/sillabo.pdf>.

Powers M., Schedl A. & Wilson Leung S. (1999), "Validating the revised Test of Spoken English against a criterion of communicative success", in *Language Testing*, 16/4, pp. 399-425.

Ross S. (1992), "Accommodative questions in oral proficiency interviews", in *Language Testing*, 9, pp. 173-186.

Van Moere A. (2006), "Validity evidence in a university group oral test", in *Language Testing*, 23, pp. 411-440.

APPENDICE

Si presenta di seguito un adattamento delle griglie proposte in O'Sullivan (2012a: 49-52) per la stesura delle specifiche di un test del parlato. Le griglie riguardano in particolare il design generale del test (Tabella 4) e alcune considerazioni sulla scrittura delle prove (Tabella 5).

Tabella 4. *Il design generale del test*

IL DESIGN GENERALE DEL TEST	
<i>Scopo</i>	Descrizione dello scopo generale del test (per esempio, se si tratta di un test di ingresso per l'università, un test di uscita, un test di piazzamento, ecc.).
<i>Destinatari</i>	Descrizione dei destinatari del test, il più possibile dettagliata.
<i>Stakes</i>	Che impatto avrà il test su chi lo sosterrà?
<i>Formato di risposta</i>	Scelta multipla; sequenza di brevi domande e risposte; ecc.
<i>Numero di esaminatori</i>	Dichiarare qual è il numero totale di esaminatori richiesti e il loro ruolo (interlocutore, valutatore, ecc.).
<i>Numero di candidati</i>	Numero totale di persone che sosterranno il test in ciascuna sessione. Quanti candidati verranno esaminati per volta?
<i>Numero di task/item</i>	Indicare il numero di task previsto.
<i>Ordine dei task</i>	Indicare qui se è richiesto seguire un ordine specifico nella somministrazione degli item.
<i>Peso dei task / item</i>	I task/gli item avranno lo stesso peso/valore?

<i>Valutazione</i>	Chiave di risposte per i test a scelta multipla; scala di valutazione/descrittori: di tipo olistico o analitico?
<i>Modalità di comunicazione</i>	Giudizi singoli o profilo (punteggio per sotto-competenze individuali)?
INFERENZE RIGUARDO LA POPOLAZIONE CHE SI SOTTOPORRÀ AL TEST	
<i>Competenze pregresse</i>	In quale misura si conoscono queste competenze? Come queste informazioni possono influire nella presa di decisioni sui task, i temi e il formato del test?
<i>Competenza linguistica</i>	Dichiarare il livello atteso dei candidati che si sottoporranno al test (QCER)
RISORSE NECESSARIE	
<i>Risorse per lo sviluppo del test</i>	Sulla base delle decisioni prese, quali risorse potrebbero essere necessarie per andare avanti? (risorse umane, tecnologia, spazi, ecc.)
<i>Risorse per la somministrazione</i>	Quali risorse saranno necessarie per la somministrazione del test? (persone, produzione di materiali, somministrazione, post-somministrazione, ecc.)

Tabella 5. *Alcune considerazioni per la scrittura delle prove*

PARAMETRI PER LO SVILUPPO DEI TASK	
<i>Scopo</i>	Requisiti del task; consente ai candidati di scegliere le strategie più appropriate e determinare quali informazioni “attivare” nei task produttivi.
<i>Formato di risposta</i>	Formato di risposta previsto.
<i>Criteri conosciuti</i>	Far sapere ai candidati come verrà valutata la loro performance; quali mezzi vengono usati per informare sui criteri di valutazione.
<i>Peso</i>	La definizione degli obiettivi può avere un impatto negativo se i candidati sono informati della valutazione che attribuisce pesi diversi ai task prima della somministrazione.
<i>Ordine degli item</i>	Di solito fissato nei test di produzione orale.

<i>Limiti di tempo</i>	Riferito sia ai tempi di pianificazione/preparazione, sia ai tempi per le risposte.
<i>Operazioni richieste</i>	Descrizione del tipo di forme/funzioni richieste per l'esecuzione del task.
IL TASK (RIGUARDA IL LINGUAGGIO NELL'INPUT E QUELLO DELL'OUTPUT PREVISTO)	
<i>Canale</i>	In termini di input: scritto, visivo (foto), grafico, o orale (dalla voce di un esaminatore, o registrato). In termini di output: varia a seconda dell'abilità che si intende valutare.
<i>Discorso</i>	Include le categorie di genere, task retorico, pattern di esposizione.
<i>Lunghezza del testo</i>	Quantità di input/output.
<i>Relazione tra parlanti</i>	Stabilire diversi tipi di relazione può influire sulla performance (per esempio, rispondere a un superiore vs. ad un compagno/coetaneo; interazione simmetrica/asimmetrica)
<i>Natura dell'informazione</i>	Grado di astrattezza; studi dimostrano che la concretezza degli argomenti ha una correlazione con la facilità di esecuzione/risposta al task.
<i>Familiarità con l'argomento</i>	Maggiore familiarità sui temi tende a portare a risultati di performance migliori. Questo è un fattore da tenere in considerazione per la valutazione di tutte le sotto abilità.
PARAMETRI LINGUISTICI	
<i>Ampiezza lessicale</i>	Riguardano il linguaggio usato nell'input (di solito stabilito ad un livello inferiore rispetto il livello da testare) e il linguaggio previsto nell'output.
<i>Ampiezza strutturale</i>	
<i>Ampiezza funzionale</i>	
INTERLOCUTORE	
<i>Velocità di eloquio</i>	L'output deve riflettere le norme della L1. L'input può essere arrangiato a seconda del livello dei candidati. Va però notato che distorcere il ritmo naturale di eloquio può essere pericoloso perché può introdurre una variabile di irrilevanza nel costrutto.
<i>Varietà regionali</i>	Può essere definita nel costrutto (per esempio, quando viene descritta una varietà regionale) e/o dal contesto (per esempio, quando una particolare varietà è dominante nel contesto educativo in oggetto).

<i>Conoscenza</i>	La performance dei candidati è migliore quando interagiscono con un compagno/amico (anche se questo può dipendere da cultura a cultura).
<i>Numero</i>	Legato alle caratteristiche dei candidati – il carattere (introverso, estroverso) influisce sul tipo di performance del candidato con diverse persone.
<i>Genere</i>	I candidati tendono a fare una performance migliore quando sono intervistati da una donna, in generale il genere dell'interlocutore può avere un effetto sulla performance del candidato.