LANGUAGE TESTING ONLINE: SPERIMENTAZIONI SULLA LINGUA ITALIANA

Letizia Cinganotto[1]

1. INTRODUZIONE

Partendo da brevi cenni alla letteratura, il presente contributo focalizza l'attenzione sul *language testing* online in riferimento alla lingua italiana come L2/LS, riportando i risultati preliminari di alcuni studi condotti presso l'Università per Stranieri di Perugia, sulla scia dell'emergenza pandemica, che ha accelerato il processo di revisione e aggiornamento delle pratiche di insegnamento e di valutazione linguistica alla luce dei rapidi sviluppi tecnologici di questi ultimi anni. Nell'ambito della ricerca in corso, si presentano due studi, che seppure con i limiti legati al numero ridotto di informanti e alla validazione ancora in fieri, restituiscono le percezioni e reazioni complessivamente positive dei partecipanti rispetto alla somministrazione di un test online su piattaforma multimediale (Moodle) e su piattaforma potenziata dall'Intelligenza Artificiale (Diffit), nonostante la necessità di opportune modifiche e adattamenti, demandati necessariamente al docente e all'esaminatore.

Gli studi descritti in questo contributo sottolineano altresì la necessità di ulteriori indagini, sperimentazioni e validazioni, anche in considerazione della recente accelerazione tecnologica legata all'Intelligenza Artificiale.

2. CENNI SUL LANGUAGE TESTING ONLINE

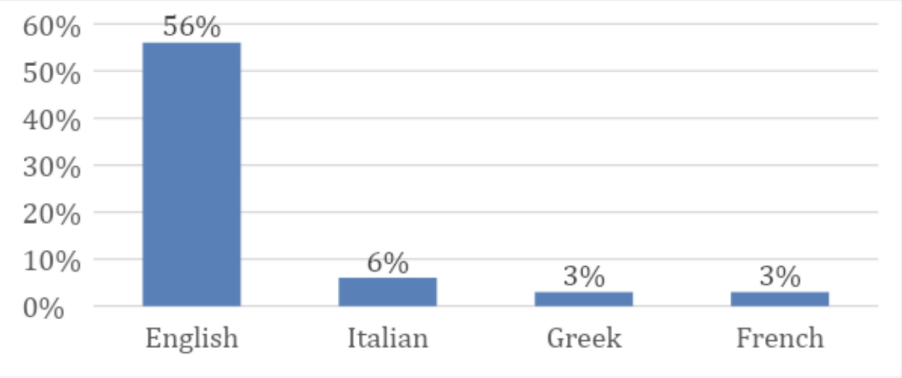
Sin dagli anni Novanta, l'ampia letteratura sul *language testing* si è concentrata su innumerevoli linee di ricerca, tra cui tre aree principali: teorica, metodologica e analitica. Dal punto di vista teorico, molti studi hanno focalizzato l'attenzione sulla comprensione del costrutto di competenza linguistica. Dal punto di vista metodologico, numerose ricerche hanno concentrato l'interesse sui test di prestazione linguistica e sulla promozione di standard professionali per lo sviluppo e l'uso dei test. Dal punto di vista analitico, innumerevoli indagini hanno focalizzato l'attenzione sull'implementazione della teoria IRT (*Item Response Theory*), sulla G-theory e sulla comprensione delle molteplici fonti di varianza nelle prestazioni dei test (Salmani Nodoushan, 2020).

Molto ampia è la letteratura di settore sul tema del *language testing* online, anche in considerazione della notevole rapidità degli sviluppi tecnologici, attualmente dominati dall'incredibile sopravvento dell'Intelligenza Artificiale, entrata ormai inevitabilmente a far parte di ogni ambito del sapere.

[1] Università per Stranieri di Perugia. Si ringrazia il Gruppo di Ricerca del CVCL formato da: M. Valentina Marasco, Danilo Rini, Roberta Rondoni, Nicoletta Santeusanio. Si ringrazia inoltre Sabrina Cittadini. Un particolare ringraziamento a Giovanna Scocozza, Direttrice del CVCL.

Vassiliou *et al.* (2023) hanno svolto una revisione sistematica della letteratura dal 2000 al 2023 sull'uso delle tecnologie per la valutazione formativa degli apprendimenti linguistici in una lingua straniera e, sulla base di una serie di criteri per l'individuazione del campione, hanno enucleato 34 pubblicazioni scientifiche, la maggior parte delle quali affrontano aspetti più tradizionali della valutazione delle singole abilità linguistiche, tra cui il 41% sulla scrittura, il 6% sulla lettura, il 6% sul parlato, il 15% sull'ascolto.

Figura 1. *Numero di pubblicazioni sul language testing online in varie lingue straniere tra il 2000 e il 2023 (Vassiliou et al., 2023: 56)*



La ricerca mostra dunque, come la tendenza sia soprattutto verso un approccio tradizionale alla valutazione delle abilità linguistiche in modo separato. Gli studi presi in esame hanno evidenziato il valore aggiunto delle tecnologie nel *language testing* per il feedback formativo, l'aumento della motivazione e della partecipazione, il miglioramento delle competenze linguistico-comunicative degli apprendenti.

Delle pubblicazioni esaminate solo il 6% riguardava la lingua italiana, dato che induce sicuramente a riflettere sulla necessità di approfondire ulteriormente questo ambito (Figura 1).

Proprio in riferimento alla lingua italiana, Newhouse e Cooper (2013) hanno svolto uno studio di tre anni all'interno di un corso di "Italian studies" presso delle scuole secondarie della Western Australia, che ha rilevato le potenzialità della valutazione online attraverso sistemi di registrazione audio-visiva della performance degli studenti, anche se la percezione era soprattutto in favore dell'uso del digitale per la verifica formativa, ma non per quella sommativa.

Un'altra revisione sistematica della letteratura effettuata da Muzaffar *et al.* (2021) ha identificato 53 ricerche pubblicate tra il 2016 e il 2020 relative all'uso delle tecnologie per la valutazione online e ha categorizzato 21 strumenti per il testing online, come indicato nella Figura 2 alla pagina che segue.

I ventuno strumenti individuati adottano in modo variabile, diverse categorie di funzionalità previste per una piena efficacia e efficienza dei test online, nello specifico:

- *Verification* & *Abnormal Behavior*;
- *Security*;
- *Question Bank Generation* & *Evaluation.*

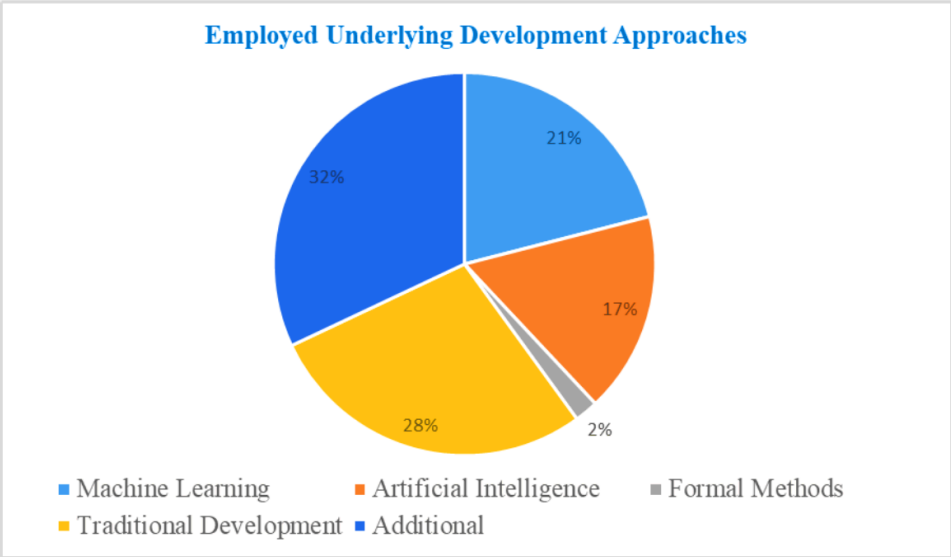
La revisione della letteratura in questione ha inoltre dimostrato che la maggior parte degli strumenti non sono *open source* e quindi non disponibili liberamente. Ciò riduce in modo significativo i vantaggi e rende difficile effettuare ulteriori studi e ricerche comparative, che potrebbero risultare particolarmente utili per gli sviluppi futuri.

Figura 2. *Strumenti per il testing online e rispettive funzionalità (Muzzafar et al., 2021: 32701)*

No.	Tool Name	Supported Features			Availability	Relevant Study
		Verification & Abnormal Behavior	Security	Question Bank Generation & Evaluation		
1	Secure Exam Environment (SEE)	Yes	Yes	Yes	N-A	[43]
2	Unified E-Examination Solution	Yes	Yes	Yes	N-A	[44]
3	Examination Management System (EMS)	Yes	Yes	Yes	N-A	[61]
4	Continuous Online Authentication System	Yes	Yes		N-A	[14]
5	Online Exam Proctoring (OEP) system	Yes	Yes		N-A	[16]
6	Secure E-Learning System	Yes	Yes		N-A	[52]
7	Online Exam Authentication System	Yes	-	-	N-A	[12]
8	Prototype Online Exam App	Yes			N-A	[17]
9	FLEXauth	Yes			N-A	[37]
10	Secure Online Examination System		Yes		N-A	[26]
11	Online Item Exam System			Yes	N-A	[15]
12	Exam Wizard			Yes	N-A	[23]
13	Online Descriptive Answer Marking System			Yes	N-A	[30]
14	Clever Testing System			Yes	N-A	[31]
15	MoLearn System			Yes	N-A	[32]
16	Snaptron			Yes	N-A	[35]
17	ViLLE			Yes	Public	[36]
18	Simple and Dynamic Examination System (SDES)			Yes	N-A	[39]
19	Automatic Evaluation System			Yes	N-A	[45]
20	e-Testing System			Yes	N-A	[46]
21	Online Examination System			Yes	N-A	[49]

Tra i principali metodi adottati negli studi presi in esame, oltre ai metodi formali e tradizionali, sono citati anche il *machine learning* e l'Intelligenza Artificiale per una percentuale complessiva del 38%, come illustrato nella Figura 3.

Figura 3. *Principali metodi adottati nel testing online citati negli studi tra il 2016 e il 2020 (Muzzafar et al., 2021: 32708)*



3. L'INTELLIGENZA ARTIFICIALE PER IL *LANGUAGE TESTING*

In quest'ultimo periodo l'Intelligenza Artificiale si è insediata prepotentemente in ogni ambito del sapere e dell'istruzione, inclusa anche la valutazione in generale e il *language testing* in particolare, aprendo la strada a nuovi scenari per la ricerca (Breck *et al.*, 2017; Zhang *et al.*, 2022).

Myllyaho *et al.* (2021) hanno effettuato una revisione sistematica della letteratura sui metodi di validazione dei sistemi di IA, identificando 90 studi, che classificano e descrivono i metodi utilizzati in contesti realistici per garantire l'affidabilità dei sistemi di IA.

È stata elaborata una tassonomia fondata su quattro metodi di validazione: prova (*trial*), validazione centrata sul modello (*model-centred*), simulazione (*simulation*) e opinione di esperti (*expert opinion*). La tassonomia potrebbe rappresentare una base di partenza per ulteriori ricerche sulla validazione dei sistemi di IA anche nel campo del *language testing*.

Nello specifico, particolarmente interessante è l'*expert opinion*, l'opinione di esperti, un metodo di validazione, che in base alla succitata ricerca, sembra non essere ancora molto diffuso. La caratteristica principale di questo metodo è che l'esperienza necessaria per condurre la validazione non viene affidata ad un esperto informatico, ma ad un esperto informato del dominio applicativo, come utente finale. L'opinione dell'esperto consente di acquisire dati dagli utenti stessi. Questo metodo potrebbe essere utile nell'ambito del *language testing*, proponendo la consultazione di studiosi o esperti in questo campo, chiamati a valutare e validare i sistemi di IA nel processo di costruzione, somministrazione e valutazione dei test linguistici.

Figura 4. *Metodi di validazione (Myllyaho et al., 2021: 9)*

Validation methods used in papers.

Validation method	Number of papers	Percentage
Expert opinion	2	2.2%
Simulation	20	22.2%
Model-centred	13	14.4%
Trial	42	46.7%
Multiple	13	14.4%

4. LE ATTIVITÀ DEL CVCL

Il Centro per la Valutazione e la Certificazione Linguistica (CVCL) dell'Università per Stranieri di Perugia[2] ha una lunga tradizione nella progettazione, costruzione e somministrazione di esami e nel rilascio di certificazioni rivolte principalmente a due diverse tipologie di target: la certificazione delle competenze linguistiche in Italiano L2/LS denominata CELI e la certificazione delle competenze glottodidattiche in Italiano come L2/LS, denominata DILS-PG (Santeusanio, 2014).

[2] Il CVCL è membro dell'Associazione CLIQ (*Certificazione Lingua Italiana di Qualità*), che comprende le quattro istituzioni ufficialmente riconosciute dai Ministeri italiani per il rilascio di certificati di Italiano come lingua straniera o lingua seconda (Università per Stranieri di Perugia, Università per Stranieri di Siena, Università Roma Tre, Società Dante Alighieri). Il CVCL è anche *full member* dell'ALTE (*Association of Language Testers in Europe*) e partecipa attivamente a tutte le diverse attività e iniziative promosse a livello internazionale.

La certificazione CELI, allineata ai livelli del *Quadro Comune Europeo di Riferimento per le Lingue, Volume Complementare* (QCERVC), comprende una parte scritta (comprensione della lettura, produzione scritta, comprensione orale e uso della lingua), valutata centralmente dal CVCL, e una parte orale, basata su input testuali, visivi e grafici per compiti comunicativi autentici, valutata dagli esaminatori a Perugia e nei vari centri convenzionati in tutto il mondo.

Gli item delle prove d'esame sono progettati in base al *Profilo della lingua italiana* (Spinelli, Parizzi, 2010), al Corpus di Perugia (Spina, 2014) e al corpus CELI (Spina *et al.*, 2002). Le prove si ispirano inoltre ai principi dell'Approccio Orientato all'Azione (Piccardo, North, 2019; Cinganotto, 2023a) e si basano sulla definizione di scenari di apprendimento autentici, proponendo compiti reali per un uso significativo della lingua italiana. Le prove sono costruite in base ad una combinazione tra *discrete-point testing* e *integrative testing*, alternando item a risposta chiusa a prove di produzione scritta.

5. Le sperimentazioni online

In considerazione del ruolo sempre più cruciale rivestito dalle tecnologie soprattutto nell'era post-pandemica, cosiddetta "new normal" (Cinganotto, 2023b; Malagnini, Cinganotto, 2023), che ha indotto scuole, università e istituzioni formative a ripensare i processi di insegnamento e di valutazione alla luce delle potenzialità del digitale, il gruppo di ricerca del CVCL, formato da docenti, ricercatori e CEL (Collaboratori Esperti Linguistici), sotto la guida della Direttrice, del Comitato Direttivo e del Consiglio Scientifico di cui l'autrice fa parte, ha intrapreso in questi ultimi anni, alcuni studi e sperimentazioni sulla digitalizzazione del processo di *language testing*[3], in quanto l'attività di ricerca rappresenta una parte importante delle attività del CVCL stesso.

In questa sede si riporteranno i risultati preliminari di due sperimentazioni, attualmente ancora in corso di validazione, che testimoniano il forte impegno del CVCL sul fronte della ricerca, dell'innovazione e della digitalizzazione.

Le sperimentazioni sono state precedute da una prima fase di approfondimento della recente letteratura in materia di valutazione (Serragiotto, 2016; Masillo, 2019; Barni, 2023; Machetti, Vedovelli, 2024), nonché da una rilettura e valorizzazione della lunga esperienza del CVCL in materia di valutazione e certificazione.

6. La sperimentazione sulla piattaforma Moodle

Una prima sperimentazione, condotta nell'a.a. 2022-23 ha riguardato la digitalizzazione di una prova adattata dal test CELI 1 di livello A2 sulla piattaforma Moodle di Ateneo. La prova è stata somministrata ad un campione formato da 32 studenti non italofoni frequentanti un corso di lingua e cultura italiana presso l'Università per Stranieri di Perugia. L'indagine mirava a comprendere le reazioni e le attitudini degli studenti sulla prova in formato digitale, rispetto a quella in formato cartaceo, raccogliendo le loro testimonianze attraverso un questionario online e alcune interviste informali. I dati emersi sono stati poi esaminati in base ai principi della *Framework Analysis* (Ritchie, Lewis, 2003), che consente di raggruppare i vari commenti in categorie tematiche, sulla base dell'individuazione di analogie e elementi comuni.

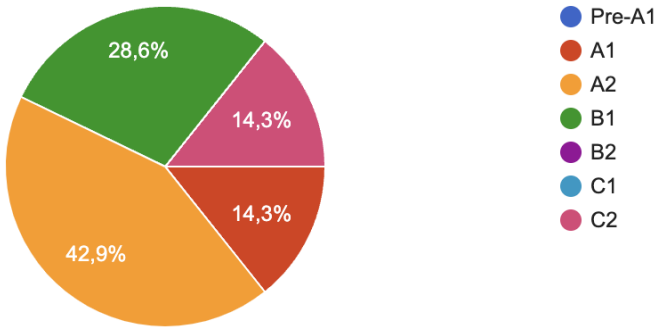
[3] Il processo di digitalizzazione del *language testing* del CVCL in corso riguarda la revisione e l'aggiornamento delle procedure di somministrazione, correzione e valutazione dei test attraverso la sperimentazione di varie soluzioni digitali.

Il campione, formato principalmente da apprendenti tra i 18 e i 30 anni, è caratterizzato da un repertorio marcatamente plurilingue: il 42,9% parla due lingue oltre alla propria lingua di provenienza, il 28,6% ne conosce tre e il 28,6% una.

La domanda relativa all'autovalutazione delle proprie competenze linguistiche ha restituito dati variegati, collocando il 42,9% dei rispondenti nel livello oggetto dell'esame CELI in questione, cioè il livello A2.

Una percentuale non irrisoria (il 28,6%) si posiziona sul livello immediatamente superiore (B1), mentre il 14,3% si identifica nel livello A1. Interessante il restante 14,3% posizionatosi sul C2. Probabilmente la distribuzione tra i vari livelli del QCERVC non è sempre ben chiara tra gli apprendenti, soprattutto per la fascia di livello elementare.

Figura 5. *Autovalutazione del livello linguistico dei partecipanti*



Il 42,9% dei rispondenti conosce già la piattaforma Moodle, grazie al corso di lingua e cultura italiana dell'Università per Stranieri di Perugia cui sono iscritti.

6.1. *Le prove*

Il test somministrato nell'ambito della sperimentazione, realizzato sulla piattaforma Moodle, era un adattamento mirato alla valutazione di competenze parziali mediante item esclusivamente oggettivi, tratti dal test CELI 1 (livello A) ed era composto da una prova di ascolto e una prova di comprensione della lettura.

La piattaforma Moodle permette la costruzione e la valutazione di un'ampia gamma di tipologie di prove, favorendo il tracciamento, il feedback automatico e il calcolo immediato dei punteggi nei quesiti a risposta chiusa, rendendo possibile ipoteticamente una restituzione immediata degli esiti.

Pertanto, facendo riferimento alle funzionalità del testing online descritte nel paragrafo 1, nella piattaforma Moodle si possono individuare tutte le categorie menzionate, nello specifico:

- *Verification* & *Abnormal Behavior*: Moodle permette la verifica a diversi livelli, attraverso criteri e punteggi predefiniti, nonché l'individuazione delle devianze e degli errori, per un feedback correttivo immediato.
- *Security*: la sicurezza dei dati e della valutazione è garantita da un'ampia gamma di protocolli, che prevedono l'impostazione di password, di parametri temporali di apertura e chiusura del test ecc.
- *Question Bank Generation* & *Evaluation*: Moodle consente di creare una banca dati di domande di diversa tipologia, con i rispettivi criteri di valutazione, da cui attingere

attraverso una modalità randomizzata, che permette di costruire test con item diversi, limitando le probabilità del *cheating*.

La prova di lettura è stata svolta da quasi tutti i partecipanti, eccetto uno. Il punteggio medio della prova conseguito dai rispondenti è 12,46/30, come si evince dal grafico (Figura 6) sottostante, che raccoglie i punteggi registrati da tutti i partecipanti per ciascuna batteria di item (da D.1 a D.5).

Figura 6. *Punteggi della prova di lettura CELI 1 (A2) sulla piattaforma Moodle*

Valutazione/30,00	**D. 1 /4,29**	**D. 2 /5,36**	**D. 3 /7,50**	**D. 4 /7,50**	**D. 5 /5,36**
5,36	2,14	1,07	1,07	6,43	-5,36
23,57	4,29	5,36	5,36	7,50	1,07
7,50	2,14	1,07	2,14	3,21	-1,07
9,64	2,14	1,07	3,21	4,29	-1,07
21,43	2,14	3,21	7,50	7,50	1,07
17,14	2,14	5,36	2,14	6,43	1,07
16,07	0,00	5,36	3,21	6,43	1,07
22,50	4,29	3,21	6,43	7,50	1,07
3,21	0,00	1,07	3,21	2,14	-3,21
8,57	0,00	1,07	4,29	4,29	-1,07
18,21	4,29	5,36	2,14	5,36	1,07
6,43	0,00	3,21	1,07	1,07	1,07
2,14	0,00	5,36	1,07	1,07	-5,36
4,29	0,00	1,07	3,21	1,07	-1,07
24,64	2,14	5,36	6,43	7,50	3,21
25,71	4,29	5,36	7,50	7,50	1,07
20,36	4,29	3,21	4,29	5,36	3,21
0,00	-	-	-	-	-
13,93	2,14	1,07	4,29	5,36	1,07
17,14	2,14	3,21	6,43	6,43	-1,07
9,64	4,29	-1,07	3,21	4,29	-1,07
-2,14	-2,14	1,07	2,14	2,14	-5,36
17,14	4,29	1,07	5,36	5,36	1,07
1,07	2,14	1,07	1,07	2,14	-5,36
6,43	0,00	3,21	3,21	3,21	-3,21
0,00	0,00	-1,07	2,14	2,14	-3,21
19,29	2,14	1,07	7,50	7,50	1,07
19,29	2,14	3,21	5,36	7,50	1,07

Valutazione/30,00	**D. 1 /4,29**	**D. 2 /5,36**	**D. 3 /7,50**	**D. 4 /7,50**	**D. 5 /5,36**
30,00	4,29	5,36	7,50	7,50	5,36
-1,07	-2,14	-1,07	2,14	3,21	-3,21
22,50	4,29	1,07	6,43	5,36	5,36
8,57	-2,14	3,21	2,14	4,29	1,07
Media generale 12,46	**1,74**	**2,44**	**3,85**	**4,72**	**-0,30**

La batteria di item che ha registrato un punteggio più basso è D.5, riportata in Figura 7, che propone scenari di apprendimento alquanto comuni, tipici della vita quotidiana.

La descrizione degli scenari di questa batteria risulta più lunga e articolata rispetto alle batterie precedenti, caratterizzate da abbinamenti e domande a scelta multipla più brevi e immediate: probabilmente la percezione del livello di difficoltà di questa batteria di item è risultata più alta.

Figura 7. *Item della prova di lettura con punteggio medio più basso*

Parte A.5

Leggi i testi da 24 a 28 e scegli la risposta corretta.

24 Il FRU è il Festival delle Radio Universitarie, organizzato da Radiophonica. Durante la manifestazione tante radio universitarie italiane si troveranno nel centro storico di Perugia, dove organizzeranno concerti e incontri sul tema della comunicazione radiofonica. Parteciperanno numerosi artisti e personaggi famosi del mondo dello spettacolo.

Il FRU

- ○ dà informazioni sull'università
- ○ propone musica e dibattiti
- ○ sceglie la migliore radio universitaria

25 Inviare denaro all'estero via web e via cellulare è comodo, veloce e sicuro. Per farlo via web è sufficiente avere un conto Bancoposta online, mentre da cellulare basta avere un numero di telefono di PosteMobile. Fino al 30 aprile 2012 per ogni invio di denaro ricevi un euro di ricarica: il bonus arriva sul tuo numero entro il 15 del mese.

Questo annuncio è rivolto a chi vuole

- ○ spedire soldi in un altro Paese
- ○ aprire un conto corrente online
- ○ ricaricare il proprio cellulare dall'estero

26 ContoOro, in occasione del suo decimo anniversario, vuole festeggiare offrendo ai suoi clienti l'opportunità di vivere una crociera MSC. È sufficiente mostrare la propria carta di credito ContoOro in un'agenzia di viaggi per avere sconti fino a 300 euro su una fantastica crociera, anche nei mesi estivi di luglio e agosto.

Grazie a ContOro la crociera MSC

- ○ costerà ai clienti solo 300 euro a persona
- ○ si potrà prenotare tutto l'anno a prezzi scontati
- ○ sarà più economica per chi pagherà con carta di credito

La prova di ascolto viene sicuramente percepita come più sfidante e complessa della prova di lettura: 14 partecipanti su 32 avviano l'ascolto del file audio, ma abbandonano subito la prova, senza svolgerla. L'impossibilità di manovrare il file audio e di metterlo in pausa nei punti di difficile comprensione rappresenta uno dei principali ostacoli riportati dai partecipanti. Tuttavia, questa caratteristica è molto comune nella prova di ascolto su computer e sicuramente necessita di un allenamento e di esercizi specifici.

Il punteggio totale medio è di 6,60/30, come riportato nella Figura 8 di seguito, che contiene i punteggi registrati per ciascuna batteria di item (da D1 a D4). Si tratta di un punteggio di gran lunga inferiore rispetto a quello della prova di lettura.

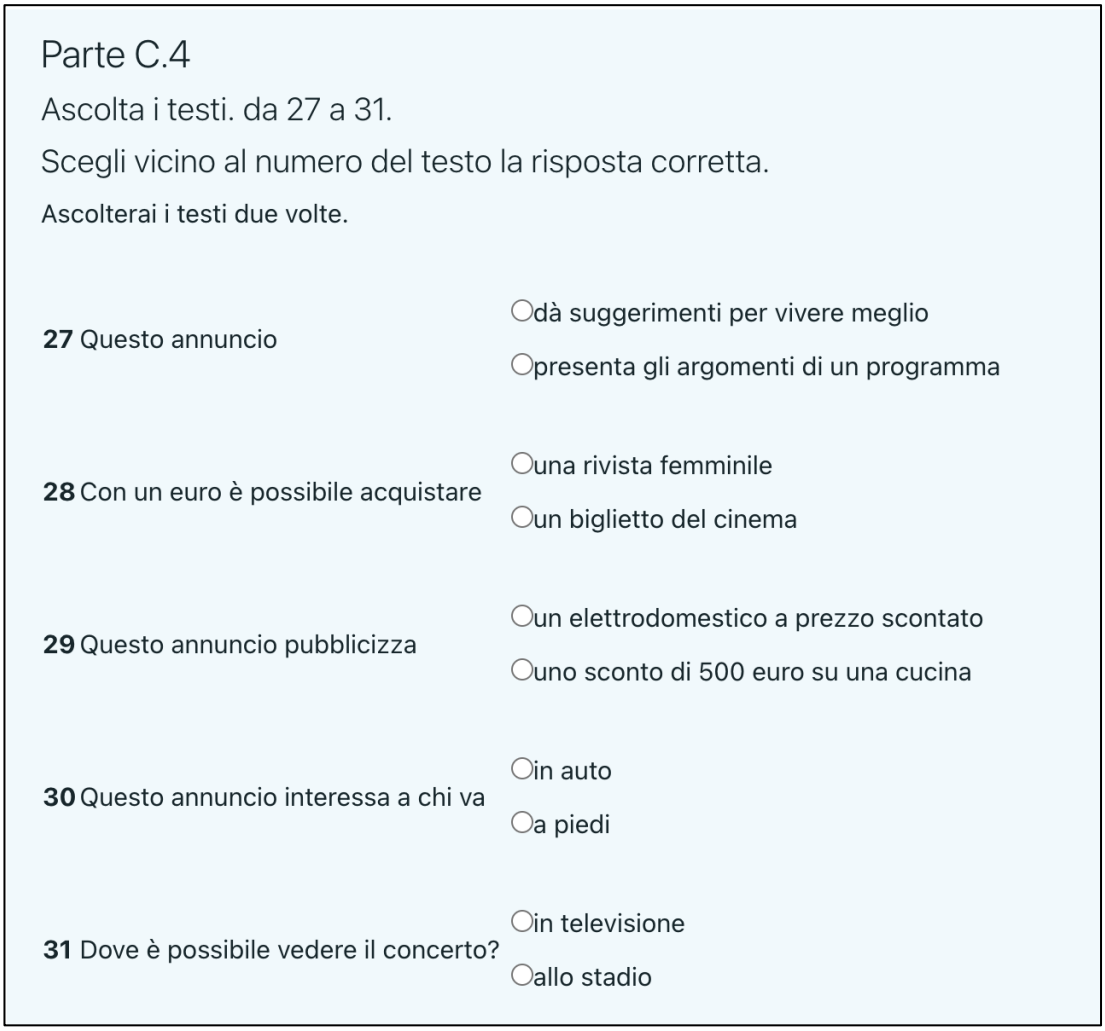
Figura 8. *Punteggi della prova di ascolto CELI 1 (A2) sulla piattaforma Moodle*

10,5	**D. 1 /8,82**	**D. 2 /7,94**	**D. 3 /8,82**	**D. 4 /4,41**
0,00	-	-	-	-
0,00	-	-	-	-
0,00	-	-	-	-
5,29	1,76	2,65	0,00	0,88
0,88	0,88	-	-	-
15,88	3,53	7,94	5,29	-0,88
22,94	3,53	7,94	7,06	4,41
0,00	-	-	-	-
0,00	-	-	-	-
22,06	8,82	0,00	8,82	4,41
0,00	-	-	-	-
0,00	-	-	-	-
0,00	-	-	-	-
4,41	1,76	0,00	1,76	0,88
3,53	1,76	0,88	0,00	0,88
24,71	7,06	7,94	7,06	2,65
3,53	1,76	0,88	0,00	0,88
0,00	5,29	0,88	-3,53	-2,65
1,76	0,00	0,88	0,00	0,88
0,00	-	-	-	-
22,06	8,82	-	8,82	4,41
0,00	-	-	-	-
22,06	8,82	0,00	8,82	4,41
0,00	-	-	-	-
0,00	-	-	-	-
22,94	3,53	7,94	7,06	4,41

10,5	**D. 1 /8,82**	**D. 2 /7,94**	**D. 3 /8,82**	**D. 4 /4,41**
5,29	1,76	2,65	0,00	0,88
1,76	-1,76	-	-0,88	4,41
28,24	8,82	7,94	8,82	2,65
0,00	-	-	-	-
0,00	-	-	-	-
10,59	1,76	4,41	7,06	-2,65
Media generale 6,60	**2,06**	**1,60**	**2,01**	**0,94**

La batteria di item D. 4 appare la più complessa, registrando il punteggio più basso. Probabilmente anche gli scenari previsti da questi item appaiono più articolati e complessi rispetto ai precedenti, che richiedono task di tipo cognitivo e linguistico evidentemente percepiti come meno sfidanti.

Figura 9. *Item della prova di ascolto con punteggio medio più basso*

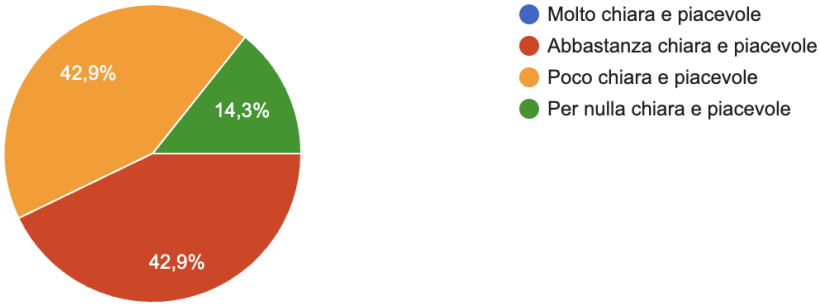


6.2. *Principali risultati*

Al completamento della prova, ai partecipanti è stato chiesto di rispondere alle domande di un questionario online, che mirava a indagare la loro percezione su vari aspetti della prova stessa.

Una domanda specifica era focalizzata sulla grafica del test, ritenuta abbastanza chiara e piacevole dal 42,9% dei rispondenti, ma poco chiara e piacevole da una percentuale equivalente di partecipanti.

Figura 10. *Percezioni dei partecipanti sulla grafica del test*



In generale, il test online viene considerato più facile e pratico del test cartaceo dal 42,9% dei partecipanti e uguale al test cartaceo dal 28,6%, restituendo una percezione comunque positiva e incoraggiante della digitalizzazione della prova.

Figura 11. *Percezioni dei partecipanti sulla prova online*

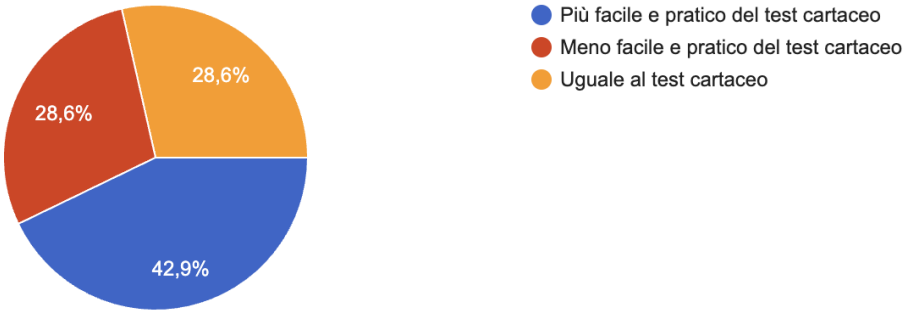
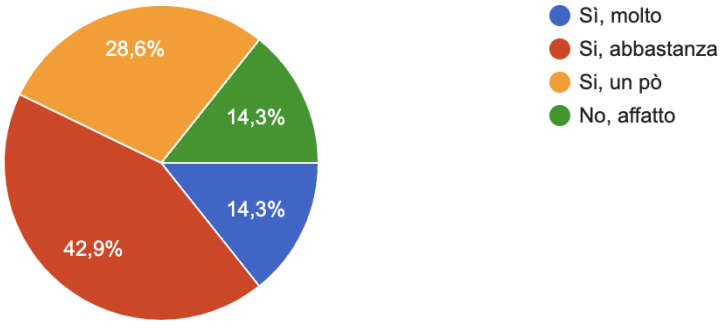


Figura 12. *Ansia dei partecipanti per lo scorrere del tempo visibile sullo schermo*



Il 57,1 dei partecipanti ha trovato più difficile la prova di ascolto rispetto a quella della lettura, dato peraltro confermato dai punteggi registrati in piattaforma.

In linea generale, il 42,9% dei rispondenti preferirebbe sostenere la prova vera e propria in formato digitale, confermando dunque, l'esigenza di proseguire verso questa direzione.

Unitamente al questionario online, sono state organizzate delle interviste in presenza con gli studenti, con l'obiettivo di acquisire ulteriori informazioni in merito alle loro reazioni e percezioni.

Interessante l'uso di un'app per la traduzione automatica dell'intervista usata da alcuni studenti cinesi per facilitare la comprensione delle domande e l'elaborazione delle risposte.

Gli studenti hanno confermato di aver riscontrato più problemi nella parte di ascolto che in quella di comprensione della lettura, probabilmente a causa delle impostazioni tecniche della piattaforma, dell'ansia per lo scorrere del tempo visibile sullo schermo e della modalità di autoconsegna del test.

Inoltre, molti rispondenti hanno trovato le immagini troppo piccole e poco chiare per essere abbinate al testo e il testo stesso non fruibile in modo ottimale.

Le osservazioni dei partecipanti sul layout e la visualizzazione del testo sullo schermo inducono a pensare che sia necessario un ripensamento generale della prova e dei singoli item, piuttosto che una trasposizione *sic et simpliciter* dal formato cartaceo a quello digitale della piattaforma Moodle: la resa grafica, la lettura delle immagini e la distribuzione del testo sullo schermo non sono molto apprezzati dai partecipanti, che ne hanno messo in luce la fruizione non pienamente adeguata.

Allo stesso tempo, i partecipanti hanno trovato questa modalità utile e comoda per l'esercizio, l'autoapprendimento e l'autovalutazione.

Di seguito alcuni dei commenti raccolti:

> Le immagini sono troppo piccole ed è difficile individuare la risposta giusta.
> La grafica e la dimensione delle immagini non sono buone.
> È comodo per gli studenti che vogliono studiare e valutare da soli.
> Si può fare in qualsiasi momento e ovunque e il test non sarà mai danneggiato.
> È conveniente, ma può causare un po' di ansia.
> È più immediato.

In base alla *Framework Analysis* di seguito le principali categorie tematiche identificate:

- *Necessità di riformulare e ripensare la costruzione delle prove e degli item.*
  Il formato online della prova, dovrebbe essere rivisto e rimodellato: non è possibile utilizzare lo stesso formato e gli stessi item della versione cartacea: il *language testing* online dovrebbe prevedere non solo l'esame e lo studio delle procedure e delle funzionalità tecniche, ma un ripensamento del format e del *test design in* generale.
- *Maggiore immediatezza e rapidità del testing online.*
  Il test online è percepito come più rapido e immediato, anche in linea con altre tipologie di certificazioni, test e corsi di formazione, sempre più orientati verso il digitale.
- *Percezione dell'impostazione del timer della prova sullo schermo come stressante e ansiogena.*
  Lo scorrere del tempo visualizzato sullo schermo, può causare ansia e stress e inficiare i risultati; probabilmente sono necessarie esercitazioni e simulazioni ad hoc anche su questo aspetto del testing online.

– *Percezione dell'utilità del test online per l'esercizio e lo studio individuale.*
  Il test online, soprattutto se fruibile anche sul cellulare, che è l'unico device di cui quasi tutti gli studenti non italofoni internazionali o con background migratorio sono dotati, viene considerato come uno strumento molto utile per l'autoapprendimento, soprattutto con feedback immediato.

Quest'ultima osservazione risulta in linea con quanto rilevato in letteratura (cfr. paragrafo 1): il *language testing* online è percepito dagli studenti come più utile per l'auto-apprendimento e per la verifica formativa che per quella sommativa a fini certificatori, come nel caso dell'esame CELI.

7. Una sperimentazione con l'Intelligenza Artificiale

Nell'ambito della ricerca sulla digitalizzazione del *language testing*, un approfondimento sull'uso dell'Intelligenza Artificiale nel testing dell'Italiano L2/LS è stato condotto all'interno dell'insegnamento di Didattica Digitale e data driven learning di un corso di laurea dell'Università per Stranieri di Perugia nell'a.a. 2023-24.

Tra le varie attività previste all'interno del corso, gli studenti sono stati guidati nella scoperta e nella sperimentazione di alcuni tool potenziati dall'Intelligenza Artificiale, in considerazione dell'importanza sempre crescente di questi strumenti all'interno del processo di insegnamento, apprendimento e valutazione delle competenze linguistiche, pur tenendo in considerazione tutti i dovuti *caveat* e le questioni etiche legate alle sperimentazioni e validazioni ancora in fieri, nonché ai rapidi sviluppi di questa recente tecnologia, con cui è difficile stare al passo.

Nello specifico, si è preso in esame il tool Diffit[4], già sperimentato con successo in base ad uno studio sull'apprendimento della lingua indonesiana (Etikasri, Andayani, 2023), che, grazie alle funzionalità dell'Intelligenza Artificiale, permette di inserire dei prompt specifici in varie lingue, sulla base dei quali si ottiene un output molto articolato, formato da un testo adattivo, un riassunto, parole chiave, una prova in formato misto (item a risposta chiusa, item a risposta aperta breve, item a risposta aperta lunga), con le relative risposte corrette o suggerite. Tutto il contenuto generato può essere modificato o integrato, sempre grazie agli strumenti dell'IA.

Diffit può essere utilizzato anche per tradurre il testo nelle varie lingue di origine degli studenti e *elaborare* domande di comprensione. Il test generato può essere scaricato e trasferito su un'ampia gamma di piattaforme, tra cui un Google Moduli, che viene creato automaticamente dal sistema, dando la possibilità agli apprendenti, di rispondere direttamente online con feedback automatico.

Le impostazioni sono basate sul sistema anglosassone K-12, nonostante la piattaforma sia multilingue. Appare dunque, difficile effettuare una comparazione precisa con i livelli del QCERVC.

È stato proprio questo il punto di partenza dello studio in questione: gli studenti sono stati chiamati a confrontare in gruppo, i test generati da Diffit per i vari *grade* di riferimento del sistema K-12, valutandone l'attendibilità e l'appropriatezza per studenti non italofoni in base ai livelli del QCERVC, a partire dall'analisi degli Inventari del *Profilo della lingua italiana*.

In base alla revisione della letteratura citata nel paragrafo 2, si è ritenuto utile, come esercitazione, adottare il punto di vista degli studenti del corso di laurea per la validazione dello strumento in modalità "expert opinion".

[4] https://beta.diffit.me/#topic.

Di seguito alcune riflessioni, frutto del lavoro del gruppo di studenti impegnati nella creazione con Diffit, di un test di livello A1 e A2, partendo dai *grade* 1-3, su un tema liberamente concordato dal gruppo: la famiglia in vacanza a Roma.

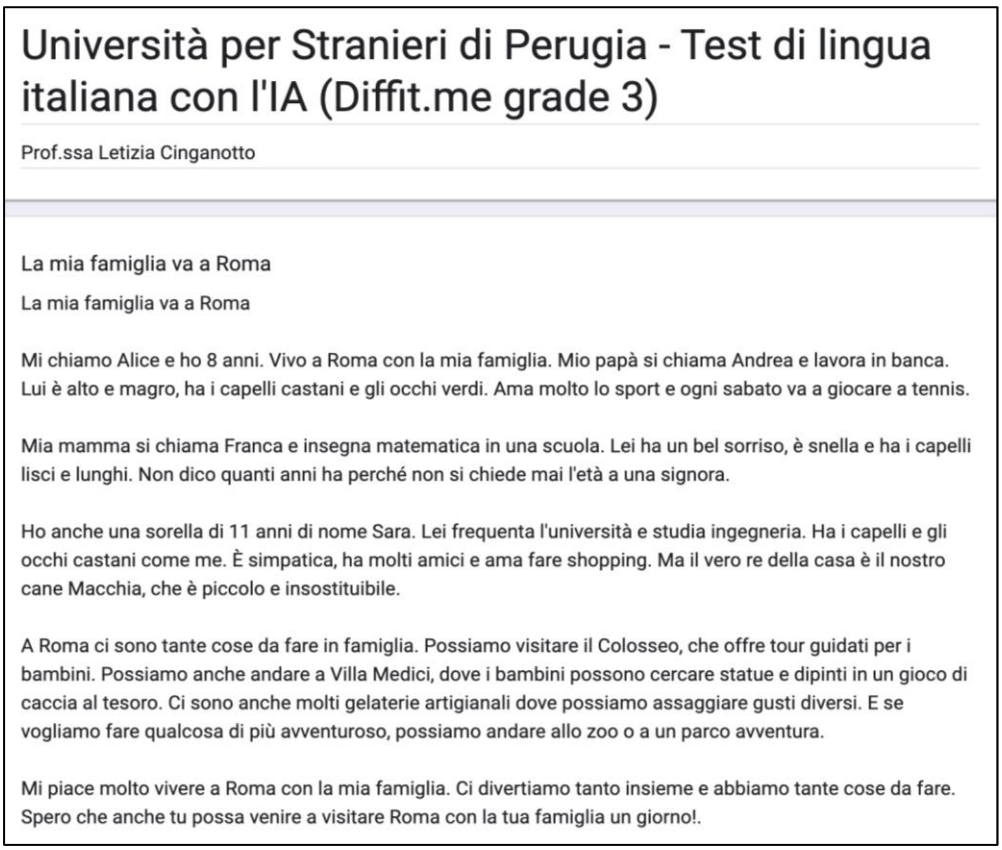
> Ri: A1-A2
> Di S.I. - giovedì, 7 dicembre 2023, 18:05
> Possiamo confrontare il livello A1 con il 2nd grade, invece il livello A2 può essere da 3rd a 4th grade, a volte 5th.
> Il problema di questa IA è che, non seguendo le indicazioni del QCER, i testi per A1 possono contenere, ad esempio, i verbi al futuro che, in teoria, lo studente di livello A1 non ha ancora affrontato. In questi casi c'è bisogno dell'intervento del docente.
> Altri esempi: il testo per A2 contiene il passato prossimo, l'imperfetto, frasi più complesse dal punto di vista della loro lunghezza, si aggiungono i connettivi, il superlativo degli aggettivi…

In base all'approfondimento e alla comparazione con gli Inventari del Profilo della lingua italiana, gli studenti hanno ipotizzato che il *grade* 3 fosse il più vicino al livello A2 del QCERCV, anche se l'analisi si è concentrata soprattutto sugli elementi morfo-sintattici e grammaticali.

Il test è stato dunque trasferito automaticamente dal tool stesso in un Google Moduli, che è stato successivamente somministrato ad un campione di 21 studenti iscritti ad un corso di lingua e cultura italiana di livello A2 dell'Università per Stranieri di Perugia.

Di seguito alcune parti del test generato dal tool e trasferito direttamente su Google Moduli.

Figura 13. *Schermata della prova realizzata da Diffit*



Università per Stranieri di Perugia - Test di lingua italiana con l'IA (Diffit.me grade 3)

Prof.ssa Letizia Cinganotto

La mia famiglia va a Roma

La mia famiglia va a Roma

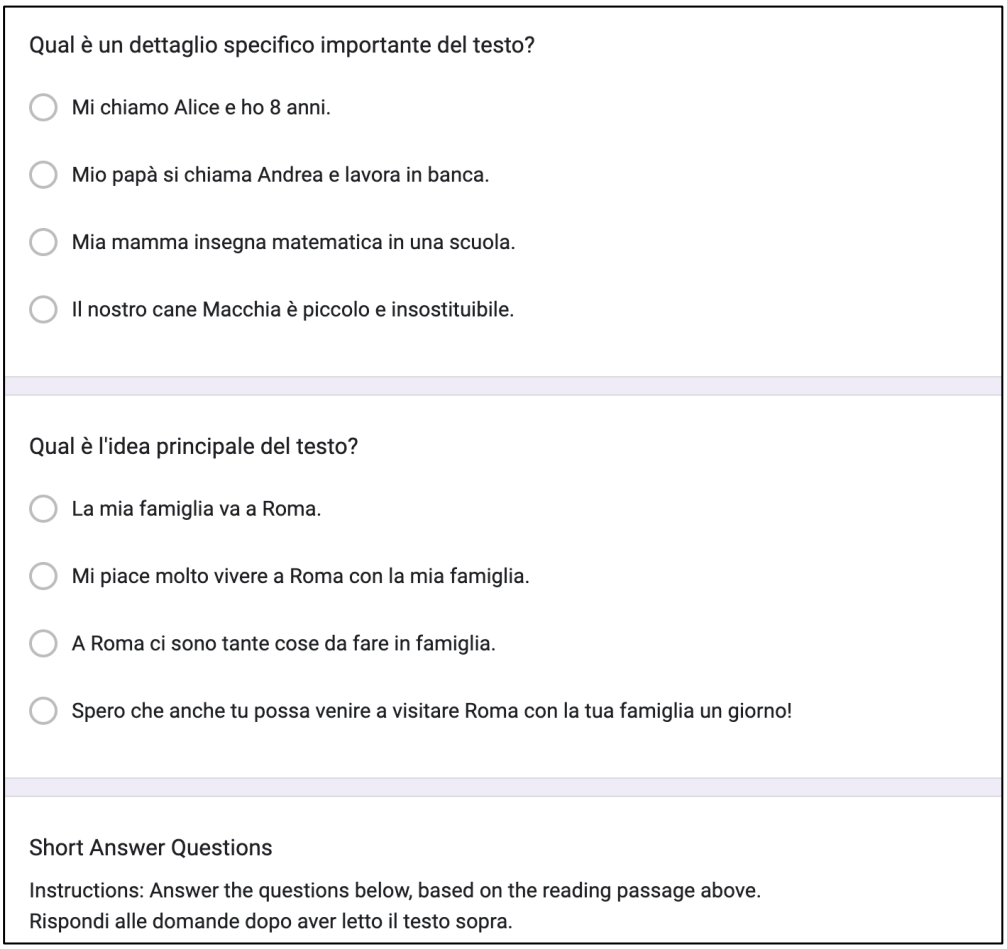
Mi chiamo Alice e ho 8 anni. Vivo a Roma con la mia famiglia. Mio papà si chiama Andrea e lavora in banca. Lui è alto e magro, ha i capelli castani e gli occhi verdi. Ama molto lo sport e ogni sabato va a giocare a tennis.

Mia mamma si chiama Franca e insegna matematica in una scuola. Lei ha un bel sorriso, è snella e ha i capelli lisci e lunghi. Non dico quanti anni ha perché non si chiede mai l'età a una signora.

Ho anche una sorella di 11 anni di nome Sara. Lei frequenta l'università e studia ingegneria. Ha i capelli e gli occhi castani come me. È simpatica, ha molti amici e ama fare shopping. Ma il vero re della casa è il nostro cane Macchia, che è piccolo e insostituibile.

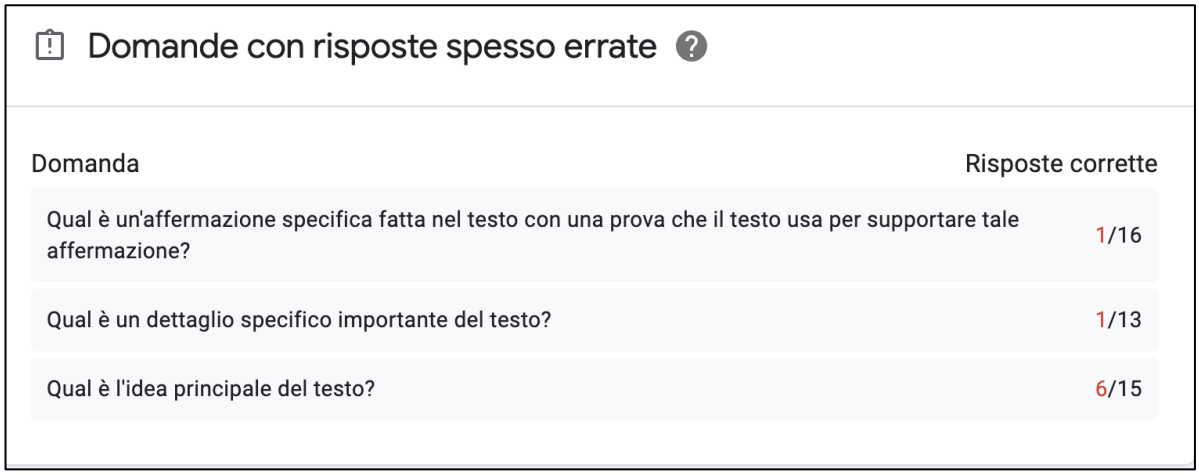
A Roma ci sono tante cose da fare in famiglia. Possiamo visitare il Colosseo, che offre tour guidati per i bambini. Possiamo anche andare a Villa Medici, dove i bambini possono cercare statue e dipinti in un gioco di caccia al tesoro. Ci sono anche molti gelaterie artigianali dove possiamo assaggiare gusti diversi. E se vogliamo fare qualcosa di più avventuroso, possiamo andare allo zoo o a un parco avventura.

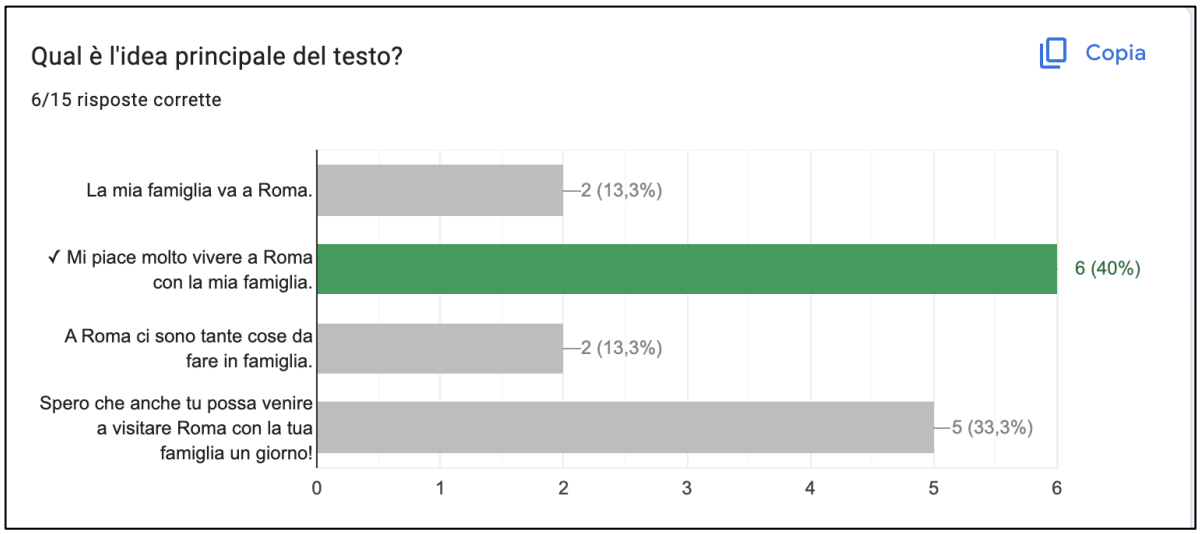
Mi piace molto vivere a Roma con la mia famiglia. Ci divertiamo tanto insieme e abbiamo tante cose da fare. Spero che anche tu possa venire a visitare Roma con la tua famiglia un giorno!.



Dai *Learning Analytics* restituiti dal Google Moduli, è evidente come le domande sottostanti presenti nel test siano percepite come particolarmente difficili dalla maggior parte dei rispondenti. È necessario un intervento massiccio da parte del docente per calibrare e adattare i contenuti della prova allo specifico target di apprendenti.

Figura 14. *Domande più comunemente errate*





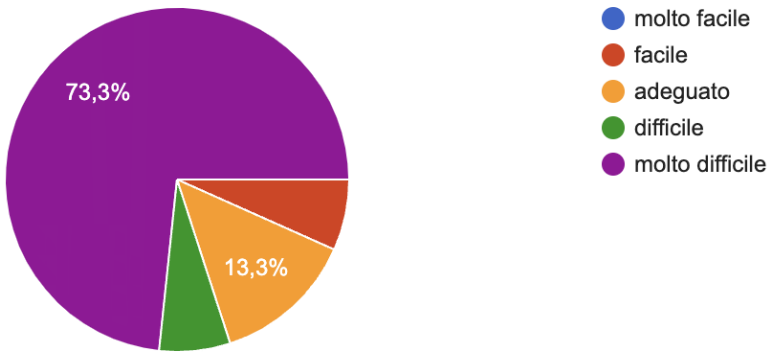
Le domande a risposta aperta richiedono la correzione e valutazione del docente, in quanto la macchina in questo caso, non è addestrata per intervenire. Tuttavia, la correzione e il feedback valutativi da parte del docente possono essere sicuramente più agevoli in questa modalità digitale rispetto al formato cartaceo. Il sistema può favorire la raccolta immediata di un'ampia quantità di testi scritti, risorsa preziosa per la ricerca.

L'esercitazione ha fornito agli studenti del corso di laurea l'opportunità di riflettere sugli errori più frequenti degli apprendenti non italofoni, come nella produzione di uno studente sinofono riportata di seguito, in cui si rilevano criticità legate all'uso degli articoli, al genere e al numero, tipici di questa tipologia di apprendenti.

> Descrivi la tua famiglia. Chi sono i membri della tua famiglia? Cosa fanno nella vita? Come sono fisicamente? Cosa vi piace fare insieme?
>
> *Nel mia famiglia soli due membri: *mi e la mia mamma.
> Nella mia famiglia ci sono *la mia madre, *il mio padre e il mio fratello.
> *Mia papà.
> Mia mamma *chiama Wangbei, Lei è un medico e ama ascoltare *musica.
> Ho un cagnolino di nome Wanzi, che è molto carino. Ci piace andare a vedere *film insieme.
> *C'è Franca, Andrea, Sara, Alice e Macchia.
> Mio papa è alto e magro, mia madre è bassa e un po' grassa, mi chiamo Dario, sono studente dell'università, ho 23 anni, sono basso e magro, mi piace giocare a biliardo. Viviamo in Cina a XiAn.

Al termine del test è stato chiesto agli informanti di valutarne il grado di complessità, in rapporto al corso di lingua di livello A2 frequentato presso l'università. Il giudizio converge sull'eccessiva difficoltà, percepita dal 73,3% dei rispondenti (Figura 15).

Figura 15. *Percezione della difficoltà del test realizzato con Diffit*



I commenti liberi dei rispondenti confermano l'eccessiva complessità di alcune strutture e la presenza di molte parole sconosciute, come evidenziato dall'*expert opinon* degli studenti del corso di laurea. Come ulteriore elemento di complessità si rileva anche la presenza delle istruzioni in inglese, che evidentemente contribuiscono a confondere ancora di più gli studenti, non essendo anglofoni e non conoscendo la lingua inglese, che altrimenti, potrebbe fungere da lingua ponte.

La piattaforma è probabilmente pensata per apprendenti alloglotti e i *grade* si riferiscono a parlanti nativi, non ad apprendenti di una lingua straniera.

Una opinione comune a tutti i rispondenti si ricollega a quanto emerso anche in riferimento all'altra sperimentazione citata in questo contributo: questo strumento può essere molto utile in ottica formativa, per l'esercizio e l'autoapprendimento individuale, in preparazione per esempio, di una prova CELI, pur considerando la difficoltà di allineamento ai livelli del QCERCV identificata come criticità di base.

Dal punto di vista del docente e dell'esaminatore è necessario tenere in debita considerazione tutte le questioni etiche legate all'IA, nonché le problematiche correlate all'autenticità e autorevolezza delle fonti su cui questi strumenti si basano, di cui non è sempre possibile verificare precisamente la provenienza.

8. Conclusioni

Il contributo ha inteso riportare i risultati preliminari di alcuni studi svolti presso l'Università per Stranieri di Perugia, nell'ambito del processo di digitalizzazione del *language testing* avviato in quest'ultimo periodo presso il CVCL, sulla scia dell'emergenza pandemica, che ha reso ancora più cogente la necessità di operare un rinnovamento delle attuali pratiche di valutazione e certificazione linguistica.

A tal fine, sono in corso studi e ricerche sul *language testing online* a partire dalla revisione della letteratura e dalla valorizzazione dell'expertise e dell'esperienza pregressa del CVCL in ambito di valutazione e di certificazione linguistica.

Un primo studio, condotto dal gruppo di ricerca del CVCL, si è concentrato sulla piattaforma Moodle, utilizzata per la digitalizzazione di un adattamento della prova CELI 1 (livello A2), percepita come piacevole e di facile utilizzo, ma con notevoli difficoltà nella prova di ascolto. Sicuramente, come peraltro rilevato dagli studenti stessi, la piattaforma Moodle può rappresentare un valido strumento per la valutazione formativa, l'autovalutazione e l'esercizio autonomo; tuttavia sarebbe necessario un più ampio

ripensamento di tutto il formato e gli item della prova, in funzione delle caratteristiche della piattaforma stessa, in quanto la semplice trasposizione del formato cartaceo in quello digitale non è ritenuta di ottimale fruizione.

Una seconda indagine, condotta come esercitazione nell'ambito di un corso di laurea dell'Università per Stranieri di Perugia ha messo in luce le potenzialità dell'Intelligenza Artificiale, che può sicuramente facilitare il lavoro di progettazione e creazione dei test, restituendo testi adattivi validi e appropriati alla tipologia di prompt inserito, con un'ampia gamma di item a risposta aperta e chiusa, per i quali è comunque insostituibile il lavoro del docente e dell'esaminatore. Soprattutto per il necessario allineamento ai livelli del QCERCV e al Profilo della lingua italiana, l'intervento del docente appare cruciale, anche in considerazione dei problemi legati alla difficoltà di verificare l'autorevolezza e l'attendibilità delle fonti da cui attingono questi strumenti potenziati dall'IA.

Si rilevano comunque, le enormi potenzialità dell'IA per l'esercizio e lo studio autonomo, in preparazione, per esempio, di una prova come un esame CELI, nonché per la riflessione e l'autovalutazione dei progressi dello studente, aspetti fondamentali per l'attivazione della metacognizione e della metariflessione.

Le piattaforme come Moodle, nonché l'ampia e sempre crescente gamma di strumenti potenziati dall'Intelligenza Artificiale, di cui Diffit rappresenta solo un esempio, potrebbero agevolare e facilitare il feedback automatico, la correzione e valutazione dei quesiti a risposta chiusa, agendo come scaffolding nei progressi dell'apprendente. Diverso potrebbe risultare il feedback sulla produzione scritta, che, anche quando affidata esclusivamente all'IA, come nel caso di molti dei più recenti strumenti, necessita spesso di un intervento o di una conferma da parte del docente o dell'esaminatore, soprattutto per verificarne il pieno allineamento ai livelli del QCERVC.

Gli studi presentati in questo contributo, seppure con le limitazioni legate al numero ridotto di informanti, confermano l'esigenza di esplorare ulteriormente questo campo di indagine e di ricerca, soprattutto in considerazione dei sempre più rapidi sviluppi tecnologici legati all'Intelligenza Artificiale.

RIFERIMENTI BIBLIOGRAFICI

Barni M. (2023), *Valutare le competenze nelle L2*, Carocci, Roma.

Breck E., Cai S., Nielsen E., Salib M., Sculley D. (2017), "The ML test score: A rubric for ML production readiness and technical debt reduction", in *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA, pp. 1123-1132.

Cinganotto L. (2023a), "L'Approccio Orientato all'Azione del QCER di Riferimento per le Lingue, Volume Complementare nella didattica digitale dell'Italiano L2/LS: scenari e attività didattiche nella percezione degli student", in *Italiano LinguaDue*, 15, 1, pp. 915-928: https://riviste.unimi.it/index.php/promoitals/article/view/20444.

Cinganotto L. (2023b), "Learning technologies for ELT during the pandemic in Italy: Teachers' attitudes", in Kourieos S., Evripidou D. (eds.), *Language Teaching and Learning during the COVID-19 Pandemic: A Shift to a New Era*, Cambridge Scholars Publishing, Cambridge, pp. 37-56.

Cinganotto L., Benedetti F., Langé G., Lamb T. (2022), *A survey of language learning/teaching with an overview of activities in Italy during the COVID-19 pandemic*, INDIRE: https://www.indire.it/wp-content/uploads/2022/02/Report-Survey-on-languages-13.02.2022.pdf.

Etikasari D., Andayani R. (2023), "Aplikasi Diffit sebagai Alternatif Media Pembelajaran Bahasa Indonesia", in *Jurnal Pendidikan Sekolah Dasare*, 2, 1, pp. 16-25.
Machetti, Vedovelli (2024), *Manuale della certificazione dell'Italiano L2*, Roma, Carocci.
Malagnini F., Cinganotto L. (2023), "Nuove opportunità del digitale nell'era del *new normal*", in Michelini M., Perla L. (a cura di), *Strategie per lo sviluppo della qualità nella didattica universitaria*, Bari, Pensa Editore, pp. 293-300.
Masillo P. (2019), *La valutazione linguistica in contesto migratorio: il test A2*, Pacini, Pisa.
Muzaffar A. W., Muhammad T., Anwar M. W., Chaudry Q., Mir S .R., Rasheed Y. (2021), "A Systematic Review of Online Exams Solutions in E-Learning: Techniques, Tools, and Global Adoption," in *IEEE Access*, 9, pp. 32689-32712: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9357335.
Myllyaho L., Raatikainen M., M.nnist T., Mikkonen T., Nurminen J. K. (2021), "Systematic literature review of validation methods for AI systems", in *The Journal of Systems & Software*, 181: https://doi.org/10.1016/j.jss.2021.111050.
Newhouse C.P., Cooper M. (2013), "Computer-based oral exams in Italian language studies", in *ReCALL*, 25, 3, pp. 321-339.
Piccardo E., North B. (2019), *The Action-oriented Approach: A Dynamic Vision of Language Education*, Multilingual Matters, Bristol UK.
Ritchie J., Lewis J. (2003), *Qualitative research practice: a guide for social science students and researchers*, Sage, London.
Salmani Nodoushan M. A. (2020), "Language assessment: Lessons learnt from the existing literature", in *International Journal of Language Studies*, 14, 2, pp. 135-146.
Santeusanio N. (2014), *Prepararsi alla DILS-PG,* Loescher, Torino.
Serragiotto G. (2016), *La valutazione degli apprendimenti linguistici,* Loescher, Torino.
Spina S., Fioravanti I., Forti L., Santucci V., Scerra A., Zanda F. (2022), "Il corpus CELI: una nuova risorsa per studiare l'acquisizione dell'italiano L2", in *Italiano LinguaDue,* 14, 1, pp. 116- 138: https://riviste.unimi.it/index.php/promoitals/article/view/18161.
Spina S. (2014), "Il Perugia Corpus: una risorsa di riferimento per l'italiano. Composizione, annotazione e valutazione", in Basili R., Lenci A., Magnini B. (a cura di), *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it*, vol. 1, Pisa University Press, Pisa, pp. 354-359.
Spinelli B., Parizzi F. (2010), *Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2. Per le Scuole superiori,* La Nuova Italia, Firenze.
Vassiliou S., Papadima-Sophocleous S., Giannikas C.N. (2023), "Technologies in Second Language Formative Assessment: A Systematic Review", in *Research Papers in Language Teaching and Learning*, 13, 1, pp. 50-63.
Zhang J. M., Harman M., Ma L., Liu Y. (2022), "Machine Learning Testing: Survey, Landscapes and Horizons", in *IEEE Transactions on Software Engineering*, 48, 1, pp. 1-36.