



TLT 2018

TRENTINO LANGUAGE TESTING

SECONDO PROFILO DELLE COMPETENZE LINGUISTICHE DEGLI STUDENTI TRENTINI

A cura di Luciano Covi e Mattia Oliviero



IPRASE – Istituto Provinciale per la Ricerca e la Sperimentazione Educativa

via Tartarotti 15 – 38068 Rovereto (TN) – C.F. 96023310228

tel. 0461 494500 – fax 0461 499266 – 0461 494399

iprase@iprase.tn.it, iprase@pec.provincia.tn.it – www.iprase.tn.it

Comitato tecnico-scientifico

Mario G. Dutto (Presidente)

Roberto Ceccato

Michael Schratz

Laura Zoller

Direttore

Luciano Covi

© Editore Provincia autonoma di Trento – IPRASE

Tutti i diritti riservati

Prima pubblicazione marzo 2019

Realizzazione grafica e stampa

La Grafica – Mori

ISBN 978-88-7702-468-8

Il volume è disponibile all'indirizzo web: www.iprase.tn.it alla voce risorse-pubblicazioni

Questa iniziativa è realizzata nell'ambito del Programma operativo FSE 2014-2020 della Provincia autonoma di Trento grazie al sostegno finanziario del Fondo sociale europeo, dello Stato italiano e della Provincia autonoma di Trento.

AZIONI A SUPPORTO DEL PIANO "TRENTINO TRILINGUE"

Sviluppo delle risorse professionali e predisposizione di strumenti di apprendimento e valutazione
CUP C79J15000600001 codice progetto 2015_3_1034_IP.01

La Commissione europea e la Provincia autonoma di Trento declinano ogni responsabilità sull'uso che potrà essere fatto delle informazioni contenute nei presenti materiali.



TLT 2018
TRENTINO LANGUAGE TESTING
SECONDO PROFILO
DELLE COMPETENZE LINGUISTICHE
DEGLI STUDENTI TARENTINI

A cura di Luciano Covi e Mattia Oliviero

MARZO 2019

Ringraziamenti

Molte sono le persone che a vario titolo hanno reso possibile questa ricerca. Si ringraziano gli studenti, i docenti, i referenti della somministrazione, gli assistenti di laboratorio, i dirigenti scolastici e il personale di segreteria per la collaborazione alla realizzazione della rilevazione linguistica.

Un particolare ringraziamento a Roberta Bisoffi, Simone Virdia e Martino Bernardi per il prezioso contributo.

IPRASE PER L'AMBIENTE



Questo documento è stampato interamente su carta certificata FSC® (Forest Stewardship Council®), prodotta con cellulosa proveniente da foreste gestite in modo responsabile, secondo rigorosi standard ambientali, sociali ed economici.



Indice

- 7 **Introduzione**
Luciano Covi
- 19 **Esiti dei profili linguistici degli studenti trentini**
Mattia Oliviero
- 45 **Le determinanti della performance linguistica degli studenti trentini**
Mattia Oliviero
- 61 **Il campionamento di Try-out 2017 e TLT 2018**
Maurizio Carpita
- 75 **Caratteristiche delle prove**
Davide Defant, Michela Chicco
- 83 **Analisi psicometrica delle prove TLT 2018**
Maurizio Carpita
- 105 **Sviluppo di un sistema automatico per la valutazione delle abilità linguistiche**
Giuseppe Falavigna, Roberto Gretter, Marco Matassoni, Nadia Mana, Ornella Mich
- 115 **Appendice**

Introduzione

Luciano Covi¹

Cos'è il programma TLT – Trentino Language Testing

Con Deliberazione n. 1458 del 31 agosto 2015, la Giunta provinciale ha deciso di varare un programma di verifica pluriennale dei livelli di competenza in lingua inglese e in lingua tedesca degli studenti trentini, affidandone la realizzazione a IPRASE. Di qui, la progettazione e l'avvio del *TLT – Trentino Language Testing*, un'azione sistematica, a livello provinciale, di valutazione del profilo linguistico degli studenti dei diversi ordini e gradi di scuola.

La necessità di accertare il livello di competenza degli studenti delle scuole provinciali in inglese e in tedesco rappresentava in effetti da tempo un'urgenza da affrontare in modo strutturale e compiuto. Non che non esistessero precedenti esperienze ed esplorazioni, realizzate sempre da IPRASE a inizi anni duemila², ma riguardavano principalmente iniziative limitate nella loro rappresentatività del sistema educativo provinciale e nel loro collegamento con il Quadro comune europeo di riferimento per le lingue (QCER). L'opportunità di colmare questo vulnus informativo risultava per altro via via più evidente a fronte di un numero crescente di iniziative di valutazione/certificazione messe in atto direttamente dalle singole scuole e, contemporaneamente, dalla crescente mole di dati forniti dalle prove standardizzate a livello nazionale (Invalsi) e internazionale (Pisa, Timss, Pirls...) riferite ad altre discipline (lingua italiana, matematica, scienze).

¹ Luciano Covi è Direttore di Iprase.

² Si vedano ad esempio i lavori *"Indagini sugli apprendimenti della lingua tedesca e della lingua inglese in quinta elementare e terza media"* a cura di D. Zuccarelli, Iprase, 2004.

Con il citato provvedimento normativo e l'avvio del *TLT* (acronimo che di seguito verrà usato per indicare il Trentino Language Testing), si è quindi inteso colmare il gap esistente, in termini di informazioni a livello provinciale, sui livelli di apprendimento tra alcune discipline oggetto di periodiche valutazioni nazionali e internazionali e il settore delle lingue straniere. Le finalità richiamate a fondamento di tale azione programmatica sono a seguito riassunte:

1. fornire agli insegnanti strumenti di verifica standardizzata per il miglioramento metodologico e didattico dell'insegnamento delle lingue e il raggiungimento degli obiettivi prefissati;
2. rafforzare le strategie di scuola per un efficace sviluppo dell'insegnamento delle lingue, anche con una sua rivisitazione metodologica lungo il corso degli anni (dalla scuola dell'infanzia alla conclusione del secondo ciclo);
3. favorire la diffusione di una cultura della valutazione delle competenze linguistiche e di una maggior consapevolezza della necessità di rapportare l'attività didattica ai risultati che si ottengono, avendo sempre in mente l'orizzonte ampio dell'apprendimento delle lingue a scuola;
4. dare agli studenti la possibilità di mettersi alla prova con un accertamento rigoroso delle competenze raggiunte e poter conseguire, al termine del percorso scolastico compiuto, certificazioni di competenze riconosciute a livello internazionale;
5. garantire ai genitori l'impegno della scuola per raggiungere obiettivi di competenza definiti e informazioni puntuali sulle performance dei singoli studenti;
6. dare ai decisori politici e all'amministrazione scolastica termini di riferimento per valutare, orientare, modificare le strategie messe in atto.

Parallelamente, i requisiti assunti nel processo di realizzazione delle prove standardizzate sono sintetizzabili nei seguenti elementi:

- a) allineamento e coerenza con il *Quadro comune europeo di riferimento per le lingue* (QCER)
- b) possibilità di confronto con i risultati ottenuti dagli studenti di altri paesi;
- c) assetto campionario e non censuario, avendo a riferimento un numero di studenti statisticamente rappresentativo dell'universo studentesco provinciale nei diversi ordini e gradi di scuola;
- d) attenzione a specifiche fasce d'età di studenti, e in particolare studenti del quinto anno di scuola primaria, del terzo anno delle secondarie di primo grado, del secondo anno delle secondarie di secondo grado e del terzo anno della leFP – Istruzione e formazione professionale;

- e) avendo a riferimento tutte le quattro abilità previste dal QCER (comprensione orale e scritta, produzione orale e scritta) per le due principali lingue comunitarie d'insegnamento (Inglese e Tedesco);
- f) con modalità di erogazione *computer-based*, secondo le metodologie avanzate esistenti e seguendo le migliori pratiche adottate nelle indagini valutative nazionali e internazionali per evitare ogni discriminazione nei confronti di alunni con bisogni educativi speciali;
- g) con stretto collegamento a materiali di sostegno e di preparazione per studenti e per docenti.

A seguito di tale provvedimento, per la prima volta nell'aprile 2016 è stata realizzata una rilevazione rappresentativa a livello provinciale di tali competenze linguistiche, con l'obiettivo di fornire un quadro conoscitivo di fondo dei livelli posseduti dalla popolazione studentesca trentina compresa tra i 10 e i 15 anni. In particolare, l'indagine ha riguardato un campione di 3000 studenti frequentanti l'ultimo anno della scuola primaria e della scuola secondaria di primo grado, il secondo anno delle secondarie di secondo grado e il terzo anno dell'istruzione e formazione professionale. Tutti i risultati della survey sono stati raccolti nel volume "*Rapporto TLT 2016 – Trentino Language Testing. Esiti delle rilevazioni delle competenze linguistiche degli studenti trentini*", pubblicato da Iprase nel 2017 e presentato in occasione della II Edizione del *Festival delle Lingue* di Rovereto.

Da un certo punto di vista, la prima iniziativa del programma trentino TLT è stata precorritrice dell'introduzione nell'a.s. 2017/18 della prova di Inglese da parte di INVALSI, una significativa novità a livello nazionale che ha visto, per la prima volta, tutti gli studenti italiani del grado 5 (quinta primaria) e del grado 8 (terza secondaria di primo grado) cimentarsi con un test in lingua Inglese, articolato in due parti: una di comprensione dell'ascolto (listening) e una di comprensione della lettura (reading). La restituzione dei risultati è avvenuta anche in questo caso secondo la descrizione del Quadro europeo delle lingue (QCER), con livelli attesi A2 per la terza secondaria di primo grado e A1 per la quinta primaria, ma con la possibilità di posizionare le competenze di ciascuno studente anche su livelli inferiori (rispettivamente livello A1 e livello pre-A1).

Pure per questa ragione, nel novembre 2017, con deliberazione n. 1794, la Giunta provinciale ha confermato l'opportunità di proseguire con gli interventi di verifica del profilo linguistico degli studenti trentini, invitando IPRASE alla "*continuazione del programma biennale di rilevazione delle competenze linguistiche TLT – Trentino*

Language Testing, iniziato nel 2016, così da fornire approfonditi e puntuali elementi a supporto e orientamento delle scelte e dell'implementazione della programmazione educativo-linguistica a livello provinciale". Ne è seguita la realizzazione, nel corso della primavera 2018, di una seconda survey, che ha evidentemente potuto contare sull'esperienza maturata con la prima indagine *TLT*.

L'impianto complessivo di questo secondo testing si è contraddistinto per essere in piena continuità con la precedente indagine del 2016, fatte salve alcune novità introdotte al fine di migliorare la qualità e la rispondenza degli esiti alle esigenze espresse dai diversi operatori impegnati nelle tradizionali valutazioni interne alle singole classi e scuole. In particolare, i nuovi accorgimenti inseriti nell'indagine *TLT 2018* sono stati:

- 1) l'affinamento delle prove utilizzate nella prima rilevazione da parte di un team di docenti dei diversi ordini e gradi di scuola coinvolti, sulla base delle indicazioni emerse da un'attenta analisi psicometrica;
- 2) la realizzazione di una fase preventiva di *try-out* (ottobre 2017), ovvero la somministrazione, in un numero congruo di classi, di un pre-test di verifica della validità dello stesso, così da addivenire a una versione definitiva opportunamente "validata";
- 3) una diversa articolazione del campione di studenti coinvolti, con una decisa sovrarappresentazione degli studenti del secondo anno delle secondarie di secondo grado e del terzo anno dell'istruzione e formazione professionale al fine di poter disaggregare gli esiti anche nei diversi indirizzi di studio;
- 4) l'impiego di un sistema di valutazione automatica aggiuntiva per le due abilità di produzione (scritta e orale), al fine di produrre in maniera automatica le valutazioni di alcuni indicatori significativi (proprietà lessicali, pronuncia, fluenza dell'eloquio, tempi di reazione, pertinenza delle risposte, capacità descrittive-narrative e comunicative...);
- 5) la definizione, a partire dagli esiti della somministrazione, di una "*Piattaforma demo*" da mettere a disposizione di tutti gli operatori delle istituzioni scolastiche e formative provinciali, al fine di favorire lo sviluppo di prassi e competenze docimologiche in lingua inglese e in lingua tedesca.

L'articolazione del Rapporto *TLT 2018*

La documentazione dell'intero processo di sviluppo del *TLT 2018* e dei risultati relativi alle competenze degli studenti delle scuole provinciali viene raccolta in questo rapporto, permettendo

così di consolidare e ampliare il primo profilo degli studenti trentini già messo a disposizione nel Rapporto *TLT 2016* innanzi citato. A ogni singola scuola sono stati già restituiti i risultati relativi alle singole classi e pertanto in questa sede si fornirà un quadro complessivo sui profili delle competenze linguistiche degli studenti a livello provinciale.

Nel primo capitolo si dà conto in modo dettagliato degli esiti delle prove standardizzate somministrate in questo secondo appuntamento riferito al programma *TLT – Trentino Language Testing*, così da mettere a disposizione di docenti e di addetti alla valutazione un ulteriore punto di osservazione che può essere messo in rapporto agli esiti del 2016, secondo uno sguardo diacronico. Nella fattispecie, vengono presentati i risultati dei test impiegando una valutazione automatica delle risposte per le abilità di ascolto e lettura (abilità ricettive) e di una valutazione esperta per le abilità produttive (produzione orale e scritta). Dopo un'attenta comparazione dei test utilizzati nei due momenti di rilevazione (2016 e 2018), al fine di identificare il raggiungimento dei livelli di abilità previsti è stata individuata una soglia di abilità minima del 60% nel 2018 (a fronte del 70% impiegato nel 2016).

Nel capitolo secondo si presenta l'analisi delle determinanti dei risultati ottenuti dagli studenti nelle diverse prove, utilizzando alcune informazioni socio-anagrafiche disponibili grazie al questionario sociolinguistico somministrato durante l'esecuzione della rilevazione. Per mezzo di un modello di regressione binaria, questa sezione mira a identificare quali caratteristiche ascrivibili agli studenti (genere, origine, carriera scolastica, background familiare, ecc.) possano risultare un fattore "di protezione" o, al contrario, un fattore "di rischio" rispetto al possedere o meno adeguate competenze linguistiche.

Nel terzo capitolo, si presentano i piani di campionamento utilizzati per la strutturazione dei gruppi di studenti che hanno partecipato sia alle rilevazioni della primavera 2018, sia ai try-out 2017 per la misurazione dei tre livelli (A1, A2 e B1) per tedesco e inglese.

Nel quarto capitolo sono richiamate le modalità utilizzate per l'impostazione e la somministrazione delle prove di competenza in lingua inglese e tedesca riferibili al QCER in funzione delle quattro abilità (ascolto, parlato, lettura e scrittura) e dei diversi livelli testati (A1, A2 e B1). Come già ricordato, la novità sostanziale che ha riguardato la strutturazione delle prove utilizzate in questa rilevazione è consistita nell'affinamento degli strumenti utilizzati nel *TLT 2016* in base sia a specifiche analisi di tipo psicometrico, sia a *try-out* preventivi realizzati nell'autunno 2017.

Il quinto capitolo propone invece un approfondimento psicometrico sui risultati della rilevazione che, come già accennato in precedenza, ha potuto beneficiare sia dell'esito delle rilevazioni *TLT 2016*, ma anche delle evidenze emerse a seguito del test di try-out condotto nell'autunno 2017.

Infine, nell'ultimo capitolo, si presenta un lavoro sperimentale e innovativo di valutazione automatica delle abilità linguistiche, nota anche come "*Call – Computer assisted language learning*", finalizzato a individuare parametri in grado di andare al di là del solo possesso o meno del livello di abilità QCER (A1, A2 o B1), quali ad esempio il *lessico* (o vocabolario), la scioltezza (*fluency*) del parlato, il numero medio di parole che compongono le frasi, la loro durata, ecc.

Alcune considerazioni ed evidenze complessive fornite dal TLT 2018

La rilevazione *TLT 2018* rappresenta un ulteriore tassello nella direzione, avviata nel 2016, di rafforzare *know-how*, competenze e strumenti di testing (quali ad esempio la "piattaforma demo") degli apprendimenti linguistici a livello provinciale. Ciò al fine di mettere tutte le istituzioni scolastiche trentine in condizione di migliorare la didattica e le abilità linguistiche delle nuove generazioni di studenti, individuandone punti di forza e criticità alla luce di evidenze empiriche fondate e mirate.

Lo sforzo messo in campo in tal senso è stato certamente ragguardevole e ha visto coinvolti numerosi esperti di settore, anche di livello internazionale, diversi ricercatori e operatori interni a IPRASE³, oltre 130 tra docenti e assistenti di laboratorio di 60 istituzioni scolastiche e formative provinciali e ben 2955 studenti di scuola primaria, secondaria di primo grado e di secondo grado. Certamente l'esperienza maturata nella prima rilevazione TLT del 2016 ha rappresentato una solida base di riferimento, da cui partire per migliorare di molto l'efficacia, la stabilità e la significatività degli esiti ottenuti.

Nel complesso, i profili linguistici degli studenti trentini ottenuti ai diversi gradi di scuola (quinto, ottavo, decimo e undicesi-

³ Si ringraziano in particolare Michela Chicco e Roberta Bisoffi per il prezioso e accurato lavoro di coordinamento di tutto il complesso processo riguardante la rilevazione TLT 2017/2018

mo per l'ambito della leFP) si confermano più che positivi e sulla scia di quanto riscontrato nella prima rilevazione TLT del 2016.

Nello specifico, per ciascun livello, è possibile richiamare le seguenti evidenze:

A) in corrispondenza del livello A1:

- i dati evidenziano in media una percentuale più alta di studenti in possesso delle abilità ricettive (ascolto e lettura) rispetto alle abilità produttive (parlato e scrittura);
- si rileva una bassa correlazione tra le due lingue inglese e tedesco nelle singole abilità, con una variabilità negli apprendimenti che riflette il diverso periodo di esposizione alle due lingue da parte degli alunni;
- si riscontrano differenze nei punteggi medi di classe tra inglese e tedesco, in particolare per le abilità produttive;
- è maggiore la correlazione tra tedesco e inglese di quella opposta, ovvero classi che ottengono buoni esiti in tedesco hanno un'alta probabilità di avere risultati simili in inglese, mentre non accade il contrario.

B) in corrispondenza del livello A2:

- si registra un sostanziale equilibrio nelle abilità tra inglese e tedesco, fatta eccezione per la scrittura, in cui i ragazzi evidenziano abilità maggiori per l'inglese rispetto al tedesco;
- esiste minore variabilità di risultati nella lingua tedesca che in quella inglese; probabilmente ciò è riconducibile anche al fatto che gli apprendimenti linguistici extrascolastici sono più comuni per quest'ultimo idioma che per il primo, creando maggiori differenze nei livelli di competenza;
- la variabilità dei risultati riguarda anche in questo caso più le abilità produttive di quelle ricettive, a conferma di quanto detto al punto precedente.

C) in corrispondenza del livello B1 – Biennio Scuola secondaria di secondo grado:

- in generale ascolto e scrittura risultano più sviluppate di parlato e lettura;
- rispetto ai diversi indirizzi, a parte il linguistico, che evidenzia evidentemente livelli di performance di gran lunga superiori alla media, ottengono buoni risultati anche il liceo classico e l'indirizzo economico;
- a parte il classico e il linguistico, si riscontra una certa variabilità di risultato tra le due lingue, soprattutto in corrispondenza della produzione orale e scritta.

D) in corrispondenza del livello B1 – Terzo anno della leFP:

- si registrano buoni risultati in inglese, più nell'ascolto e nella scrittura che nella produzione orale e nella lettura;

- a parte l'ascolto, i ragazzi dell'Istruzione e Formazione Professionale registrano risultati contenuti nella lingua tedesca e comunque inferiori rispetto all'inglese.

I buoni risultati raggiunti dagli studenti trentini nella rilevazione TLT 2018 trovano peraltro ulteriore conferma dagli esiti ottenuti nelle prove INVALSI 2018, che per la prima volta hanno testato anche le competenze in lingua inglese in quinta primaria e all'ultimo anno delle secondarie di primo grado. In questa disciplina la scuola trentina si posiziona tra le regioni con i risultati più alti del Paese, conseguendo risultati di vera eccellenza. Oltre il 90% degli allievi della V primaria raggiunge il livello A1 nelle competenze di lettura (reading). Inoltre Trento, insieme a Bolzano, è l'unica area del Paese in cui oltre il 90% di questi alunni raggiunge il livello A1 nelle competenze di ascolto (listening). Anche al termine del primo ciclo d'istruzione gli allievi trentini raggiungono ottimi risultati in Inglese sia in lettura (reading) sia in ascolto (listening), posizionandosi tra gli studenti migliori d'Italia. A Trento si riscontra una delle percentuali più alte d'Italia di allievi che raggiungono il livello A2 nel listening, dato di quasi 20 punti percentuali superiore alla media nazionale. Anche nel reading gli allievi trentini che raggiungono il livello A2 superano di quasi 10 punti percentuali il valore nazionale e si collocano tra gli studenti con i risultati migliori.

Tali risultati appaiono ancora più significativi se si considera che in provincia la lingua di riferimento non è solo l'inglese ma anche il tedesco. Inoltre, poiché le prove proposte da INVALSI sono il frutto di una collaborazione internazionale, i risultati così buoni degli allievi trentini acquisiscono un ulteriore valore, ossia testimoniano livelli di competenza (lettura e ascolto) in lingua inglese che pongono gli allievi della scuola del Trentino al livello delle migliori esperienze europee.

Figura 1 - Esiti prove INVALSI 2018 in inglese per livello al grado 5 (listening e reading).

■ Pre A1 ■ A1

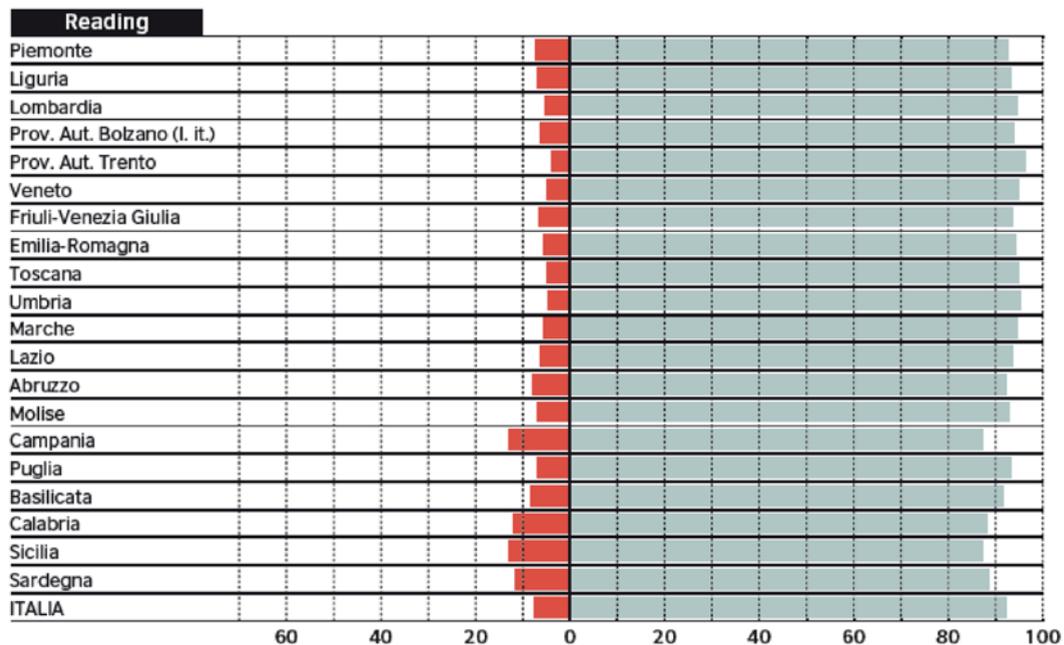
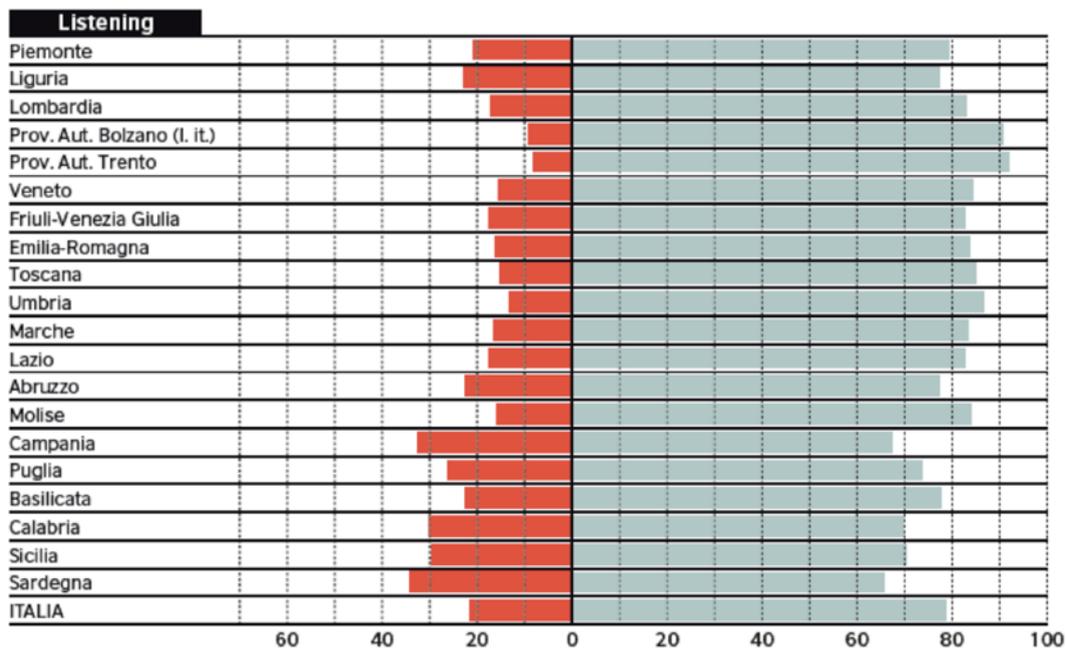
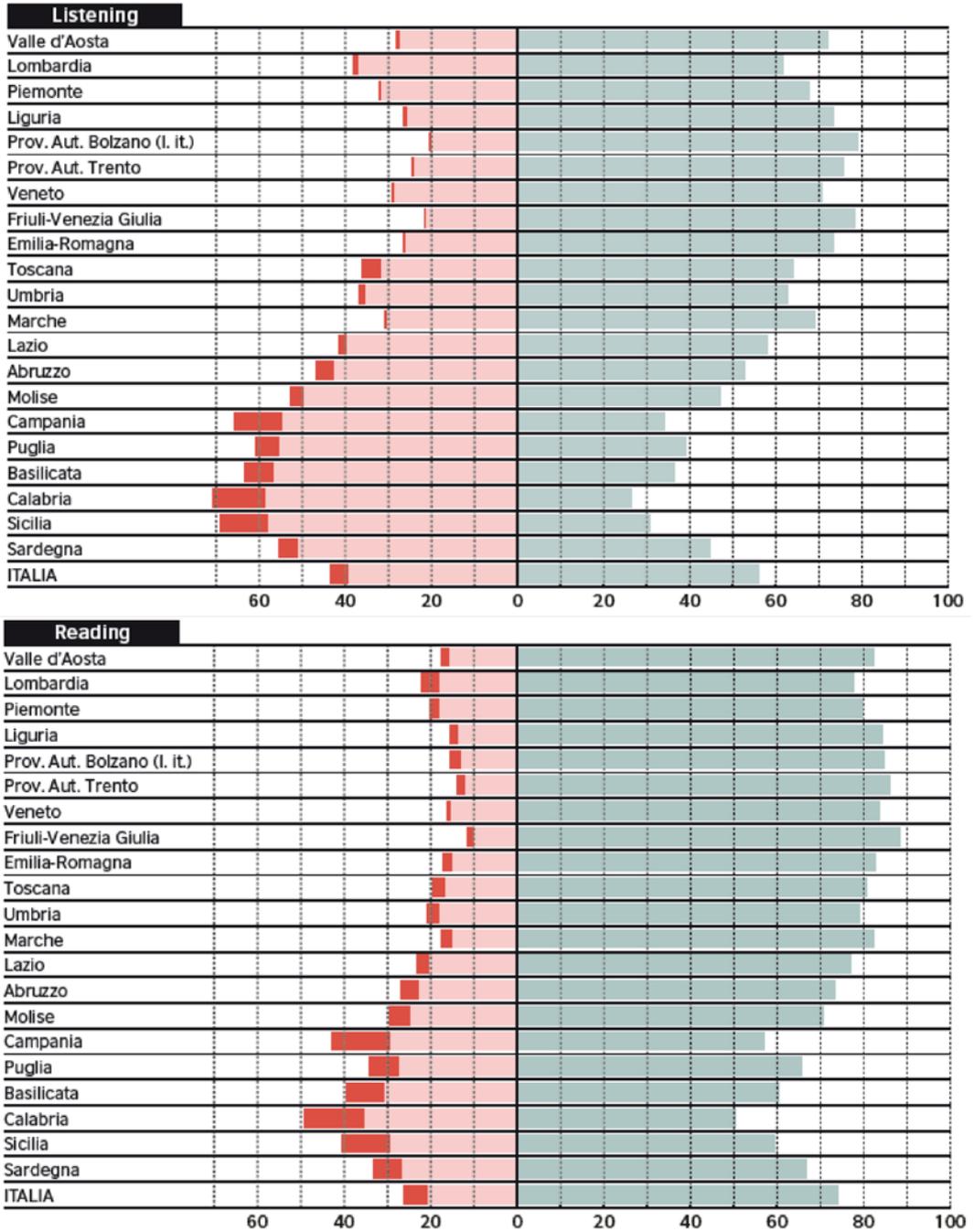


Figura 2 - Esiti prove INVALSI 2018 in inglese per livello al grado 8 (listening e reading).

■ Pre A1 ■ A1 ■ A2



A fronte dei buoni risultati emersi, gli esiti del TLT 2018 non mancano di evidenziare situazioni che necessitano di maggior attenzione e interventi di miglioramento. Ad esempio:

- la produzione orale (parlato) risulta l'abilità meno sviluppata per tutti i livelli e per entrambe le lingue; certamente si tratta dell'abilità più complessa ed esigente in termini di tempo richiesto per esercitarla e per rilevarla, ma si tratta anche dell'abilità che richiede metodologie, strumenti e modalità di insegnamento delle lingue più innovativi;
- le abilità produttive (produzione orale e scritta) sono quelle dove la variabilità tra le due lingue è più alta e dove è più necessario lavorare;
- esiste una certa variabilità dei risultati tra le diverse classi, specie in corrispondenza delle abilità produttive e per i livelli A1 e A2 rispetto alla quale è necessario dedicare maggiore attenzione.

Dal punto di vista delle indicazioni che la rilevazione TLT 2018 fornisce al sistema educativo provinciale, si conferma l'opportunità di considerare, a tutti i livelli scolastici, la dimensione dell'oralità (ascolto e parlato) come componente essenziale, quanto la comprensione del testo e la scrittura. In quest'ottica, potrebbero risultare particolarmente importanti e utili l'adozione di specifiche soluzioni funzionali (anche di tipo tecnologico), mirati rapporti di collaborazione con altre realtà o sistemi scolastici a elevata performance, ovvero esperienze e progetti di mobilità all'estero appositamente strutturate a tal fine. Lo sviluppo delle competenze comunicative e di interazione in lingua potrebbe essere inoltre favorito da una rivisitazione dei criteri e delle modalità di valutazione interna alle scuole, nonché da una rivisitazione dell'insegnamento delle lingue in linea con le indicazioni suggerite dalla crescente diffusione di nuove modalità di valutazione standardizzata (sulla falsariga del DSD⁴ o della stessa piattaforma Demo TLT di Iprase).

Così come nel 2016, anche nel 2018 la variabilità dei risultati tra le classi è apparsa piuttosto elevata, a testimonianza che le variabili di contesto assumono un certo peso esplicativo rispetto alle abilità linguistiche dei giovani studenti trentini. Inoltre, come avviene per le altre aree disciplinari, piuttosto significativa è l'influenza del livello di istruzione dei genitori: le aspettative, la con-

⁴ Il DSD Deutsches Sprachdiplom è un diploma di lingua tedesca rilasciato dalla Conferenza dei Ministri dell'Istruzione degli Stati federali della Repubblica Federale Germania. Certifica la competenza nelle quattro abilità linguistiche di livello A2/B1 del QCER per il diploma DSD I e di livello B2/C1 del QCER per il diploma DSD II.

sapevolezza dell'importanza delle lingue, le esperienze di esposizione linguistica possono essere molto diverse in rapporto al contesto culturale della famiglia.

Il fatto che le competenze linguistiche si rafforzino per accumulazione negli anni, come dimostrano i risultati ottenuti ai diversi gradi scolastici sia per l'inglese che per il tedesco, conferma l'importanza di un approccio per obiettivi, che debbono essere coerenti e connessi lungo l'intero percorso scolastico e di crescita degli studenti. Ciò soprattutto al fine di evitare "ripartenze" che possono rappresentare gravi dispendi di tempo ed energie, sia per i docenti che per gli alunni.

La definizione degli obiettivi scolastici dovrebbe possibilmente tener conto anche delle occasioni di utilizzo attivo delle lingue in ambito extrascolastico. In quest'ottica, gioca un ruolo certamente importante il differenziale esistente nelle possibilità di uso quotidiano dell'inglese e del tedesco: mentre la pervasività dell'uso della lingua inglese appare molto significativa, anche per via delle nuove tecnologie dell'informazione e della comunicazione, non così avviene nel caso della lingua tedesca, il cui utilizzo rimane purtroppo eccessivamente circoscritto a specifici ambiti territoriali o economico-produttivi (come ad esempio il turismo, i trasporti o alcuni settori manifatturieri).

1 | Esiti dei profili linguistici degli studenti trentini

Mattia Oliviero⁵

L'obiettivo di questo capitolo è di fornire un quadro generale sui profili linguistici degli studenti trentini. La rilevazione delle competenze linguistiche si è svolta su base campionaria e ha visto coinvolti 2.526 studenti che hanno sostenuto la prova in inglese e 2.181 che hanno sostenuto quella in tedesco suddivisi in 60 istituti e 149 classi⁶ (cfr. Tabella 1.1). Nel primo ciclo la prova del livello A1 è stata svolta al termine della scuola primaria mentre per il livello A2 al termine della Secondaria di primo grado; per il secondo ciclo invece la presenza del livello B1 è stata rilevata al termine del biennio della Secondaria di secondo grado e del triennio per l'Istruzione e Formazione Professionale.

I test standardizzati di cui si riportano gli esiti hanno rilevato in modo disgiunto tutte le abilità linguistiche e sono stati somministrati attraverso l'utilizzo di una piattaforma informatica (*computer-based*). La valutazione delle prove si è svolta in due modalità distinte: *automatica* per le abilità di ascolto e lettura – effettuata direttamente dal sistema – ed *esperta* per le abilità di produzione orale e scritta – realizzata da valutatori esperti⁷.

Al fine di identificare il raggiungimento dei livelli di abilità previsti è stata individuata una soglia di abilità minima del 60%. La Tabella 1.2 mostra la percentuale di studenti in possesso della specifica competenza linguistica per livello previsto (rispettivamente A1, A2 e B1) e per lingua (Inglese e Tedesco), ovvero la percentuale di studenti che nelle singole abilità ha ottenuto un punteggio di almeno il 60% del punteggio totale.

⁵ Mattia Oliviero è collaboratore di ricerca presso Iprase.

⁶ Per un approfondimento sui metodi di campionamento e sulle caratteristiche del campione si veda il Capitolo 3.

⁷ I dettagli sull'impostazione, la rilevazione e la valutazione delle prove possono essere approfonditi nel Capitolo 4.

Tabella 1.1 - Numero di istituti, classi e studenti che hanno partecipato alla rilevazione TLT 2018.

	INGLESE	TEDESCO
<i>V Primaria (A1)</i>		
Istituti	21	21
Classi	30	30
Studenti	502	496
<i>III sec. I grado (A2)</i>		
Istituti	16	16
Classi	30	30
Studenti	584	585
<i>II sec. II grado (B1)</i>		
Istituti	20	20
Classi	57	56
Studenti	1059	1021
<i>III leFP (B1)</i>		
Istituti	10	6
Classi	27	6
Studenti	381	79
<i>Totale</i>		
Istituti	67*	63*
Classi	144	123
Studenti	2526	2181

* le somme indicate non coincidono con il totale degli istituti coinvolti nel TLT 2018 poiché nella maggior parte di questi sono state svolte le prove in entrambe le lingue.

In linea generale ciò che emerge sono i buoni livelli di abilità raggiunti dagli studenti trentini per entrambe le lingue soprattutto per quanto riguarda l'ascolto, la lettura e la produzione scritta. La produzione orale invece risulta essere l'abilità meno sviluppata.

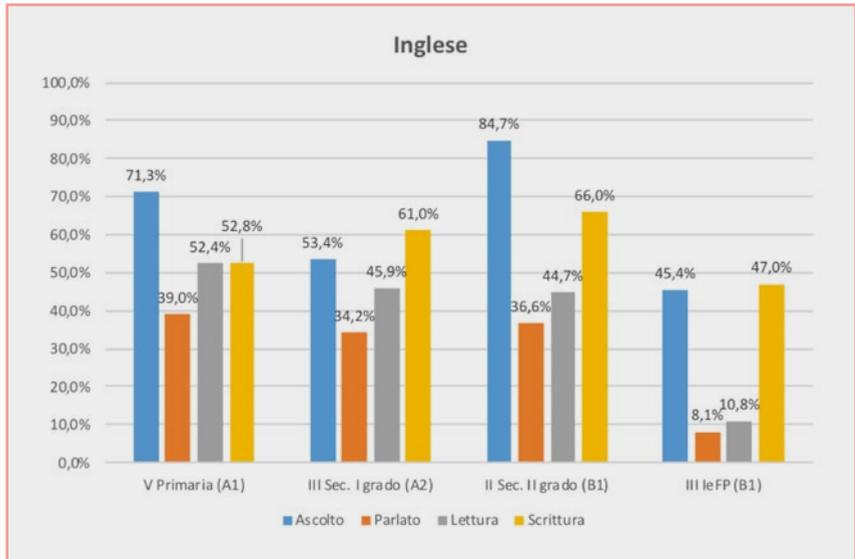
Nello specifico, il primo elemento da considerare è che l'ascolto in inglese risulta l'abilità più sviluppata per il livello A1 – con il 71,3% degli studenti che supera la soglia – e per il livello B1 (Secondaria di secondo grado) – con l'84,7% degli studenti che possiede la competenza (cfr. Figura 1.1).

Tabella 1.2 - Studenti al di sopra della soglia di abilità minima (60%) per livello, lingua e abilità linguistica (valori percentuali %).

		ASCOLTO	PARLATO	LETTURA	SCRITTURA
V Primaria (A1)	Inglese	71,3	39,0	52,4	52,8
	Tedesco	53,4	27,4	74,0	40,9
III Secondaria I grado (A2)	Inglese	53,4	34,2	45,9	61,0
	Tedesco	51,8	28,4	48,2	48,4
II Secondaria II grado (B1)	Inglese	84,7	36,6	44,7	66,0
	Tedesco	72,6	36,0	49,5	47,9
III IeFP (B1)	Inglese	45,4	8,1	10,8	47,0
	Tedesco	27,8	0,0	7,6	12,7

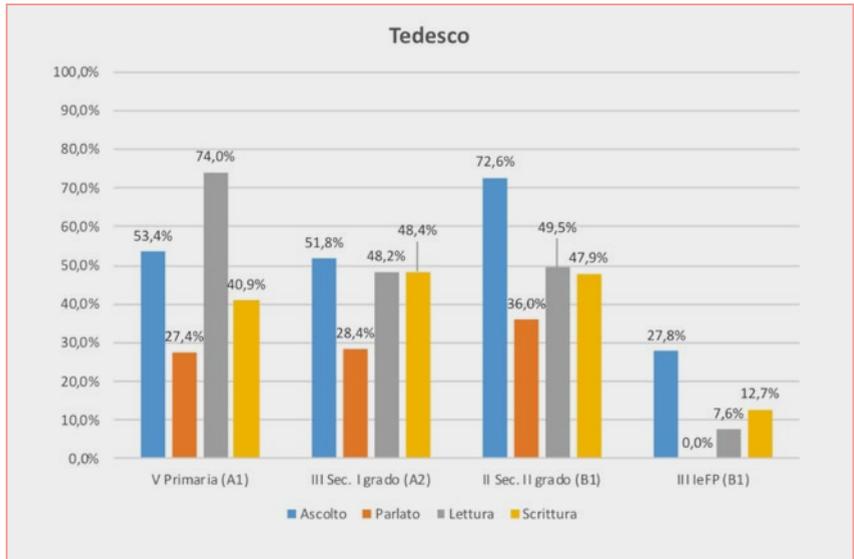
Esito questo che può trovare un'interpretazione anche nei numerosi stimoli che, attraverso l'ascolto della musica o la visione di film e serie tv in lingua originale, permettono soprattutto agli adolescenti di esercitare in misura maggiore l'ascolto della lingua inglese. Per quanto riguarda invece il livello A2, l'abilità posseduta in modo più diffuso dagli studenti risulta essere la produzione scritta (61,0%) che supera di poco l'ascolto (53,4%). In media, per tutti i livelli considerati, l'abilità di produzione orale risulta essere posseduta da circa uno studente su tre confermandosi trasversalmente per entrambe le lingue il profilo linguistico maggiormente deficitario. Se da un lato questo risultato propone l'abilità del parlato come l'area con maggiori margini di miglioramento, dall'altro la produzione orale si configura come una delle competenze più complesse e contemporaneamente più esigenti in termini di tempo richiesto per esercitarla e che richiede strumenti *ad hoc* per una sua pratica diffusa e costante (cfr. paragrafo 4.6 Piattaforma Demo). Un ulteriore elemento da considerare è di tipo tecnico ed è legato alla modalità con la quale è stato svolto il test (*computer-based*; cfr. Capitolo 4). Infatti, l'interazione in fase di valutazione della produzione orale diventa fondamentale e aver interagito con un computer può aver rappresentato un ostacolo in particolar modo per gli studenti più giovani.

Figura 1.1 - Percentuale di studenti al di sopra della soglia di abilità minima (60%) per livello, lingua e abilità linguistica - Lingua inglese.



Per quanto riguarda la lingua tedesca (cfr. Figura 1.2) si evidenzia un sostanziale equilibrio tra i vari livelli per le abilità di ascolto, lettura e scrittura. Infatti, la percentuale di studenti sopra la soglia minima di abilità per queste competenze si attesta intorno al 50%. L'eccezione in positivo è rappresentata dalla lettura per il livello A1 e l'ascolto per il livello B1 – Secondaria di secondo grado nelle quali la proporzione di studenti in possesso della competenza si attesta rispettivamente a 74,0% e 72,6%. Così come sottolineato per l'inglese, anche per il tedesco la produzione orale risulta l'abilità meno sviluppata, in misura maggiore per i primi due livelli (27,4% per l'A1 e 28,4% per l'A2) rispetto al livello B1 (36,0%).

Figura 1.2 - Percentuale di studenti al di sopra della soglia di abilità minima (60%) per livello, lingua e abilità linguistica - Lingua tedesca.

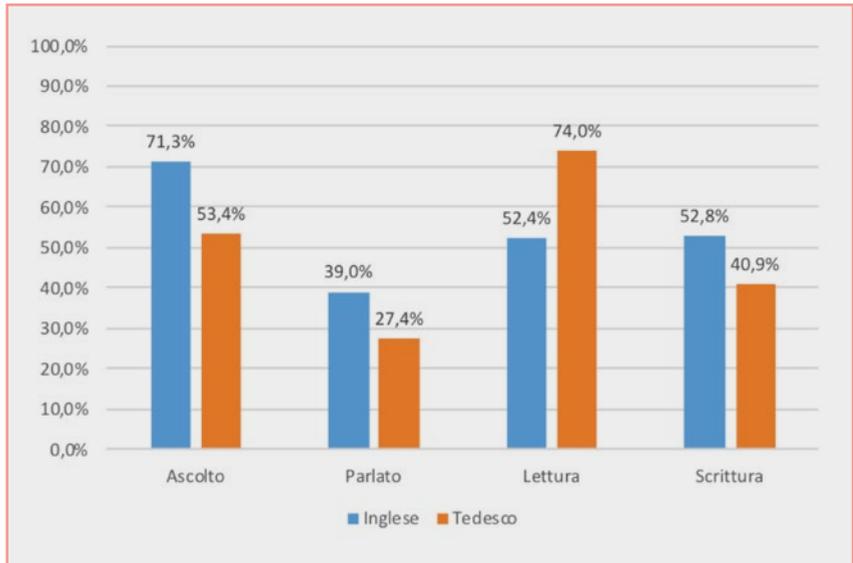


1.1 Confronto tra lingue

1.1.1 A1 – V Primaria

A livello generale, il primo aspetto rilevante che emerge nella comparazione tra lingua inglese e tedesca per il livello A1 indica che le *abilità ricettive* – ascolto e lettura – mostrano in media una percentuale più alta di studenti sopra la soglia rispetto alle *abilità produttive* – parlato e scrittura (cfr. Figura 1.3). In un quadro generale in cui le competenze in lingua inglese risultano essere più sviluppate rispetto a quelle in lingua tedesca, una particolarità è rappresentata dalla lettura in lingua tedesca nella quale tre studenti su quattro sono in possesso dell'abilità risultando la competenza più sviluppata per il livello A1 superando, seppur di poco, l'ascolto in lingua inglese (74,0% e 71,3% rispettivamente). Il secondo aspetto da considerare, a cui si è già fatto riferimento in precedenza, sottolinea come l'abilità del parlato risulti il profilo linguistico meno sviluppato in particolare per la lingua tedesca dove solo uno studente su quattro risulta in possesso di tale competenza.

Figura 1.3 - Percentuale di studenti al di sopra della soglia di abilità minima (60%) per lingua e abilità linguistica - A1 Primaria.



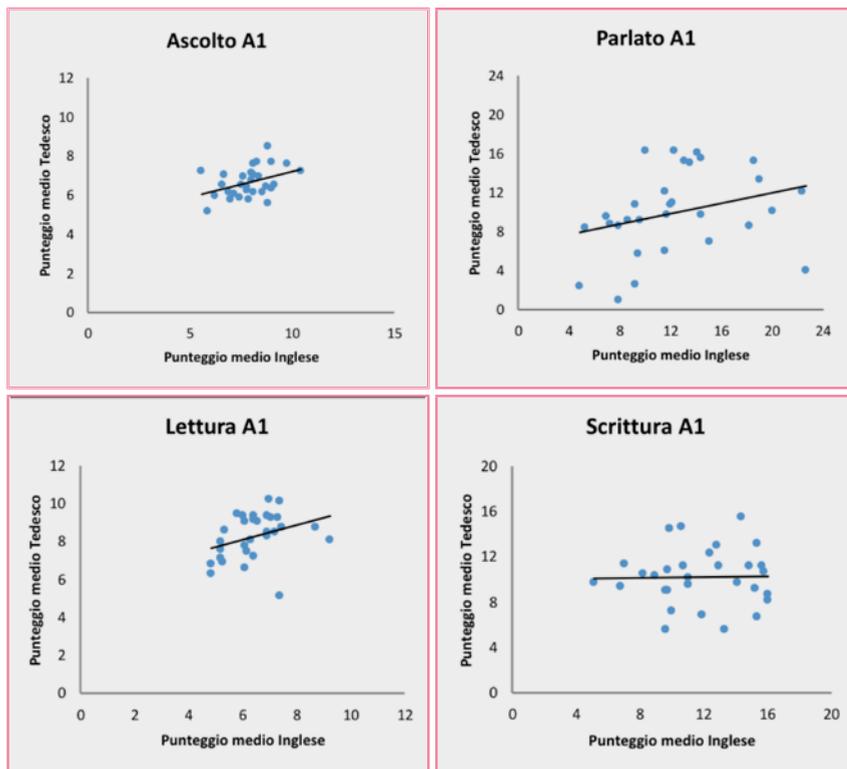
Le differenze riscontrate nei profili linguistici degli studenti tra le due lingue sono supportate anche dalle correlazioni tra i punteggi individuali degli studenti nelle quattro abilità. I coefficienti di correlazione, mostrati in Tabella 1.3, indicano la misura con la quale i punteggi individuali ottenuti in un'abilità variano al variare dei punteggi ottenuti in un'altra. In linea generale, i coefficienti di correlazione tra abilità si attestano mediamente su valori medio-bassi – dove 0 indica l'assenza di correlazione mentre -1 e +1 una correlazione perfetta (negativa o positiva) – nelle quattro abilità di entrambe le lingue. Quello che emerge è che si possono osservare coefficienti di correlazione più alti per l'inglese tra le abilità di ascolto e lettura da un lato (0.53), e tra parlato e scrittura dall'altro (0.49). Questo indica che, seppur in un quadro di coefficienti di correlazione moderati, tendenzialmente gli studenti che hanno ottenuto un punteggio alto (o basso) in ascolto/inglese, tenderanno ad avere un punteggio alto (o basso) anche in lettura/inglese e viceversa. Lo stesso discorso si può applicare per le due abilità produttive. Per il tedesco questa relazione sembra essere più debole tra le due abilità ricettive rispetto a quanto riscontrato per la lingua inglese. I valori di correlazione particolarmente bassi tra le due lingue nella stessa abilità, indicano la presenza di variabilità negli apprendimenti che nello specifico della scuola primaria possono essere interpretati alla luce dell'eterogeneità in termini di anni di apprendimento tra le due lingue.

Tabella 1.3 - Correlazione tra punteggi individuali nelle varie competenze e tra lingue - A1 Primaria.

INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Ascolto	1			
Parlato	0.4349	1		
Lettura	0.5253	0.3513	1	
Scrittura	0.485	0.4942	0.4781	1
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Ascolto	1			
Parlato	0.2692	1		
Lettura	0.3735	0.3982	1	
Scrittura	0.4195	0.5111	0.416	1
CORRELAZIONE INGLESE – TEDESCO				
	Ascolto	Parlato	Lettura	Scrittura
	0.3919	0.3211	0.4077	0.3319

Questo concetto si ripresenta in maniera più netta se osserviamo la correlazione tra i punteggi medi a livello di classe nelle due lingue (cfr. Figura 1.4). In particolare, la distribuzione dei punteggi medi a livello di classe suggerisce come vi sia una maggiore variabilità dei risultati nelle prove delle abilità produttive – parlato e scrittura – rispetto alle abilità ricettive – ascolto e lettura – nelle quali la dispersione risulta meno pronunciata. Ad esempio, per la produzione scritta il coefficiente di correlazione è prossimo allo zero (linea quasi orizzontale). Questo sta a indicare un'eterogeneità molto forte in questa abilità tra le due lingue. Risulta importante sottolineare nuovamente e avere presente in sede di interpretazione come il confronto tra le due lingue nella scuola primaria debba tenere conto della differenza per quanto riguarda gli anni di esposizione alle due lingue. Infatti, come verrà sottolineato nelle sezioni successive, vi è una tendenza a un aumento dei coefficienti di correlazione tra uguali abilità nelle due lingue al crescere degli anni di apprendimento. Più in concreto, il punteggio ottenuto dagli studenti di V Primaria nell'ascolto per l'inglese ad esempio, tenderà a essere più simile all'ascolto in tedesco nel livello A2 rispetto all'A1, e questo si presenterà in misura ancora maggiore per il livello B1.

Figura 1.4 - Correlazione Inglese-Tedesco tra i punteggi medi di classe nelle quattro abilità - A1 Primaria.

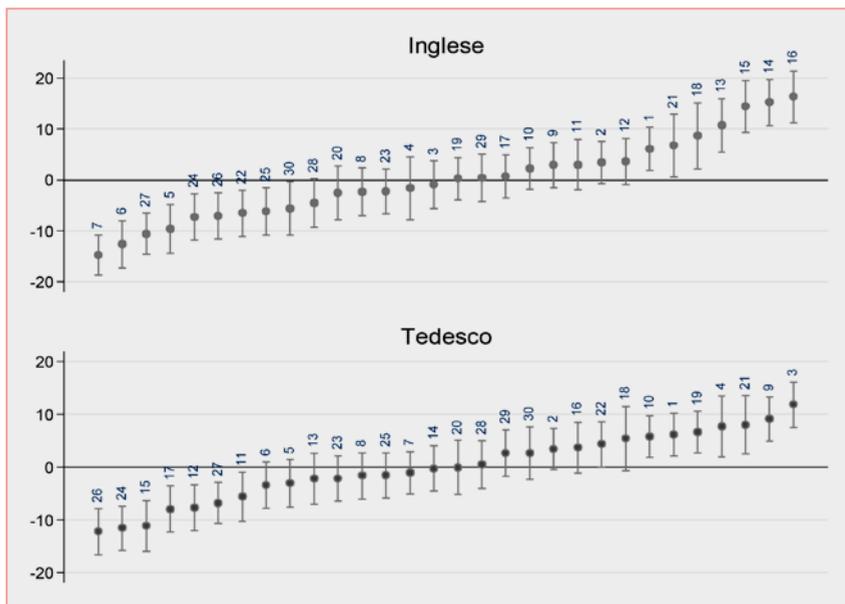


Anche per quanto riguarda la distribuzione dei punteggi totali medi per classe ottenuti nel test nelle due lingue (Figura 1.5) possiamo notare come vi siano alcune classi che mostrano punteggi significativamente superiori alla media (linea orizzontale) ma anche classi con valori inferiori. Inoltre, a ulteriore supporto del fatto che la bassa correlazione nei punteggi tra la lingua inglese e tedesca per la V Primaria possa riflettere i diversi anni di apprendimento delle due lingue, le classi che mostrano i punteggi più alti in inglese in linea generale tendono a mostrare punteggi non altrettanto alti in tedesco (in media o inferiori) – è il caso ad esempio delle classi indicate con i numeri 13, 14, 15 e 16⁸ – mentre risulta meno frequente il meccanismo opposto. Appare frequente infatti che una classe con un punteggio alto in tedesco mostri un punteggio in media o superiore anche per quanto riguarda l'inglese. È necessario specificare tuttavia che l'intento di questo

⁸ I numeri assegnati alle classi nelle Figure 1.5 e 1.8 sono attribuiti casualmente e hanno il solo scopo di supportare la presentazione dei risultati.

confronto vuole essere quello di identificare dei pattern generali a supporto della didattica e non di stilare una graduatoria degli esiti delle varie classi.

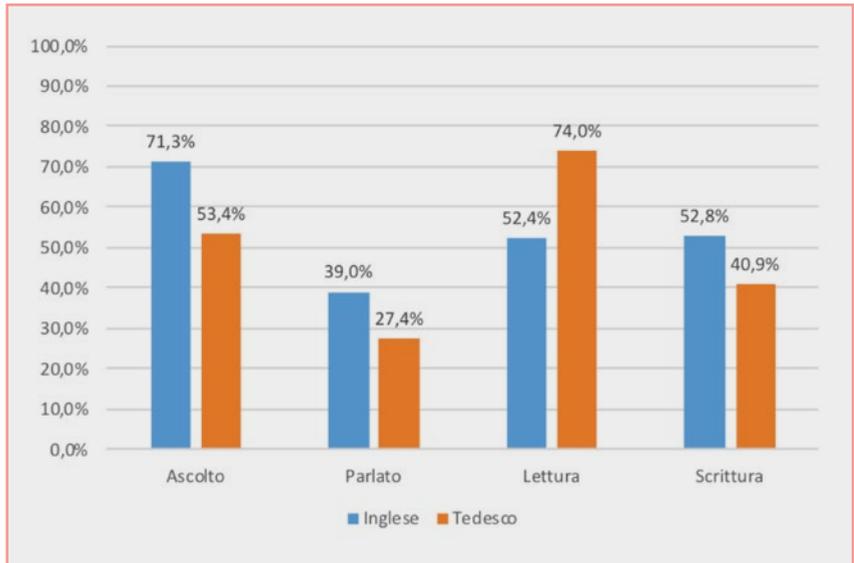
Figura 1.5 - Variabilità punteggi medi per classe in inglese e tedesco rispetto alla media generale - A1 Primaria.



1.1.2 A2 – III Secondaria di primo grado

Il meccanismo per il quale al crescere degli anni di esposizione alle lingue tendono a diminuire le differenze in termini di profili linguistici tra inglese e tedesco è supportato anche dai risultati dell'A2 – III Secondaria di primo grado. Infatti, osservando la Figura 1.6 – in cui viene posto l'accento sul confronto tra inglese e tedesco nelle singole competenze per il livello A2 – risulta evidente come le differenze tra le lingue si siano ridotte rispetto a quanto visto per il livello A1 (cfr. Figura 1.3). Segnale questo che la progressiva uniformizzazione del monte ore di insegnamento delle due lingue ha come risultato quello di produrre un equilibrio maggiore nelle abilità linguistiche in inglese e tedesco. Nello specifico, l'ascolto e la lettura mostrano risultati pressoché identici (con circa il 50% di studenti in possesso dell'abilità), mentre per il parlato la differenza tra le due lingue risulta modesta. Un discorso diverso emerge per la produzione scritta nella quale permane un moderato vantaggio della lingua inglese (61%) sulla lingua tedesca (48,4%).

Figura 1.6 - Percentuale di studenti al di sopra della soglia di abilità minima (60%) per lingua e abilità linguistica - A2 Secondaria I grado.



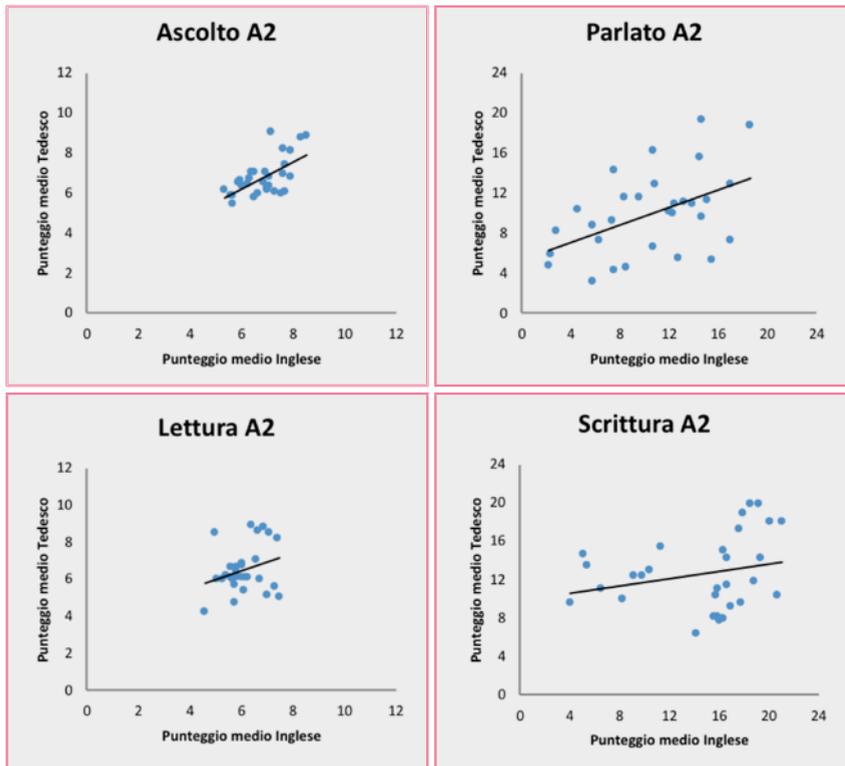
La correlazione tra le quattro competenze risulta più alta nel tedesco rispetto all'inglese (cfr. Tabella 1.4), il che potrebbe indicare che gli studenti nel tedesco sono soggetti a una didattica più uniforme che non tende a prediligere una competenza piuttosto di un'altra. A differenza del livello A1, per il livello A2 non è possibile identificare un pattern di similitudine tra le abilità ricettive da una parte e le abilità produttive dall'altra. Quello che però si evidenzia è che vi è una tendenza, come accennato nella sezione precedente, a un aumento della correlazione tra lingue nelle varie abilità. Segnale questo che l'accumulo nel numero di anni di apprendimento in entrambe le lingue tende a rendere più omogenei i punteggi individuali degli studenti nelle due lingue.

Anche per il livello A2, la distribuzione dei punteggi medi delle classi in inglese e tedesco (cfr. Figura 1.7) risulta essere meno omogenea nelle prove relative alle abilità produttive (molto forte nel caso della produzione scritta) rispetto alle abilità ricettive, sintomo questo di un alto grado di eterogeneità tra le classi nella produzione orale e scritta tra lingua inglese e tedesca.

Tabella 1.4 - Correlazione tra punteggi individuali nelle quattro competenze - A2 Secondaria I grado.

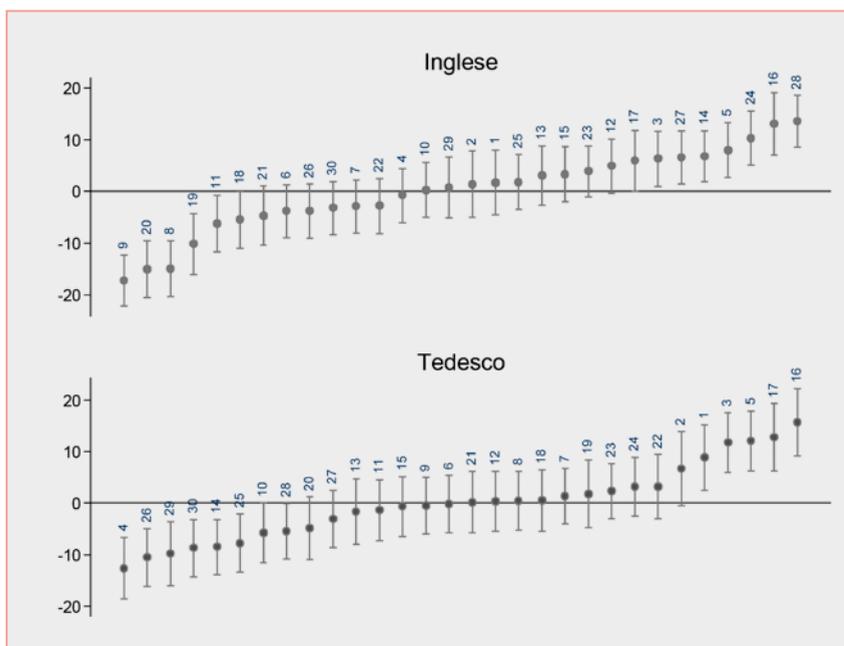
INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Ascolto	1			
Parlato	0.2938	1		
Lettura	0.4236	0.3152	1	
Scrittura	0.4613	0.4687	0.4892	1
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Ascolto	1			
Parlato	0.4549	1		
Lettura	0.5905	0.4768	1	
Scrittura	0.57	0.5754	0.6087	1
CORRELAZIONE INGLESE – TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
	0.4792	0.4016	0.4411	0.4316

Figura 1.7 - Correlazione Inglese-Tedesco tra i punteggi medi di classe nelle quattro abilità - A2 Secondaria I grado.



Per quanto riguarda la distribuzione dei punteggi medi delle classi emerge come vi siano per l'inglese alcune classi con punteggi decisamente inferiori alla media generale, mentre nella prova di tedesco le classi con punteggio inferiore alla media pur essendo in numero maggiore mostrano divari di minore entità; le classi che invece di collocano al di sopra della media si distribuiscono in modo numericamente simile tra inglese e tedesco (cfr. Figura 1.8).

Figura 1.8 – Distribuzione variazione punteggio medio a livello classe rispetto alla media e intervalli di confidenza – A2 Secondaria I grado.



1.1.3 B1 – Il Secondaria di secondo grado

Con la numerosità del campione del TLT 2018 individuata in corrispondenza del livello B1, è possibile effettuare un'analisi più dettagliata dei singoli indirizzi in modo da provare a identificare alcuni pattern. Ciò che emerge dall'analisi dei risultati mostrati in Tabella 1.5 è un sostanziale primato dell'indirizzo linguistico in tutte e quattro le abilità e per entrambe le lingue. Solo nel caso della produzione scritta il liceo linguistico viene avvicinato in termini di esiti dall'indirizzo classico, sia per l'inglese che per il tedesco.

Tabella 1.5 - Studenti al di sopra della soglia di abilità minima (60%) per indirizzo - B1 Secondaria II grado (valori percentuali %).

	INGLESE			
INDIRIZZO	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Classico	97,5	61,3	70,0	98,8
Scientifico	93,4	29,8	56,2	57,4
Linguistico	100,0	78,5	86,9	99,1
Scienze Umane	77,7	14,4	23,0	64,7
Artistico	64,6	14,6	16,7	37,5
Economico	93,1	38,6	43,4	76,2
Tecnologico	74,1	37,5	30,6	55,1
Professionale	28,9	5,3	0,0	10,5

	TEDESCO			
INDIRIZZO	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Classico	83,8	61,3	63,8	93,8
Scientifico	79,6	39,6	52,5	51,7
Linguistico	93,5	78,5	80,4	96,3
Scienze Umane	72,1	26,9	39,4	38,5
Artistico	50,0	14,0	24,0	24,0
Economico	82,0	48,5	60,8	63,4
Tecnologico	53,2	5,1	32,4	5,1
Professionale	30,0	0,0	3,3	3,3

Il liceo scientifico ottiene buoni risultati in corrispondenza delle due abilità ricettive (ascolto e lettura), e minori nelle abilità produttive (parlato e scrittura). Nella lingua inglese (cfr. Figura 1.9), i tecnici si confermano a livelli sostanzialmente simili e in alcuni casi superiori agli indirizzi scienze umane e artistico. Con riferimento al tedesco, il tecnico tecnologico soffre maggiormente rispetto a quanto rilevato nella prova di inglese. Buoni appaiono invece i risultati nel caso dell'indirizzo economico, i cui risultati si pongono su livelli simili o superiori a quelli dell'indirizzo scientifico – nello specifico nella produzione scritta in inglese e in modo trasversale per la lingua tedesca (cfr. Figura 1.10).

Figura 1.9 – Percentuale di studenti sopra la soglia del 60% in inglese per abilità e indirizzi – B1 Secondaria II grado.

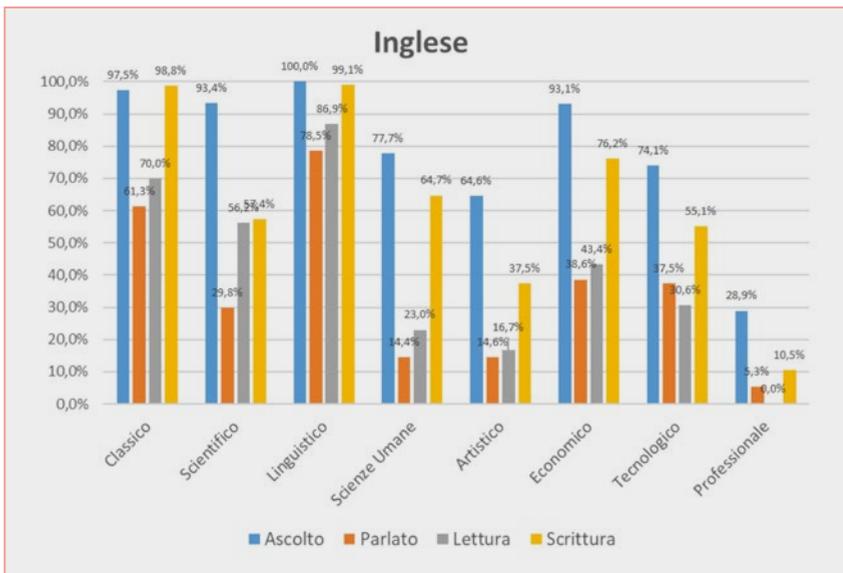
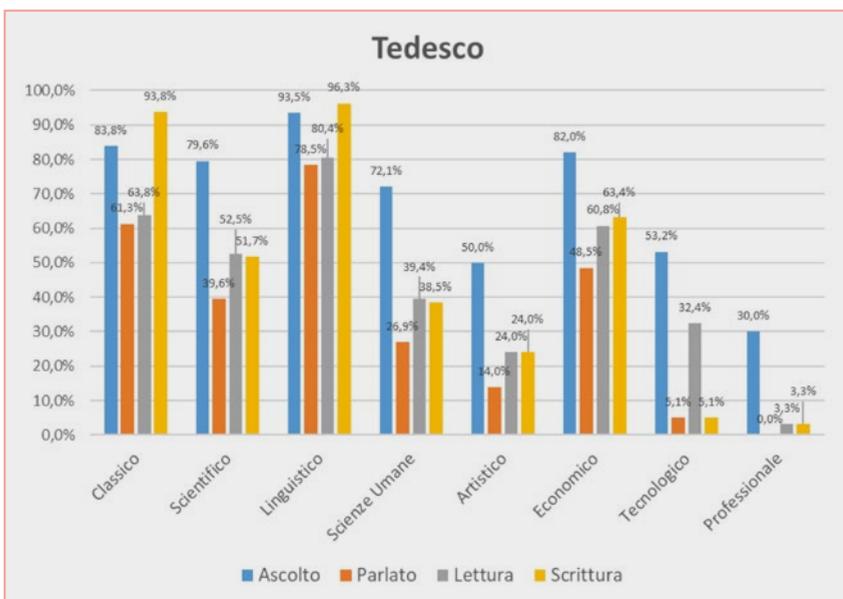


Figura 1.10 – Percentuale di studenti sopra la soglia del 60% in tedesco per abilità e indirizzi – B1 Secondaria II grado.



Come per il livello A2, i valori di correlazione tra i punteggi individuali degli studenti di II Secondaria di secondo grado risultano essere più forti in tedesco che in inglese (cfr. Tabella 1.6). Segnale che anche gli studenti della secondaria di secondo grado sono sottoposti per il tedesco a una didattica più omogenea che non

favorisce un'abilità rispetto a un'altra. Quello che risulta interessante è notare come la correlazione tra lingue diverse per la stessa abilità registri coefficienti più forti rispetto al livello A2 e ancor più rispetto al livello A1. Questo indica che, al crescere degli anni di apprendimento delle due lingue, aumenta anche l'omogeneità tra inglese e tedesco nelle varie competenze. Va precisato che anche se i coefficienti di correlazione osservati indicano legami moderati, appare comunque utile indicare il trend osservato tra i vari livelli.

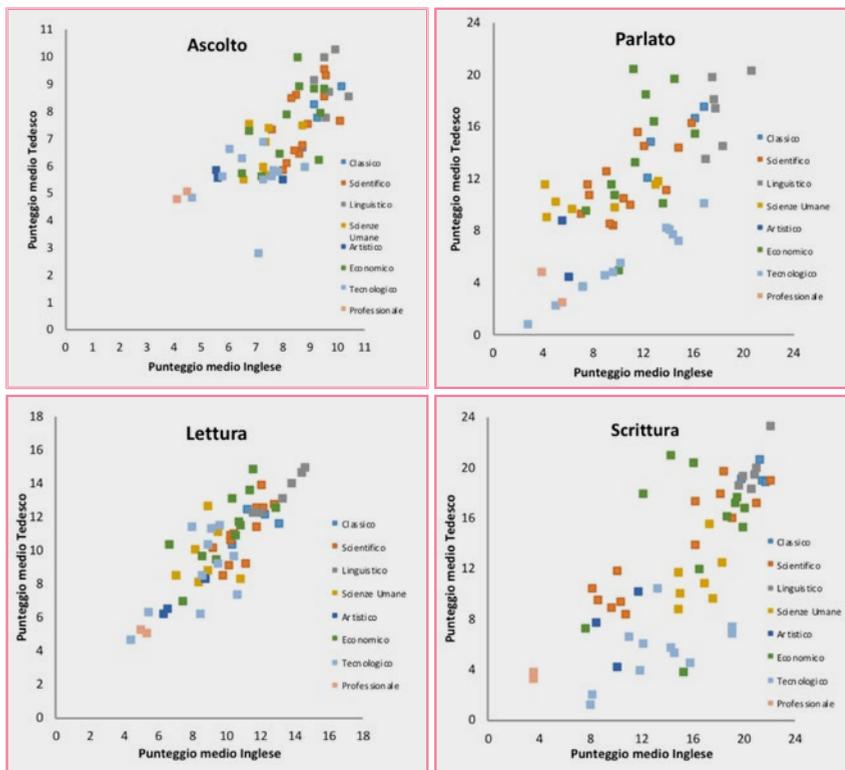
Tabella 1.6 - Correlazione tra punteggi individuali nelle quattro competenze - B1 Secondaria II grado.

INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Ascolto	1			
Parlato	0.4192	1		
Lettura	0.6017	0.4225	1	
Scrittura	0.5324	0.463	0.5156	1
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Ascolto	1			
Parlato	0.4912	1		
Lettura	0.5494	0.4784	1	
Scrittura	0.5321	0.6352	0.5926	1
CORRELAZIONE INGLESE – TEDESCO				
	Ascolto	Parlato	Lettura	Scrittura
	0.4488	0.4346	0.5197	0.5398

Per quanto riguarda invece la correlazione tra i punteggi medi delle classi nelle due lingue si ripresenta una maggiore eterogeneità nei punteggi per le abilità produttive rispetto a quelle ricettive. Osservando infatti la Figura 1.11 si può notare come ad esempio per l'ascolto le varie classi si collochino su di una linea immaginaria obliqua che va dall'angolo basso a sinistra verso l'alto a destra (equivalente a una correlazione fortemente positiva e prossima al valore 1), il che indica una significativa omogeneità tra classi nella padronanza delle singole abilità nelle due lingue. Nello specifico, sono gli indirizzi del classico e in particolare del linguistico che mostrano una correlazione positiva trasversalmente a tutte le abilità. Sintomo questo di uno sviluppo omogeneo della

abilità nelle due lingue. Per gli altri indirizzi si rilevano comportamenti invece più differenziati ed etogenei tra le diverse classi.

Figura 1.11 – Correlazione Inglese–Tedesco tra i punteggi medi di classe nelle quattro abilità – B1 Secondaria II grado.



1.1.4 B1 – Istruzione e Formazione Professionale

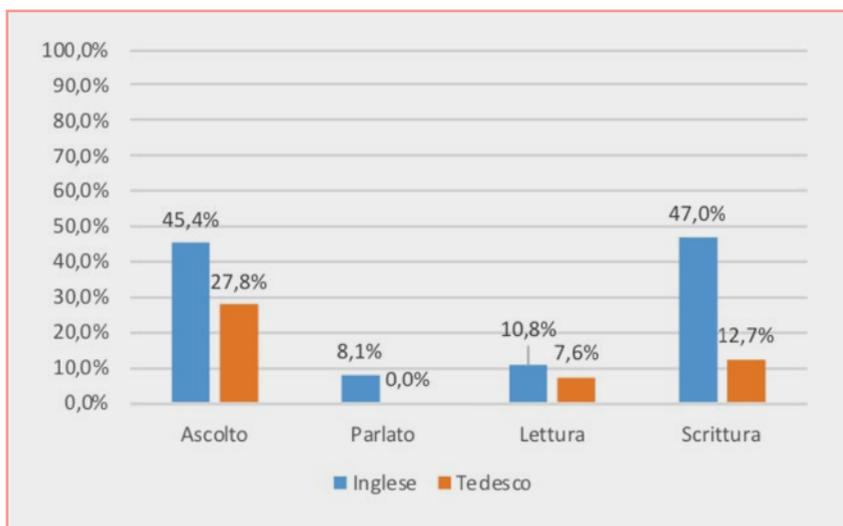
Il confronto tra lingua inglese e tedesca per il livello B1 – Istruzione e Formazione Professionale risulta meno consistente alla luce della ridotta numerosità del campione di studenti che ha svolto la prova in lingua tedesca (79 studenti). Di seguito vengono comunque presentati sinteticamente i risultati al fine di fornire un quadro completo degli esiti sulle competenze linguistiche per tutti i livelli coinvolti (cfr. Figura 1.12).

Così come per la secondaria di secondo grado, anche per l'Istruzione e Formazione Professionale i profili linguistici più solidi risultano essere quelli relativi all'ascolto e alla produzione scritta. Tuttavia, le differenze che si riscontrano tra le lingue sono più ampie di quanto rilevato per gli studenti della secondaria di secondo grado. Infatti, la percentuale di studenti che raggiunge la soglia minima

per l'abilità di ascolto in inglese (45,4%) risulta essere maggiore rispetto a quanto riscontrato per la lingua tedesca (27,8%). Per la produzione scritta il divario risulta ancora più pronunciato: gli esiti mostrano infatti che la distanza tra la proporzione di studenti in possesso dell'abilità di scrittura in inglese (47,0%) e in tedesco (12,7%) si attesta nell'ordine dei 35 punti percentuali.

Per quanto riguarda le abilità di parlato e lettura, in linea generale si evidenzia una percentuale di studenti in possesso di tali competenze di molto inferiore rispetto a quanto visto per le abilità di ascolto e scrittura. Nello specifico, la lettura si posiziona su livelli simili sia per l'inglese (10,8%) che per il tedesco (7,6%); mentre gli esiti per la produzione orale mostrano come, a fronte di un 8,1% di studenti in possesso della competenza per l'inglese, nessuno studente risulti in possesso della soglia minima di competenza per la lingua tedesca.

Figura 1.12 – Percentuale di studenti al di sopra della soglia di abilità minima (60%) per lingua e abilità linguistica – B1 IeFP.



1.2 Approfondimento qualitativo sulla produzione orale e scritta⁹

Di seguito si riportano gli esiti dei risultati emersi dallo sviluppo di un sistema di valutazione automatica (cfr. Capitolo 6).

⁹ Questa parte è stata sviluppata utilizzando i risultati raccolti da FBK nell'ambito dello sviluppo di un sistema automatico per la valutazione delle competenze (cfr. Capitolo 6).

L'elaborazione dei dati ha consentito di fornire alcune statistiche "qualitative" relative alla produzione orale e scritta del test 2018. L'obiettivo di questa analisi è presentare alcuni dati complementari rispetto a quelli già presentati in precedenza che possano contribuire a fornire un quadro più completo dei profili linguistici degli studenti trentini.

Nello specifico, i grafici riportati di seguito mostrano i risultati relativi alla durata media delle frasi registrate (cfr. Figura 1.13) e il numero medio di parole che compongono le frasi (cfr. Figura 1.14). Dall'analisi delle prove di parlato emerge chiaramente come al crescere del livello, cresca anche la capacità degli studenti di comporre e pronunciare frasi di durata maggiore. Infatti, se per il livello A1 la durata media di una frase è di circa due secondi, per il livello A2 il valore medio aumenta a circa 9-10 secondi per raggiungere poi i 12 secondi per il livello B1 – Secondaria di secondo grado in inglese e quasi 15 invece per il tedesco. Sulla stessa linea si collocano i risultati relativi al numero medio di parole che compongono le frasi pronunciate. In linea generale emerge come per il parlato un secondo di registrazione corrisponda mediamente alla pronuncia di circa due parole. Per il livello A1 quindi, il numero medio di parole pronunciate per frase si attesta su un valore di circa 4, mentre risulta maggiore per il livello A2 (circa 18) e il livello B1 (circa 27).

Figura 1.13 – Durata media delle frasi in secondi suddivise per livello e lingua – Parlato.

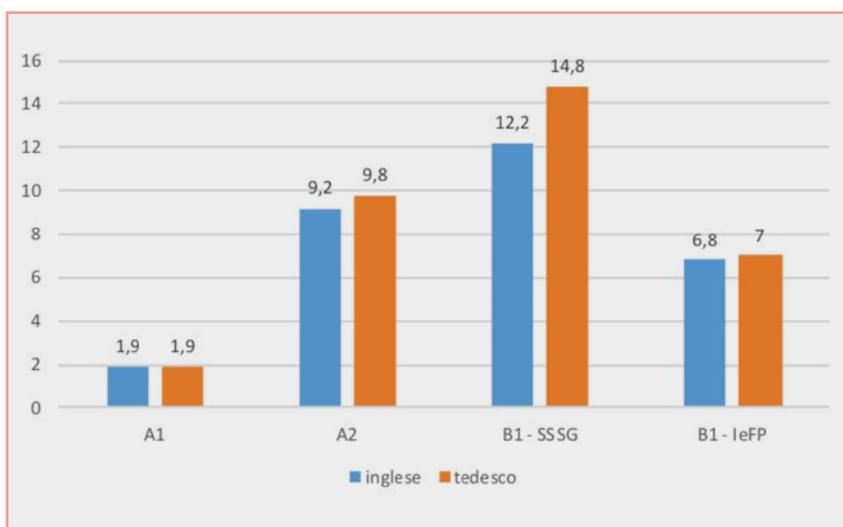
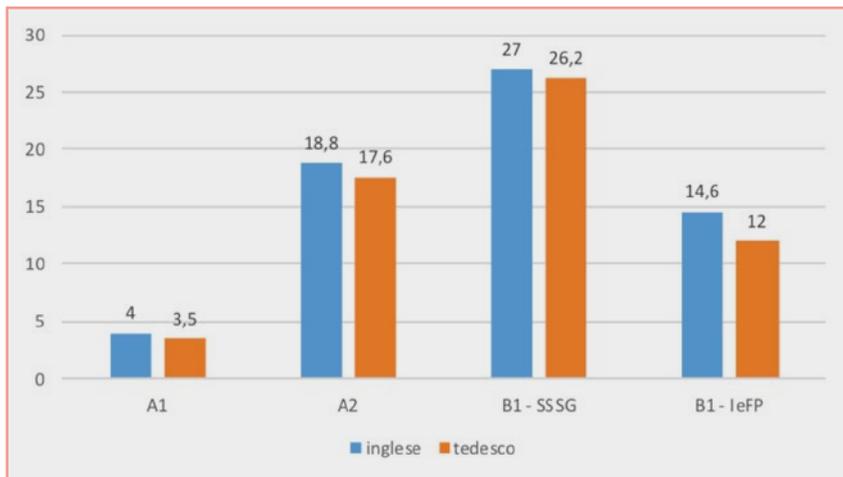
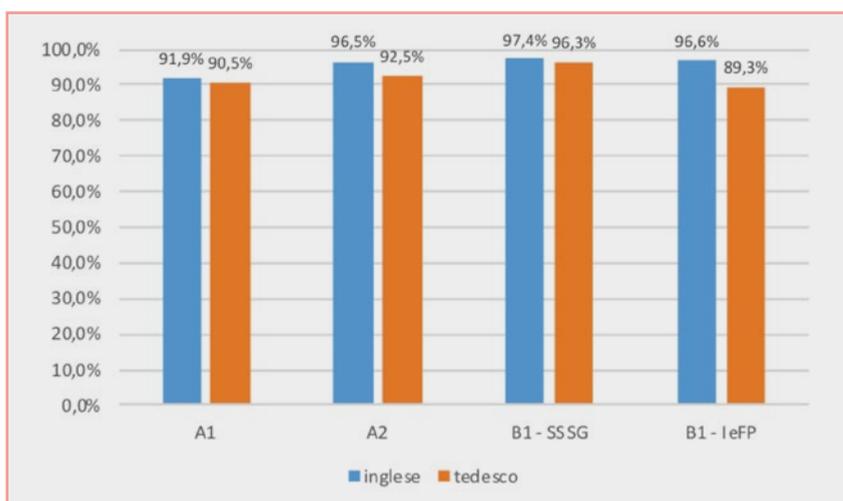


Figura 1.14 - Numero medio di parole per frase suddivise per livello e lingua - Parlato.



Di seguito vengono riportati i dati relativi alla percentuale di parole corrette nelle prove di produzione scritta, considerando come errori sia le parole scritte sbagliate sia le parole scritte in una lingua diversa da quella attesa. Il maggior numero di anni di esposizione alle lingue sembra influenzare positivamente la percentuale di parole scritte correttamente per i vari livelli. Infatti, la Figura 1.15 mostra che per il livello A1 la percentuale di parole corrette si attesta intorno al 91-92% in entrambe le lingue, mentre aumenta a circa il 97% se si considera il livello B1-SSSG. I risultati delle analisi mostrano anche che il livello B1-leFP è quello nel quale la differenza tra le due lingue risulta più ampia: 96,6% per l'inglese e 89,3% per il tedesco.

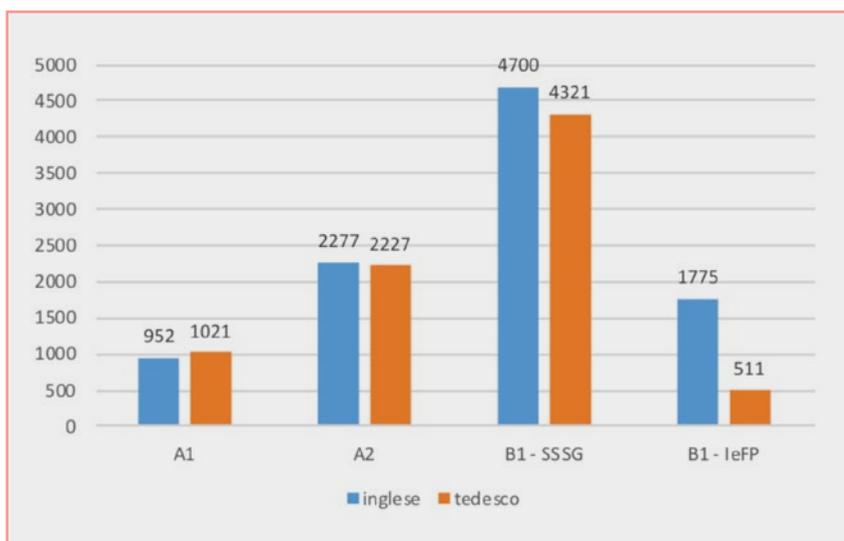
Figura 1.15 – Percentuale di parole scritte correttamente per livello e lingua.



Un parametro interessante da esaminare in un'ottica di confronto tra i vari livelli di una lingua è il *lessico* (o vocabolario) che solitamente viene utilizzato per indicare il complesso di parole e delle locuzioni di una lingua. In questa sede, il lessico misura la quantità di parole differenti che sono state utilizzate nelle abilità di produzione scritta e orale per le due lingue. Nei grafici che seguono questo indicatore viene presentato separatamente per lo scritto e per il parlato suddiviso per i vari livelli e lingue.

La Figura 1.16 nello specifico mostra come il lessico della produzione orale tenda ad aumentare al crescere degli anni di apprendimento delle due lingue. Se per il livello A1 (circa 1000 parole) e A2 (circa 2000) emerge un sostanziale equilibrio tra le due lingue, il livello B1 mostra invece delle differenze rilevanti. Per la secondaria di secondo grado il lessico in inglese risulta moderatamente maggiore rispetto al tedesco – rispettivamente 4700 e 4321 parole – mentre per l'Istruzione e Formazione Professionale il divario risulta molto pronunciato sia tra le due lingue ma anche nel confronto con la secondaria di secondo grado.

Figura 1.16 – Numero di parole differenti utilizzate per livello e per lingua nella produzione orale.

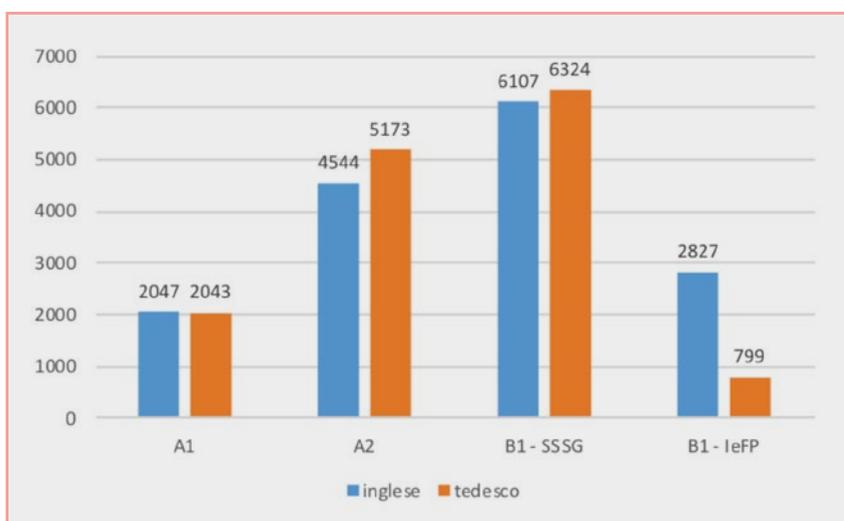


In linea generale emerge come il lessico per la produzione scritta (cfr. Figura 1.17) risulti mediamente superiore rispetto al lessico del parlato, fino al doppio per il livello A1 e A2. Questo può essere spiegato dal fatto che in una lingua il numero di parole utilizzate nello scritto risulta in genere molto più ampio rispetto

al parlato in quanto raramente sono presenti ripetizioni che sono spesso sostituite da sinonimi o perifrasi.

Inoltre, un'altra distinzione che emerge rispetto ai dati sul parlato è che la scrittura in tedesco mostra un lessico maggiore sia per il livello A2 che per il livello B1-SSSG, mentre per il livello A1 si nota un sostanziale equilibrio tra le due lingue. Per il livello B1-leFP si ripropone il trend visto per l'analisi del parlato: il lessico in inglese risulta decisamente maggiore rispetto al lessico in tedesco ma di molto inferiore rispetto a quanto riscontrato per la secondaria di secondo grado.

Figura 1.17 – Numero di parole differenti utilizzate per livello e per lingua nella produzione scritta.



Infine, le tabelle A1.1, A1.2, A1.3 e A1.4 in appendice riportano le 20 parole più frequenti suddivise nei diversi casi (scritto vs. parlato, inglese vs. tedesco) e per livello di proficiency. Per ogni parola sono riportati il numero di occorrenze e la percentuale sul totale di parole utilizzate per livello.

1.3 L'analisi della varianza negli esiti

Alla luce dei risultati fin qui presentati, diventa fondamentale approfondire in quale misura i profili linguistici degli studenti varino tra scuole, tra classi di una stessa scuola e tra studenti in ciascuna classe. In altre parole, l'analisi della varianza presen-

tata di seguito ci consente di scomporre la variabilità totale dei risultati nelle varie competenze linguistiche secondo tre parametri: i) varianza between istituti – varianza tra le scuole; ii) varianza between classi – varianza tra classi; e iii) varianza within classi – varianza all’interno di ciascuna classe tra studenti.

Ciò che emerge da questa analisi (Tabella 1.7) è che in V Primaria si possono osservare due pattern distinti rispettivamente per le competenze di ascolto e lettura da una parte, e le abilità di produzione orale e scritta dall’altro. Nelle prime due una quota molto consistente della variabilità totale, tra l’84% e il 95% a seconda della lingua e della competenza, è ascrivibile a differenze degli studenti all’interno delle classi (quindi a caratteristiche “individuali”); per le altre due competenze invece la quota complessiva di variabilità spiegata da quale scuola o classe si frequenta incide per una percentuale che oscilla tra il 41% e il 29%. Questo indica che per le abilità produttive frequentare una determinata scuola/classe può equivalere a differenze rilevanti in termini di raggiungimento delle abilità di produzione orale e scritta.

Tabella 1.7 - Scomposizione della varianza nei punteggi - A1 Primaria (valori percentuali %).

INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	14.4	34.1	7.0	34.5
varianza between classi	1.5	7.0	6.6	0.0
varianza within classi	84.0	58.9	86.5	65.5
varianza complessiva	100.0	100.0	100.0	100.0
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	3.8	21.1	2.4	16.6
varianza between classi	1.6	19.2	11.0	12.2
varianza within classi	94.6	59.7	86.6	71.2
varianza complessiva	100.0	100.0	100.0	100.0

Le differenze osservate tra le competenze ascolto e lettura da una parte e scritto e parlato dall'altro si ripropongono in maniera del tutto simile nelle prove in III secondaria I grado. Emergono in particolare l'assenza totale di variabilità tra le scuole nelle prove in inglese di ascolto e lettura (0% e 0%) da una parte e l'importanza di frequentare una scuola piuttosto che un'altra per la prova in inglese di scrittura (36% della variabilità totale spiegata da differenze tra scuole).

Tabella 1.8 - Scomposizione della varianza nei punteggi - A2 Secondaria I grado (valori percentuali %).

INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	0.0	26.5	0.0	36.3
varianza between classi	6.9	7.4	9.5	8.6
varianza within classi	93.1	66.1	90.5	55.1
varianza complessiva	100.0	100.0	100.0	100.0
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	4.6	18.8	7.2	13.6
varianza between classi	5.6	9.3	7.7	9.8
varianza within classi	89.8	71.9	85.1	76.6
varianza complessiva	100.0	100.0	100.0	100.0

La scomposizione della varianza dei risultati per il livello B1 è stata svolta distinguendo tra gli studenti dei licei e quelli dei tecnici; nonostante questo accorgimento ci si aspetta comunque a priori una quota significativa della variabilità spiegata da differenze tra le diverse scuole, questo per la differenziazione dei curricula esistente nelle macro-categorie di licei e tecnici e per la diversa composizione socioculturale degli studenti presente nei diversi indirizzi di studio.

Dalla lettura complessiva dei dati per i licei emerge come lo scritto, sia in inglese che in tedesco, risulti la prova nella quale le differenze tra classi contano più delle differenze tra le scuole, oltre che la prova nella quale la variabilità totale spiegata da differenze tra studenti all'interno delle classi raggiunge appena il 50%.

I risultati per i tecnici invece suggeriscono come l'insegnamento dell'inglese e del tedesco abbiano un diverso "trattamento" nelle diverse scuole, dal momento che per l'inglese il grosso della variabilità osservata dipende da differenze all'interno delle classi mentre per il tedesco differenze tra scuole o tra classi nelle scuole contano molto, fino o oltre il 50% nel caso delle prove di parlato e scritto.

Tabella 1.9 - Scomposizione della varianza nei punteggi - B1 Licei (valori percentuali %).

INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	21.4	19.0	17.0	16.4
varianza between classi	8.9	18.5	19.8	33.5
varianza within classi	69.7	62.5	63.2	50.1
varianza complessiva	100.0	100.0	100.0	100.0
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	20.4	17.5	10.6	18.9
varianza between classi	6.9	10.4	21.3	28.6
varianza within classi	72.7	72.1	68.2	52.5
varianza complessiva	100.0	100.0	100.0	100.0

Tabella 1.10 - Scomposizione della varianza nei punteggi - B1 Tecnici (valori percentuali %).

INGLESE				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	6.1	14.3	0.0	13.5
varianza between classi	18.7	8.2	31.7	20.8
varianza within classi	75.2	77.6	68.3	65.7
varianza complessiva	100.0	100.0	100.0	100.0
TEDESCO				
	ASCOLTO	PARLATO	LETTURA	SCRITTURA
varianza between istituti	13.4	22.6	5.5	22.7
varianza between classi	23.5	25.6	31.3	36.8
varianza within classi	63.1	51.8	63.2	40.5
varianza complessiva	100.0	100.0	100.0	100.0

Riassumendo, i risultati a livello di classi V della primaria mostrano una discreta eterogeneità nei risultati medi tra scuole per due delle quattro competenze misurate: nel parlato ma soprattutto nello scritto vi sono Istituti Comprensivi con risultati molto buoni e altri con risultati decisamente più bassi. A livello di classi III della Secondaria di primo grado si ha una maggiore omogeneità dei risultati complessivi a livello scuola o classe mentre si osserva una fortissima eterogeneità tra i risultati di ragazzi di una stessa classe; l'unica eccezione sembra essere nuovamente la prova di scrittura nella quale la quota di devianza attribuibile a differenze tra scuole e tra classi nelle scuole arriva al 25-35% a seconda della prova in Inglese o Tedesco. A livello di classi II della Secondaria di secondo grado la maggior eterogeneità nei risultati tra scuole è un risultato atteso per via della natura dei diversi percorsi scolastici che i ragazzi intraprendono al momento dell'iscrizione.

L'analisi della varianza appena presentata ha restituito un quadro secondo il quale l'eterogeneità negli esiti riscontrata tra studenti all'interno delle classi è risultata molto elevata. Questo ci porta nel capitolo successivo a indagare le caratteristiche individuali degli studenti associate al possesso delle varie abilità linguistiche al fine di individuare quali siano le determinanti individuali delle competenze linguistiche.

2 | Le determinanti della performance linguistica degli studenti trentini

Mattia Oliviero¹⁰

Dopo aver presentato gli esiti della rilevazione sulle competenze linguistiche (cfr. Capitolo 1), l'obiettivo di questo capitolo è di esaminare che tipo di relazione esiste tra le caratteristiche individuali degli studenti – come l'età, il genere, la carriera scolastica o il background familiare – e la probabilità di “possedere” le varie abilità linguistiche. L'analisi presentata di seguito è stata condotta utilizzando modelli di regressione binaria¹¹ per ognuna delle quattro competenze misurate durante la rilevazione suddivise per livello e lingua. Nello specifico, la variabile dipendente si presenta come una *dummy*, vale a dire che assume un valore uguale a 0 nel caso in cui lo studente non abbia ottenuto il 60% del punteggio totale nelle singole abilità e uguale a 1 se invece ha raggiunto o superato tale soglia.

In appendice vengono presentate le statistiche descrittive delle caratteristiche individuali utilizzate come variabili esplicative nel modello (cfr. Tabelle A2.1, A2.2, A2.3 e A2.4) e gli output dei modelli di regressione binaria per i tre livelli di indagine differenziati per lingua (cfr. Tabelle A2.5, A2.6 e A2.7). Per ogni abilità sono state previste due specificazioni del modello. La differenza tra le due è che nella seconda è stata aggiunta l'informazione relativa alla scuola frequentata dagli studenti – chiamata anche *effetti fissi* della scuola. In termini pratici, l'inserimento degli *effetti fissi* della scuola consente di controllare per tutte quelle caratteristiche a livello di scuola che influenzano l'apprendimento delle lingue ma che non sono direttamente misurabili.

¹⁰ Mattia Oliviero è collaboratore di ricerca presso Iprase.

¹¹ Il modello utilizzato è denominato *probit*, modello di regressione non lineare utilizzato quando la variabile dipendente è binaria (con valori pari a 0 o 1).

I coefficienti sono presentati sotto forma di effetti marginali, i quali indicano l'incremento o la diminuzione della probabilità di possedere la relativa competenza associata alla variazione di una certa caratteristica, come ad esempio l'essere una ragazza rispetto all'essere ragazzo. Due sono i tipi di variabile che si possono incontrare nei modelli e le rispettive possibili interpretazioni. Nel caso in cui si consideri una variabile *continua* – come il numero di ore di insegnamento in CLIL – l'effetto marginale va interpretato come la variazione percentuale (positiva o negativa) nella probabilità di possedere o meno una competenza che si verifica a un aumento *unitario* della variabile considerata, al netto di tutti gli altri fattori considerati. Nel caso invece in cui si prenda in considerazione una variabile *dicotomica* o *categoriale*, come ad esempio il genere o il titolo di studio dei genitori, l'effetto marginale corrisponde alla variazione percentuale nel passare dalla categoria di riferimento alla categoria considerata. Inoltre, la significatività statistica – che misura la precisione e l'accuratezza dell'effetto e quindi la sua attendibilità – è indicata dagli asterischi posti a fianco all'effetto marginale. Maggiore è il numero di asterischi, maggiore sarà anche l'attendibilità statistica dell'effetto¹². Nel caso di nessun asterisco non vi è un'attendibilità statistica del valore.

Al fine di rendere più immediata l'interpretazione, i risultati delle stime vengono presentati nel testo in formato grafico per lingua e per livello di indagine. A fianco a ogni variabile esplicativa la presenza del segno + o del segno – indica una maggiore o minore propensione a possedere la relativa competenza. L'assenza di segno indica che la variazione della specifica variabile esplicativa non è associata a differenze statisticamente significative in relazione alla probabilità di possedere una determinata competenza linguistica.

2.1 A1 – Primaria

Le Tabelle 2.1 e 2.2 mostrano i risultati relativi al livello A1 – V Primaria rispettivamente per l'inglese e per il tedesco. La prima caratteristica presa in considerazione è il genere dello studente. Da uno sguardo d'insieme emerge come il fatto di essere una ragazza sia associato a una maggiore propensione a possedere

¹² Nello specifico: tre asterischi *** significatività all'1%, due asterischi ** significatività al 5%, un asterisco * significatività al 10%.

una determinata competenza linguistica solamente per la lingua tedesca. Nello specifico infatti, l'essere ragazza è associato ad una maggiore propensione a possedere le competenze della lettura e della scrittura (rispettivamente +11,7%*** e +11,8%***) e in misura minore per quanto riguarda la competenza dell'ascolto (+7,5%*). Emerge invece che alla competenza del parlato tedesco non vi sia nessuna differenza statisticamente significativa tra ragazze e ragazzi.

A livello di carriera scolastica, gli studenti in ritardo di un anno non si comportano in maniera statisticamente differente rispetto agli studenti in regola con gli studi. È tuttavia necessario specificare che a livello di V Primaria il fenomeno dei ritardi è molto limitato in quanto solamente il 2% del campione si trova in questa condizione; oltretutto in buona parte si tratta di studenti stranieri di 1^a generazione (studenti stranieri arrivati in Italia in età scolare, inseriti in classi non corrispondenti alla loro età anagrafica, caratteristica che nel modello viene controllata).

Anche la provenienza degli studenti sembra incidere in modo marginale, almeno a livello della V Primaria. Essere studenti stranieri di 2^a generazione (studenti stranieri nati in Italia) o di 1^o generazione rispetto all'esser studenti italiani è debolmente associato a un effetto positivo o negativo in buona parte delle specificazioni dei modelli: vi è un lieve vantaggio degli studenti stranieri di 2^a generazione rispetto agli italiani nella prova di scrittura in inglese (+19,6%***) e nel parlato di tedesco (+23,4%***) mentre gli studenti stranieri di 1^a generazione mostrano comportamenti del tutto simili agli italiani.

Il background socioeconomico e culturale della famiglia è generalmente tenuto in considerazione includendo nel modello il titolo di studio "più alto" posseduto dai genitori. Tuttavia, conoscere quale titolo di studio possiedono i genitori è per i ragazzi di quest'età un'informazione ancora vaga, come dimostrato dall'alta proporzione di "non so" (vicina al 60%) alle due domande del questionario. Per ovviare a questa problematica, nel questionario era stata preventivamente predisposta una domanda relativa alla quantità di libri presenti in casa¹³. Questa variabile viene spesso utilizzato come indicatore del contesto socioeconomico e culturale dei ragazzi in questa fascia di età.

¹³ La domanda è: "Quanti libri ci sono all'incirca a casa tua?". Le modalità di risposta vanno da "nessuno o pochi" (0-10 libri) a "Abbastanza da riempire tre o più scaffali" (più di 200 libri). Per facilitare la risposta a fianco di ciascuna modalità di risposta sono stati inseriti dei disegni ad indicare visivamente la quantità come riferimento.

Tabella 2.1 - INGLESE: probabilità di possedere ciascuna competenza - A1 Primaria.

VARIABILI ESPLICATIVE	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere (rif. Maschio)				
Femmina				
Anni di ritardo (rif. In regola)				
1 o più anni di ritardo				
Origine (rif. Italiano)				
Straniero di 2^a generazione				+
Straniero di 1^a generazione				
Libri a casa (rif. >200 libri)				
0–10 libri	-	-	-	-
11–25 libri	-		-	-
26–100 libri	-			
101–200 libri				
Ore settimanali CLIL	+	+	+	
Ordine delle prove (rif. 1 ^a)				
2^a				
Prova “speciale” (rif. No)				
Sì	-	-		-

Gli effetti della variabile con informazioni sul numero di libri a casa risultano coerenti e più efficaci rispetto all’inserimento del titolo di studio dei genitori. In particolare emerge lo “svantaggio” degli studenti che riportano di possedere nessuno o pochi libri a casa rispetto a coloro che riportano di averne più di 200 (categoria di riferimento) in termini di probabilità di possedere una determinata competenza. Questo risultato si presenta con caratteristiche trasversali sia tra competenze sia per quanto riguarda le lingue e può essere quantificato in un effetto negativo tra il -23%** per il parlato in inglese e il -34%*** per la scrittura in tedesco. Al crescere del numero di libri posseduti a casa, diminuisce anche la differenza rispetto alla categoria di riferimento (coloro che a casa posseggono più di 200 libri) fino ad azzerarsi per coloro che riportano di possedere tra i 101 e 200 libri. In linea generale, le competenze dove le differenze in termini di background socioeconomico e culturale sono maggiormente rilevanti e significative risultano essere l’ascolto per l’inglese e il parlato per il tedesco; mentre il background socioculturale dei genitori sembra avere

meno influenza nella competenza del parlato in inglese e della lettura in tedesco.

Nei modelli è stata anche inclusa una variabile relativa al numero di ore di didattica CLIL svolte alla settimana. Tuttavia, le risposte fornite dagli studenti risultano poco coerenti tra di loro, dal momento che per moltissime classi nel campione degli studenti all'interno della stessa sezione si sono registrate risposte molto diverse alla domanda del questionario. Per questo motivo si è optato per utilizzare il valore modale – valore che si ripete con maggiore frequenza – di queste risposte a livello di classe. Dalle stime sembra esserci una relazione positiva tra il numero di ore svolte in CLIL e la probabilità di possedere una determinata competenza. Nello specifico, l'associazione risulta positiva in lingua inglese per l'ascolto (+9,3%*** per ogni ora in più alla settimana), il parlato (+14,0**) e la lettura (7,4%*); mentre per il tedesco risulta statisticamente significativa solo per l'ascolto (+ 3,5%***).

È necessario specificare che l'intento di questa indagine non è di identificare un nesso causale tra lo svolgimento della didattica CLIL e il possesso delle varie competenze linguistiche. Il focus di questo capitolo sono le performance linguistiche, e la variabile sulla didattica in CLIL viene considerata in termini di associazione tra il numero di ore svolte e la probabilità di possedere una specifica abilità linguistica senza ambizione di identificare un legame di *causa-effetto*.

Infine, nel modello sono state incluse informazioni anche sull'ordine di svolgimento delle prove (inglese-tedesco o tedesco-inglese, senza stime statisticamente significative) e sugli studenti che hanno svolto i test con strumenti compensativi. Per questi ultimi i risultati mostrano una probabilità inferiore rispetto agli studenti senza bisogni educativi speciali di possedere tutte le competenze sia in inglese che in tedesco (a eccezione della lettura inglese la quale pur presentando un effetto negativo questo non risulta statisticamente significativo) con effetti particolarmente elevati per quanto riguarda la produzione scritta (-34,0%*** per l'inglese e -36,7% per il tedesco).

Prima di presentare i risultati per la III Secondaria di primo grado si segnala come l'indice R^2 (il quale è in sintesi una misura di quanto le variabili utilizzate in un modello sono in grado di "spiegare" la variabilità della variabile dipendente) cresca molto nel passaggio dalla specificazione senza a quella con gli effetti fissi scuola per le prove di parlato e scritto (cfr. Tabella A2.5 in appendice), a prova del fatto che per queste due competenze la scuola frequentata conti molto e forse più delle caratteristiche degli studenti, indizio del fatto che la didattica sia molto eterogenea sul territorio.

Tabella 2.2 - TEDESCO: probabilità di possedere ciascuna competenza - A1 Primaria.

VARIABILI ESPLICATIVE	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere (rif. Maschio)				
Femmina	+		+	+
Anni di ritardo (rif. In regola)				
1 o più anni di ritardo				
Origine (rif. Italiano)				
Straniero di 2^a generazione		+		
Straniero di 1^a generazione				
Libri a casa (rif. >200 libri)				
0–10 libri	–	–	–	–
11–25 libri	–	–		–
26–100 libri		–		
101–200 libri				
Ore settimanali CLIL	+			
Ordine delle prove (rif. 1 ^a)				
2^a				
Prova “speciale” (rif. No)				
Si	–	–	–	–

2.2 A2 – III Secondaria di primo grado

L'analisi delle determinanti delle performance linguistiche all'ultimo anno della scuola secondaria di primo grado (cfr. Tabelle 2.3 e 2.4) presenta delle similitudini con gli esiti relativi alla scuola primaria. Per quanto riguarda il divario di genere, i risultati delle stime presentano un quadro perfettamente sovrapponibile a quello della scuola primaria. Se da un lato infatti per le competenze in lingua inglese non si rintracciano differenze significative a eccezione della produzione scritta (+9,6%** per le ragazze rispetto ai ragazzi), dall'altro l'essere una ragazza è associato a una maggiore probabilità di possedere le competenze del parlato (+8,1%**) e della scrittura (+16,7%***) in tedesco.

Così come per la V Primaria, anche per la III Secondaria di primo grado la regolarità degli studi non sembra presentare un vantaggio rispetto agli studenti che hanno accumulato un ritardo scolastico. Inoltre, l'origine degli studenti non costituisce né un vantaggio né uno svantaggio per la propensione a possedere

le abilità linguistiche in inglese e tedesco, mostrando da questo punto di vista un profilo di maggiore equità rispetto a quanto visto per la primaria.

Per quanto riguarda il background familiare, per gli studenti della III Secondaria di primo grado abbiamo applicato lo stesso ragionamento della V Primaria alla luce dell'alta percentuale di "non so" alla domanda sul titolo di studio dei genitori (circa il 17% del campione). Il quadro che emerge rispecchia l'andamento riscontrato negli studenti della V Primaria. Nello specifico, possedere nessuno o pochi libri a casa è associato a una probabilità minore – variabile da -23,8%*** del parlato in inglese a -33,7%*** della scrittura in tedesco – rispetto a coloro che hanno riportato di avere a casa più di 200 libri. L'effetto del background socioeconomico e culturale non influenza tutte le competenze ed entrambe le lingue in modo lineare ma si possono comunque ricavare indicazioni abbastanza chiare sull'importanza del ruolo del background familiare da cui provengono i ragazzi per le loro performance linguistiche.

Rispetto alle stime per la V Primaria, per la III Secondaria di primo grado si è utilizzata una dummy 0/1 a seconda che la classe non abbia svolto/abbia svolto didattica CLIL durante l'anno al posto del numero di ore di CLIL alla settimana; questa scelta è dovuta al fatto che nel campione di studenti e di classi coloro che hanno svolto didattica CLIL in inglese e/o tedesco è pari rispettivamente a circa il 5% e il 10%, mentre per la V primaria era la totalità. Con questi valori è sufficientemente informativo per la stima distinguere tra coloro che hanno o non hanno avuto questa esperienza scolastica. Aver svolto didattica CLIL in inglese dà una maggiore probabilità di possedere le competenze di ascolto (+32,9%***) e scrittura (+28,9 %***); per il tedesco invece i risultati sono poco concludenti, dal momento che gli effetti marginali non sono significativi o sono contrastanti nel momento in cui si inseriscono gli effetti fissi scuola. Tuttavia dai risultati emerge come l'abilità del parlato in tedesco sia associata negativamente allo svolgimento del CLIL (-18,5%***).

Anche in questo caso è importante specificare che il focus di questa analisi è di esaminare l'effetto delle determinanti sulle performance linguistiche e non quello di stimare un effetto causale tra didattica CLIL e profili linguistici. In questa sede infatti ci limitiamo a rilevare il grado di associazione tra coloro che hanno svolto didattica CLIL e coloro che non l'hanno svolta senza nessuna pretesa di individuare un nesso di *causa-effetto*.

Gli studenti con bisogni educativi speciali (BES) che hanno svolto la prova con strumenti compensativi mostrano una minor

propensione al superamento della soglia minima di abilità per tutte le abilità a eccezione della produzione orale in inglese nella quale non si rilevano differenze statisticamente significative.

Nel passaggio dalla specificazione senza a quella con gli effetti fissi l' R^2 del modello cresce molto nel caso del parlato/inglese e tedesco e scrittura/inglese, segno di come l'insegnamento della lingua inglese per queste competenze sia variegato e spieghi una parte molto consistente delle differenze tra studenti rilevate nei test.

Tabella 2.3 - INGLESE: probabilità di possedere ciascuna competenza - A2 Secondaria I grado.

VARIABILI ESPLICATIVE	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere (rif. Maschio)				
Femmina				+
Anni di ritardo (rif. In regola)				
1 o più anni di ritardo				
Origine (rif. Italiano)				
Straniero di 2^a generazione				
Straniero di 1^a generazione				
Libri a casa (rif. >200 libri)				
0–10 libri	–	–	–	–
11–25 libri	–		–	–
26–100 libri				–
101–200 libri				
Attività CLIL in settimana (rif. No)				
Sì	+			+
Ordine delle prove (rif. 1 ^a)				
2^a				
Prova "speciale" (rif. No)				
Sì	–		–	–

Tabella 2.4 - TEDESCO: probabilità di possedere ciascuna competenza - A2 Secondaria I grado.

VARIABILI ESPLICATIVE	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere (rif. Maschio)				
Femmina		+		+
Anni di ritardo (rif. In regola)				
1 o più anni di ritardo		-		
Origine (rif. Italiano)				
Straniero di 2^a generazione				
Straniero di 1^a generazione				
Libri a casa (rif. >200 libri)				
0-10 libri	-	-	-	-
11-25 libri	-	-		-
26-100 libri				
101-200 libri				
Attività CLIL in settimana (rif. No)				
Si		-		
Ordine delle prove (rif. 1 ^a)				
2^a				-
Prova "speciale" (rif. No)				
Si	-	-	-	-

2.3 B1 – Secondaria di secondo grado

Il sovracampionamento per le secondarie di secondo grado a cui si fa riferimento nel capitolo 3, ci ha consentito di svolgere delle analisi più dettagliate sull'effetto dell'indirizzo di studi rispetto a quanto era stato possibile fare per il TLT 2016.

Per la II Secondaria di secondo grado (cfr. Tabelle 2.5 e 2.6) il genere ha un effetto più diffuso e trasversale rispetto alla primaria e alla secondaria di secondo grado. Infatti, le studentesse mostrano una maggior propensione rispetto ai ragazzi a possedere l'abilità del parlato e scrittura in inglese (+9,1%** e +6,9%*) e tutte e quattro le abilità in tedesco (+5,9%*, +11,1%***, +15,8%*** e +20,8%***). La maggiore propensione osservata nella lingua tedesca è anche la più robusta dal punto di vista della precisione nelle stime, a comprovare con molta sicurezza che per questa lingua i ragazzi soffrono di un gap importante e strutturale.

L'essere in ritardo di uno o più anni rispetto a un percorso di studi senza bocciature è un fattore "di rischio" statisticamente significativo in tutte le abilità per la lingua tedesca; mentre per l'inglese i risultati non mostrano differenze significative tra gli studenti regolari e gli studenti che hanno avuto esperienza di bocciature. Questo risultato però non sorprende dal momento che una buona parte dell'effetto di questa variabile viene incorporato dalla presenza del corso di studi frequentato: in molti casi infatti gli studenti in ritardo di uno o più anni non si distribuiscono casualmente in tutti i corsi ma sono prevalenti in alcuni.

Per misurare il background culturale della famiglia per la secondaria di secondo grado è stato utilizzato, a differenza della primaria e secondaria di primo grado, il titolo di studio "più alto" posseduto dai genitori in quanto la percentuale delle risposte "non so" è ridotto rispetto agli altri due livelli (circa l'8%). I risultati mostrano che l'effetto del background socioeconomico e culturale della famiglia ha un impatto limitato sulla probabilità di possedere o meno una competenza ma non perché non sia un fattore cruciale negli apprendimenti scolastici, quanto piuttosto per il fatto che buona parte del suo impatto è assorbito dagli effetti fissi maturità. Così come per gli studenti che hanno accumulato anni di ritardo, studenti con un diverso background culturale tendono a scegliere percorsi scolastici specifici piuttosto che altri. Un fenomeno questo di (auto)selezione all'interno dei diversi indirizzi di studio: questo fa sì che una parte dell'effetto del background culturale sugli apprendimenti venga assorbito dall'effetto fisso maturità il quale, come vedremo in seguito, risulta spesso determinante. Fatta questa opportuna premessa, emerge comunque un'evidenza parziale di uno svantaggio di ragazze e ragazzi provenienti da famiglie con un'istruzione inferiore rispetto a coloro che provengono da famiglie con genitori laureati (categoria di riferimento per il calcolo degli effetti marginali): così, ad esempio, gli studenti i cui padri e madri possiedono al più una licenza media o una qualifica professionale hanno una minore probabilità di possedere la competenza ascolto/inglese (-12,6%***), lettura/inglese (-13,5%**), parlato/tedesco (-11,1%*) e scrittura/tedesco (-16,5%***).

Come si può osservare con chiarezza nelle Tabelle 2.5 e 2.6, a parità di altre condizioni (genere, origine, titolo di studio, ecc.), studenti nei diversi indirizzi di studio mostrano profonde differenze nella propensione a possedere o meno una competenza. I coefficienti marginali mostrano la minore o maggiore probabilità di superare la soglia minima di punteggio degli studenti dei vari indirizzi di studio rispetto a coloro i quali frequentano classi a

indirizzo linguistico¹⁴, categoria che è stata scelta come riferimento, e la quasi totalità è di segno negativo e statisticamente molto significativo, cosa che ci permette una lettura dei dati più agevole e di esser ragionevolmente sicuri delle tendenze che emergono dai dati. Commentare ogni singolo effetto marginale tuttavia non è utile in una fase di lettura complessiva della situazione, possiamo però individuare delle macro-tendenze. La prima è che, con l'eccezione del classico in alcuni casi, gli studenti degli indirizzi liceali ottengono risultati peggiori dei loro compagni dei licei linguistici e secondo un gradiente negativo liceo scientifico → scienze umane → artistico. La seconda è che tra gli indirizzi tecnici quello economico sembra andare sistematicamente meglio di quello tecnologico; in molti casi gli studenti dell'economico hanno una probabilità di possedere le diverse competenze maggiore rispetto agli indirizzi liceali, con l'eccezione di linguistico e classico.

Per quanto riguarda l'aver svolto o meno ore di didattica CLIL durante l'anno i risultati delle stime mostrano pochi e sparuti effetti, spesso di segno negativo come nel caso del parlato e scrittura in inglese e del parlato/tedesco.

Tratti distintivi, rispetto ai livelli A1 e A2, vengono rilevati per coloro che hanno svolto il test con una prova speciale. Infatti per quanto riguarda la lingua inglese si evidenzia uno svantaggio solamente nell'abilità del parlato (-22,7%***) mentre non vi sono differenze significative per le altre abilità. Per la lingua tedesca invece il gap risulta più evidente e con stime robuste per la produzione orale (-28,9%***), la lettura (-24,9%***) e soprattutto la produzione scritta (-35,6%***).

¹⁴ A eccezione dell'ascolto/inglese nel quale la categoria di riferimento è rappresentata dal liceo scientifico. Questo è dovuto al fatto che la totalità degli studenti iscritti all'indirizzo linguistico ha superato la soglia del 60% e quindi tutti risultano in possesso della competenza. Questo porta il modello di regressione binaria a escludere dall'analisi le osservazioni che nelle varie categorie hanno sempre valore 0 oppure 1 (come in questo caso).

Tabella 2.5 - INGLESE: probabilità di possedere ciascuna competenza - B1 Secondaria II grado.

VARIABILI ESPLICATIVE	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere (rif. Maschio)				
Femmina		+		+
Anni di ritardo (rif. In regola)				
1 o più anni di ritardo				
Origine (rif. Italiano)				
Straniero di 2^a generazione				+
Straniero di 1^a generazione				
Titolo di studio dei genitori (rif. Laurea)				
Licenza elementare o media				
Qualifica professionale	-		-	
Diploma di maturità				
Non so	-			
Indirizzo di studi (rif. Linguistico)	*			
Classico		-	-	
Scientifico	rif.	-	-	-
Scienze Umane	-	-	-	-
Artistico	-	-	-	-
Economico		-	-	-
Tecnologico	-	-	-	-
Professionale	-	-	-	-
Attività CLIL in settimana (rif. No)				
Sì		-		-
Ordine delle prove (rif. 1 ^a)				
2^a				
Unica	-		-	
Prova "speciale" (rif. No)				
Sì		-		

Tabella 2.6 - TEDESCO: probabilità di possedere ciascuna competenza - B1 Secondaria II grado.

VARIABILI ESPLICATIVE	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere (rif. Maschio)				
Femmina	+	+	+	+
Anni di ritardo (rif. In regola)				
1 o più anni di ritardo	-	-	-	-
Origine (rif. Italiano)				
Straniero di 2^a generazione		-		
Straniero di 1^a generazione				
Titolo di studio dei genitori (rif. Laurea)				
Licenza elementare o media		-		
Qualifica professionale				-
Diploma di maturità				
Non so			-	
Indirizzo di studi (rif. Linguistico)				
Classico		-		
Scientifico	-	-	-	-
Scienze Umane	-	-	-	-
Artistico	-	-	-	-
Economico	-	-	-	-
Tecnologico	-	-	-	-
Professionale	-	-	-	-
Attività CLIL in settimana (rif. No)				
Si		-		
Ordine delle prove (rif. 1 ^a)				
2^a				
Prova "speciale" (rif. No)				
Si		-	-	-

2.4 B1 – Istruzione e Formazione Professionale

Il campione inizialmente previsto per gli studenti iscritti a un percorso di Istruzione e Formazione Professionale (596 studenti; cfr. Capitolo 3) è stato predisposto al fine di permettere un'analisi multivariata sulle determinanti individuali delle performance linguistiche anche per il livello B1 – Istruzione e Formazione Professionale. Tuttavia, la percentuale di studenti che ha effettivamente svolto la prova si è rivelata di molto inferiore rispetto al campione iniziale¹⁵ e questo non ha consentito di poter stimare con sufficiente precisione il contributo di ciascun fattore alle performance misurate nelle quattro competenze. A ciò si è aggiunto il fatto che alle prove in lingua tedesca ha partecipato un numero esiguo di ragazzi¹⁶, questo per il combinato disposto della decisione di non far partecipare studenti che non avessero il tedesco come materia di insegnamento nell'anno scolastico in corso e delle assenze registrate tra coloro che invece avrebbero dovuto partecipare.

In assenza di stime robuste per l'inglese si è scelto di presentare semplici descrittive, a seconda di alcune caratteristiche degli studenti, della percentuale di studenti che in ciascuna prova ha raggiunto il punteggio necessario per possedere il livello B1 (cfr. Tabella 2.7); queste non hanno valenza di carattere causa-effetto ma vanno considerate come semplici fotografie, da diversi punti di vista, del fenomeno indagato.

Per quanto riguarda il genere, le abilità di ascolto e scrittura sono possedute in misura maggiore dalle ragazze (56,7% e 53,7%) rispetto ai ragazzi (39,0% e 43,1%). Nella lettura invece si registra un sostanziale equilibrio con circa uno studente su dieci in possesso della competenza sia tra i ragazzi che tra le ragazze. Discorso diverso per la produzione orale nella quale vi è una percentuale maggiore di ragazzi che raggiunge la soglia del 60% rispetto alle ragazze (rispettivamente 10,2% e 4,5%).

Dall'analisi dei dati non emergono differenze rilevanti tra studenti in regola con gli studi e coloro che hanno accumulato un ritardo scolastico così come tra studenti con origini straniere e studenti italiani. Tuttavia, per quanto riguarda la produzione orale si segnala una percentuale leggermente superiore di studenti in possesso dell'abilità tra gli stranieri (13,5% per le prime generazioni e 16,7% per le seconde generazioni) rispetto agli studenti

¹⁵ Circa il 70% degli studenti ha svolto effettivamente la prova sul campione totale.

¹⁶ Gli studenti totali che hanno sostenuto la prova in tedesco sono 79. Questa ridotta numerosità non consente di svolgere analisi suddivise per le caratteristiche degli studenti per la lingua tedesca.

italiani (7,1%); mentre per la lettura sono gli studenti italiani che mostrano una percentuale di raggiungimento della soglia più alta (12,0%) rispetto agli studenti stranieri (5,4% per le prime generazioni e 0,0% per le seconde generazioni).

I dati relativi al background socioeconomico e culturale presentano in linea generale un quadro simile a quanto visto per gli altri livelli: al crescere del titolo di studio posseduto dai genitori cresce anche la percentuale di studenti in possesso dell'abilità. Tuttavia, una chiara direzione in questo senso può essere identificata solamente per le abilità di produzione orale e di lettura, mentre per l'ascolto e la produzione scritta le differenze seguono un andamento meno netto e omogeneo. Infatti, gli studenti con genitori laureati risultano essere in possesso delle competenze di ascolto e scrittura in percentuale minore rispetto agli studenti i cui genitori sono in possesso di un titolo di studio inferiore. Non potendo però controllare per altre caratteristiche, non siamo in grado di affermare se questi risultati riflettano un pattern reale oppure sono dovuti, ad esempio, al fatto che tra i figli di laureati vi siano più ragazzi che, come emerge dai risultati, mostrano percentuali di ottenimento delle abilità di ascolto e scrittura inferiori rispetto alle ragazze.

Tra i due macro-indirizzi considerati non si rilevano differenze rilevanti se non in relazione all'abilità dell'ascolto nella quale la percentuale di studenti in possesso della competenza risulta maggiore per gli iscritti all'indirizzo dei Servizi (56,2%) rispetto all'indirizzo Industria e Artigianato (30,4%).

Infine, gli studenti con disturbi specifici dell'apprendimento che hanno svolto la prova con strumenti compensativi mostrano percentuali inferiori di raggiungimento della soglia minima nelle varie competenze. Queste differenze risultano particolarmente pronunciate per l'ascolto e la produzione scritta.

Tabella 2.7 - INGLESE: percentuale di studenti in possesso del livello per abilità - B1 leFP.

	ASCOLTO	PARLATO	LETTURA	SCRITTURA
Genere				
Maschio	39,0	10,2	10,2	43,1
Femmina	56,7	4,5	11,9	53,7
Ritardo				
In regola	45,5	8,5	11,2	45,2
1 o più anni di ritardo	45,5	7,9	10,5	48,7
Origine				
Italiano	45,8	7,1	12,0	47,7
Straniero di 2^a generazione	38,9	16,7	0,0	33,3
Straniero di 1^a generazione	43,2	13,5	5,4	45,9
Titolo di studio dei genitori				
Licenza elementare o media	39,4	6,1	4,5	47,0
Qualifica professionale	43,5	4,3	4,3	42,4
Diploma di maturità	53,0	11,3	15,9	52,3
Laurea	41,7	11,1	16,7	44,4
Non so	34,5	6,9	10,3	37,9
Indirizzo di studio				
Servizi	56,2	8,2	10,5	49,3
Industria e Artigianato	30,4	8,1	11,2	43,5
Attività CLIL in inglese				
No	40,8	9,5	10,2	46,3
Sì	60,5	3,5	12,8	48,8
Prova "speciale"				
No	48,9	9,5	11,6	50,2
Sì	22,6	0,0	5,6	26,4

3 | Il campionamento di Try-out 2017 e TLT 2018

Maurizio Carpita¹⁷

Introduzione

In questo capitolo si presentano i criteri utilizzati per la progettazione dei campioni di studenti che hanno partecipato ai due test linguistici effettuati nel biennio 2017-2018 per la misurazione dei tre livelli (A1, A2 e B1) considerati da IPRASE per tedesco e inglese.

In particolare, nell'ottobre del 2017 è stata effettuata una rilevazione pilota su scala ridotta (Try-Out o TO) con l'obiettivo di verificare lo strumento (Test) dal punto di vista psicometrico, in previsione della seconda e ben più ampia rilevazione delle competenze linguistiche degli studenti trentini (Trentino Language Testing o TLT) che sarebbe stata effettuata a marzo 2018 e che seguiva quella sperimentale svolta due anni prima con il TLT 2016.

Poiché il TO 2017 è stato somministrato all'inizio dell'anno scolastico (ottobre), allo scopo di fornire un'indicazione comparabile del livello degli studenti, la prova è stata svolta in classi immediatamente successive a quelle di riferimento dei livelli linguistici considerati da IPRASE; ad esempio il livello A1, rilevato con il TLT 2016 e 2018 sugli studenti della V primaria, con il TO 2017 è stato invece somministrato agli studenti della I classe della Secondaria di primo grado.

Il piano di campionamento per il TLT 2018, progettato dopo vari approfondimenti dei risultati della sperimentazione condotta con il TLT 2016 e alcune verifiche tecniche di fattibilità operativa (che, ad esempio, hanno portato a confermare la scelta del 2016 di somministrare i test di entrambe le lingue agli stessi studenti), ha tenuto conto sia degli obiettivi statistici dell'indagine (predi-

¹⁷ Maurizio Carpita è Professore ordinario di Statistica presso l'Università degli Studi di Brescia.

sporre un test solido dal punto di vista psicometrico e ottenere un campione rappresentativo della popolazione di studenti trentini), sia delle complessità organizzative che ovviamente tali rilevazioni comportano. Inoltre, l'obiettivo è rilevare differenze di una certa dimensione, poiché la presenza di "effetti deboli" può essere in gran parte attribuibile all'eterogeneità degli studenti che compongono le classi e alle modalità di insegnamento. Questi due limiti (risultati riferiti al sistema nel suo complesso e individuazione di soli effetti rilevanti) insieme alla considerazione che comunque la popolazione di riferimento è relativamente omogenea (studenti di età analoghe che vivono nello stesso contesto provinciale), hanno portato come si vedrà a ridurre in modo ragionevole la dimensione campionaria necessaria per ottenere i risultati voluti. Sulla base di queste premesse, nei due paragrafi che seguono si presentano le caratteristiche dei piani di campionamento definiti per i TO 2017 e TLT 2018.

3.1 Il piano di campionamento per il TO 2017

Come anticipato nell'introduzione, il principale obiettivo dei campioni utilizzati per il Try-Out (TO) effettuato a ottobre 2017 è stato quello di permettere di verificare le caratteristiche statistiche delle nuove versioni dei test linguistici di inglese e tedesco, ottenute da quelli utilizzati da IPRASE nel TLT 2016 per la verifica del possesso dei livelli A1, A2 e B1. Le modifiche effettuate, suggerite dall'analisi psicometrica, hanno portato alla sostituzione o revisione di alcune domande. Inoltre, al fine di favorire la valutazione dell'effettiva difficoltà dei quesiti rispetto al livello di riferimento di un test (ad esempio A2), sono stati aggiunti dei quesiti di livello inferiore (A1) e di livello superiore (B1): dal punto di vista psicometrico, ci si attende che quesiti di livello inferiore e superiore siano rispettivamente più facili e più difficili di quelli del livello intermedio che si vuole effettivamente misurare, permettendo quindi una migliore "calibrazione" dei livelli di competenze linguistiche effettivamente raggiunti dagli studenti.

Per verificare se le nuove domande messe a punto dagli esperti linguistici funzionavano adeguatamente e per ottenere una buona calibrazione del test, i campioni di studenti sono stati scelti in modo tale da presentare nel limite del possibile abilità variabili su tutti i livelli di difficoltà delle domande: ad esempio, nei campioni del TO 2017 delle classi delle SSPG avrebbero dovuto essere presenti studenti in grado di rispondere o meno in modo corretto – oltre ovviamente ai quesiti del livello di riferimento

(A2) – anche ai quesiti dei livelli inferiore (A1) e superiore (B1). Per raggiungere questo obiettivo sono stati considerati i risultati del TLT 2016, i quali hanno mostrato che la maggiore variabilità delle conoscenze linguistiche è attribuibile agli studenti che si trovano nelle stesse classi (percentuali di varianza tra conoscenze linguistiche degli studenti dentro le classi: Primaria 52%-93%; Secondaria I grado 63%-95%; Secondaria II grado 35%-91%) e tra le scuole (percentuali di varianza tra conoscenze linguistiche tra classi: Primaria 5%-35%; Secondaria I grado 2%-27%; Secondaria II grado 0%-61%). Per questo motivo, nella fase di estrazione dei campioni si è quindi seguita l'indicazione che le classi scelte dovevano essere preferibilmente e per quanto possibile di istituti diversi. Sempre per aumentare l'eterogeneità delle abilità rilevate nella fase di try-out, si è inoltre ritenuto opportuno campionare classi in strati con diversi livelli attesi di competenze linguistiche: ad esempio, nel caso della scuola secondaria di secondo grado, campionando classi tra quelle dei licei linguistici e tra quelle degli istituti tecnici. È importante sottolineare che, per evitare problemi organizzativi e di privacy, tutti gli studenti di ogni classe del campione sono stati invitati a partecipare al test.

Per quanto riguarda la scelta della dimensione dei campioni, ovvero la numerosità dei tre gruppi di studenti ai quali somministrare i tre livelli del test, la letteratura psicometrica suggerisce che, per ottenere una calibrazione stabile delle domande di $\pm 0,5$ logit (che è una misura statistica dell'abilità) con un livello di confidenza del 95% (che è una misura statistica del margine di certezza del risultato dovuto al campionamento), si dovrebbero considerare almeno di 100 studenti, mentre per il livello del 99% è necessario un campione di almeno 150 studenti (Linacre J.M. 1994. *Sample size and item calibration stability*. *Rasch Measurement Transactions*, 7:4, 328).

Rispetto a queste indicazioni minime generali, per i campioni delle scuole secondarie di primo e secondo grado si è deciso di scegliere numerosità intermedie a quelle suggerite dai due livelli di confidenza, mentre per la scuola primaria si è deciso di raddoppiare il campione degli studenti di V Primaria con riferimento al livello di significatività più elevato (99%): il motivo di questa scelta è legato al fatto che il primo livello di formazione scolastica non è di solito considerato nei test che rilevano le conoscenze linguistiche e può essere quindi utile disporre di un campione più ampio per verificare l'effettiva "tenuta" dello strumento.

Come già spiegato nell'introduzione, poiché il try-out è stato effettuato all'inizio del nuovo anno scolastico (ottobre 2017), per avere un campione rappresentativo delle conoscenze linguistiche

alla fine della V Primaria, III Secondaria di primo grado e II Secondaria di secondo grado come previsto dai TLT 2016 e 2018, gli studenti sono stati scelti tra i neoiscritti degli anni successivi.

La Tabella 3.1 riporta quindi la proposta operativa delle tre dimensioni campionarie programmate per il TO 2017, ottenute a seguito delle considerazioni precedenti e tenuto conto della dimensione media delle classi. Il fatto che ci fosse la possibilità di non coinvolgere tutti gli studenti di una classe (si consideri la presenza di studenti con BES e un tasso fisiologico di non partecipazione del 10% verificatosi con il Test 2016), ha consigliato di aumentare il numero delle classi programmato inizialmente da 12 a 15 per il primo livello di scuola e da 6 a 7 per i due livelli di scuola successivi.

Tabella 3.1 - Dimensioni programmate dei campioni per il TO 2017.

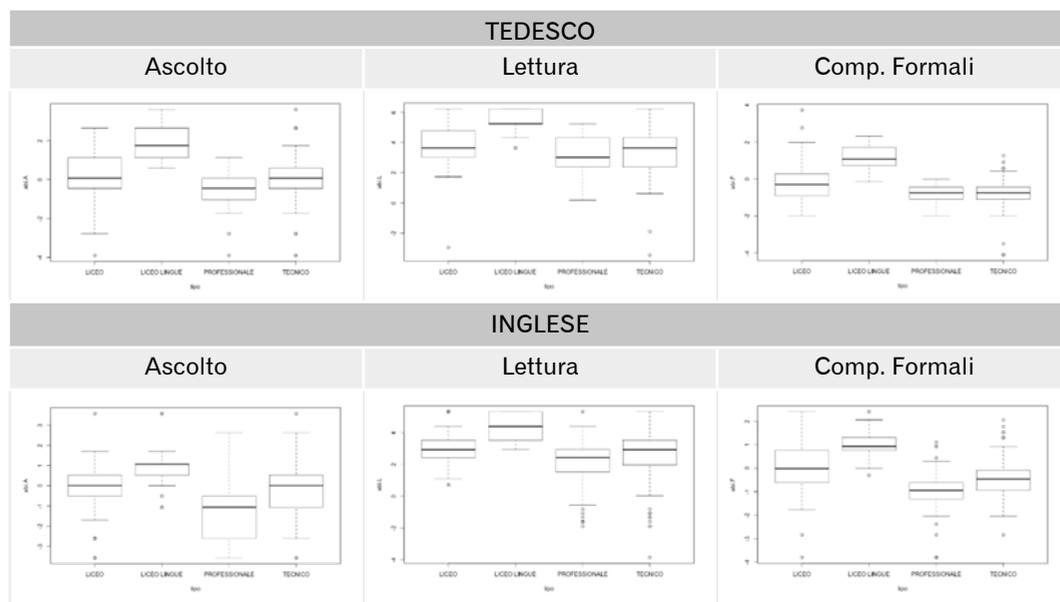
SCUOLA	NUMERO DI STUDENTI ATTESI	DIMENSIONE MEDIA CLASSI	NUMERO DI CLASSI DA CAMPIONARE IN SCUOLE DIVERSE
I Secondaria I grado	300	21	15
I Secondaria II grado	140	20	7
III Secondaria II grado	130	19	7

3.2 Il piano di campionamento per il TLT 2018

Il principale obiettivo che ci si è posti per la rilevazione della seconda edizione del Trentino Language Testing (TLT), svolta nel 2018 e che seguiva la prima effettuata nel 2016, è stato quello di permettere di verificare sulla base di campioni rappresentativi il livello di conoscenza delle lingue inglese e tedesco della popolazione degli studenti trentini. Il secondo obiettivo della rilevazione condotta per il TLT 2018 è stato quello di ottenere una maggiore rappresentatività dei livelli di conoscenza delle lingue nelle scuole superiori del Trentino: si è ritenuto utile e opportuno perseguire questo obiettivo, considerata l'eterogeneità dell'offerta formativa di questo livello di istruzione, che si rivolge a studenti che studiano le lingue già da alcuni anni e in alcuni casi con obiettivi specifici (si pensi ad esempio al liceo linguistico e alla formazione professionale orientata al turismo). Tale eterogeneità è stata verificata anche con i dati, considerando i risultati del TLT 2016. La Figura 3.1 rappresenta graficamente con i diagrammi scatola-baffi o *box-plot* (la linea in grassetto che rappresenta la mediana e il

rettangolo contiene il 50% degli studenti con abilità centrali, tra il 1° e il 3° quartile) le distribuzioni delle tre abilità linguistiche (misurate con il modello di Rasch) rilevate da IPRASE con il primo test di tedesco e inglese (Ascolto, Lettura e Competenze Formali) somministrato agli studenti delle scuole superiori raggruppati per tipologia di istituto (Liceo, Liceo Linguistico, Istituti Professionali e Istituti Tecnici): come si può osservare, le distribuzioni delle abilità risultano in diversi casi piuttosto diverse, in particolare più elevata e meno variabile quella del Liceo Linguistico.

Figura 3.1 - Box-Plot delle distribuzioni delle tre abilità degli studenti delle scuole superiori per il TLT 2016.



Dal punto di vista statistico, riguardo alla relazione tra dimensione campionaria e precisione delle stime, sono state considerate le seguenti regole generali (Cohen J. 1988. *Statistical power analysis for the behavioral sciences*. 2nd Edition, LEA Publ.):

- 1) l'errore standard della stima di una percentuale (ad esempio la "% di studenti che hanno il livello A1 di inglese") con un campione casuale di 500 osservazioni non supera i 2 punti percentuali. In pratica, se si vuole costruire un intervallo di confidenza al 95% per tale proporzione, significa che questo intervallo avrà una semi-ampiezza di meno di 4 punti percentuali rispetto alla stima (se la stima fosse 45%, l'intervallo di confidenza al livello del 95% sarebbe al massimo 41%-49%);

- 2) per valutare la significatività della differenza tra due valori medi (ad esempio la “differenza tra i punteggi medi nel test di tedesco tra le scuole primarie di Trento e Rovereto e le altre scuole primarie”) con un effetto di media ampiezza (*effect size* = 0,5) e un test t di Student al livello di significatività del 1% con una potenza del 90%, bastano due campioni casuali ognuno di 120 osservazioni;
- 3) per verificare se, in un modello di regressione lineare con 10 regressori, ulteriori 5 regressori (variabili di controllo) hanno un effetto anche minimo sulla variabile dipendente (*effect size* = 0,1), con un livello di significatività del 1% e una potenza del 90%, è sufficiente un campione casuale di 250 osservazioni.

Si noti che in nessuno dei tre esempi si fa riferimento alla dimensione della popolazione. Si tratta di una proprietà delle stime statistiche, che si verifica se si usano campioni del tutto casuali: la dimensione dell'errore di stima non dipende dalla dimensione della popolazione (che può essere anche nell'ordine di centinaia di milioni o miliardi) ma solo dalla dimensione del campione, che però deve essere estratto in modo assolutamente casuale. Ad esempio, l'errore standard del campione casuale di 500 osservazioni al punto 1) rimane sempre al massimo di 2 punti percentuali, anche se la popolazione è composta da 1 miliardo di persone.

Alle considerazioni di carattere statistico si affiancano poi quelle di tipo organizzativo e di budget; queste ultime, in particolare modo, hanno difatti permesso di andare oltre la dimensione campionaria minima definita seguendo le regole generali e coinvolgere un maggior numero di studenti, destinando le risorse economiche a disposizione a un campionamento “articolato” per la II Secondaria di secondo grado e per la III leFP.

La Tabella 3.2 riporta quindi la proposta operativa delle dimensioni campionarie per il TLT 2018, ottenuta a seguito delle considerazioni precedenti.

Tabella 3.2 - Dimensioni programmate dei campioni per il TLT 2018.

SCUOLA	NUMERO DI STUDENTI ATTESI	DIMENSIONE MEDIA CLASSI	NUMERO DI CLASSI DA CAMPIONARE
V Primaria	500	17	30
III Secondaria I grado	600	21	30
II Secondaria II grado	1.180	18–21	58
III leFP (Ing)	500	19	26
III leFP (Ted)	120	20	6

Come si può osservare, se la numerosità campionaria programmata per i primi due livelli di istruzione considerati è in linea con la regola generale al punto 1) e quindi soddisfa pienamente anche quelle dei punti 2) e 3), per il terzo livello di istruzione considerato (II Secondaria di secondo grado e III Istruzione e Formazione Professionale) la numerosità del campione programmata è stata più che raddoppiata, permettendo così di tener maggiormente conto di quanto detto in precedenza sull'eterogeneità delle scuole superiori e delle evidenze della Figura 3.1. In questo modo, la somma delle dimensioni campionarie programmate per TO 2017 e TLT 2018 risulta pari a circa 3.000, numerosità massima limite di riferimento per garantire alla rilevazione il rispetto delle tempistiche di somministrazione del test.

3.3 La procedura di campionamento per TO 2017 e TLT 2018

La strategia di campionamento in ottica di TLT 2018 per la V Primaria e la III Secondaria di primo grado ha previsto, per ciascuno dei due ordini, la selezione di un campione probabilistico di studenti a due stadi: il primo stadio a livello di plessi/scuole e il secondo stadio a livello di sezioni. Questa tecnica di campionamento risulta particolarmente adatta alla natura "gerarchica" delle unità di campionamento, con gli studenti raggruppati in classi e queste ultime raggruppate in scuole.

Per entrambi gli ordini non si applica nessuna stratificazione esplicita ma l'elemento territoriale viene tenuto in considerazione ordinando le scuole per comunità di valle (stratificazione implicita) oltre che per dimensione complessiva¹⁸.

Al primo stadio si seleziona senza re-immissione e con probabilità variabile un campione di scuole s_{sc} ; ciascuna scuola i viene estratta con una probabilità di inclusione π_i proporzionale al numero totale dei suoi studenti iscritti nelle sezioni della V Primaria o della III Secondaria di primo grado; ad esempio, se una generica scuola A ha il doppio degli studenti di una generica scuola

¹⁸ L'ordinamento è stato fatto seguendo il metodo del *serpentine approach*: le unità di campionamento (scuole) del primo stadio sono ordinate dalla più piccola alla più grande nel primo strato implicito (comunità di valle), dalla più grande alla più piccola nel secondo strato implicito, di nuovo dalla più piccola alla più grande nel terzo strato implicito e così via. Questo accorgimento permette di definire l'insieme delle scuole *sostitutive* selezionando le unità di campionamento con il numero di studenti più simile a quello di ciascuna scuola estratta nel campione.

B, allora la scuola A avrà il doppio delle probabilità di essere selezionata rispetto alla scuola B. La probabilità π_i è definita in questo modo:

$$\pi_i = s \times \alpha_i = s \times \frac{\ddot{A}_i}{\sum_{i=1}^s \ddot{A}_i}$$

dove

$$\ddot{A}_i = A_i \times \frac{2}{s_i} = A_i \times \frac{2}{\min(2, S_i)}$$

e dove s è il numero totale di classi estratte, A_i è il numero di studenti della scuola i e S_i è il numero di classi nella scuola i .

Il rapporto $\frac{2}{\min(2, S_i)}$ è uguale a 1 per le scuole con almeno 2 sezioni e a 2 per le scuole con una sola sezione; esso aumenta la probabilità di selezione per le piccole scuole, assicurando tuttavia *un'uguale (o molto simile) probabilità di inclusione finale a tutti gli studenti della popolazione di riferimento*, a prescindere dalla dimensione di ciascuna scuola.

Al secondo stadio, per ciascuna scuola S_i selezionata nel primo stadio, viene estratto un campione di classi a probabilità uguali e senza re-immissione: esso sarà composto da 1 o 2 classi a seconda che la scuola selezionata abbia rispettivamente una sola sezione o almeno due sezioni¹⁹. La probabilità (condizionata) di estrazione è quindi definita da:

$$\pi_{j|i} = \frac{S_j}{S_i}$$

e tutti gli studenti A_{ij} appartenenti alle classi estratte sono sottoposti alla prova.

Nel caso della II Secondaria di secondo grado e della III Istruzione e Formazione Professionale (Test di Inglese) si è utilizzata una strategia di campionamento del tutto simile, introducendo però

¹⁹ La scelta di campionare fino a due classi per scuola, e non fino a tre così come fatto per il TLT 2016, è stata presa alla luce dei risultati osservati nell'indagine del 2016, dalla quale era emerso che la variabilità nei risultati degli studenti era in misura prevalente tra scuole e non tra classi all'interno della stessa scuola.

l'elemento della stratificazione, con la quale si costruisce un campione nel quale la proporzione tra il numero di studenti di ciascuno strato (nel nostro caso abbiamo considerato i diversi indirizzi di studio come singoli strati, quindi liceo classico, liceo scientifico, liceo delle scienze umane, ecc.) è il più simile possibile a quella esistente a livello di popolazione di riferimento; la Tabella 3.3 riporta la composizione, in termini di indirizzi di studio, della popolazione di studenti della II Secondaria di secondo grado e della III leFP.

Tabella 3.3 - Popolazione studenti Secondaria II grado e Istruzione e Formazione Professionale.

INDIRIZZI DI STUDIO	STUDENTI NELLA POPOLAZIONE	% RISPETTO ALLA POPOLAZIONE
Secondaria II grado		
Classico	190	4.8
Scientifico	1030	26.0
Linguistico	375	9.5
Scienze Umane	520	13.1
Artistico	286	7.2
Economico	640	16.2
Tecnologico	816	20.6
Professionale	103	2.6
Totale	3960	100
Istruzione e Formazione professionale (Test ING)		
Servizi	1057	58.2
Industria e artigianato	760	41.8
Totale	1817	100
Istruzione e Formazione professionale (Test TED)		
Servizi	710	100

Una volta stabilita a priori la dimensione di ciascun sotto-campione (applicando le percentuali viste nella terza colonna della Tabella e alla dimensione complessiva del campione per la II Secondaria di secondo grado, 1180, e per la III leFP, 500) si procede alla definizione di questi mediante una selezione a due stadi,

applicata di volta in volta alla popolazione di studenti iscritti in ciascuno degli strati/indirizzi di studio considerati²⁰.

Per ogni strato, al primo stadio vengono estratti gli Istituti Superiori (IS) o Centri di Formazione Professionale (CFP) con probabilità proporzionale alla loro dimensione²¹. Nel secondo stadio invece all'interno degli IS estratti al primo stadio viene estratto un campione di 1, 2, o 3 classi a probabilità uguali e senza re-immissione: in particolare, il numero di classi estratte è pari a 1 per IS/CFP aventi 1 classe in totale, 2 per IS/CFP con 2 classi e 3 per IS/CFP con 3 o più classi.

Nel caso della III Istruzione e Formazione Professionale (Test di Tedesco) si è invece utilizzata la strategia campionaria a uno stadio, estraendo 6 classi in 6 scuole diverse.

3.4 I campioni effettivi dei TO 2017 e TLT 2018

Gli studenti che hanno effettivamente partecipato al TO 2017 sono stati in totale 596, coinvolgendo in totale 16 istituti e 29 classi. A marzo 2018 si è svolta la seconda rilevazione linguistica con il *TLT 2018*: complessivamente il campione teorico iniziale comprendeva 2.955 studenti di cui 2.526 ragazzi hanno effettivamente sostenuto la prova in inglese e 2.181 hanno sostenuto quella in tedesco suddivisi in 70 istituti e 149 classi (cfr. Tabella 3.4).

Tabella 3.4 - Numero di istituti, classi e studenti che hanno partecipato alle rilevazioni TO 2017 e TLT 2018.

	TO 2017		TLT 2018	
	INGLESE	TEDESCO	INGLESE	TEDESCO
V Primaria (A1)				
Istituti	8	8	21	21
Classi	15	15	30	30
Studenti	329	329	502	496

²⁰ La scelta del campionamento a due stadi anziché uno solo (in questo caso si sarebbero estratte in ciascuno strato il numero di classi necessarie) è dovuta al fatto che dalle simulazioni con quest'ultimo metodo si ottenevano campioni nei quali spesso un Istituto Superiore aveva fino a 5/6 classi estratte, minando la necessità di assicurare variabilità tra scuole e rendendo molto complessa l'individuazione delle scuole/classi sostituite in caso di rifiuto a partecipare da parte dell'IS.

²¹ Per ogni strato vengono considerati soltanto gli Istituti Superiori (o Centri di Formazione Professionale nel caso della IeFP) e le classi afferenti all'indirizzo di studio per il quale si sta definendo il sotto-campione.

	TO 2017		TLT 2018	
	INGLESE	TEDESCO	INGLESE	TEDESCO
III Sec. I grado (A2)				
Istituti	7	7	16	16
Classi	7	7	30	30
Studenti	138	138	584	585
II Sec. II grado (B1)				
Istituti	7	7	20	20
Classi	7	7	57	56
Studenti	130	130	1059	1021
III IeFP (B1)				
Istituti	–	–	10	6
Classi	–	–	27	6
Studenti	–	–	381	79
Totale				
Istituti	16*	16*	67**	63**
Classi	29	29	144	123
Studenti	597	597	2526	2181

* la somma degli istituti non coincide con il totale poiché le classi relative ai livelli A2 e B1 appartengono ai medesimi Istituti.

** le somme indicate non coincidono con il totale degli istituti coinvolti nel TLT 2018 poiché nella maggior parte di questi sono state svolte le prove in entrambe le lingue.

Per quanto riguarda il TLT 2018, nello specifico 57 classi di 20 istituti secondari di secondo grado hanno partecipato alla rilevazione delle competenze linguistiche suddivisi per indirizzo, come evidenziato nella Tabella 3.5.

Tabella 3.5 - Dettaglio campione teorico scuola secondarie di II grado per indirizzo.

INDIRIZZO	ISTITUTI*	CLASSI	STUDENTI
Classico	2	4	85
Scientifico	7	13	254
Linguistico	2	6	115
Scienze umane	4	8	159
Artistico	1	3	59
Economico	5	11	210
Tecnologico	4	10	235
Professionale	1	2	46

* il totale degli istituti è maggiore dei 20 indicati più sopra in quanto alcuni istituti offrono più indirizzi.

La Tabella 3.6 indica invece i centri di formazione professionale suddivisi per settore di competenza: sono 12 per un totale di 32 classi.

Tabella 3.6 - Dettaglio campione teorico CFP per indirizzo.

INDIRIZZO	CENTRI*	CLASSI	STUDENTI
Settore servizi	10	20	394
Settore Industria e Artigianato	4	9	166
Settore Agricoltura e Ambiente	1	3	37

* il totale è maggiore dei 12 indicati più sopra in quanto alcuni CFP offrono più indirizzi.

Nelle tabelle 3.7a e 3.7b vengono elencati gli istituti che hanno partecipato rispettivamente al TO 2017 e al TLT 2018, e in Figura 3.2 la loro distribuzione territoriale.

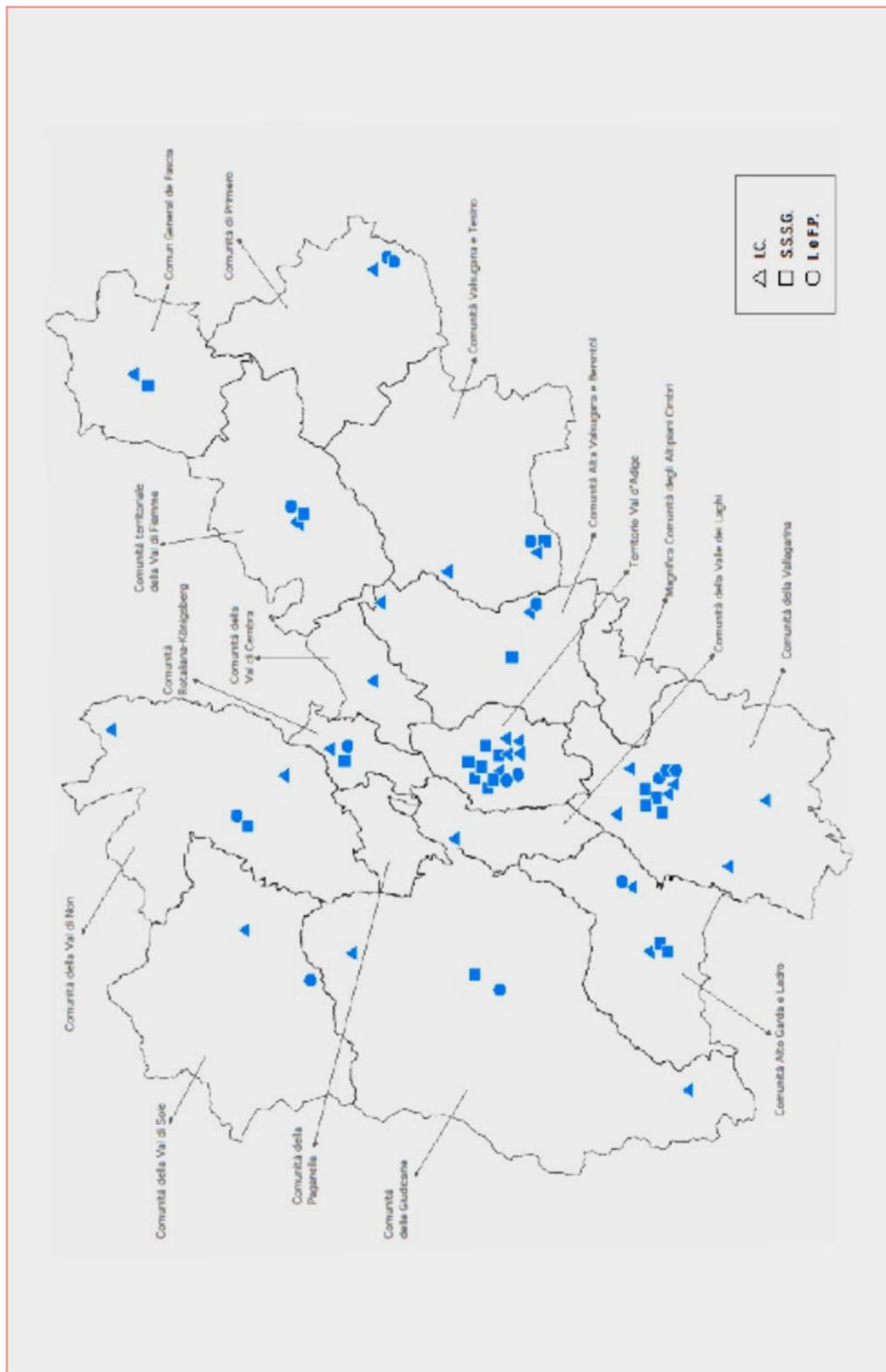
Tabella 3.7a - Elenco degli istituti che hanno partecipato al try out 2017.

ISTITUTI COMPRESIVI	SCUOLE SECONDARIE II GRADO E IeFP
IC Aldeno Mattarello	CFP Enaip – Riva del Garda
IC Folgaria, Lavarone, Luserna	Istituto di Istruzione “Don Milani” – Rovereto
IC Giudicarie esteriori	Istituto di Istruzione “La Rosa Bianca” – Cavalese
IC Lavis	Istituto di Istruzione “M. Martini” – Mezzolombardo
IC Pergine 2	Istituto di Istruzione “M. Curie” – Pergine
IC Riva 1	ITET “A. Pilati” – Cles
IC Trento 4	Liceo Linguistico “S.M. Scholl” – Trento
IC Villa Lagarina	Liceo Scientifico “G. Galilei” – Trento

Tabella 3.7b - Elenco degli istituti campionati per il TLT 2018.

Istituti comprensivi	Scuole secondarie di II grado	IeFP
IC Ala	ITE "Tambosi-Battisti" – Trento	CFP Barelli Rovereto
IC Aldeno Mattarello	ITET "F. e G. Fontana" – Rovereto	CFP ENAIP Borgo Valsugana
IC Alta Vallagarina	ITET "G. Floriani" – Riva del Garda	CFP ENAIP Ossana
IC Altopiano di Piné	ITT "Buonarroti-Pozzo" – Trento	CFP ENAIP Villazzano
IC Arco	ITT "Marconi" – Rovereto	CFP ENAIP Arco
IC Bassa Val di Sole	Ist. di Istruzione "A. Degasperì" – Borgo Valsugana	CFP ENAIP Primiero
IC Cavalese	Ist. di Istruzione "La Rosa Bianca" – Cavalese	CFP ENAIP Tesero
IC Cembra	Ist. di Istruzione "Don Milani" – Rovereto	CFP ENAIP Tione di Trento
IC Centro Valsugana	Ist. di Istruzione "L. Guetti" – Tione	CFP "G. Veronesi" – Rovereto
IC del Chiese	Ist. di Istruzione "M. Curie" – Pergine	CFP UPT Cles
IC Fondo-Revò	Ist. di Istruzione "M. Martini" – Mezzolombardo	CFP UPT Trento
IC Levico Terme	Istituto delle Arti – Trento	Fondazione "E. Mach" – S. Michele A/A
IC Mezzolombardo-Paganella	Liceo "A. Maffei" – Riva del Garda	IFP Alberghiero Rovereto e Levico Terme
IC Riva 1	Liceo "A. Rosmini" – Rovereto	
IC Rovereto Est	Liceo "A. Rosmini" – Trento	
IC Rovereto Nord	Liceo "B. Russell" – Cles	
IC Rovereto Sud	Liceo scientifico "Da Vinci" – Trento	
IC Strigno e Tesino	Liceo classico "G. Prati" – Trento	
IC Taio	Liceo linguistico "S.M. Scholl" – Trento	
IC Trento 4	Scuola ladina di Fassa	
IC Trento 6		
IC Trento 7		
IC Val Rendena		
IC Valle dei Laghi-Dro		
IC Villa Lagarina		
IC di SP e SSPG del Primiero		
Scuola ladina di Fassa		

Figura 3.2 - Distribuzione delle scuole partecipanti al TLT 2018 nel territorio provinciale.



4 | Caratteristiche delle prove

Davide Defant²², Michela Chicco²³

4.1 Impostazione delle prove

La rilevazione linguistica ha visto la predisposizione di prove di competenza in lingua inglese e tedesca relativamente ai livelli A1, A2 e B1, sviluppate in coerenza con quanto definito dal QCER (Quadro Comune Europeo di Riferimento per le lingue). Come accennato nei capitoli precedenti, le prove sono state somministrate in formato computer-based attraverso l'utilizzo di una piattaforma informatica nel corso del mese di marzo 2018.

Il test ha voluto rilevare in modo disgiunto tutte le abilità linguistiche. Le domande, somministrate in modo automatico, prevedevano una risposta ai quesiti attraverso una singola scelta/click oppure interagendo con il sistema attraverso la tastiera o il microfono. Ogni test è stato strutturato per valutare le abilità linguistiche seguendo l'ordine ascolto, parlato, lettura e scrittura.

Per ogni abilità sono state previste due sezioni distinte, caratterizzate da esercizi diversi tra una sezione e l'altra. Le domande sono state sviluppate in relazione ai descrittori riportati nel QCER e relativi a ogni singola abilità.

In relazione alla durata complessiva del test, si riepilogano nella seguente tabella le durate parziali relative a ogni abilità unitamente alla durata d'insieme del test stesso.

²² Davide Defant è Direttore di ISIT, Istituto Accademico per Interpreti e Traduttori, Trento.

²³ Michela Chicco è docente esperta presso Iprase.

Tabella 4.1 - Durata del test per abilità e livello.

ABILITÀ	LIVELLO A1	LIVELLO A2	LIVELLO B1
Ascolto	10 minuti	10 minuti	20 minuti
Parlato	5 minuti	8 minuti	10 minuti
Lettura	15 minuti	25 minuti	30 minuti
Scrittura	20 minuti	30 minuti	30 minuti
Durata complessiva	50 minuti	73 minuti	90 minuti

4.2 Definizione e format degli item

Per la definizione dei contenuti della rilevazione e del format degli item, IPRASE ha individuato un pool di dodici docenti (sei per la lingua inglese e sei per la lingua tedesca) i quali, suddivisi a coppie per lingua e per livello, hanno elaborato i quesiti proposti durante la rilevazione. Trasversalmente alle quattro abilità sopra riportate, l'interazione con il sistema è avvenuta attraverso alcune tipologie di domande o esercizi, tra cui:

- *Ascolto e scelta multipla*: il sistema proponeva l'ascolto di un documento audio, in relazione al quale venivano proposti dei quesiti a risposta multipla;
- *Lettura e selezione*: il sistema proponeva al candidato una domanda, un testo, un'etichetta, ecc., e lo studente doveva indicare la parola, la frase o il periodo corretto, su una selezione di quattro risposte;
- *Drag & Drop* (solo per il livello A1): allo studente si chiedeva di abbinare per trascinamento alcune informazioni (ad esempio immagine di animale/nome, nome del colore/colore, ecc.);
- *Breve testo scritto*: lo studente era chiamato a comporre un testo su un argomento assegnato (il numero massimo di caratteri era evidenziato a video);
- *Breve interazione orale*: allo studente si chiedeva di interagire con il sistema rispondendo a domande o di produrre un testo a voce su un argomento dato.

Per le abilità ricettive (Ascolto e Lettura) sono stati previsti item a stimolo "chiuso", ovvero a ogni quesito venivano attribuite quattro possibili risposte, di cui una sola corretta. Per le abilità di produzione (Parlato e Scrittura), invece, i quesiti erano a risposta "aperta", per cui lo studente poteva rispondere in maniera personalizzata all'interno di spazi e tempi predefiniti direttamente dal sistema oppure, solo per uno dei due esercizi previsti per il livello

A1, allo studente veniva chiesto di creare associazioni trascinandolo con il mouse alcune etichette già predisposte (Drag & Drop). Per rendere più familiare la rilevazione, per tutte le sezioni del test è stato realizzato del materiale autentico, opportunamente adattato al livello degli studenti come immagini di contesto, avatar e notizie o eventi aggiornati, ecc. In questo modo gli studenti hanno affrontato la rilevazione in modo più motivante perché più vicino alla lingua reale.

4.2 Valutazione della prova

A ogni sezione della prova sono stati attribuiti dei punteggi, come riportato in tabella 4.2.

Tabella 4.2 - Attribuzione punteggio per abilità e livello.

ABILITÀ	LIVELLO A1	LIVELLO A2	LIVELLO B1
Ascolto	12	12	12
Parlato	24	24	24
Lettura	12	12	12
Scrittura	20	24	24
Punteggio complessivo	68	72	72

La valutazione per l'individuazione del punteggio analitico è stata contraddistinta da due modalità: *automatica* ed *esperta*. Il sistema ha corretto automaticamente le sezioni dedicate ad Ascolto e Lettura, mentre gli elaborati relativi alla produzione orale e a quella scritta sono stati valutati da esaminatori, che hanno assegnato un punteggio a seconda di una griglia di valutazione appositamente predisposta.

A seconda del livello considerato gli indicatori utili alla valutazione della prova di produzione orale hanno riguardato: la pertinenza della risposta, l'adeguatezza delle competenze formali, la proprietà lessicale, la pronuncia, la fluenza, l'efficacia comunicativa. Per la valutazione delle prove di scrittura invece sono stati considerati l'adempimento alla consegna, la correttezza morfo-sintattica, l'appropriatezza lessicale, coerenza e coesione, l'efficacia e la ricchezza comunicativa e infine l'ortografia e la punteggiatura.

4.4 Il coinvolgimento delle scuole

Al fine di garantire un'efficace erogazione dell'intera fase di testing, ogni realtà scolastica coinvolta ha individuato dei *supervisors* con il compito di sovrintendere le attività e di accompagnare gli studenti durante la fase di test. Nel corso del mese di febbraio 2018, sono state proposte delle sessioni di formazione territoriale, ognuna della durata di tre ore ciascuna, in cui si sono approfondite le caratteristiche delle prove e le modalità di esecuzione. A questi incontri sono stati invitati anche gli assistenti di laboratorio, figure di supporto previste per garantire la funzionalità tecnica delle aule.

In occasione di questi incontri è stato consegnato a ogni referente scolastico copia del materiale informativo predisposto, composto da una serie di guide al test (Guida docente, Guida tecnica e Guida studente) e un documento FAQ (Frequently Asked Questions), appositamente creato per facilitare la risoluzione delle problematiche più comuni relative alla fase di erogazione della rilevazione linguistica. Il materiale andava ad approfondire in dettaglio le finalità generali dell'intervento, le caratteristiche e la struttura dei test, i criteri di valutazione, le modalità di somministrazione, la loro durata, le situazioni di criticità nonché le informazioni circa il servizio di help desk.

Il calendario delle prove è stato predisposto lasciando alle scuole l'individuazione della giornata e dell'orario di svolgimento tra alcune date proposte. Nelle giornate di erogazione delle prove sono state effettuate delle visite alle scuole per monitorare l'andamento della rilevazione e per risolvere eventuali criticità.

4.5 Le strutture tecnologiche necessarie: verifica e supporto

Per garantire il corretto svolgimento dell'intera rilevazione, prima dell'inizio delle attività di testing, sono stati completati tutti i sopralluoghi presso gli Istituti scolastici oggetto del campione per verificare la dotazione hardware, software, nonché qualità e potenza della connessione internet dei laboratori informatici, aspetto questo di fondamentale importanza, poiché tutta la piattaforma, essendo online e caratterizzata dallo scambio di file audio, necessitava di una buona connessione internet per poter operare correttamente. A ogni sopralluogo si è verificato che tutti i PC avessero almeno le specifiche tecniche riportate in tabella 4.3.

Tabella 4.3 - Caratteristiche tecniche dei PC necessarie per lo svolgimento dei test.

SPECIFICA TECNICA HARDWARE – REQUISITO MINIMO	SPECIFICA TECNICA SOFTWARE – REQUISITO MINIMO
Processore minimo: Intel® Core™ 2 Duo 2.66GHz o equivalente	Sistema Operativo: Windows 7 con SP1 32 o 64 bit
Memoria RAM minima: 2 GB	Browser internet: Chrome: indicativamente versione 36 o superiore Mozilla Firefox: indicativamente versione 35 o superiore MS Edge; Opera: indicativamente versione 22 o superiore
Spazio disco libero minimo disponibile: 1 GB	
Scheda video con risoluzione minima di 1024x768 con 65536 colori.	Plugin: JavaScript abilitato
Monitor 17" o superiore	
Scheda audio: 44.1 KHz (o migliore) frequenza di campionamento, full-duplex	Internet: ADSL o superiore

Al termine di ogni sopralluogo è stato compilato un *Rapporto di visita* all'interno del quale sono state riportate le caratteristiche del laboratorio informatico, come sopra indicato. Per ognuna delle situazioni critiche emerse, sono state messe in atto misure volte a rendere possibile lo svolgimento delle prove. Di seguito si riportano le criticità di maggiore rilevanza emerse e le soluzioni adottate.

Laboratori sprovvisti di cuffie. Viste le caratteristiche del test (prove di ascolto e di parlato) ogni PC doveva necessariamente essere dotato di cuffie con microfono per l'ascolto della parte audio e per la registrazione della voce dello studente. In tutti quei casi in cui il sopralluogo ha evidenziato come la strumentazione presente nelle scuole non fosse adeguata rispetto alle specifiche minime richieste, è stato messo a disposizione un numero adeguato di cuffie dotate di microfono, idonee a garantire la corretta realizzazione del test. Tale dotazione tecnica è rimasta a disposizione degli Istituti.

Laboratori con struttura hardware/software non sufficiente. In alcuni casi si è cercata una soluzione "interna" al problema verificando se era possibile far accedere le classi al laboratorio di uno dei plessi del medesimo Istituto comprensivo. Laddove questa soluzione non era praticabile si sono messi a disposizione degli Istituti coinvolti dei PC portatili: due laboratori mobili, ognuno composto da quindici postazioni.

Laboratori con connessione a internet debole o praticamente assente. In alcuni casi laddove il collegamento internet di è dimostrato assente o non idoneo per la realizzazione della rilevazione,

si è messa a disposizione una connessione internet basata su tecnologia 4G.

4.6 Piattaforma Demo

Allo scopo di favorire le possibilità di verifica delle competenze linguistiche degli studenti, parallelamente alla piattaforma online dedicata alla rilevazione linguistica TLT 2018, è stato creato e messo a disposizione di tutte le istituzioni scolastiche provinciali uno strumento di testing online in base al quale le scuole possono verificare autonomamente il livello di apprendimento linguistico, sia in inglese che in tedesco, degli studenti e delle studentesse. La piattaforma Demo offre la possibilità, attraverso apposite credenziali di accesso assegnate a tutte le istituzioni scolastiche e formative provinciali, di svolgere prove di lingua inglese e tedesca appositamente create. Le prove, riguardanti le quattro abilità (ascolto, lettura, parlato e scrittura), sono relative a quattro dei livelli di competenza linguistica individuati dal QCER: A1, A2, B1 e B2. La piattaforma, inizialmente riservata alle classi coinvolte nella rilevazione linguistica di marzo 2018, rimarrà a disposizione di tutti gli Istituti del territorio provinciale per complessivi tre anni scolastici.

La piattaforma Demo presenta possibilità di utilizzo con finalità didattiche diversificate. Ogni prova, una per lingua e per livello, può essere utilizzata come prova unitaria di competenza linguistica tra classi parallele di uno stesso istituto oppure come test d'entrata/uscita per la definizione del livello linguistico in possesso degli studenti. La predisposizione della piattaforma, che permette al docente di ascoltare i file audio registrati dagli studenti e leggere le loro produzioni scritte, può risultare utile per una valutazione continuativa delle abilità produttive; infine utilizzando materiali autentici forniti dai docenti, gli studenti possono utilizzare la piattaforma per salvare e inviare al proprio docente i file audio e di scrittura.

Di seguito si riportano i tempi relativi alle singole parti delle prove, suddivisi per abilità e livello, caricate in piattaforma.

Tabella 4.4 - Piattaforma DEMO: durata delle prove per abilità e livello.

PROVE	LIVELLO A1	LIVELLO A2	LIVELLO B1	LIVELLO B2
Ascolto 1	5 minuti	5 minuti	9 minuti	10 minuti
Ascolto 2	5 minuti	5 minuti	11 minuti	10 minuti
Parlato 1	2 minuti	4 minuti	5 minuti	5 minuti
Parlato 2	3 minuti	4 minuti	5 minuti	5 minuti
Lettura 1	6 minuti	10 minuti	15 minuti	15 minuti
Lettura 2	9 minuti	15 minuti	15 minuti	15 minuti
Scrittura 1 (max. parole)	5 minuti (D&D)	12 minuti (60)	20 minuti (150)	25 minuti (200)
Scrittura 2 (max. parole)	15 minuti	18 minuti (120)	10 minuti (50)	15 minuti (100)
Durata complessiva	50 minuti	73 minuti	90 minuti	100 minuti

5 Analisi psicometrica delle prove TLT 2018

Maurizio Carpita²⁴

Introduzione

In questo capitolo si presenta una sintesi dei risultati ottenuti dall'analisi psicometrica effettuata utilizzando i dati raccolti con i sei test IPRASE per tre livelli linguistici (A1, A2 e B1) e due lingue (tedesco e inglese), predisposti per la seconda rilevazione delle competenze linguistiche degli studenti trentini ed effettuata nella primavera del 2018 (TLT – Trentino Language Testing 2018). Come nel caso dell'analisi svolta in occasione della prima esperienza pilota condotta con il TLT 2016²⁵, obiettivo della presente analisi è stato di verificare le proprietà statistiche delle prove linguistiche IPRASE del 2018.

La progettazione del nuovo TLT 2018 ha potuto beneficiare delle evidenze emerse a seguito della prima esperienza del 2016, nonché dei risultati ottenuti con il Test di try-out effettuato nell'autunno del 2017. Poiché gli approfondimenti e le riflessioni emerse nel primo biennio di attività del progetto TLT hanno portato alla revisione di parte dei contenuti linguistici di alcuni quesiti, le aspettative riguardo ai risultati che si sarebbero ottenuti del TLT 2018 erano positive.

Come nel caso del TLT 2016, anche per la nuova rilevazione condotta nel 2018 sono state utilizzate solo le risposte delle prove riconducibili a *test di prestazione massima standardizzati*, ovvero i quesiti di ASCOLTO e LETTURA, i quali richiedono agli studenti (per omogeneità non sono stati considerati quelli con bisogni

²⁴ Maurizio Carpita è Professore ordinario di Statistica presso l'Università degli Studi di Brescia.

²⁵ Carpita M., L'Analisi Psicometrica dei Test, in Covi L. e Dutto M.G. (2017). Rapporto TLT 2016 – Trentino Language Testing, Capitolo 5, Editore Provincia Autonoma di Trento – IPRASE.

educativi speciali – BES) di scegliere la risposta corretta tra un insieme di alternative.

L'analisi psicometrica di un test di prestazione massima, somministrato in forma standardizzata a un gruppo di studenti, viene normalmente condotta utilizzando un'ampia varietà di tecniche e di modelli statistici, che permettono di valutare la *coerenza* dei dati raccolti rispetto agli obiettivi e alla struttura della prova. Questo tipo di analisi si basa su una ben precisa *ipotesi fondamentale*: ogni studente possiede un certo livello di *abilità* (conoscenze e/o competenze) che si può osservare – seppur in modo impreciso – grazie alle risposte date alle domande formulate con diversa *difficoltà* inserite nel test. Se tale assunzione è vera, buona parte degli studenti meno abili sarà in grado di rispondere esattamente solo alle domande più facili sbagliando quasi certamente quelle più difficili, mentre buona parte degli studenti più abili sarà in grado di rispondere esattamente anche alle domande più difficili. È evidente che la situazione appena descritta si potrà verificare solo se il test è stato opportunamente *calibrato*, ovvero se le domande selezionate sono coerenti con l'ipotesi fondamentale: ad esempio, se tutte le domande del test sono troppo facili, tutti gli studenti saranno in grado di rispondere correttamente e non sarà quindi possibile valutarne l'abilità.

La teoria moderna che riguarda lo studio delle proprietà statistiche delle domande che compongono un test (IRT: *Item Response Theory*) suggerisce di riferirsi al *modello di Rasch*; tale modello, noto per la sua semplicità ed efficacia nel descrivere un *test unidimensionale* (ovvero che misura un'unica abilità), viene utilizzato da tempo nelle rilevazioni internazionali OCSE-PISA, nelle rilevazioni nazionali INVALSI ed è stato impiegato anche in occasione della rilevazione TLT 2016 di IPRASE. Ovviamente quello di Rasch non è l'unico modello possibile, non c'è garanzia che i dati siano conformi alle sue ipotesi e per questo è stato a volte criticato: l'analisi statistica delle risposte agli item che compongono il test rappresenta quindi un passaggio molto importante per capire quanto i suoi risultati siano coerenti con il modello.

Questo capitolo è così organizzato. Nel primo paragrafo si descrivono brevemente gli strumenti psicometrici utilizzati per valutare le caratteristiche dei Test TLT 2018 delle abilità linguistiche in inglese e tedesco degli studenti trentini. Nelle sei sezioni del secondo paragrafo (una per ogni livello di istruzione e per ogni lingua) si mostrano e si commentano i risultati ottenuti con tali tecniche statistiche applicate alle risposte date dagli studenti alle domande dei test IPRASE. Il capitolo si conclude con alcune considerazioni conclusive.

5.1 Gli strumenti di analisi psicometrica utilizzati

Gli indici e i modelli psicometrici utilizzati per verificare le caratteristiche statistiche delle risposte date alle domande (in seguito si usa il sinonimo *item*) di un *test di prestazione massima standardizzato*, si basano sul numero totale di risposte esatte (*punteggio*) fornite da ogni singolo studente (utile per valutare la sua *abilità*) e sul numero totale di risposte esatte fornite a ogni singola domanda (utile per valutare la sua *difficoltà*). Se le domande sono state formulate in modo coerente con gli obiettivi del test, si ritiene ragionevole che studenti con il punteggio più basso (alto) rispondano generalmente esattamente alle domande più facili (difficili). In questo paragrafo si descrivono brevemente gli strumenti utilizzati per valutare in che misura i test TLT 2018 predisposti da IPRASE per misurare le abilità linguistiche degli studenti trentini soddisfano le proprietà psicometriche desiderabili.

I due indici più semplici, che permettono di valutare in via preliminare la coerenza delle risposte date a ogni singolo item rispetto ai risultati del test nel suo complesso (punteggi ottenuti escludendo l'item considerato), sono chiamati *indici di difficoltà e discriminatività*. Per il calcolo di tali indici si considera solo la metà degli studenti appartenenti ai due gruppi che hanno ottenuto nel test il 25% dei risultati più alti (gruppo A) e il 25% dei risultati bassi (gruppo B) e sono così definiti e interpretati:

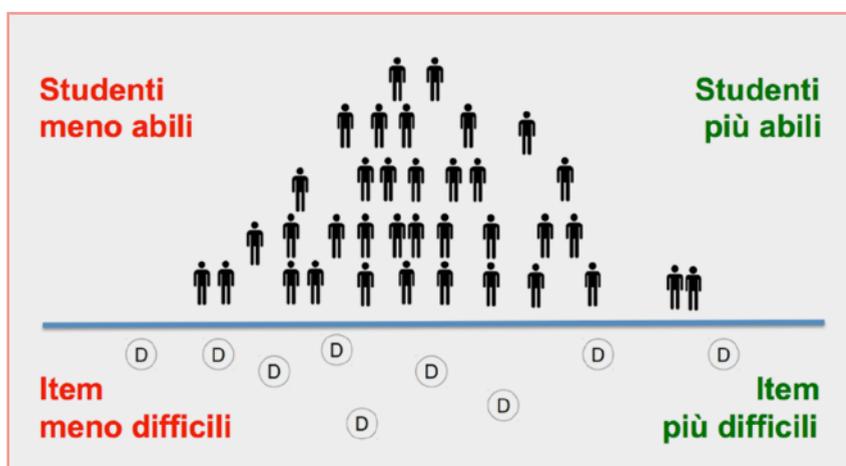
- **Indice DIF (*Difficoltà*):** è la percentuale degli studenti dei gruppi A e B che hanno sbagliato la risposta dell'item considerato. Si considera l'item troppo facile se l'indice DIF è inferiore al 10%, con il valore ottimale che si colloca tra il 25% e il 75%;
- **Indice DIS (*Discriminatività*):** è la differenza tra le percentuali di studenti dei gruppi A e B che hanno risposto esattamente all'item considerato. Generalmente si attribuisce all'item una insufficiente capacità discriminante se l'indice DIF è inferiore a 20 punti percentuali, una sufficientemente capacità discriminante se è compreso tra 20 e 30 punti e una buona capacità discriminante se è superiore a 30 punti.

Nel paragrafo successivo dedicato ai risultati, per offrire una lettura congiunta e visiva dei valori degli indici DIF e DIS, per ogni test TLT 2018 è stato creato appositamente un grafico che evidenzia con diversi colori i livelli di difficoltà e discriminatività degli item.

Come già spiegato nell'introduzione, il modello di Rasch è molto utilizzato nell'ambito dei test psicometrici, perché descrive in modo rigoroso le proprietà desiderabili di un *test unidimensionale*. Affinché un test sia conforme al modello di Rasch, le *difficoltà*

dei suoi item devono disporsi in modo ordinato sul *continuum* che descrive le *abilità* degli studenti, come rappresentato schematicamente nella Figura 5.1: ad esempio, i due studenti a sinistra sono i meno abili perché hanno una probabilità superiore al 50% di rispondere esattamente alle due domande meno difficili, mentre i due studenti a destra sono i più abili perché hanno una probabilità superiore al 50% di rispondere esattamente a quasi tutte le domande più difficili (fa eccezione l'ultima domanda, che è un po' troppo difficile per tutti gli studenti). In generale, un test è ben calibrato se abilità degli studenti e difficoltà degli item si collocano in un intervallo di variazione della scala simmetrico, da -3 a +3 *logit* (unità di misura di Rasch)²⁶.

Figura 5.1 – Mappa delle abilità degli studenti e delle difficoltà degli item conforme al modello di Rasch.



Va considerato che ognuno dei test TLT 2018 è stato progettato per rilevare un solo livello di competenza linguistica (ad esempio

²⁶ Il modello di Rasch può essere esteso in varie direzioni, in particolare inserendo per ogni item oltre al parametro di *difficoltà* anche il parametro di *discriminatività*: tale possibilità sembra utile nel caso dei Test TLT 2018 perché, come si vedrà nel paragrafo successivo, alcuni item mostrano indici DIS molto diversi dagli altri. Le verifiche effettuate hanno però sconsigliato l'utilizzo del modello a 2 parametri, sia perché gli item che mostrano DIS molto diversa sono pochi (non più di 2-3 per ogni test) e quindi l'ordinamento delle difficoltà degli item si modifica raramente, sia perché le stime delle probabilità di risposta corretta non cambiano.

il livello A2 in inglese), quindi gli item di ogni singolo test sono stati scelti con difficoltà abbastanza simili e con l'obiettivo di individuare la soglia di raggiungimento del livello considerato. Per questo motivo oltre al fatto che il campo di variazione della misura di Rasch risulta generalmente inferiore a quello usuale (6 *logit*), l'individuazione del numero di risposte corrette necessario al raggiungimento del livello di competenza linguistica considerato è stato individuato in corrispondenza del primo item che supera la difficoltà zero.

Oltre all'analisi della conformità dei test al modello di Rasch (che ha permesso di individuare eventuali item problematici), per ogni domanda è stata effettuata anche l'analisi dei cosiddetti *distrattori*, ovvero delle risposte inserite nel test insieme a quella corretta: dei risultati ottenuti con tale analisi si dirà brevemente nel paragrafo successivo.

Per valutare la precisione e l'accuratezza dei test linguistici IPRASE sono stati considerati due indici di *affidabilità (reliability)*:

- Indice REL.12: si basa sulla correlazione tra i due punteggi ottenuti dividendo il test in due parti (*metodo split-half*). Nel caso dei Test TLT 2018, la correlazione è stata calcolata tra i punteggi dei due esercizi previsti per ogni singola abilità, considerando come valore adeguato un indice REL.12 superiore al 50%;
- Indice REL.CO: è un'estensione del classico indice *Alpha di Cronbach*, che si usa quando non si assume che gli item siano *omogenei*. Nel caso dei test TLT 2018, si considera come valore adeguato un indice REL.CO superiore al 70%.

Infine, per valutare in che misura il modello di Rasch riesce a prevedere la risposta (corretta o errata) data da uno studente, sono stati utilizzati tre indici:

- Indice ACC (*accuratezza*): percentuale di classificazioni corrette del modello di Rasch per *tutte le risposte*;
- Indice SEN (*sensibilità*): percentuale di classificazioni corrette del modello di Rasch per le sole *risposte corrette*;
- Indice SPE (*specificità*): percentuale di classificazioni corrette del modello di Rasch per le sole *risposte errate*.

Per ulteriori approfondimenti e dettagli sulla metodologia adottata si rinvia al Capitolo 5 del precedente Rapporto TLT 2016 di IPRASE²⁷.

²⁷ Carpita M., *L'Analisi Psicometrica dei Test*, in Covi L. e Dutto M.G. (2017). Rapporto TLT 2016 – Trentino Language Testing, Capitolo 5, Editore Provincia Autonoma di Trento – IPRASE.

5.2 I risultati dell'analisi psicometrica

Nelle successive sei sezioni di questo paragrafo si presentano brevemente i risultati ottenuti dall'analisi psicometrica svolta sulle risposte fornite alle domande di ASCOLTO e LETTURA per le lingue INGLESE e TEDESCO contenute nei sei test TLT 2018 effettuati da IPRASE nella primavera dello scorso anno. Per ogni test, nella figura sono rappresentati i risultati ottenuti per ASCOLTO (a sinistra) e LETTURA (a destra); il primo grafico rappresenta gli indici DIFxDIS (congiuntamente) di ogni item, il secondo è la mappa abilità-difficoltà di Rasch e l'ultimo mostra gli indici di affidabilità e predittività.

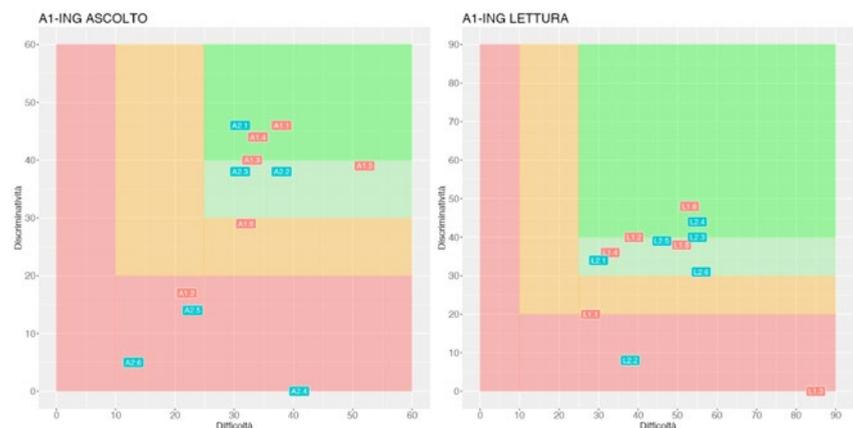
5.2.1 Risultati per il Test A1 di INGLESE

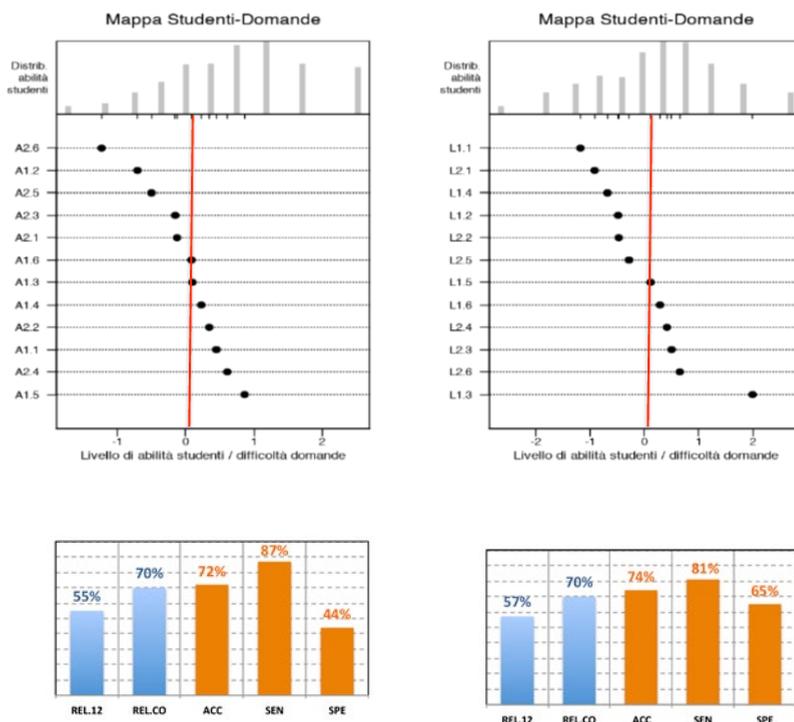
Il numero di studenti considerati è 477, con una media di risposte mancanti per item del 2% e un massimo del 7%, quindi influenti sugli esiti dell'analisi psicometrica.

Con riferimento all'abilità ASCOLTO, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 10% e il 50%, mentre in termini di *discriminatività* si evidenziano due item problematici (A2.4 con DIS = 0 e A2.6 con DIS = 5), due item con discriminatività sufficiente (A1.2 e A2.5) e i rimanenti 8 item con buona discriminatività, compresa tra 30 e 45 punti percentuali. La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una buona distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* (per brevità non riportata) segnala come problematico il solo item A2.4. L'*analisi dei distrattori* utilizzati nei singoli item (anche questa non riportata per brevità) segnala l'anomalia di A2.4, che mostra la presenza di un distrattore distorto per abilità medio-basse degli studenti. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (A1 e A2) risulta adeguata (REL.12 = 55%) mentre la coerenza complessiva raggiunge la soglia minima richiesta (REL.CO = 70%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 72% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 87%) rispetto alle risposte errate (SPE = 44%). La *soglia di abilità del test di ASCOLTO* per il livello A1 di INGLESE, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso.

Con riferimento all'abilità *LETTURA*, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 30% e il 60%, con l'unica eccezione di L1.3 che ha un livello di difficoltà molto elevato (85%); in termini di *discriminatività*, si evidenziano due item problematici (L1.3 con DIS = 0 e L2.2 con DIS = 9), un item con discriminatività sufficiente (L1.1) e i rimanenti 9 item con buona discriminatività, compresa tra 30 e 50 punti percentuali. La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una buona distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* conferma come problematici gli item L1.3 e L2.2. L'*analisi dei distrattori* utilizzati nei singoli item segnala l'anomalia di L1.3, che non mostra alcuna relazione tra abilità degli studenti e frequenze associate alle 4 risposte previste. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (L1 e L2) risulta adeguata (REL.12 = 57%) mentre la coerenza complessiva raggiunge la soglia minima richiesta (REL.CO = 70%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 74% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 81%) rispetto alle risposte errate (SPE = 65%). La *soglia di abilità del test di LETTURA* per il livello A1 di INGLESE, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso.

Figura 5.2 - Analisi psicometrica degli item - A1 INGLESE.





5.2.2 Risultati per il Test A1 di TEDESCO

Il numero di studenti considerati è 471, con una media di risposte mancanti per item del 2% e un massimo del 5%, quindi influenti sugli esiti dell'analisi psicometrica.

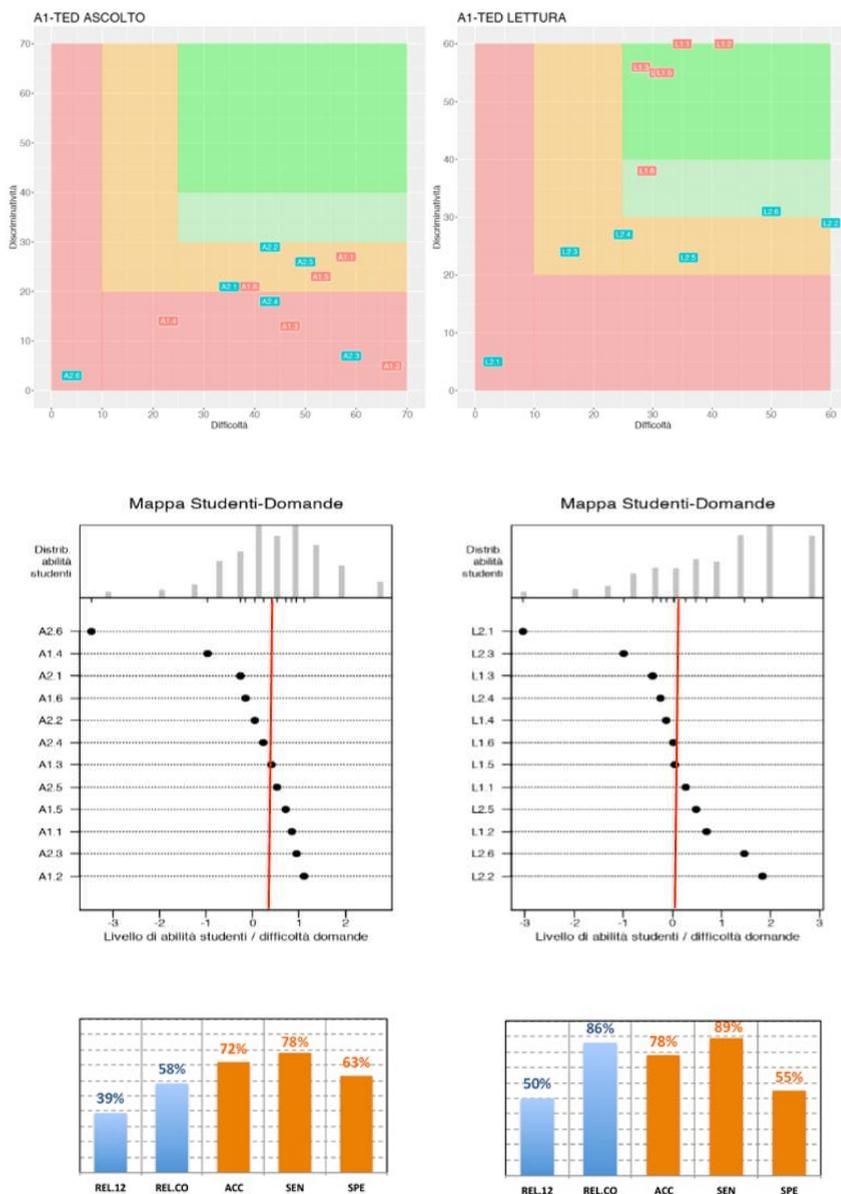
Con riferimento all'abilità *ASCOLTO*, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 20% e il 70%, con l'eccezione di A2.6 che ha difficoltà praticamente nulla; in termini di *discriminatività* si evidenziano tre item problematici (ancora A2.6 con DIS = 0, A1.2 con DIS = 5 e A2.3 con DIS = 8), tre item con discriminatività comunque insufficiente (A1.3, A1.4 e A2.4) e i rimanenti 6 item con sufficiente discriminatività, compresa tra 20 e 30 punti percentuali. La *mapa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) con l'unica eccezione di A2.6 che è evidentemente troppo facile e una buona distribuzione delle abilità degli studenti (barre in grigio); comunque, l'*analisi di adattamento del modello* (per brevità non riportata) non segnala item problematici. L'*analisi dei distrattori* utilizzati nei singoli item (anche questa non riportata per brevità)

non segnala anomalie. Per quanto riguarda *l'affidabilità del test*, la coerenza tra le due parti (A1 e A2) risulta sufficiente (REL.12 = 39%) mentre la coerenza complessiva non raggiunge la soglia minima richiesta (REL.CO = 58%). La *capacità predittiva del modello* è comunque discreta: l'esito delle singole risposte alle domande del test è previsto correttamente nel 72% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 78%) rispetto alle risposte errate (SPE = 63%). La *soglia di abilità del test di ASCOLTO* per il livello A1 di TEDESCO, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso²⁸.

Con riferimento all'abilità *LETTURA*, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 15% e il 60%, con l'unica eccezione di L2.1 che ha un livello di difficoltà prossimo a zero; in termini di *discriminatività*, solo l'item L2.1 è problematico (DIS = 5), mentre i rimanenti item hanno almeno sufficiente e in alcuni casi ottima discriminatività, compresa tra 30 e 60 punti percentuali. La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) con l'eccezione di L2.1 che è evidentemente troppo facile e una distribuzione delle abilità degli studenti (barre in grigio) che tende a concentrarsi su valori elevati; *l'analisi di adattamento del modello* segnala come problematici gli item L2.2, L2.5 e L2.6. *L'analisi dei distrattori* utilizzati nei singoli item segnala l'anomalia di L2.2, che ha un distrattore distorto per abilità medio-basse degli studenti. Per quanto riguarda *l'affidabilità del test*, la coerenza tra le due parti (L1 e L2) risulta adeguata (REL.12 = 50%) mentre la coerenza complessiva è molto buona (REL.CO = 86%). Anche la *capacità predittiva del modello* è buona ma sbilanciata: l'esito delle singole risposte alle domande del test è previsto correttamente nel 78% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 89%) rispetto alle risposte errate (SPE = 55%). La *soglia di abilità del test di LETTURA* per il livello A1 di TEDESCO, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso.

²⁸ Vista l'eccessiva facilità dell'item A2.6, la soglia è stata spostata dalla difficoltà dell'item A2.4 a quella dell'item A1.3. Comunque, rimane uguale il numero di studenti che supera la soglia (possiede il livello A1).

Figura 5.3 – Analisi psicometrica degli item del Test TLT 2018 – A1 TEDESCO.



5.2.3 Risultati per il Test A2 di INGLESE

Il numero di studenti considerati è 529, con una media di risposte mancanti per item dell'1% e un massimo del 4%, quindi influenti sugli esiti dell'analisi psicometrica.

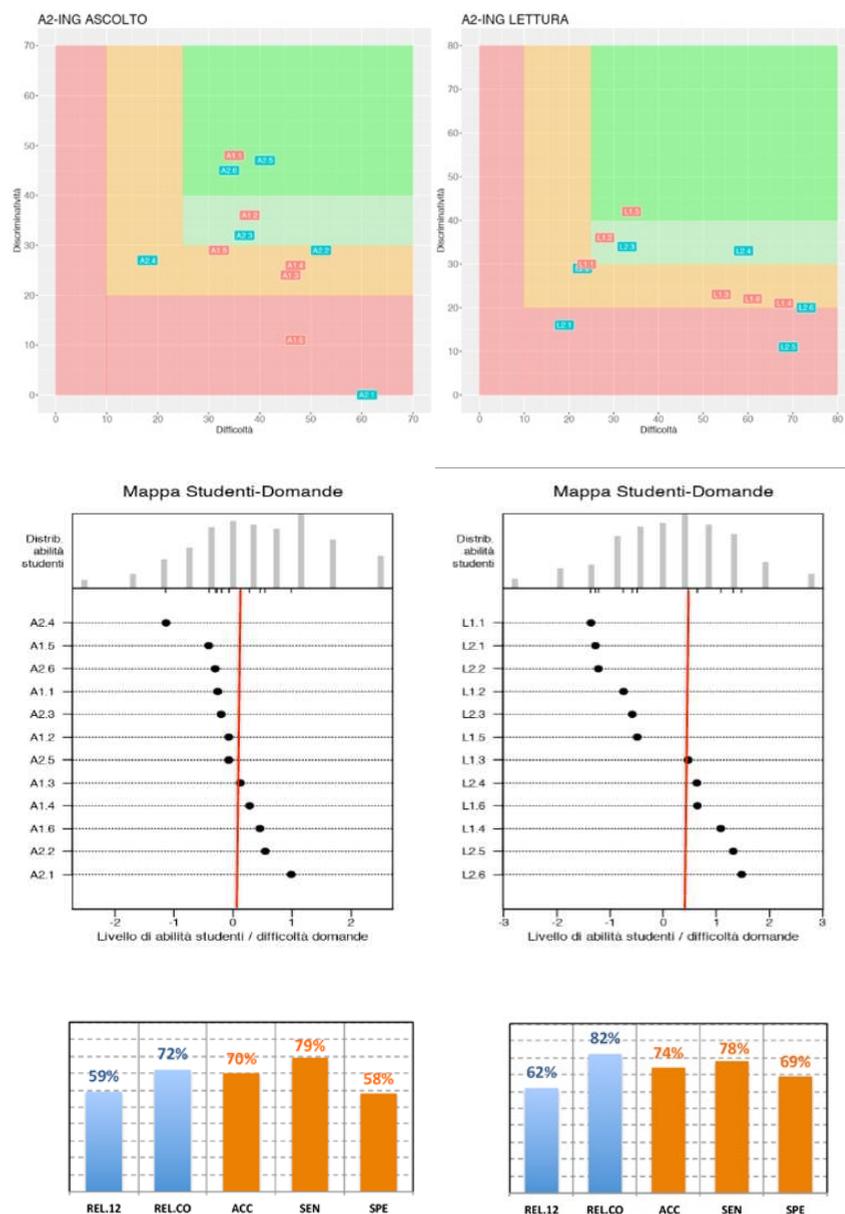
Con riferimento all'abilità ASCOLTO, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi

tra il 20% e il 50%, mentre in termini di *discriminatività* si evidenziano due item problematici (A2.1 con DIS = 0 e A1.6 con DIS = 10), 5 item con discriminatività sufficiente (20-30 punti percentuali) e i rimanenti 5 item con buona o ottima discriminatività (30-50 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una buona distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* (per brevità non riportata) conferma come problematici gli item A2.1 e A1.6. L'*analisi dei distrattori* utilizzati nei singoli item (anche questa non riportata per brevità) conferma l'anomalia di A2.1, che mostra la presenza di due distrattore con andamenti distorti e analoghi a quello della risposta corretta. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (A1 e A2) risulta adeguata (REL.12 = 59%) mentre la coerenza complessiva supera la soglia minima richiesta (REL.CO = 72%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 70% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 79%) rispetto alle risposte errate (SPE = 58%). La *soglia di abilità del test di ASCOLTO* per il livello A2 di INGLESE, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 8 come numero minimo di risposte corrette atteso.

Con riferimento all'abilità *LETTURA*, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 20% e il 70%; in termini di *discriminatività*, si evidenziano due item relativamente problematici (L2.5 con DIS = 10 e L2.1 con DIS = 18), 6 item con discriminatività sufficiente (20-30 punti percentuali) e i rimanenti 4 item con buona discriminatività (30-40 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) – con un "salto" di abilità intorno a zero – e una buona distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* non segnala item problematici. L'*analisi dei distrattori* utilizzati nei singoli item conferma l'anomalia di L2.5, con un distrattore che maschera la risposta esatta per abilità degli studenti medio-alte. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (L1 e L2) risulta adeguata (REL.12 = 50%) mentre la coerenza complessiva supera la soglia minima richiesta (REL.CO = 72%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 75% dei casi (ACC), però con un buon equilibrio tra la capacità di prevedere le risposte cor-

rette (SEN = 78%) e le risposte errate (SPE = 71%). La *soglia di abilità del test di LETTURA* per il livello A2 di INGLESE, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso.

Figura 5.4 - Analisi psicometrica degli item - A2 INGLESE.



5.2.4 Risultati per il Test A2 di TEDESCO

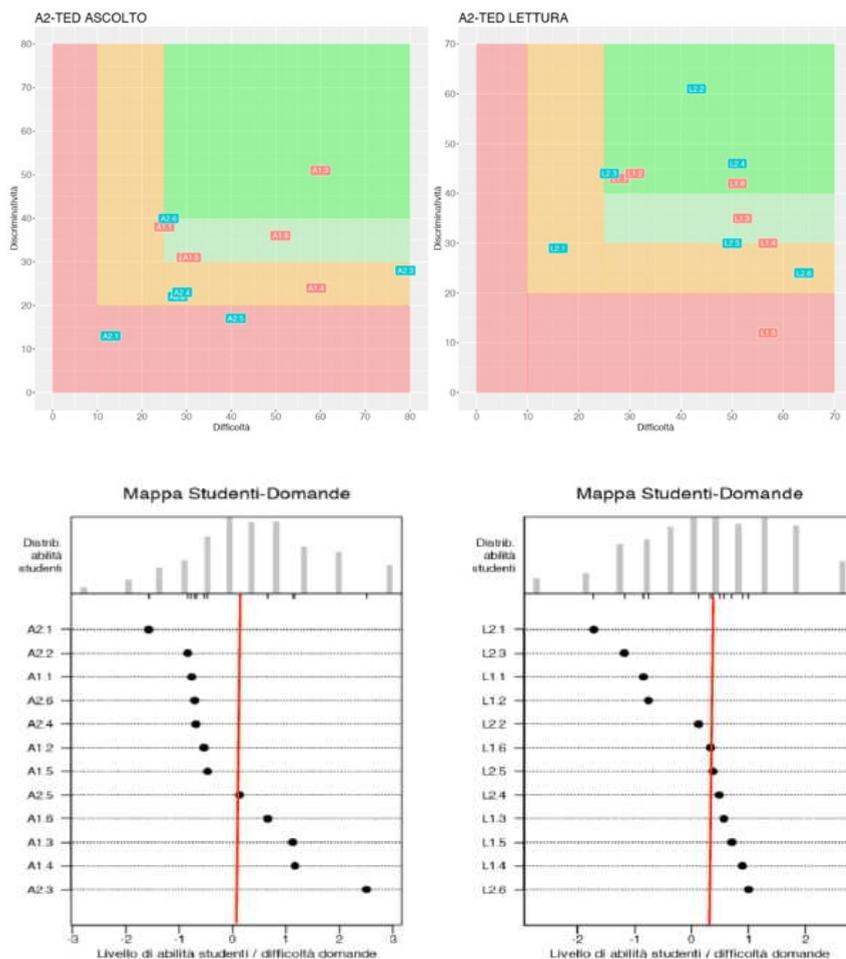
Il numero di studenti considerati è 533, con una media di risposte mancanti per item del 2% e un massimo del 4%, quindi influenti sugli esiti dell'analisi psicometrica.

Con riferimento all'abilità *ASCOLTO*, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 10% e il 60% e con l'item A2.3 molto difficile (DIF = 80%), mentre in termini di *discriminatività* si evidenziano due item relativamente problematici (A2.1 con DIS = 13 e A2.5 con DIS = 18), 4 item con discriminatività sufficiente (20-30 punti percentuali), 6 item con buona discriminatività (30-40 punti percentuali) e un item (A1.3) con ottima discriminatività (50 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una buona distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* (per brevità non riportata) conferma come problematico il solo item A2.5. L'*analisi dei distrattori* utilizzati nei singoli item (anche questa non riportata per brevità) segnala l'anomalia di A2.3, che mostra la presenza di un distrattore distorto che maschera la risposta corretta anche per abilità elevate degli studenti. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (A1 e A2) risulta adeguata (REL.12 = 57%) mentre la coerenza complessiva supera la soglia minima richiesta (REL.CO = 77%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 76% dei casi (ACC), con maggiore capacità di prevedere le risposte corrette (SEN = 84%) rispetto alle risposte errate (SPE = 66%). La *soglia di abilità del test di ASCOLTO* per il livello A2 di TEDESCO, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 8 come numero minimo di risposte corrette atteso.

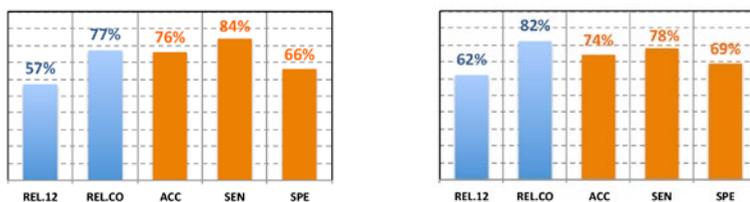
Con riferimento all'abilità *LETTURA*, il grafico DIFxDIS mostra che i 12 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 15% e il 65%; in termini di *discriminatività*, si evidenzia un solo item problematico (L1.5 con DIS = 12), due item con discriminatività sufficiente (20-30 punti percentuali) e 9 item con buona discriminatività (3 item con DIS tra 30 e 40 punti percentuali) o ottima discriminatività (6 item con DIS tra 30 e 50 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una buona distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* conferma come problematico l'item L1.5. L'*analisi dei distrattori* utilizzati nei singoli item segnala la presenza nell'item L2.6 di un distrattore distorto anche per abilità degli studenti medio-alte. Per

quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (L1 e L2) risulta adeguata ($REL.12 = 62\%$), così come adeguata la coerenza complessiva ($REL.CO = 82\%$). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 74% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette ($SEN = 78\%$) rispetto alle risposte errate ($SPE = 69\%$). La *soglia di abilità del test di LETTURA* per il livello A2 di TEDESCO, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso²⁹.

Figura 5.5 - Analisi psicometrica degli item - A2 TEDESCO.



²⁹ Poiché si evidenzia un "vuoto" tra le difficoltà dei primi 4 item e i successivi, invece della difficoltà dell'item L2.2 la soglia è stata posta in corrispondenza di quella dell'item L2.5. Comunque, rimane uguale il numero di studenti che supera la soglia (possiede il livello A2).



5.2.5 Risultati per il Test B1 di INGLESE

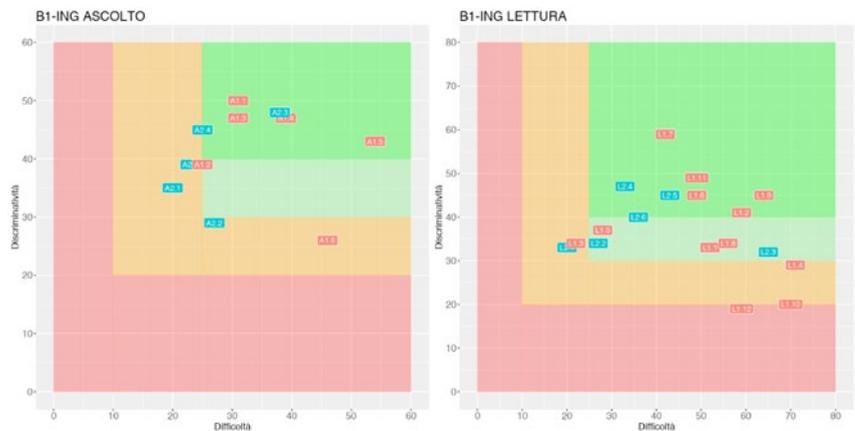
Il numero di studenti considerati è 1.336, una media di risposte mancanti per item del 3%, con massimo 18% per l'item L1.12 e anche l'item L1.11 ha una percentuale di *missing* relativamente elevata (12%): entrambi i quesiti sono della parte di test che conta ben 12 item, quindi non hanno nel complesso influenza sugli esiti dell'analisi psicometrica.

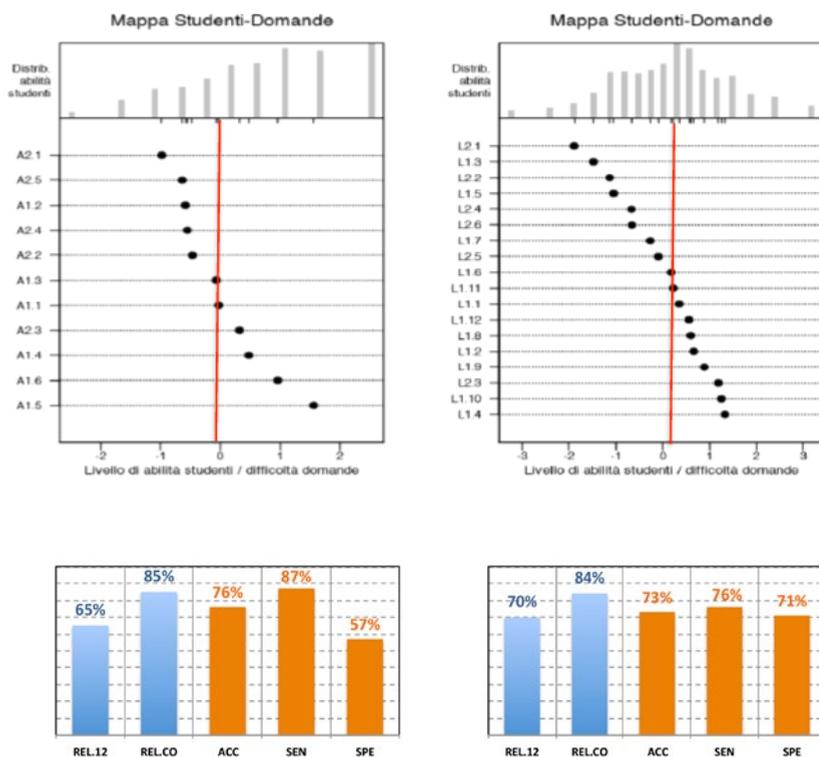
Con riferimento all'abilità *ASCOLTO*, il grafico DIFxDIS mostra che gli 11 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 20% e il 55%, mentre in termini di *discriminatività* si evidenziano 2 item con discriminatività sufficiente (20-30 punti percentuali), 3 item con buona discriminatività (30-40 punti percentuali) e i rimanenti 6 item con ottima discriminatività (40-50 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una distribuzione delle abilità degli studenti più concentrata sulle abilità elevate (barre in grigio); l'*analisi di adattamento del modello* (per brevità non riportata) conferma come problematico il solo item A1.6. L'*analisi dei distrattori* utilizzati nei singoli item (anche questa non riportata per brevità) segnala l'anomalia di A1.5, che mostra la presenza di un distrattore distorto che maschera la risposta corretta anche per abilità elevate degli studenti. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (A1 e A2) risulta adeguata (REL.12 = 65%) mentre la coerenza complessiva supera ampiamente la soglia minima richiesta (REL.CO = 85%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 76% dei casi (ACC), con maggiore capacità di prevedere le risposte corrette (SEN = 87%) rispetto alle risposte errate (SPE = 57%). La *soglia di abilità del test di ASCOLTO* per il livello B1 di INGLESE, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 7 come numero minimo di risposte corrette atteso.

Con riferimento all'abilità *LETTURA*, il grafico DIFxDIS mostra che i 18 item utilizzati per il test hanno livelli di *difficoltà* comprese

si tra il 20% e il 70%; in termini di *discriminatività*, si evidenzia un solo item sulla soglia problematica (L1.12 con DIS = 20), 2 item con discriminatività sufficiente (20-30 punti percentuali), 7 item con buona discriminatività (30-40 punti percentuali) e 8 item con ottima discriminatività (40-60 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un ottimo ordinamento delle difficoltà degli item (pallini neri) e un'ottima distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* segnala come relativamente problematici gli item L1.10 e L1.11. L'*analisi dei distrattori* utilizzati nei singoli item segnala la presenza negli item L1.4, L1.9 e soprattutto L1.10 di un distrattore distorto anche per abilità degli studenti medio-alte. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (L1 e L2) risulta adeguata (REL.12 = 70%), così come adeguata la coerenza complessiva (REL.CO = 84%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 73% dei casi (ACC), con buon bilanciamento tra capacità di prevedere le risposte corrette (SEN = 76%) e risposte errate (SPE = 71%). La *soglia di abilità del test di LETTURA* per il livello B1 di INGLESE, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 10 come numero minimo di risposte corrette atteso.

Figura 5.6 - Analisi psicométrica degli item - B1 INGLESE.





5.2.6 Risultati per il Test B1 di TEDESCO

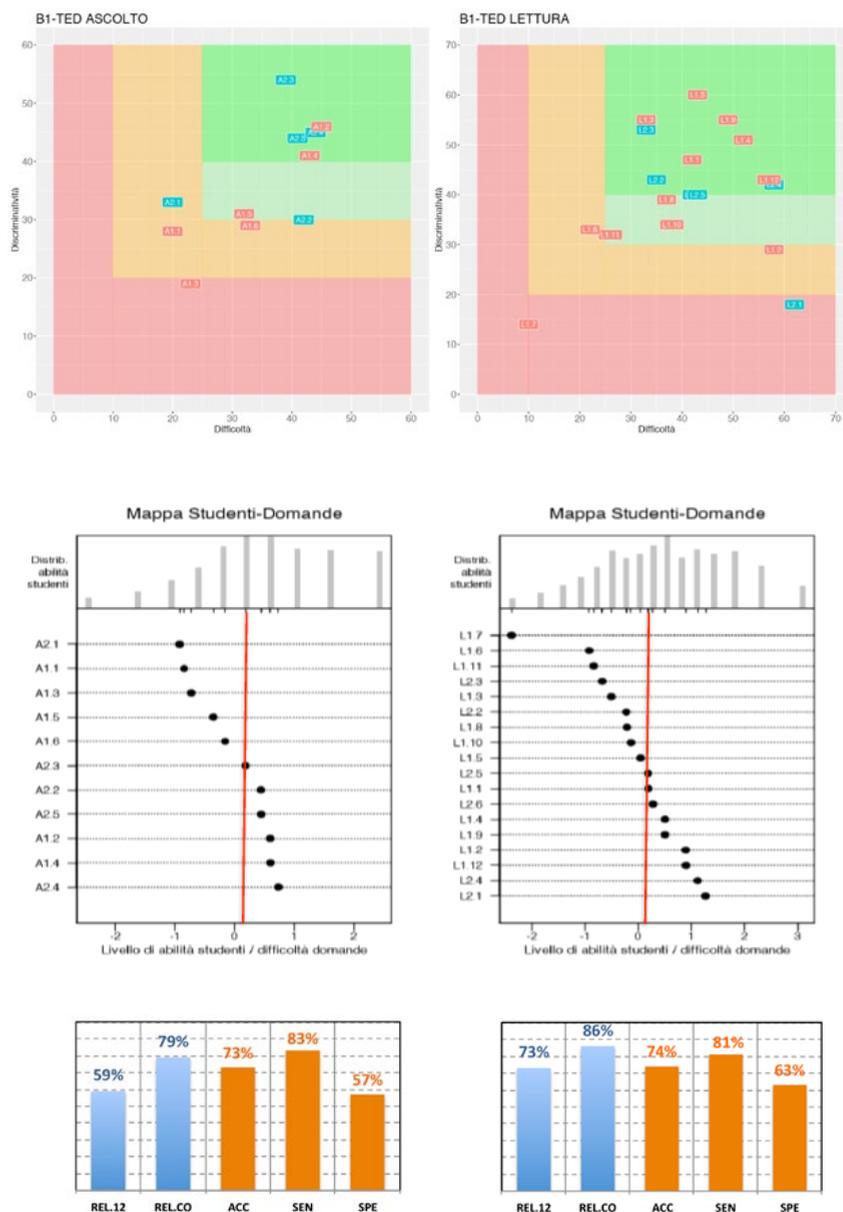
Il numero di studenti considerati è 1.046, con una media di risposte mancanti per item del 2% e un massimo del 8%, quindi influenti sugli esiti dell'analisi psicometrica.

Con riferimento all'abilità *ASCOLTO*, il grafico DIFxDIS mostra che gli 11 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 20% e il 45%, mentre in termini di *discriminatività* si evidenzia un solo item (A1.3) con discriminatività sulla soglia problematica (DIS = 18), 2 item con discriminatività sufficiente (20-30 punti percentuali), 3 item con buona discriminatività (30-40 punti percentuali) e i rimanenti 5 item con ottima discriminatività (40-55 punti percentuali). La *mappa studenti-domande* ottenuta con il modello di Rasch evidenzia un buon ordinamento delle difficoltà degli item (pallini neri) e una distribuzione delle abilità degli studenti più concentrata sulle abilità elevate (barre in grigio); l'*analisi di adattamento del modello* (per brevità non riportata) indica come problematici gli item A2.1 e A2.3. L'*analisi dei distrattori* utilizzati nei singoli item (anche questa non riportata per brevità) non segnala sostanziali anomalie. Per quanto riguarda l'*affidabili-*

tà del test, la coerenza tra le due parti (A1 e A2) risulta adeguata (REL.12 = 59%) mentre la coerenza complessiva raggiunge la soglia minima richiesta (REL.CO = 79%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 73% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 83%) rispetto alle risposte errate (SPE = 57%). La *soglia di abilità del test di ASCOLTO* per il livello B1 di TEDESCO, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 6 come numero minimo di risposte corrette atteso.

Con riferimento all'abilità *LETTURA*, il grafico DIFxDIS mostra che i 18 item utilizzati per il test hanno livelli di *difficoltà* compresi tra il 20% e il 60%, con l'item L1.7 che risulta molto facile (DIF = 10%); in termini di *discriminatività*, si evidenziano 2 item sotto la soglia problematica (oltre a L1.7 con DIS = 14 e L2.1 con DIS = 18), 1 item con discriminatività sufficiente (DIS = 28), 4 item con buona discriminatività (30-40 punti percentuali) e i rimanenti 11 item con ottima discriminatività (40-60 punti percentuali). La *mapa studenti-domande* ottenuta con il modello di Rasch evidenzia un ottimo ordinamento delle difficoltà degli item (pallini neri) – con l'eccezione del già segnalato L1.7 (troppo facile) – e un'ottima distribuzione delle abilità degli studenti (barre in grigio); l'*analisi di adattamento del modello* segnala come relativamente problematici gli item L1.2 e L2.1. L'*analisi dei distrattori* utilizzati nei singoli item segnala solo la presenza nell'item L2.1 di un distrattore distorto anche per abilità degli studenti medio-alte. Per quanto riguarda l'*affidabilità del test*, la coerenza tra le due parti (L1 e L2) risulta adeguata (REL.12 = 73%), così come più che adeguata la coerenza complessiva (REL.CO = 86%). La *capacità predittiva del modello* è buona: l'esito delle singole risposte alle domande del test è previsto correttamente nel 74% dei casi (ACC), però con una maggiore capacità di prevedere le risposte corrette (SEN = 81%) rispetto alle risposte errate (SPE = 63%). La *soglia di abilità del test di LETTURA* per il livello B1 di TEDESCO, fissata in corrispondenza del primo item che supera il livello di abilità zero stimato dal modello di Rasch, individua 11 come numero minimo di risposte corrette atteso.

Figura 5.7 - Analisi psicometrica degli item - B1 TEDESCO.



5.3 Considerazioni conclusive

In questo capitolo è stata presentata la sintesi dell'analisi psicometrica effettuata sui risultati della rilevazione dei tre livelli (A1, A2 e B1) delle due competenze linguistiche (ASCOLTO e LETTURA) inserite nei due Test TLT 2018 (TEDESCO e INGLESE). La progettazione dei test predisposti da IPRASE per la seconda

rilevazione effettuata nella primavera del 2018 ha potuto beneficiare sia dell'esito della rilevazione pilota svolta con il TLT 2016 sia delle evidenze emerse a seguito del test di try-out condotto nell'autunno del 2017: si era quindi fiduciosi che l'esperienza acquisita potesse migliorare l'esito della rilevazione, rendendola più stabile e precisa.

L'analisi psicometrica è stata condotta utilizzando alcune delle procedure già impiegate in occasione la prima rilevazione. Tali procedure permettono di valutare la coerenza dei dati raccolti rispetto all'ipotesi fondamentale che ogni studente possiede un certo livello di abilità che si può rilevare grazie alle risposte date a domande inserite nel test che presentano dal punto di vista linguistico difficoltà diverse, ma calibrate rispetto al livello di competenza linguistica considerato. Al fine della verifica della calibrazione del TLT 2018 sono stati utilizzati due indici di *item analysis* (utili per valutare difficoltà e discriminatività dei quesiti), le stime delle abilità degli studenti e delle difficoltà degli item ottenute grazie al noto *modello di Rasch*, due *indici di affidabilità* (utili per valutare la coerenza interna dei test) e tre *indici di capacità predittiva* (utili per verificare in che misura, date le abilità degli studenti, il modello di Rasch è in grado di prevedere correttamente gli esiti del test).

I risultati riportati e commentati nel terzo paragrafo per ognuno dei sei test possono essere così sintetizzati. Per quanto riguarda i due test di livello A1, considerato che sono stati somministrati a studenti della V primaria, ovvero molto più giovani di quelli usualmente considerati nei test di prestazione massima, gli esiti possono considerarsi molto buoni: a parte qualche domanda troppo facile/difficile o con scarsa capacità discriminante, complessivamente le difficoltà degli item risultano ben distribuite, con una discreta attendibilità e una buona capacità predittiva. I due test di livello A2 mostrano caratteristiche psicometriche migliori, con solo un paio di item deboli, che comunque non impediscono di ottenere una buona distribuzione delle difficoltà, con buona attendibilità e capacità predittiva. Analoghe considerazioni possono essere fatte anche per i due test di livello B1, con una attendibilità ancora superiore. Vista la buona calibrazione raggiunta dai test, la soglia necessaria per il raggiungimento del livello di abilità previsto è stata collocata vicino al centro (zero) delle difficoltà degli item stimate con il modello di Rasch: poiché in corrispondenza di tale *abilità media* il numero di risposte corrette atteso risulta 7 su 12 (solo per A2 di ASCOLTO sarebbe 8 su 12) oppure 11 su 18 (per B1 di LETTURA), la soglia percentuale di risposte corrette attese può essere ragionevolmente fissata al 60%.

In conclusione, dal punto di vista psicometrico è chiara la robustezza dei test TLT 2018, meglio bilanciati e più stabili di quelli utilizzati per la prima rilevazione effettuata due anni prima. Appare quindi evidente che l'esperienza acquisita da IPRASE, non solo con la prima rilevazione pilota ma anche grazie al test di try-out svolto nell'autunno del 2017, ha permesso di migliorare in modo significativo la precisione del livello di misurazione delle competenze linguistiche degli studenti trentini.

6 Sviluppo di un sistema automatico per la valutazione delle abilità linguistiche

Giuseppe Falavigna³⁰, Roberto Gretter, Marco Matassoni³¹,
Nadia Mana, Ornella Mich³²

Introduzione

Il problema della valutazione automatica di una seconda lingua è ampiamente studiato nella comunità scientifica nell'ambito dell'area di ricerca nota come *Computer Assisted Language Learning* (CALL). In questo contesto sono stati sviluppati, e sono tuttora argomento di studio, metodi di valutazione automatici che operano su due distinte modalità di input: scritta e parlata. In entrambi i casi vengono valutate, indipendentemente, specifiche competenze linguistiche (ad es. la correttezza grammaticale, la ricchezza lessicale, la comprensione, ecc.) degli studenti della seconda lingua (*L2 learners*) attraverso l'uso di appropriati sistemi di classificazione. Le singole competenze vengono quindi raggruppate in maniera da formare un punteggio complessivo congruente con una qualche scala di valutazione standard. Una delle scale di valutazione molto diffusa è la *Common European Framework of Reference for Languages* (CEFR, Council of Europe, 2001). La scala CEFR, utilizzata in questo lavoro, definisce 6 livelli di abilità linguistica: A1 (principiante), A2, B1, B2, C1 e C2.

Il sistema sviluppato si basa su un insieme di *features* che vengono estratte:

1. dalla voce degli studenti, in risposta alle domande del test;
2. dalle risposte scritte degli studenti alle domande del test;
3. dalla trascrizione automatica delle risposte orali.

³⁰ Daniele Falavigna è a capo dell'unità di ricerca SpeechTeK presso FBK.

³¹ Roberto Gretter e Marco Matassoni sono ricercatori del gruppo di ricerca SpeechTeK presso FBK.

³² Nadia Mana e Ornella Mich sono ricercatrici dell'unità di ricerca i3-Intelligent Interfaces and Interaction presso FBK.

Queste features, nella fattispecie degli insiemi di valori numerici, vengono elaborate da dei sistemi di classificazione basati sull'impiego di reti neurali, addestrate utilizzando come riferimento i punteggi forniti dagli esperti umani che hanno valutato manualmente ciascun indicatore di abilità linguistica.

La valutazione automatica delle rilevazioni linguistiche svolte da Iprase rappresenta una sfida tecnologica e metodologica allo stesso tempo. Infatti, bisogna considerare che i test di "proficiency" vengono svolti in aule (tipicamente le aule di informatica) dove coppie di computers sono poste sullo stesso banco e i microfoni rimangono aperti per lunghi intervalli di tempo (al momento non esiste la possibilità di interrompere la registrazione quando i ragazzi hanno finito di parlare). I problemi posti sono quindi molteplici, dal punto di vista del solo riconoscimento automatico della voce (*Automatic Speech Recognition*, ASR) le maggiori difficoltà sono dovute ai seguenti fattori:

- a) riconoscimento di voci di bambini e di parlatori non nativi, cioè studenti italiani che cercano di parlare inglese e tedesco;
- b) presenza di un numero molto elevato nelle risposte di fenomeni di parlato spontaneo (esitazioni, false partenze, frammenti di parole, ecc);
- c) presenza di parole pronunciate in lingue diverse (parole inglesi, italiane e tedesche vengono spesso pronunciate in risposta a una singola domanda);
- d) presenza di un elevato rumore di fondo dovuto al fatto che, come sopra menzionato, il microfono rimane aperto per un lungo intervallo di tempo anche dopo che la risposta è stata completata;
- e) presenza di parlatori non collaborativi (gli studenti spesso scherzano, ridono, parlano sottovoce).

Per superare e/o limitare le problematiche sopra elencate sono stati utilizzate alcune soluzioni abbastanza innovative, essenzialmente basate sull'impiego di algoritmi che consentono di "adattare" modelli standard (ad es. modelli acustici o modelli del linguaggio usati in ASR) al particolare dominio.

In particolare, si è addestrata una rete neurale (*Deep Neural Network*, DNN), per il modello acustico del ASR, utilizzando dati multilingua: inglese, tedesco e italiano. In questa maniera il modello acustico è in grado di riconoscere frasi che contengono parole pronunciate nelle tre diverse lingue. Questa modifica al modello acustico, rispetto all'approccio standard impiegato in ASR, ha anche richiesto di usare un lessico fonetico comune alle tre lingue: inglese, italiano e tedesco. Tale lessico è stato definito in termini di unità dell'alfabeto fonetico internazionale IPA (Inter-

national Phonetic Alphabet). Soprattutto, si sono utilizzate unità acustiche che rappresentano esplicitamente i fenomeni di parlato spontaneo (risate, esitazioni, ecc.) osservati nelle registrazioni audio. Ciò ha consentito di migliorare sensibilmente le prestazioni del sistema ASR e, conseguentemente, del sistema complessivo di valutazione automatica. Si è anche fatto uso di informazioni fornite dal “funzionamento interno” del sistema ASR, in particolare valori di probabilità acustiche e durate dei singoli fonemi. Infine, abbiamo usato modelli del linguaggio che consentono di riconoscere gli errori più frequenti commessi dagli studenti. Tali modelli sono stati usati sia per migliorare le prestazioni del sistema ASR, sia per estrarre le features linguistiche usate per classificare gli insiemi delle risposte scritte.

6.1 Descrizione dei dati

Complessivamente, la campagna di valutazione consiste di tre fasi, previste nel 2016, 2018 e 2020. Ciascuna di esse coinvolge circa 3000 studenti di età compresa tra i 9 ed i 16 anni, appartenenti a 4 diversi livelli scolastici e tre livelli di competenza (A1, A2, B1). La campagna 2018 è stata suddivisa in due parti: un try-out, effettuato nel 2017 su circa 500 studenti, seguito dalla valutazione del 2018 che ha coinvolto circa 2500 studenti. Per analizzare le capacità linguistiche degli studenti sono stati condotti diversi test utilizzando domande con risposte a scelta multipla, che possono essere valutate automaticamente. Naturalmente il test comprende anche produzioni scritte e parlate fornite dagli studenti in risposta a qualche stimolo, la cui valutazione richiede l'intervento di esperti. Lo studio di FBK ha considerato esclusivamente tali dati, cercando di costruire un sistema automatico capace di fornire una valutazione che si avvicini alla valutazione degli esperti.

Nelle tabelle che seguono sono riportati i dati dei ragazzi che hanno partecipato alla fase di produzione di testi, scritti e parlati, suddivisi per livello (A1, A2, B1a, B1b). Tabella 6.1 riporta il numero di studenti, mentre la Tabella 6.2 riporta il numero di frasi, scritte o parlate.

Tabella 6.1 - Numero di studenti che hanno partecipato alle sessioni scritte e parlate, per ogni campagna (2016, 2017, 2018) e per livello (A1: V Primaria; A2: III SSPG; B1a: II SSSG; B1b: III IeFP).

ANNO	SCRITTO					PARLATO				
	A1	A2	B1-SSSG	B1-IEFP	TOT	A1	A2	B1-SSSG	B1-IEFP	TOT
2016	1074	1520	376	141	3111	1074	1520	378	140	3112
2017	111	320	124	0	555	105	316	121	0	542
2018	517	614	1112	465	2708	508	593	1086	364	2551

Tabella 6.2 - Numero di frasi, scritte o pronunciate dagli studenti, per ogni campagna (2016, 2017, 2018) e per livello (A1: V Primaria; A2: III SSPG; B1a: II SSSG; B1b: III IeFP).

ANNO	SCRITTO					PARLATO				
	A1	A2	B1-SSSG	B1-IEFP	TOT	A1	A2	B1-SSSG	B1-IEFP	TOT
2016	2050	5680	1325	324	9379	4099	5694	1337	322	11452
2017	419	615	469	0	1503	1032	5440	1337	0	7809
2018	1004	2340	4179	924	8447	8750	5663	13067	1846	29326

Le tabelle 6.3 e 6.4 riportano dati relativi alla quantità di materiale raccolto durante le varie campagne, in termini di frasi scritte e di registrazioni audio.

Tabella 6.3 - Quantità di materiale pronunciato dagli studenti, per ogni campagna (2016, 2017, 2018) e per lingua.

ANNO	LINGUA	NUMERO STUDENTI	NUMERO REGISTRAZIONI AUDIO	DURATA REGISTRAZIONI (HH:MM:SS)	NUMERO DOMANDE DIVERSE
2016	inglese	2748	17462	69:03:37	85
2016	tedesco	2542	15866	60:03:01	101
2017	inglese	511	4112	16:25:45	24
2017	tedesco	478	3739	15:33:06	23
2018	inglese	2332	15770	93:14:53	24
2018	tedesco	2072	13658	95:54:56	23

Tabella 6.4 - Quantità di materiale scritto dagli studenti, per ogni campagna (2016, 2017, 2018) e per lingua.

ANNO	LINGUA	NUMERO STUDENTI	NUMERO FRASI SCRITTE	NUMERO PAROLE SCRITTE	NUMERO DOMANDE DIVERSE
2016	inglese	3074	5062	299138	20
2016	tedesco	2870	4658	192144	25
2017	inglese	533	758	37225	5
2017	tedesco	529	745	30802	5
2018	inglese	2560	4600	293958	5
2018	tedesco	2200	3889	202309	5

Dal punto di vista del riconoscimento vocale i seguenti fattori sono quelli che impattano maggiormente sulle prestazioni del sistema:

- le voci da riconoscere sono relative a ragazzi di età variabile, compresa tra i 9 e i 16 anni (i modelli acustici sono spesso addestrati su voci di adulti, i dati disponibili per ragazzi e bambini sono relativamente pochi; la voce dei ragazzi in quelle fasce di età si modifica in maniera sensibile);
- il parlato è non nativo (italiani che parlano inglese e tedesco);
- le registrazioni sono caratterizzate da un gran numero di fenomeni spontanei (esitazioni, false partenze, risate, parole interrotte, ecc.);
- si nota spesso presenza di più lingue nella stessa frase (a una risposta attesa in inglese possono comparire parole in italiano – “non so come si dice carota” – o, per errore, in tedesco);
- presenza di un livello significativo di rumore di fondo dovuto al fatto che il microfono rimane aperto per un intervallo di tempo fisso e, dopo che la frase è finita, ogni cosa detta dallo studente o dai suoi vicini viene comunque registrato e non è separabile dalla parte utile della registrazione;
- presenza di soggetti non collaborativi (gli studenti spesso scherzano, ridono, parlano sottovoce, ecc.).

Le risposte scritte sono invece caratterizzate da un alto livello di rumore dovuto a errori di ortografia, inserimento di frammenti di parole, presenza di parole appartenenti a più lingue, presenza di risposte incoerenti (ad esempio battute, commenti non relativi alle domande, sequenze strampalate di caratteri, ecc.).

Per ogni frase, scritta o pronunciata, un singolo esperto ha dato una votazione composita, tipicamente con valori 0-1-2 (scarso, medio, buono) per ogni singolo indicatore. A ogni famiglia di domande corrisponde un numero variabile di indicatori (tra 3 e 6), asso-

ciati a diverse caratteristiche (rilevanza della risposta, pronuncia, correttezza sintattica, proprietà lessicali, ecc.). La somma dei valori ottenuti per i singoli indicatori fornisce il punteggio complessivo (che assume valori tra 0 e 12, nel caso di 6 indicatori). Per i dati audio del 2016, tuttavia, nel caso in cui una sessione fosse composta da più registrazioni (tipicamente brevi), è disponibile solo la media degli indicatori calcolata sui singoli indicatori di ogni registrazione.

6.2 Sistema ASR

Il riconoscimento automatico del parlato viene investigato da molti anni e al giorno d'oggi la tecnologia si presenta matura per offrire svariati servizi, dai sistemi telefonici di risposta automatica agli assistenti vocali che stanno entrando da poco nel mercato consumer. Purtroppo realizzare un sistema di riconoscimento per il contesto applicativo in analisi presenta due problemi particolarmente complessi:

- il parlato L2: come appare evidente, un parlatore non madrelingua tende a introdurre diversi fenomeni nella propria produzione orale (errori di pronuncia, uso scorretto della grammatica, uso di parole "ibride" derivate dalla lingua madre);
- il parlato non-adulto: la voce degli studenti nella fascia 10-16 presenta caratteristiche diverse dalla voce adulta e richiede una modellizzazione specifica (ovvero richiede un addestramento con dati significativi di questa particolare popolazione).

Unitamente a questi due fattori generali abbiamo dovuto considerare alcune peculiarità per la realizzazione del sistema di riconoscimento utilizzato per le trascrizioni delle produzioni orali nel TLT 2018:

- l'elevata variabilità della qualità delle registrazioni, dovuta a differenze marcate nell'attrezzatura di registrazione, alla relativa configurazione (es: il volume di registrazione), alla presenza di rumori di fondo;
- la presenza di fenomeni *cross-language* ovvero la possibilità di trovare espressioni appartenenti non alla lingua L2 di riferimento del test ma all'altra lingua L2 (es. uso di alcune parole in tedesco in risposte a domande in inglese);
- presenza di fenomeni spontanei non pertinenti con la risposta: risate, colpi di tosse, interazioni con compagni di classe o insegnanti.

Questi fattori ci hanno suggerito di utilizzare specifici corpora vocali per l'addestramento del sistema di trascrizione; nello specifico sono stati utilizzati:

- porzioni del materiale della valutazione di prova del 2017;
- dati di studenti della fascia 6-13 acquisiti da FBK in passato presso alcune classi del Trentino;
- dati di parlatori adulti non-nativi, ad es. italiani che leggono frasi in tedesco e inglese.

Analogamente, per costruire un modello delle parole e delle frasi utilizzate nelle risposte orali degli studenti, è stato utilizzato un corpus di frasi raccolte da:

- materiale scaricato da internet relativo a corsi di lingue o testi semplici;
- materiale derivato dalle risposte scritte delle campagne TLT 2016 e 2018 dopo una fase di “pulizia” (correzioni di errori comuni o tipografici);
- materiale derivato dalla trascrizione manuale del test preliminare della valutazione TLT 2018 effettuato nel 2017 su alcune classi campione; in questo caso sono stati annotati e modellati alcuni fenomeni spontanei tipici come esitazioni, risate o suoni non pertinenti.

6.3 Sito web per la visualizzazione dei dati: progettazione e implementazione

In questa sezione viene descritto il processo che ha portato alla progettazione e implementazione di un sito web per la visualizzazione, mediante grafici dinamici, degli esiti della rilevazione delle competenze linguistiche degli studenti trentini nel 2018.

Il sito web è stato pensato principalmente a uso dei dirigenti degli istituti scolastici coinvolti nella rilevazione, i quali sono interessati ad avere viste diverse, per esempio aggregate per classe o per abilità, degli esiti della rilevazione pertinenti al proprio istituto.

Mentre una presentazione cartacea o digitale (ad es. in formato pdf) implicherebbe una visualizzazione statica dei dati, il sito web offre il vantaggio di una visualizzazione dinamica: agendo semplicemente sui bottoni della pagina web, dinamicamente il sistema mostra i grafici relativi alle scuole, classi e lingue selezionate.

L'intero processo di progettazione e implementazione si è svolto secondo tecniche di progettazione partecipata: i diversi stakeholder interessati (personale IPRASE, ricercatori e tecnici informatici di FBK) sono stati ripetutamente coinvolti nelle varie fasi del processo. Attraverso sia incontri in presenza che per via telematica sono state raccolte le informazioni utili alla progettazione del sito web, ovvero relative ai contenuti e alle esigenze di

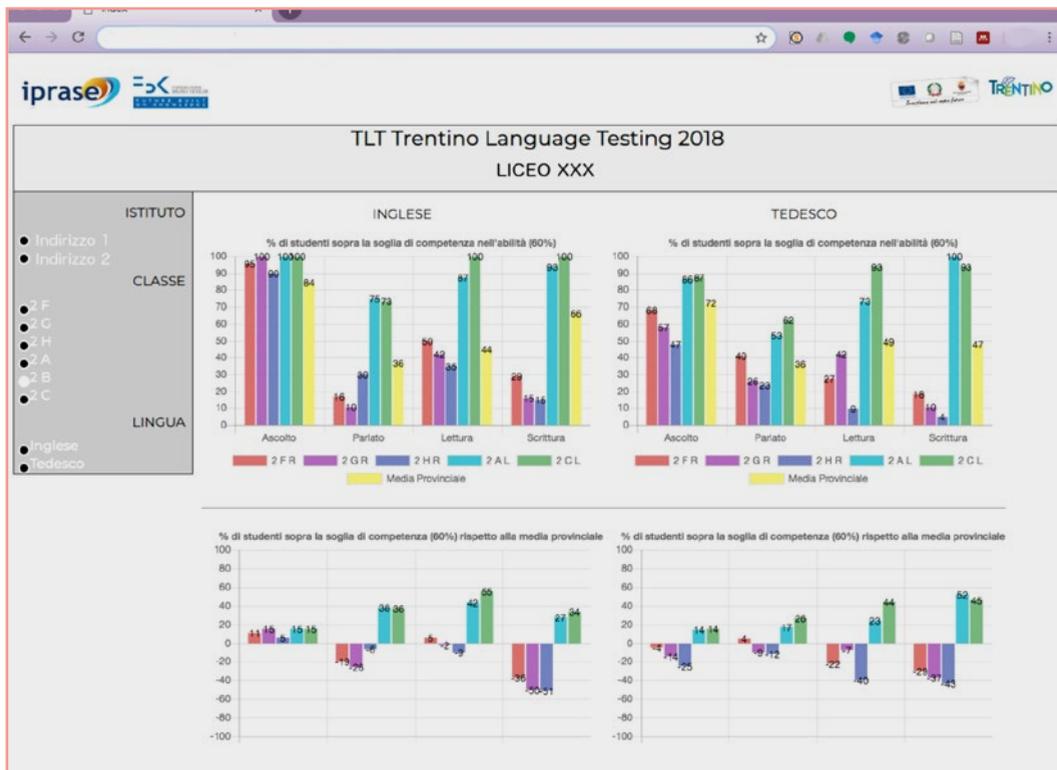
visualizzazione dei dirigenti scolastici. In questi incontri è anche stato più volte analizzato e discusso il materiale via via prodotto (mockups e prototipi). Le modifiche, realizzate sulla base di quanto discusso e concordato, ha portato al pacchetto software finale. Questo processo ha permesso la condivisione dei contenuti e della struttura finale da dare al sito.

In particolare, si è concordato di visualizzare i seguenti tipi di grafici:

- un grafico a barre che rappresenti le percentuali di studenti che superano il livello di soglia (60%) di competenza nelle quattro abilità considerate (Ascolto, Parlato, Lettura, Scrittura). Un esempio è riportato in figura 6.1, grafici in alto;
- un grafico a barre che rappresenti le percentuali di studenti che hanno superato il livello di soglia (60%) di competenza nell'abilità rispetto alla media provinciale, nel caso di istituti comprensivi e di centri di formazione professionale, e rispetto alla media di indirizzo nel caso di scuole secondarie di secondo grado; un esempio è riportato in figura 6.1, grafici in basso;
- un grafico a barre che rappresenti le percentuali di studenti che superano il livello di soglia (60%) di competenza nelle quattro abilità considerate, dove sull'asse delle ascisse sono rappresentate le diverse classi dello specifico istituto;
- un grafico a barre che rappresenti le percentuali di studenti che hanno superato il livello di soglia (60%) di competenza nell'abilità rispetto alla media provinciale nel caso di istituti comprensivi e di centri di formazione professionale, e rispetto alla media di indirizzo nel caso di scuole secondarie di secondo grado; sull'asse delle ascisse sono rappresentate le diverse classi dello specifico istituto; su quello delle ordinate la differenza rispetto alla media (provinciale o di indirizzo, secondo il tipo di scuola).

Tutti i grafici sono proposti sia per la lingua inglese che per la lingua tedesca, posti affiancati così da rendere più immediato un eventuale confronto.

Figura 6.1 - Versione finale di una pagina di visualizzazione.



Per default, quando una pagina viene aperta sono riportati i grafici relativi a tutte le classi dell'istituto o della scuola considerata, per entrambe le lingue (inglese e tedesco). Sul menu posto in una barra laterale a sinistra (Figura 6.1) all'apertura del sito web si trovano quindi tutti i bottoni, precisamente quelli relativi a Istituto, Classe e Lingua, selezionati e sulla parte a destra della barra i corrispondenti grafici. L'interfaccia permette però facilmente di deselegionare una o più scuole, e/o una o più classi, e/o una delle lingue per avere delle visualizzazioni parziali, più specifiche.

L'accesso al sito avviene mediante una finestra di autenticazione. Ciascun istituto interessato, attraverso un proprio ID e una password personale, può accedere solo ai propri dati, rispettando in questo modo la privacy dei dati relativi agli altri istituti coinvolti nella valutazione.

Conclusioni

Il progetto rappresenta una sfida tecnologica di notevole impatto applicativo. I risultati ottenuti, sia per quanto riguarda le pre-

stazioni dei sistemi ASR che la classificazione della proficiency, dimostrano che è assolutamente fattibile la realizzazione di sistemi automatici di ausilio alla valutazione delle abilità linguistiche.

Nell'ambito di futuri progetti sarà possibile utilizzare le tecnologie sviluppate sino ad ora sia per ricavare statistiche aggregate (a livello di plesso o di singolo istituto scolastico), che per monitorare (ad esempio da parte degli insegnanti) l'andamento della proficiency linguistica a livello di classe o di un singolo studente.

Infine, la possibilità di realizzare sistemi di valutazione della proficiency utilizzando architetture *cloud* consentirà di gestire una raccolta dati centralizzata che permetterà di raffinare i modelli alla base della tecnologia migliorandone le prestazioni.

Appendice

Appendice Capitolo 1

Tabella A1.1 - Parole più frequenti in tedesco per il parlato suddivise per livello. Per ogni gruppo sono riportate il numero di parole complessive; per ogni parola sono riportate la frequenza e la percentuale sul totale delle parole.

A1	A2	B1–SSG	B1–IEFP
Parole totali= 12389	Parole totali= 44951	Parole Totali= 155152	Parole totali=1554
390 3.14% wiedersehen	830 1.84% a	4041 2.60% a	91 5.85% a
368 2.97% pizza	557 1.23% esse	1174 0.75% spiele	25 1.60% non
351 2.83% cola	451 1.00% freizeit	1125 0.72% zimmer	23 1.48% hallo
302 2.43% wohne	418 0.93% jahre	1094 0.70% party	20 1.28% party
281 2.26% danke	412 0.91% alt	943 0.60% computer	18 1.15% no
260 2.09% is	406 0.90% uhr	921 0.59% mutter	13 0.83% computer
258 2.08% hobbies	374 0.83% schule	916 0.59% familie	12 0.77% la
222 1.79% lieblingsessen	362 0.80% stehe	894 0.57% vater	11 0.70% familie
208 1.67% hallo	357 0.79% spiele	883 0.56% gehe	10 0.64% spiele
208 1.67% alt	351 0.78% sieben	867 0.55% zwei	10 0.64% mag
200 1.61% jahre	335 0.74% hausaufgaben	863 0.55% mag	10 0.64% fussball
167 1.34% zehn	332 0.73% gehe	860 0.55% gibt	9 0.57% lo
166 1.34% essen	323 0.71% hallo	839 0.54% freizeit	8 0.51% spielen
110 0.88% italienisch	296 0.65% familie	821 0.52% mache	8 0.51% e
100 0.80% mag	288 0.64% fußball	706 0.45% bruder	7 0.45% null
96 0.77% tschüß	283 0.63% mache	687 0.44% sommer	7 0.45% essen
95 0.76% fußball	278 0.61% vater	658 0.42% liebe	7 0.45% boh
90 0.72% coca	273 0.60% mutter	633 0.40% schwester	6 0.38% zimmer
88 0.71% spreche	270 0.60% pizza	624 0.40% hallo	6 0.38% vater
87 0.70% deutsch	256 0.57% heiße	562 0.36% heißt	6 0.38% uhr

Tabella A1.2 - Parole più frequenti in tedesco per lo scritto suddivise per livello. Per ogni gruppo sono riportate il numero di parole complessive; per ogni parola sono riportate la frequenza e la percentuale sul totale delle parole.

A1	A2	B1–SSG	B1–IEFP
Parole totali= 15966	Parole totali= 61159	Parole Totali= 119486	Parole totali=2704
534 3.34% thomas	711 1.16% gut	2428 2.03% geld	48 1.77% euro
512 3.20% hallo	661 1.08% hallo	1720 1.43% euro	43 1.59% hallo
259 1.62% alt	504 0.82% klasse	1309 1.09% bekomme	38 1.40% marta
216 1.35% jahre	421 0.68% thomas	1264 1.05% eltern	30 1.10% pro
204 1.27% hund	376 0.61% geht's	1118 0.93% pro	30 1.10% geld
195 1.22% mag	375 0.61% gehe	946 0.79% hallo	30 1.10% eltern
192 1.20% katze	366 0.59% sympathisch	901 0.75% marta	28 1.03% bekomme
181 1.13% fußball	334 0.54% hi	898 0.75% monat	24 0.88% monat
181 1.13% eis	300 0.49% heißt	599 0.50% mutter	16 0.59% e
161 1.00% is	284 0.46% beste	529 0.44% jahre	15 0.55% kinder
134 0.83% zehn	283 0.46% geht	514 0.43% alt	14 0.51% non
109 0.68% schwester	278 0.45% jahre	500 0.41% kinder	14 0.51% mutter
99 0.62% spielen	276 0.45% gehen	493 0.41% kaufen	13 0.48% a
96 0.60% mutter	275 0.45% alt	488 0.40% gehen	12 0.44% zehn
94 0.58% bruder	264 0.43% fußball	470 0.39% k	12 0.44% kind
87 0.54% my	243 0.39% spielen	462 0.38% verdienen	12 0.44% il
86 0.53% zwei	236 0.38% freundin	457 0.38% geben	12 0.44% arbeite
81 0.50% heißt	227 0.37% heute	418 0.35% möchte	11 0.40% spielen
78 0.48% pizza	210 0.34% freunde	412 0.34% sommer	10 0.37% spiele
74 0.46% meine	207 0.33% lehrerin	408 0.34% fünf	10 0.37% sohn

Tabella A1.3 - Parole più frequenti in inglese per il parlato suddivise per livello. Per ogni gruppo sono riportate il numero di parole complessive; per ogni parola sono riportate la frequenza e la percentuale sul totale delle parole.

A1	A2	B1–SSG	B1–leFP
Parole totali= 16476	Parole totali= 50903	Parole Totali= 173353	Parole totali=20704
742 4.50% yes	1284 2.52% sport	2259 1.30% sport	343 1.65% sport
659 4.00% favourite	1083 2.12% play	1806 1.04% favourite	337 1.62% favourite
623 3.78% bye	1013 1.99% favourite	1225 0.70% book	207 1.00% book
611 3.70% pizza	728 1.43% football	1221 0.70% tv	182 0.87% tv
463 2.81% thank	532 1.04% watch	959 0.55% visit	154 0.74% channel
315 1.91% pet	451 0.88% yes	954 0.55% channel	136 0.65% prefer
303 1.83% live	361 0.70% volleyball	921 0.53% internet	132 0.63% beautiful
278 1.68% salami	339 0.66% family	867 0.50% prefer	126 0.60% internet
257 1.56% subject	318 0.62% friend	831 0.47% lot	113 0.54% visit
236 1.43% fine	276 0.54% prefer	807 0.46% play	109 0.52% play
218 1.32% ten	270 0.53% friends	728 0.42% love	101 0.48% live
201 1.22% hobbies	264 0.51% love	718 0.41% live	93 0.44% village
194 1.17% school	255 0.50% tv	715 0.41% people	89 0.43% love
194 1.17% dog	250 0.49% pizza	698 0.40% friends	81 0.39% yes
185 1.12% cola	248 0.48% sports	685 0.39% watch	78 0.37% football
163 0.98% hi	229 0.45% name	616 0.35% beautiful	75 0.36% hello
158 0.95% please	223 0.43% food	611 0.35% village	71 0.34% funny
142 0.86% hello	221 0.43% hi	577 0.33% town	66 0.31% city
132 0.80% coca	207 0.40% hello	531 0.30% city	65 0.31% watch
115 0.69% hobby	194 0.38% drink	515 0.29% time	64 0.30% hi

Tabella A1.4 - Parole più frequenti in inglese per lo scritto suddivise per livello. Per ogni gruppo sono riportate il numero di parole complessive; per ogni parola sono riportate la frequenza e la percentuale sul totale delle parole.

A1	A2	B1–SSG	B1–leFP
Parole totali= 15068	Parole totali= 77335	Parole Totali= 168817	Parole totali=34738
243 1.61% lake	892 1.15% hi	4403 2.60% language	977 2.81% language
193 1.28% hello	882 1.14% favourite	3169 1.87% learn	647 1.86% learn
192 1.27% trentino	689 0.89% bye	1591 0.94% study	482 1.38% english
178 1.18% play	632 0.81% teacher	1459 0.86% speak	449 1.29% study
178 1.18% am	632 0.81% class	1196 0.70% hi	330 0.95% speak
157 1.04% family	613 0.79% friend	1156 0.68% english	247 0.71% hi
145 0.96% ice	469 0.60% play	731 0.43% people	149 0.42% languages
139 0.92% bike	449 0.58% sorry	661 0.39% languages	144 0.41% people
136 0.90% swimming	446 0.57% fine	660 0.39% spanish	135 0.38% spanish
136 0.90% football	435 0.56% susan	653 0.38% lot	117 0.33% school
115 0.76% garda	382 0.49% tennis	571 0.33% bye	117 0.33% country
106 0.70% cream	303 0.39% funny	526 0.31% write	115 0.33% world
102 0.67% bye	302 0.39% subject	506 0.30% country	110 0.31% travel
83 0.55% hi	297 0.38% english	472 0.28% school	107 0.30% lot
81 0.53% eat	291 0.37% school	422 0.25% learning	106 0.30% read
80 0.53% whit	289 0.37% sport	413 0.24% grammar	102 0.29% start
77 0.51% hallo	269 0.34% hair	380 0.22% read	86 0.24% learning
76 0.50% trento	259 0.33% beautiful	354 0.21% try	84 0.24% book
74 0.49% swim	253 0.32% centre	339 0.20% travel	82 0.23% bye
68 0.45% mountain	252 0.32% love	336 0.19% soon	78 0.22% friend

Appendice Capitolo 2

Tabella A2.1 - Descrittive A1 Primaria.

VARIABILE	OSSERVAZIONI	MEDIA	DEV. STANDARD
<i>Genere</i>			
Ragazzo	510	0.518	0.5
Ragazza	510	0.482	0.5
<i>Anni di ritardo</i>			
0	510	0.978	0.145
1 o più	510	0.022	0.145
<i>Origine</i>			
Italiano	510	0.843	0.364
Straniero 2 ^a generazione	510	0.098	0.298
Straniero 1 ^a generazione	510	0.059	0.236
<i>Numero di libri posseduti a casa</i>			
Nessuno o pochi (0–10 libri)	510	0.116	0.320
Abbastanza da riempire una mensola (11–25 libri)	510	0.247	0.432
Abbastanza da riempire uno scaffale (26–100 libri)	510	0.331	0.471
Abbastanza da riempire due scaffali (101–200 libri)	510	0.153	0.360
Abbastanza da riempire tre o più scaffali (più di 200 libri)	510	0.145	0.353
Non so	510	0.008	0.088
<i>Attività CLIL</i>			
Ore Clil settimanali Inglese	510	4.325	0.929
Ore Clil settimanali Tedesco	510	1.898	2.094
<i>Prova "speciale"</i>			
No	510	0.947	0.224
Sì	510	0.053	0.224
<i>Ordine delle prove</i>			
Inglese–Tedesco	510	0.567	0.496
Tedesco–Inglese	510	0.433	0.496

Tabella A2.2 - Descrittive A2 Secondaria I grado.

VARIABILE	OSSERVAZIONI	MEDIA	DEV. STANDARD
<i>Genere</i>			
Ragazzo	609	0.488	0.5
Ragazza	609	0.512	0.5
<i>Anni di ritardo</i>			
0	609	0.936	0.245
1 o più	609	0.064	0.245
<i>Origine</i>			
Italiano	608	0.862	0.345
Straniero 2 ^a generazione	608	0.086	0.280
Straniero 1 ^a generazione	608	0.053	0.223
<i>Numero di libri posseduti a casa</i>			
Nessuno o pochi (0–10 libri)	609	0.143	0.350
Abbastanza da riempire una mensola (11–25 libri)	609	0.225	0.418
Abbastanza da riempire uno scaffale (26–100)	609	0.328	0.470
Abbastanza da riempire due scaffali (101–200 libri)	609	0.167	0.374
Abbastanza da riempire tre o più scaffali (più di 200 libri)	609	0.130	0.336
Non so	609	0.007	0.081
<i>Attività Clil in settimana Inglese</i>			
No	609	0.951	0.217
Sì	609	0.049	0.217
<i>Attività Clil in settimana Tedesco</i>			
No	609	0.9	0.3
Sì	609	0.1	0.3
<i>Prova “speciale”</i>			
No	609	0.905	0.294
Sì	609	0.095	0.294
<i>Ordine delle prove</i>			
Inglese–Tedesco	609	0.521	0.5
Tedesco–Inglese	609	0.479	0.5

Tabella A2.3 - Descrittive B1 Secondaria II grado.

VARIABILE	OSSERVAZIONI	MEDIA	DEV. STANDARD
<i>Genere</i>			
Ragazzo	1101	0.455	0.498
Ragazza	1101	0.545	0.498
<i>Anni di ritardo</i>			
0	1101	0.86	0.347
1 o più	1101	0.14	0.347
<i>Origine</i>			
Italiano	1101	0.944	0.231
Straniero 2 ^a generazione	1101	0.019	0.137
Straniero 1 ^a generazione	1101	0.037	0.189
<i>Titolo di studio più elevato dei genitori</i>			
Licenza elementare o media	1101	0.066	0.249
Qualifica professionale	1101	0.102	0.302
Diploma maturità	1101	0.45	0.498
Laurea o superiore	1101	0.304	0.46
Non so	1101	0.077	0.267
<i>Indirizzo di studio</i>			
Classico	1101	0.077	0.267
Scientifico	1101	0.225	0.418
Linguistico	1101	0.098	0.298
Scienze Umane	1101	0.127	0.333
Artistico	1101	0.047	0.212
Economico	1101	0.182	0.386
Tecnologico	1101	0.207	0.405
Professionale	1101	0.036	0.187
<i>Attività Clil in settimana Inglese</i>			
No	1101	0.851	0.356
Sì	1101	0.149	0.356
<i>Attività Clil in settimana Tedesco</i>			
No	1101	0.955	0.208
Sì	1101	0.045	0.208
<i>Prova "speciale"</i>			
No	1101	0.965	0.183
Sì	1101	0.035	0.183
<i>Ordine delle prove</i>			
Inglese–Tedesco	1101	0.629	0.483
Tedesco–Inglese	1101	0.351	0.478
Inglese	1101	0.02	0.14

Tabella A2.4 - Descrittive B1 Istruzione e Formazione Professionale.

VARIABILE	OSSERVAZIONI	MEDIA	DEV. STANDARD
<i>Genere</i>			
Ragazzo	442	0.649	0.478
Ragazza	442	0.351	0.478
<i>Anni di ritardo</i>			
0	441	0.501	0.501
1 o più	441	0.499	0.501
<i>Origine</i>			
Italiano	442	0.853	0.355
Straniero 2 ^a generazione	442	0.045	0.208
Straniero 1 ^a generazione	442	0.102	0.303
<i>Titolo di studio più elevato dei genitori</i>			
Licenza elementare o media	431	0.186	0.389
Qualifica professionale	431	0.237	0.426
Diploma maturità	431	0.390	0.488
Laurea o superiore	431	0.104	0.306
Non so	431	0.084	0.277
<i>Indirizzo di studio</i>			
Servizi	442	0.636	0.482
Industria e Artigianato	442	0.364	0.482
<i>Attività Clil in settimana Inglese</i>			
No	442	0.805	0.396
Sì	442	0.195	0.396
<i>Attività Clil in settimana Tedesco</i>			
No	442	0.980	0.141
Sì	442	0.020	0.141
<i>Prova "speciale"</i>			
No	442	0.871	0.336
Sì	442	0.129	0.336
<i>Ordine delle prove</i>			
Tedesco–Inglese	442	0.042	0.224
Tedesco	442	0.158	0.224
Inglese	442	0.800	0.224

Tabella A2.5 - Determinanti nella probabilità di possedere il livello - A1 Primaria.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	ascolto inglese	ascolto inglese scuole	parlato inglese	parlato inglese scuole	lettura inglese	lettura inglese scuole	scrittura inglese	scrittura inglese scuole	ascolto tedesco	ascolto tedesco scuole	parlato tedesco	parlato tedesco scuole	lettura tedesco	lettura tedesco scuole	scrittura tedesco	scrittura tedesco scuole
Effetti Marginali																
Variabili esplicative																
Prova "speciale" (rif. No)																
Si	-0.291** (0.132)	-0.334** (0.146)	-0.205** (0.094)	-0.279*** (0.072)	-0.202 (0.123)	-0.226* (0.125)	-0.340*** (0.111)	-0.427*** (0.091)	-0.193* (0.110)	-0.165 (0.128)	-0.225*** (0.073)	-0.204*** (0.027)	-0.207* (0.125)	-0.385*** (0.135)	-0.367*** (0.064)	-0.395*** (0.033)
Ordine delle prove (rif. 1 ^a)																
2 ^a	0.035 (0.068)	-0.014 (0.028)	-0.035 (0.094)	-0.019 (0.044)	0.034 (0.061)	-0.021 (0.067)	0.006 (0.112)	-0.044 (0.040)	-0.032 (0.047)	-0.019 (0.043)	-0.015 (0.102)	0.190*** (0.057)	0.019 (0.063)	0.111** (0.051)	-0.016 (0.095)	0.028 (0.082)
Genere (rif. Ragazzo)																
Ragazza	0.055 (0.045)	0.060 (0.048)	0.038 (0.048)	0.096 (0.058)	0.078 (0.050)	0.082* (0.048)	0.057 (0.040)	0.080 (0.057)	0.075* (0.042)	0.078* (0.044)	0.047 (0.042)	0.098** (0.048)	0.117*** (0.039)	0.132*** (0.046)	0.118** (0.054)	0.169*** (0.058)
Anni di ritardo (rif. No)																
Si	-0.007 (0.104)	-0.056 (0.138)	-0.033 (0.147)	0.024 (0.195)	0.131 (0.147)	0.163 (0.125)	-0.019 (0.155)	0.051 (0.127)	-0.234* (0.141)	-0.309** (0.138)	-0.061 (0.079)	-0.174*** (0.024)	0.044 (0.126)	-0.014 (0.150)	-0.213 (0.155)	-0.240** (0.122)
Origine (rif. Italiano)																
Straniero 2 ^a generazione	0.037 (0.058)	-0.060 (0.076)	0.109 (0.098)	0.054 (0.088)	0.007 (0.109)	-0.034 (0.117)	0.196** (0.080)	0.121 (0.094)	-0.070 (0.073)	-0.038 (0.079)	0.234** (0.107)	0.257*** (0.086)	0.008 (0.060)	-0.011 (0.054)	0.053 (0.098)	0.157* (0.092)
Straniero 1 ^a generazione	-0.043 (0.106)	-0.066 (0.102)	0.127 (0.170)	0.217 (0.164)	-0.004 (0.118)	0.002 (0.122)	-0.038 (0.123)	0.019 (0.112)	0.085 (0.109)	0.124 (0.107)	-0.016 (0.109)	0.050 (0.104)	0.020 (0.082)	0.005 (0.087)	0.067 (0.147)	0.240* (0.136)
Libri a casa (rif. >200 libri)																
0-10 libri	-0.253*** (0.085)	-0.238*** (0.086)	-0.229** (0.097)	-0.281*** (0.102)	-0.313*** (0.104)	-0.295*** (0.110)	-0.261*** (0.091)	-0.292** (0.114)	-0.312*** (0.101)	-0.288** (0.113)	-0.279*** (0.070)	-0.311*** (0.049)	-0.271*** (0.089)	-0.254*** (0.078)	-0.337*** (0.089)	-0.429*** (0.086)
11-25 libri	-0.135*** (0.052)	-0.134** (0.055)	0.029 (0.077)	-0.016 (0.106)	-0.163* (0.090)	-0.147 (0.095)	-0.117* (0.066)	-0.114 (0.077)	-0.154** (0.075)	-0.134 (0.083)	-0.119* (0.062)	-0.177*** (0.058)	-0.087 (0.073)	-0.122 (0.083)	-0.239*** (0.062)	-0.337*** (0.083)
26-100 libri	-0.121*** (0.042)	-0.124*** (0.044)	-0.016 (0.058)	-0.076 (0.076)	-0.001 (0.078)	0.019 (0.078)	-0.098 (0.083)	-0.112 (0.087)	-0.028 (0.071)	-0.021 (0.079)	-0.102* (0.053)	-0.135*** (0.050)	0.045 (0.076)	0.054 (0.073)	-0.062 (0.069)	-0.093 (0.075)
101-200 libri	-0.054 (0.060)	-0.027 (0.063)	0.066 (0.073)	0.074 (0.096)	0.114 (0.079)	0.151* (0.079)	0.003 (0.102)	0.104 (0.121)	0.041 (0.088)	0.052 (0.094)	-0.070 (0.069)	-0.087 (0.075)	0.090 (0.072)	0.098 (0.065)	0.026 (0.099)	0.043 (0.102)
Non so	0.050 (0.144)	-0.011 (0.197)	-0.138 (0.245)	-0.311* (0.185)	-0.261 (0.277)	-0.265 (0.247)	0.013 (0.276)	-0.141 (0.278)	-0.058 (0.241)	-0.010 (0.244)						
Ore settimanali CLIL	0.093*** (0.032)	-0.012 (0.016)	0.140*** (0.054)	0.062** (0.029)	0.074* (0.042)	0.019 (0.049)	0.090 (0.056)	-0.016 (0.022)	0.035*** (0.012)	0.170*** (0.022)	-0.016 (0.023)	-0.220*** (0.031)	0.015 (0.015)	-0.142*** (0.031)	0.024 (0.021)	0.083* (0.046)
Osservazioni	502	488	502	465	502	502	502	488	496	496	492	470	496	486	492	492
R-quadro	0.0776	0.189	0.0905	0.262	0.0817	0.133	0.0653	0.321	0.0702	0.107	0.0545	0.351	0.0893	0.210	0.0948	0.263

Errori standard in parentesi: *** p<0.01, ** p<0.05, * p<0.1

Tabella A2.6 - Determinanti nella probabilità di possedere il livello - A2 Secondaria I grado.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	ascolto inglese	ascolto inglese	parlato inglese	parlato inglese	lettura inglese	lettura inglese	scrittura inglese	scrittura inglese	ascolto tedesco	ascolto tedesco	parlato tedesco	parlato tedesco	lettura tedesco	lettura tedesco	scrittura tedesco	scrittura tedesco
	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole	scuole
Effetti Marginali																
Variabili esplicative																
Prova "speciale" (rif. No)																
Si	-0.211*** (0.056)	-0.244*** (0.061)	-0.103 (0.069)	-0.143** (0.065)	-0.244*** (0.058)	-0.273*** (0.060)	-0.211*** (0.064)	-0.379*** (0.064)	-0.308*** (0.072)	-0.324*** (0.059)	-0.181*** (0.055)	-0.226*** (0.043)	-0.290*** (0.062)	-0.309*** (0.058)	-0.337*** (0.064)	-0.354*** (0.064)
Ordine delle prove (rif. 1°)																
2°	0.076 (0.063)	0.153*** (0.039)	-0.112 (0.095)	-0.087 (0.055)	0.088 (0.062)	0.146** (0.057)	-0.083 (0.116)	0.039 (0.065)	0.014 (0.063)	0.005 (0.060)	-0.101 (0.082)	-0.081 (0.063)	-0.079 (0.070)	-0.047 (0.074)	-0.188** (0.083)	-0.213*** (0.075)
Genere (rif. Ragazzo)							blue									
Ragazza	-0.048 (0.040)	-0.047 (0.041)	0.043 (0.044)	0.021 (0.049)	0.064 (0.042)	0.077* (0.042)	0.096** (0.039)	0.144*** (0.045)	0.038 (0.050)	0.042 (0.055)	0.081** (0.032)	0.091** (0.044)	0.044 (0.050)	0.045 (0.053)	0.167*** (0.042)	0.201*** (0.048)
Anni di ritardo (rif. No)																
Si	-0.074 (0.086)	-0.062 (0.079)	-0.092 (0.094)	-0.077 (0.097)	-0.044 (0.108)	-0.001 (0.118)	-0.143 (0.088)	-0.211* (0.123)	-0.016 (0.097)	-0.019 (0.096)	-0.161** (0.073)	-0.186*** (0.066)	-0.140 (0.089)	-0.116 (0.090)	-0.058 (0.104)	-0.052 (0.107)
Origine (rif. Italiano)																
Straniero 2° generazione	0.079 (0.072)	0.094 (0.067)	-0.008 (0.074)	0.001 (0.065)	-0.061 (0.087)	-0.080 (0.083)	-0.003 (0.081)	-0.011 (0.092)	-0.076 (0.080)	-0.066 (0.069)	0.035 (0.098)	0.050 (0.117)	-0.057 (0.092)	-0.065 (0.091)	-0.060 (0.085)	-0.088 (0.079)
Straniero 1° generazione	0.044 (0.085)	0.064 (0.090)	0.013 (0.091)	0.002 (0.092)	0.007 (0.120)	0.008 (0.128)	0.055 (0.081)	0.123 (0.086)	-0.051 (0.104)	0.015 (0.103)	0.006 (0.122)	0.086 (0.140)	-0.094 (0.113)	-0.048 (0.101)	-0.149 (0.109)	-0.060 (0.108)
Libri a casa (rif. >200 libri)																
0-10 libri	-0.221*** (0.078)	-0.277*** (0.085)	-0.179* (0.095)	-0.209*** (0.080)	-0.219** (0.100)	-0.252** (0.105)	-0.267** (0.104)	-0.227** (0.109)	-0.244** (0.098)	-0.327*** (0.086)	-0.142 (0.113)	-0.344*** (0.092)	-0.165 (0.103)	-0.273*** (0.075)	-0.122 (0.100)	-0.285*** (0.077)
11-25 libri	-0.204*** (0.063)	-0.232*** (0.062)	-0.074 (0.073)	-0.051 (0.064)	-0.130* (0.076)	-0.138* (0.083)	-0.226*** (0.080)	-0.155* (0.081)	-0.178*** (0.076)	-0.200*** (0.070)	-0.137* (0.082)	-0.250*** (0.090)	-0.068 (0.076)	-0.122* (0.067)	-0.147* (0.078)	-0.227*** (0.073)
26-100 libri	-0.041 (0.052)	-0.040 (0.056)	-0.083 (0.072)	-0.100 (0.072)	-0.101 (0.079)	-0.088 (0.085)	-0.126** (0.062)	-0.102 (0.073)	-0.104 (0.065)	-0.108 (0.066)	-0.069 (0.079)	-0.169* (0.091)	-0.025 (0.053)	-0.043 (0.049)	-0.070 (0.071)	-0.114 (0.074)
101-200 libri	0.025 (0.064)	0.023 (0.069)	0.109 (0.076)	0.134* (0.081)	0.011 (0.069)	0.028 (0.078)	-0.067 (0.052)	0.009 (0.064)	0.033 (0.079)	0.003 (0.083)	-0.029 (0.080)	-0.079 (0.101)	0.034 (0.069)	-0.021 (0.068)	-0.031 (0.073)	-0.109 (0.070)
Non so	-0.003 (0.302)	0.077 (0.288)			0.096 (0.302)	0.118 (0.316)	-0.043 (0.284)	0.208 (0.178)	0.166 (0.243)	0.127 (0.279)			0.341 (0.217)	0.280 (0.278)	0.048 (0.142)	0.087 (0.217)
Attività CLIL in settimana (rif. No)																
Si	0.329*** (0.044)				-0.171 (0.127)		0.289*** (0.110)		-0.004 (0.050)	0.086 (0.087)	-0.185*** (0.061)	0.020 (0.081)	-0.078 (0.054)	0.186** (0.081)	-0.064 (0.056)	0.050 (0.099)
Osservazioni	583	583	581	554	583	583	583	564	584	584	581	514	584	584	584	584
R-quadro	0.0625	0.107	0.0439	0.235	0.0495	0.0946	0.0696	0.309	0.0568	0.124	0.0648	0.231	0.0508	0.130	0.0899	0.222

Errori standard in parentesi: *** p<0.01, ** p<0.05, * p<0.1

Figura A2.1 - Analisi dei distrattori A1 Inglese - Ascolto.

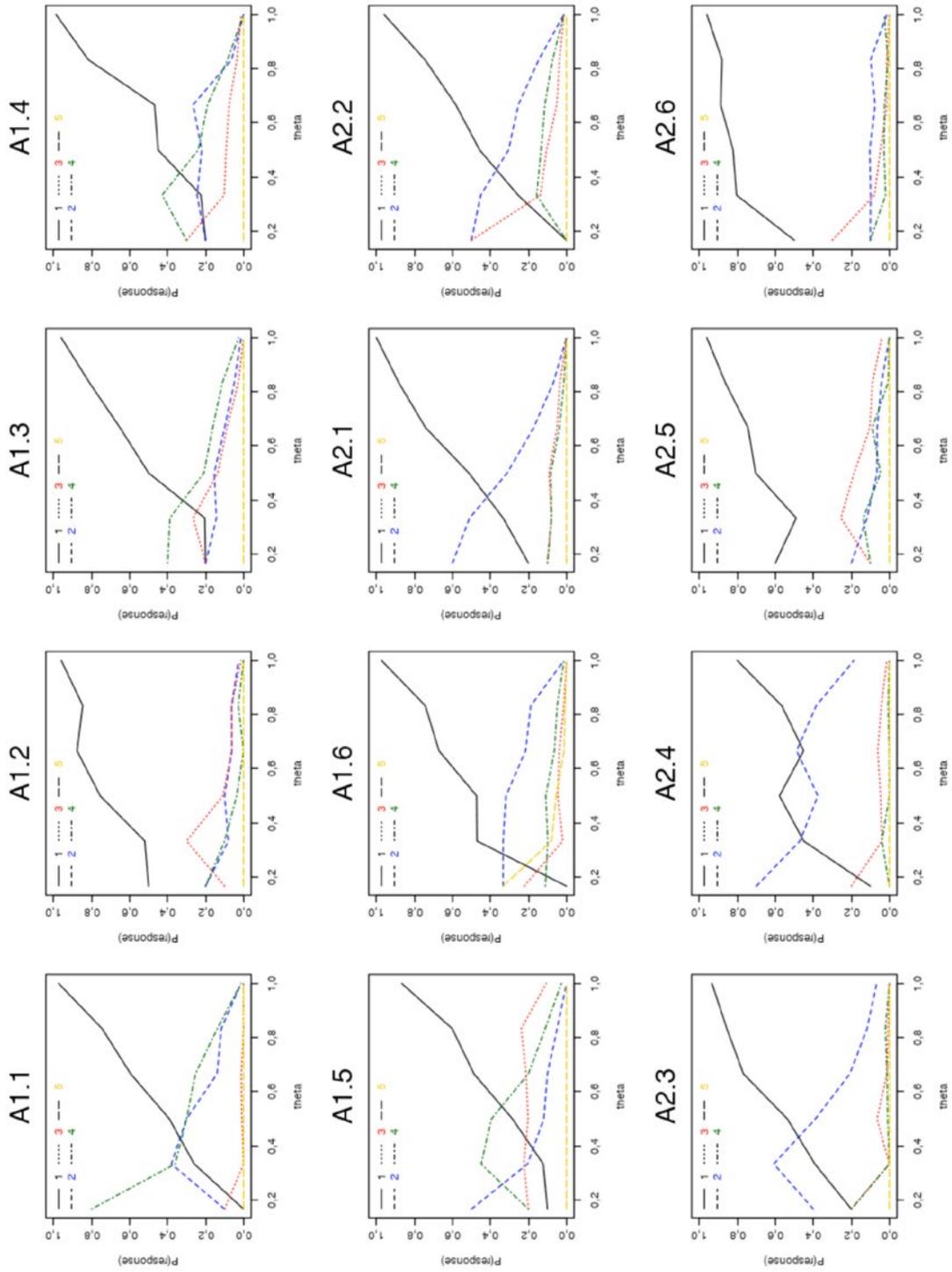


Figura A2.2 - Analisi dei distrattori A1 Inglese - Lettura.

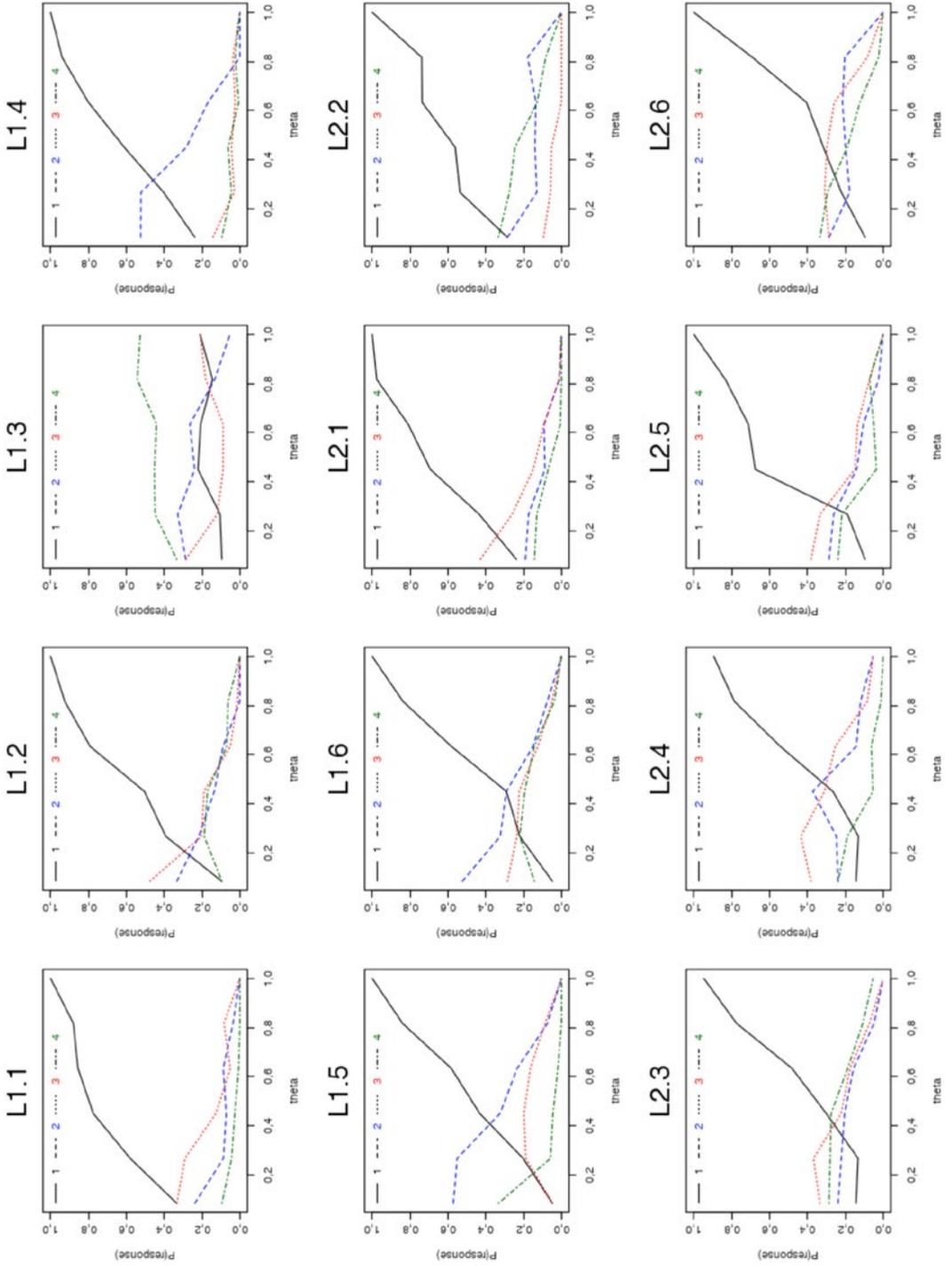


Figura A2.3 - Analisi dei distrattori A1 Tedesco - Ascolto.

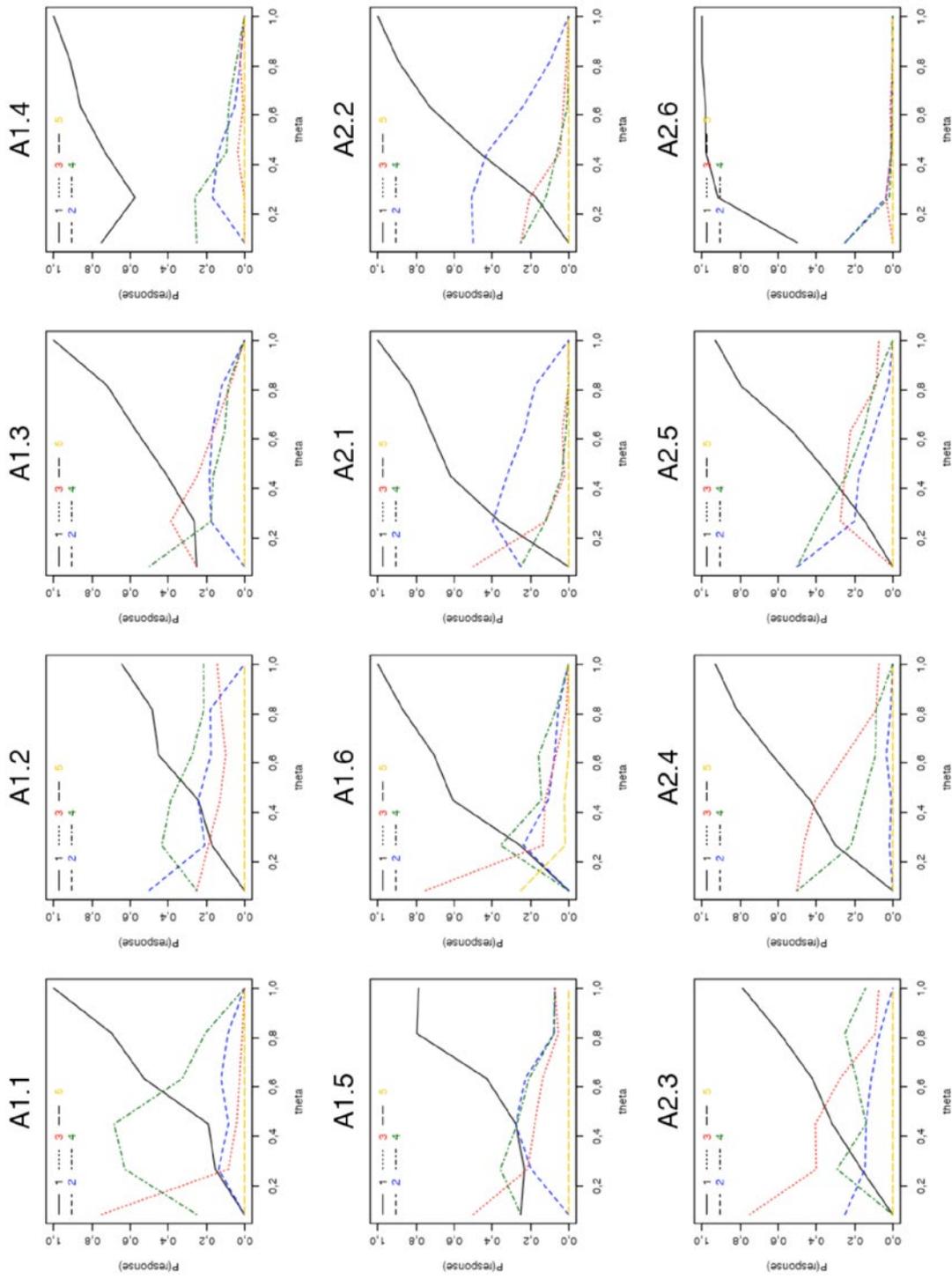


Figura A2.4 - Analisi dei distrattori A1 Tedesco - Lettura.

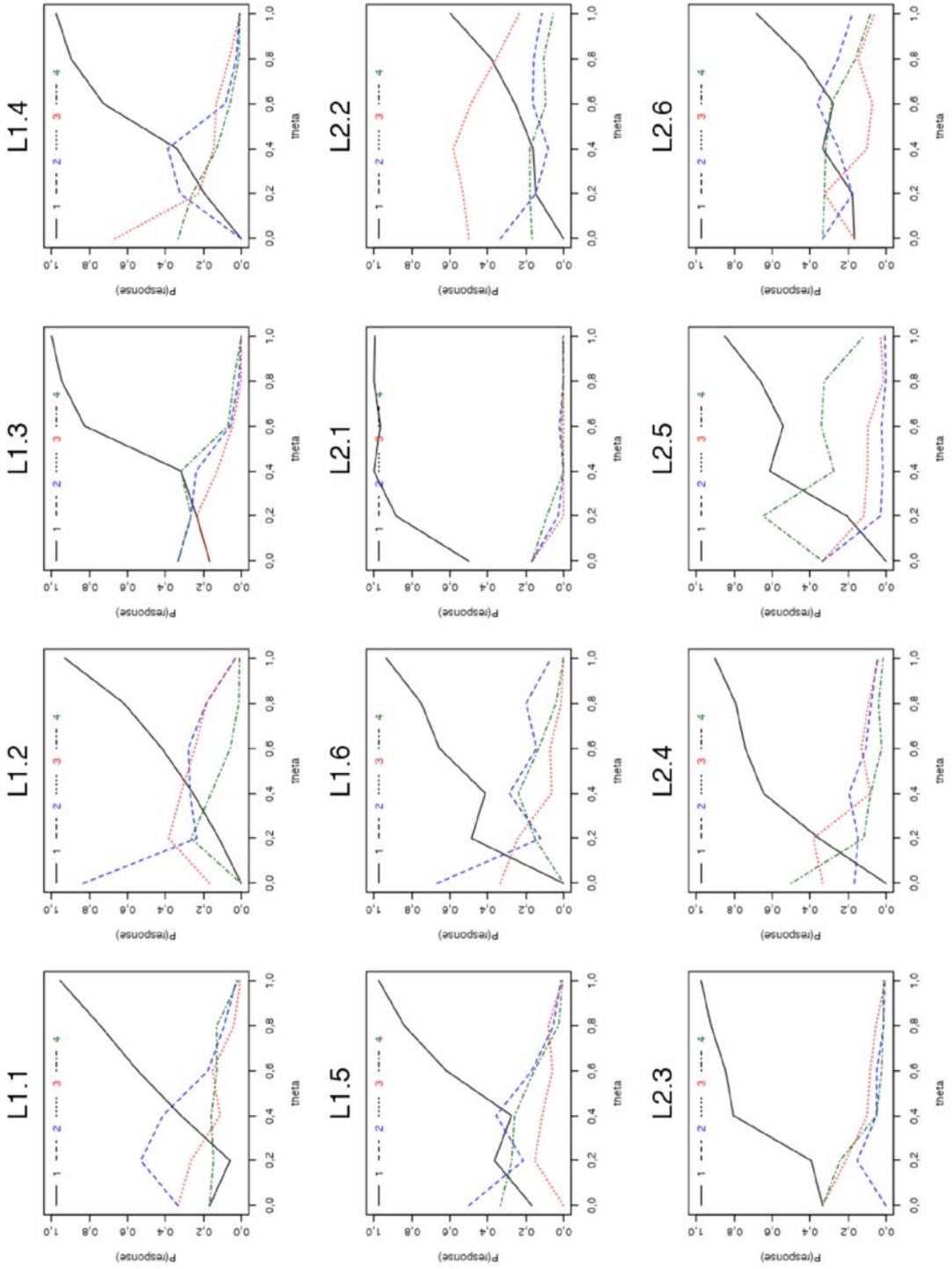


Figura A2.5 - Analisi dei distrattori A2 Inglese - Ascolto.

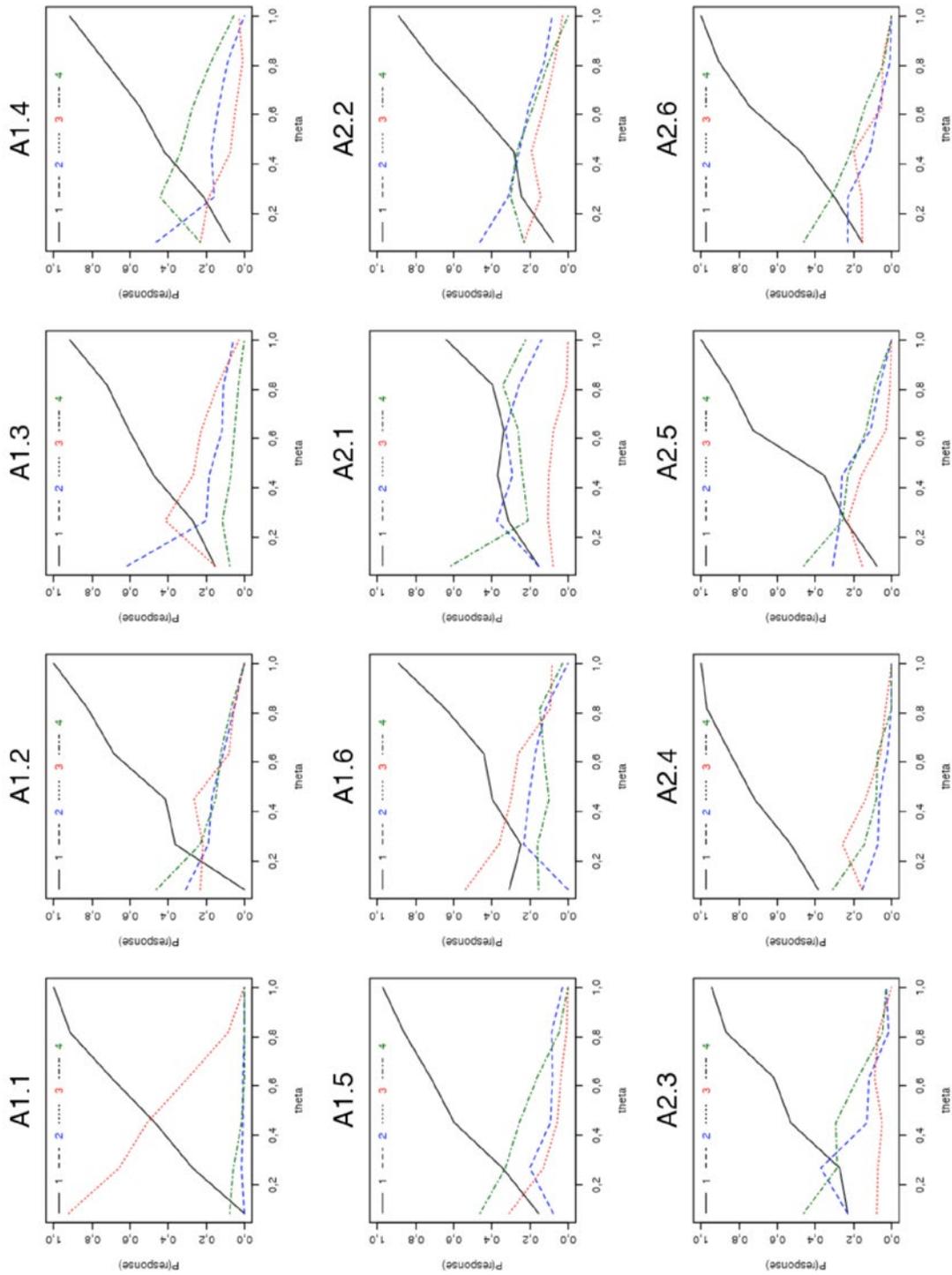


Figura A2.6 - Analisi dei distrattori A2 Inglese - Lettura.

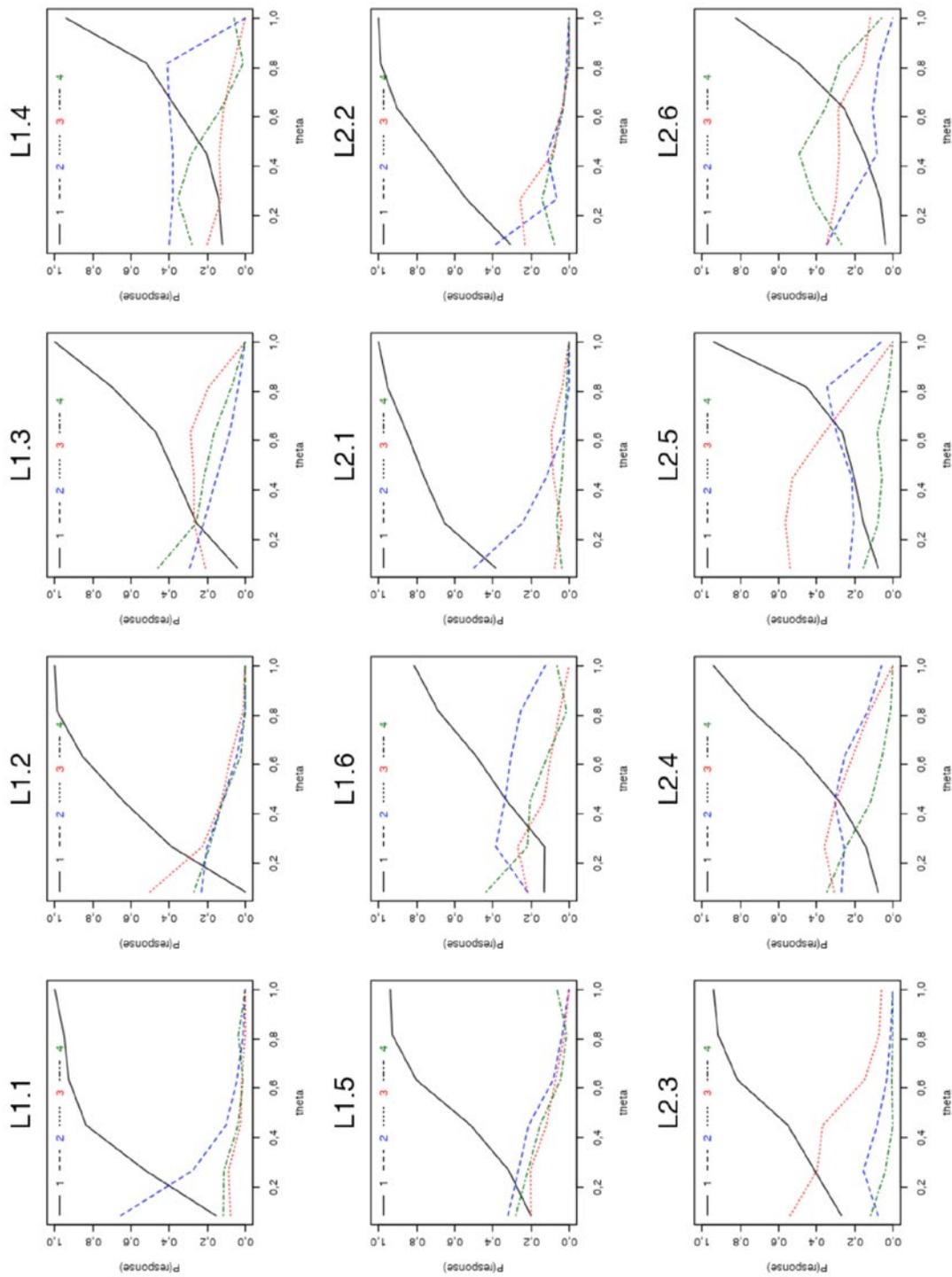


Figura A2.7 - Analisi dei distrattori A2 Tedesco - Ascolto.

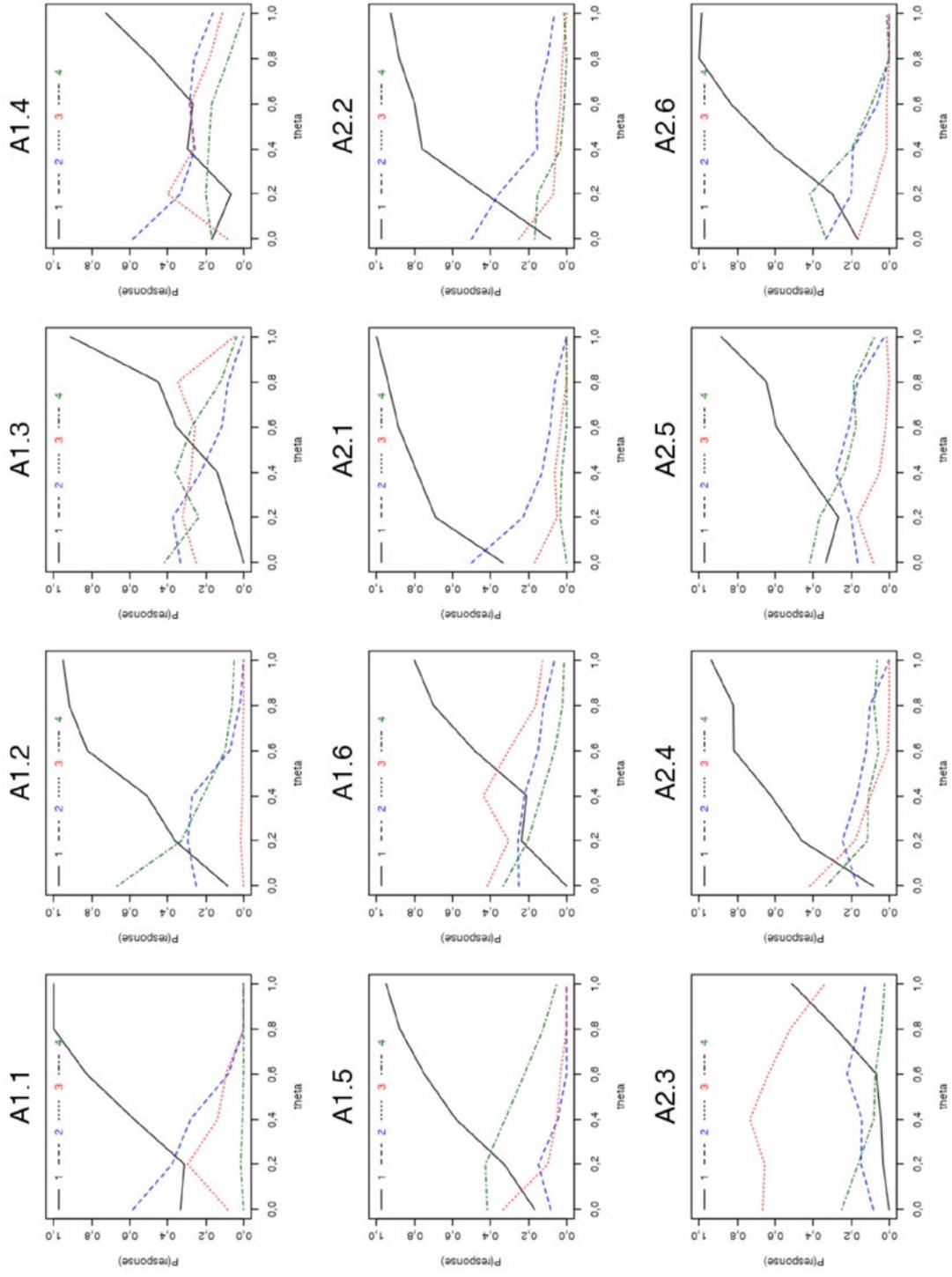


Figura A2.8 - Analisi dei distrattori A2 Tedesco - Lettura.

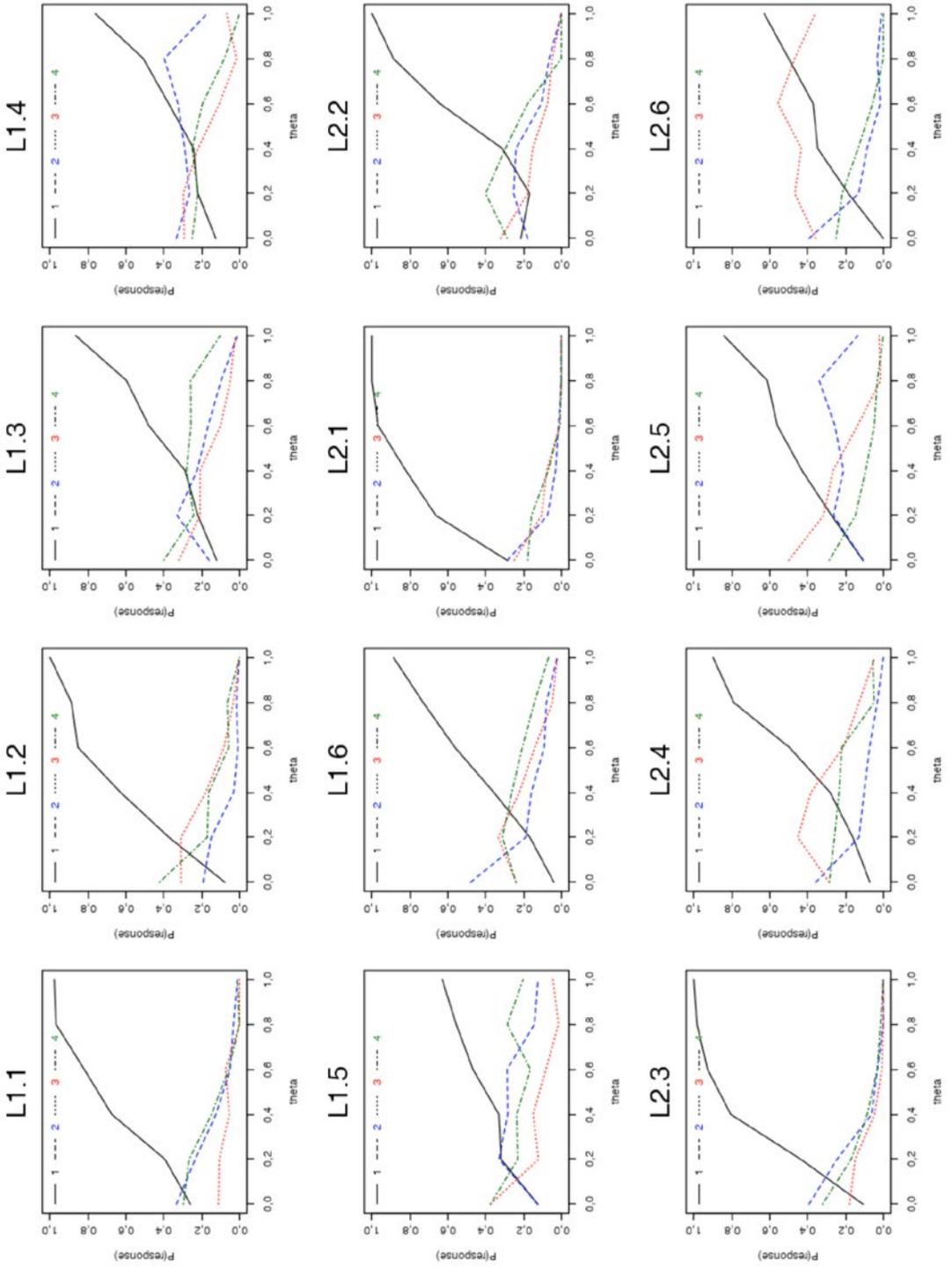


Figura A2.9 - Analisi dei distrattori B1 Inglese - Ascolto.

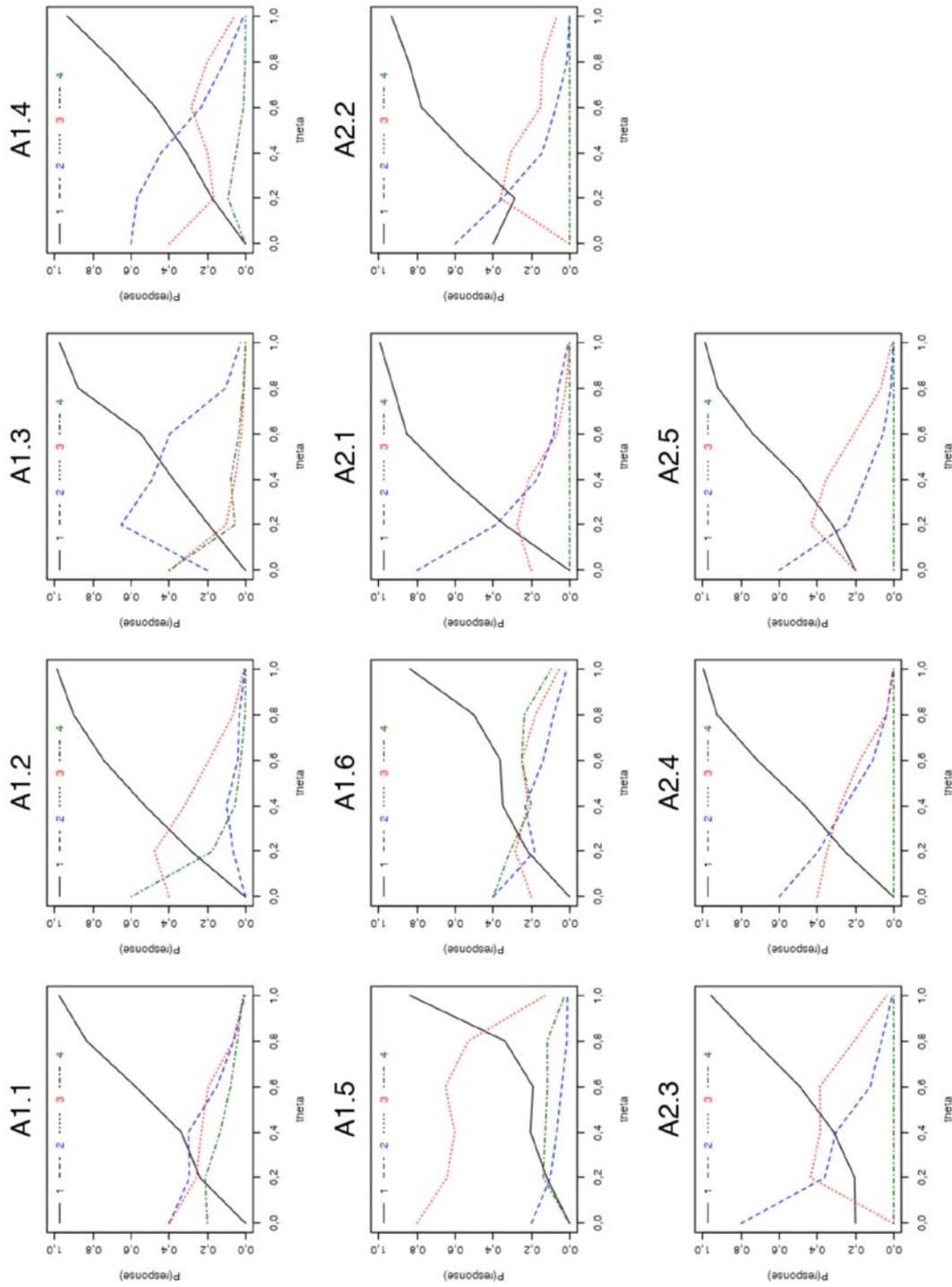


Figura A2.10 - Analisi dei distrattori B1 Inglese - Lettura (parte 1).

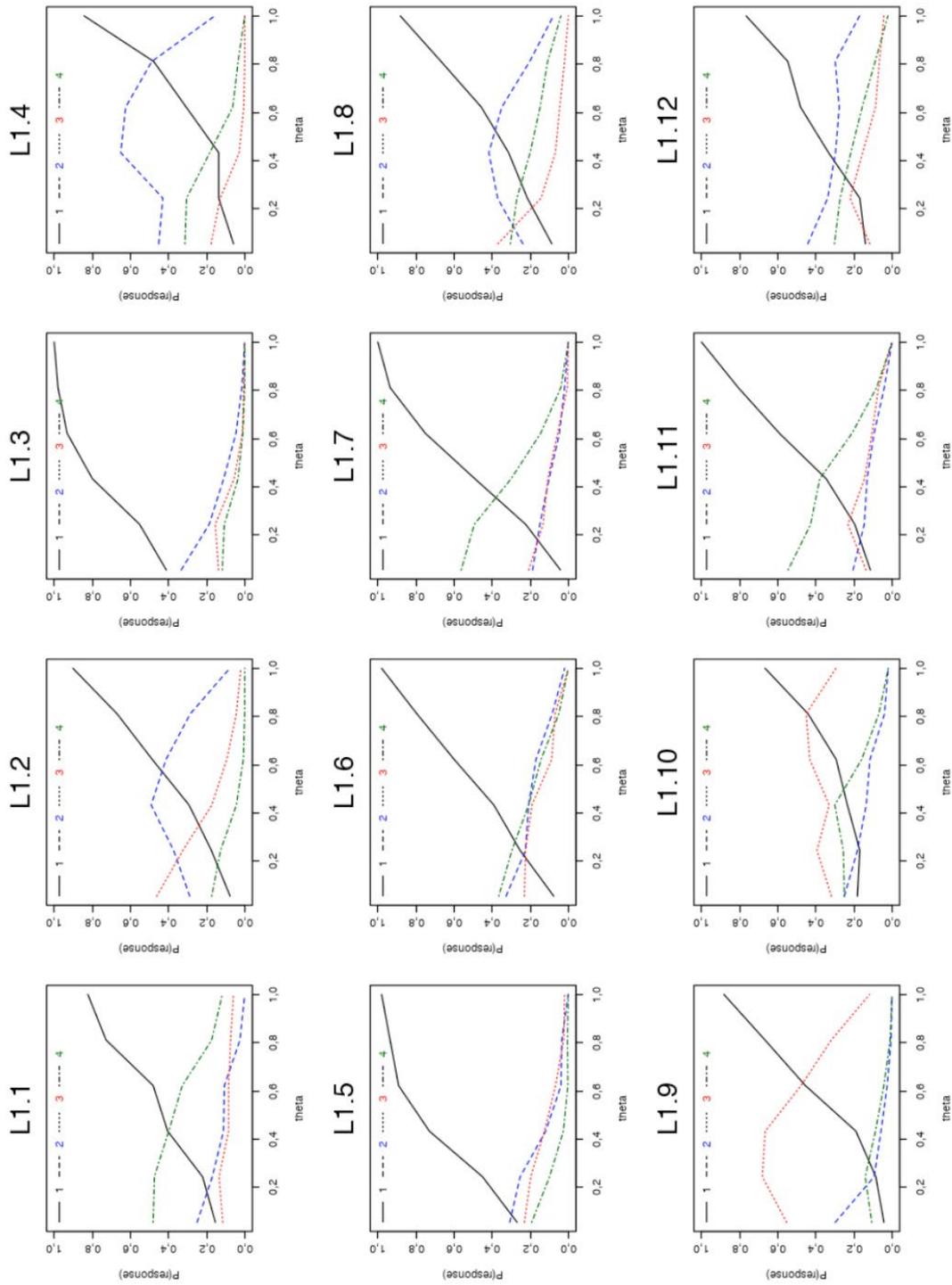


Figura A2.11 - Analisi dei distrattori B1 Inglese - Lettura (parte 2).

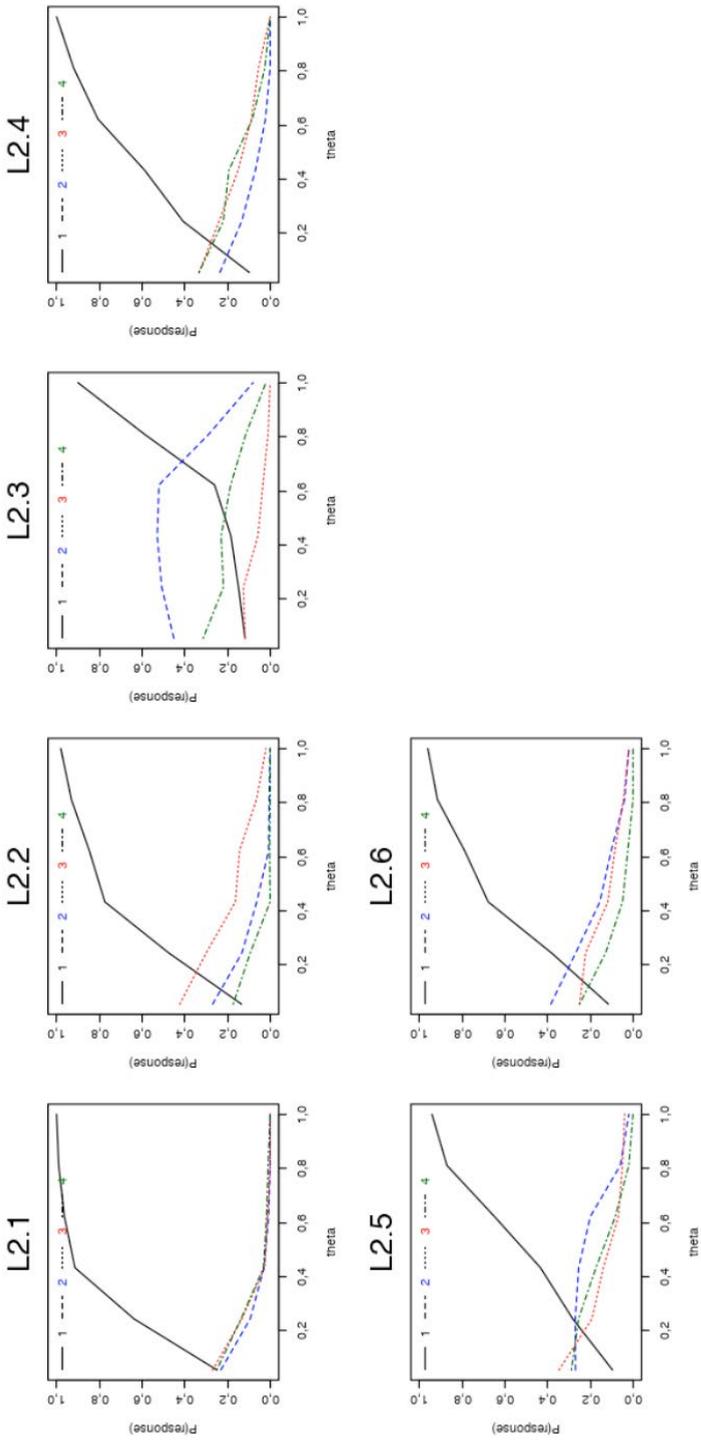


Figura A2.12 - Analisi dei distrattori B1 Tedesco - Ascolto.

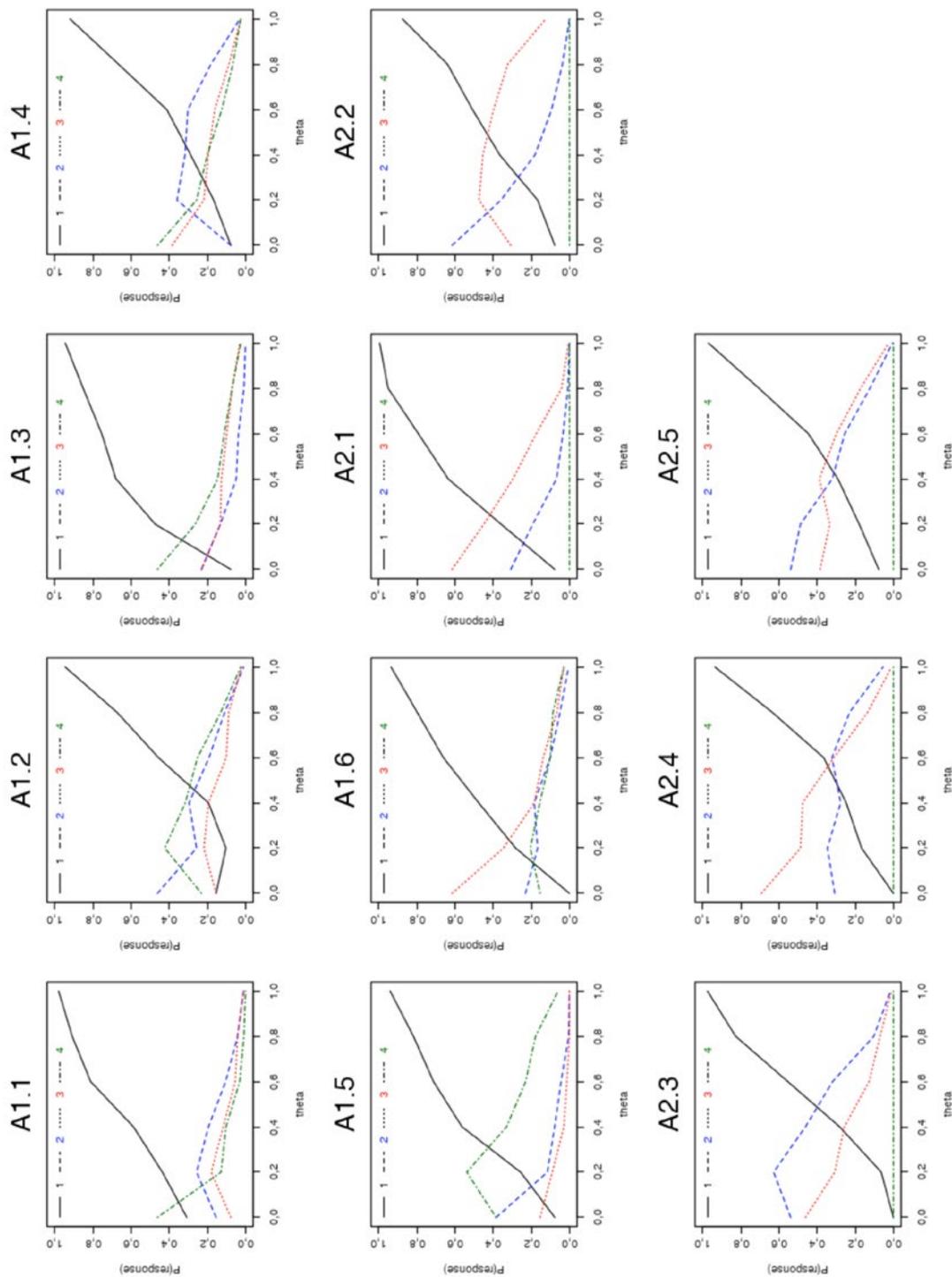


Figura A2.13 - Analisi dei distrattori B1 Tedesco - Lettura (parte 1).

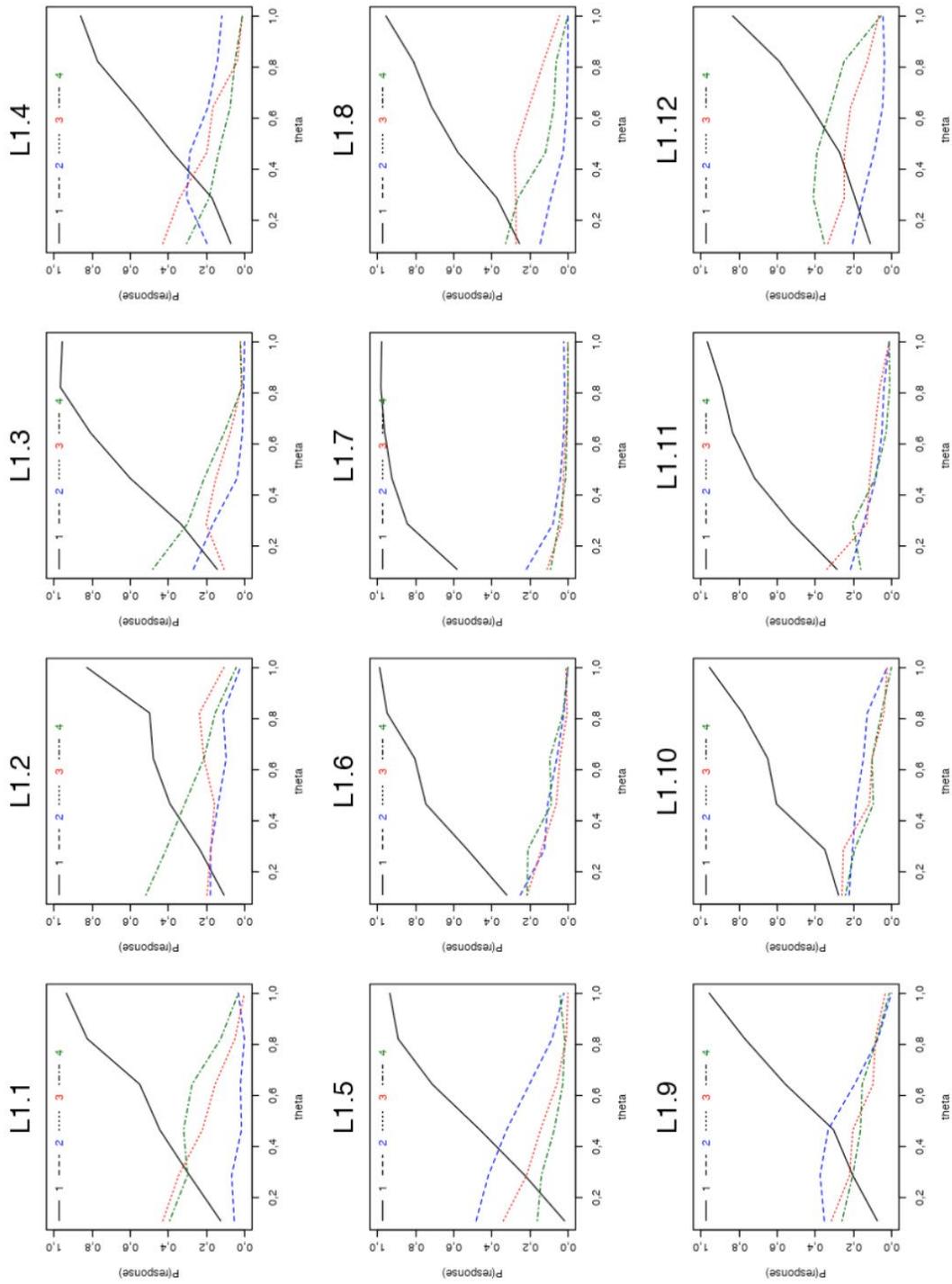
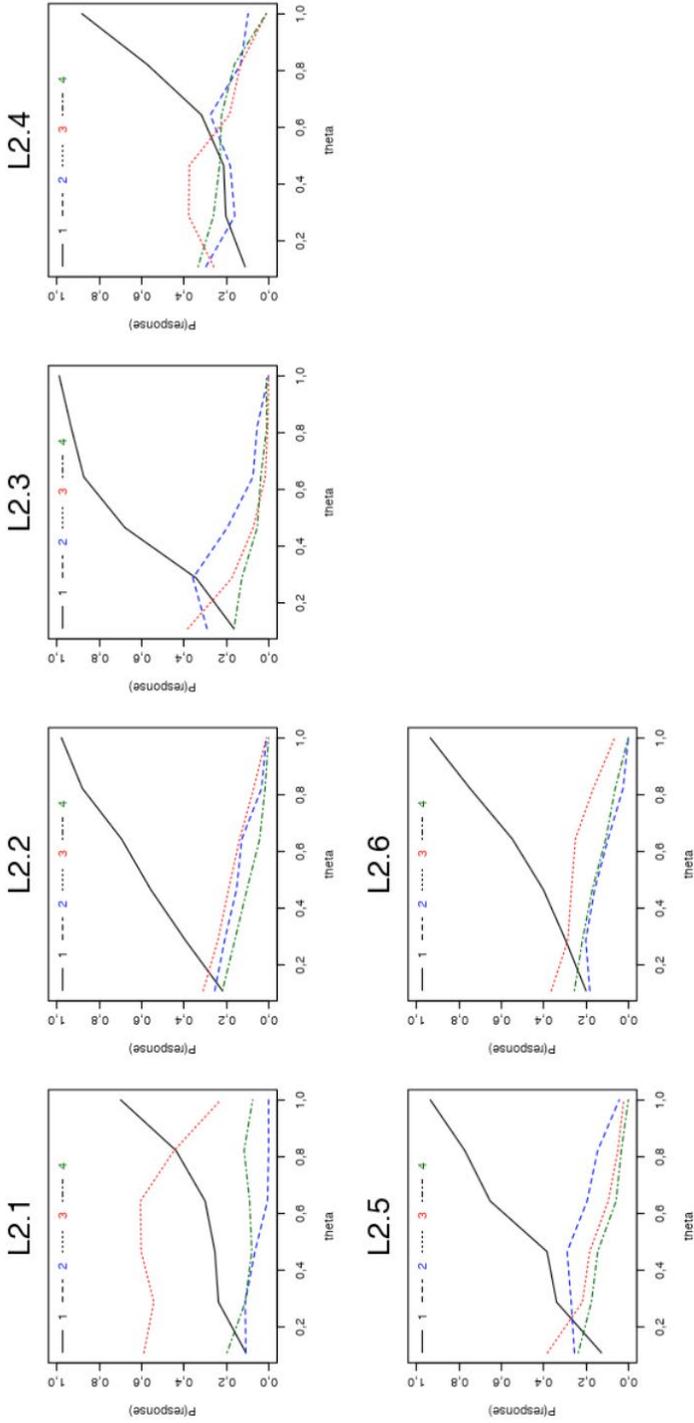


Figura A2.14 - Analisi dei distrattori B1 Tedesco - Lettura (parte 2).



Appendice Capitolo 6

Trascrizione manuale dei dati audio

Per creare dei corpora di adattamento e valutazione per il riconoscimento vocale, sono stati trascritti manualmente una parte consistente dei dati audio registrati nel 2017. Le linee guida definite per l'annotazione manuale, che sono il risultato di una serie di prove, discussioni e affinamenti, richiedevano un compromesso tra accuratezza e velocità della trascrizione. I punti più significativi rispetto a "normali" direttive per trascrizioni audio (trascrizioni ortografiche accurate di tutto quello che viene pronunciato, senza punteggiatura, tutto minuscolo) sono:

- solo il relatore principale deve essere trascritto ortograficamente; la presenza di altre voci (compagni di scuola, insegnanti) dovrebbe essere segnalata solo con l'etichetta "@voice";
- la presenza di audio sussurrato è risultata significativa, quindi dovrebbe essere esplicitamente contrassegnata con l'etichetta "()";
- le parole pronunciate male – tipicamente per ignoranza della pronuncia corretta nella lingua straniera considerata – devono essere contrassegnate da un segno "#" (senza cercare di trascrivere foneticamente i suoni pronunciati);
- le parole in una lingua diversa da quella attesa devono essere segnalate esplicitamente, come in: "i am 10 years old @it(io ho già risposto)";
- la presenza di fenomeni spontanei o rumori significativi deve essere segnalata mediante espressioni precedute da @ (esitazioni: @ehm, @mm, @ee; altri fenomeni: @cough, @laugh; rumori vari: @ns, @bkg, @voice).

La maggior parte dei dati del 2017 è stata trascritta manualmente dagli studenti di due scuole linguistiche italiane (nel seguito "C" e "S") e ricontrollata dai ricercatori di FBK. Una parte dei dati è stata trascritta in modo indipendente da coppie di studenti al fine di calcolare l'accordo tra annotatori, mostrato in tabella A6.1 in termini di Word Accuracy (WA), usando la prima trascrizione come riferimento (dopo aver normalizzato le trascrizioni, tolto le esitazioni e altre etichette relative a voci e rumori di fondo, ecc.). Il livello di accordo, piuttosto basso, riflette la difficoltà del compito, anche se si deve notare che gli stessi trascrittori erano non madrelingua inglese/tedesco. Nell'interpretare i risultati di riconoscimento del sistema sviluppato sarà quindi necessario tener presente questa incertezza.

Tabella A6.1 - Accordo inter-annotatore tra coppie di studenti, in termini di accuratezza delle parole. Gli studenti hanno trascritto prima le frasi in inglese e successivamente quelle in tedesco.

SCUOLA	LINGUA	#PAROLE TRASCritte	#PAROLE DIVERSE	ACCORDO (WA)	LINGUA	#PAROLE TRASCritte	#PAROLE DIVERSE	ACCORDO (WA)
C	<i>inglese</i>	965	237	75.44%	<i>tedesco</i>	822	139	83.09%
S	<i>inglese</i>	1370	302	77.96%	<i>tedesco</i>	1290	226	82.48%

Sia per gli esperimenti di riconoscimento vocale che per gli esperimenti di classificazione, i dati della popolazione studentesca sono stati divisi in base all'identità degli studenti in una parte di addestramento ed una di valutazione, con proporzioni di $\frac{2}{3}$ e $\frac{1}{3}$, rispettivamente, senza sovrapposizioni.

La Tabella A6.2 riporta i dati sul set di dati parlato. L'etichetta "all" identifica l'intero insieme di dati, mentre "clean" definisce il relativo sottoinsieme in cui sono state escluse le frasi contenenti voci di sottofondo, parole incomprensibili e parole frammentate. I dati etichettati come train sono stati usati per adattare i modelli acustici e linguistici del sistema, mentre i dati di test sono stati usati per calcolare le prestazioni del sistema di riconoscimento.

Tabella A6.2 - Statistiche sui dati audio (2017) trascritti manualmente.

ID	# FRASI	DURATA		# PAROLE	
		TOTALE	MEDIA	TOTALE	MEDIA
tedesco train all	1448	04:47:45	11.92	9878	6.82
tedesco train clean	589	01:37:59	9.98	2317	3.93
inglese train all	2301	09:03:30	14.17	26090	11.34
inglese train clean	916	02:45:42	10.85	6249	6.82
tedesco test all	671	02:19:10	12.44	5334	7.95
tedesco test clean	260	00:43:25	10.02	1163	4.47
inglese test all	1142	04:29:43	14.17	13244	11.60
inglese test clean	423	01:17:02	10.93	3404	8.05

Stima della proficiency

Il sistema di riconoscimento rappresenta il passo preliminare nella costruzione del classificatore che serve per assegnare un punteggio in maniera automatica alla frase da analizzare. Il

compito del sistema complessivo è dunque quello di analizzare una specifica risposta orale fornita dallo studente, di estrarre dei parametri dalla stessa (come ad es. il numero di parole utilizzate, il numero di parole corrette, la pertinenza di queste parole con la domanda) e di generare un punteggio per ciascuno degli indicatori a cui siamo interessati. Per costruire questo sistema si utilizza un approccio basato su apprendimento automatico dove i dati di riferimento sono rappresentati dai punteggi forniti dagli esperti alle frasi raccolte e analizzate nelle precedenti campagne di valutazione.

Come già chiarito precedentemente, i punteggi assegnati sono:

- 0, che rappresenta una risposta non corretta;
- 1, alla risposta parzialmente corretta;
- 2, alla risposta corretta.

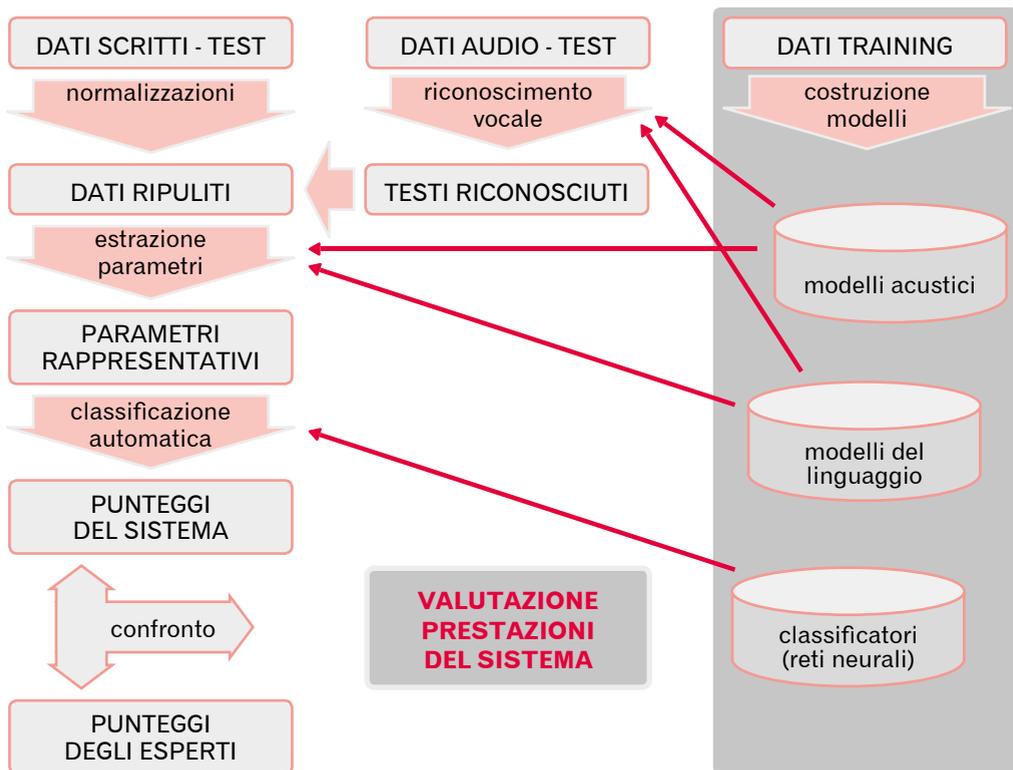
La disponibilità congiunta delle risposte degli studenti e dei relativi punteggi assegnati dai valutatori esperti ha consentito di addestrare delle reti neurali in grado di imitare il lavoro del valutatore ovvero generare un punteggio partendo da una risposta. Per dare un'idea dei parametri testuali e acustici estratti da una risposta elenchiamo alcuni valori che sono calcolati per ogni risposta e utilizzati dal classificatore per ipotizzare il corretto punteggio da assegnare:

- numeri relativi alla probabilità di osservare la sequenza di parole osservata rispetto alle tipiche risposte attese;
- numero di parole usate e numero di parole non appartenenti al dizionario atteso;
- lunghezza della frase e durata dei silenzi nella frase;
- numeri relativi alla confidenza del sistema nel riconoscere parlato madrelingua o parlato non nativo.

La frase da analizzare viene pertanto rappresentata con una serie di numeri che attraverso alcune caratteristiche descrivono il contenuto in termini di correttezza e pertinenza rispetto alla domanda; il classificatore viene quindi addestrato per imparare ad assegnare un corrispondente punteggio a questa serie sintetica di numeri.

L'intero flusso di elaborazioni effettuato dal sistema automatico di valutazione è mostrato in figura A6.1, per la classificazione sia delle risposte scritte che di quelle parlate.

Figura A6.1 - Diagramma di flusso del sistema di classificazione delle risposte scritte.



In sintesi, l'insieme delle risposte scritte dagli studenti viene inizialmente elaborato al fine di effettuare:

1. normalizzazione del testo, necessaria per le elaborazioni successive, che consiste in: *a)* codificare tutti i testi delle risposte nel formato UTF-8, *b)* espandere i numeri da cifre in lettere (ad es. "123" diventa "centoventitre"), *c)* rimuovere la punteggiatura;
2. correzione delle parole scritte erroneamente;
3. individuare le parole non conosciute, di seguito denominate <unk>;
4. individuare le parole funzionali, di seguito denominate <stop>, non di contenuto.

Ad esempio, la seguente risposta scritta: *Dir paolo I em in Amal-fi. Ai like pizza,ai like spaghetti.my name is jacopo, Amalfi is heppy.* diventa dopo la normalizzazione descritta sopra al punto 1): *dir paolo i em in amalfi ai like pizza ai like spaghetti my name is jacopo amalfi is heppy*

mentre dopo la correzione – che peraltro non riesce a correggere tutto – del passo 2) diventa: *dir paolo i am in amalfi i like pizza i like spaghetti my name is jacopo amalfi is happy*

ed infine dopo l'applicazione delle procedure definite nei punti 3) e 4), la medesima frase risulta essere: *dir paolo <stop> am <stop> amalfi <stop> <stop> pizza <stop> <stop> spaghetti <stop> name <stop> jacopo amalfi <stop> happy.*

Allo stato attuale del lavoro, l'implementazione dei punti 3) e 4) è abbastanza stabilizzata mentre il punto 2) è un modulo che necessita di alcune migliorie.

I dati scritti, elaborati come sopra visto, vengono suddivisi in due insiemi: training e test. L'insieme di training viene usato per "addestrare" dei modelli, tipicamente modelli del linguaggio (Language Models LMs), che a loro volta consentiranno di estrarre delle features dalle singole risposte scritte. L'insieme iniziale delle features con le quali abbiamo iniziato a lavorare è il seguente:

1. numero totale di parole della risposta;
2. numero di parole non presenti nel dizionario (out-of-vocabulary OOV) della lingua da classificare (parole scritte sbagliate o parole molto rare);
3. diversi valori di "perplexità" forniti da diversi LMs (cioè quanto bene la risposta viene rappresentata dai vari LMs);
4. numero di back-off effettuati dai vari LMs (quante volte nella risposta si osservano sequenze di parole mai viste dei dati);
5. numero totale di parole di contenuto (tolte le parole funzionali).

Si vuole far notare che l'insieme delle features sopra definite è tuttora oggetto di studio e che buona parte del lavoro futuro consisterà nell'implementare e valutare features più efficaci.

Una volta estratte le features dalle risposte scritte che formano l'insieme di training, si addestrano dei classificatori – basati sull'impiego di reti neurali – che hanno il compito di imparare i punteggi degli esperti. Tali classificatori vengono poi applicati all'insieme delle frasi di test e forniranno un punteggio per ognuna di esse, che dovrà assomigliare il più possibile a quello dato dagli esperti.

Come già anticipato, la procedura sopra descritta è stata applicata ai dati scritti facenti parte di tutti e tre i corpora (2016, 2017, 2018). I risultati ottenuti, per quanto molto preliminari, rispecchiano abbastanza bene quelli forniti dalla valutazione esperta.

Analisi delle risposte orali

Il flusso di elaborazione per l'analisi delle produzioni orali è riportato in figura A6.1, relativamente alle risposte scritte. La differenza principale sta nel fatto che i testi cui va applicato il

sistema di classificazione vengono prodotti da un riconoscitore automatico della voce. Una seconda differenza sta nel fatto che le features da classificare vengono ottenute anche dalle risposte audio, cioè dai rispettivi segnali, non solamente dalle loro trascrizioni automatiche. In particolare si sono utilizzate features ottenute tramite:

1. valori di probabilità forniti dal sistema ASR in corrispondenza alle parole riconosciute;
2. numero e relative durate delle esitazioni presenti nella trascrizione automatica delle frasi pronunciate dagli studenti.

Inizialmente, al fine di ottenere un primo insieme di modelli acustici di partenza, da usarsi nel sistema ASR, abbiamo preferito lavorare con registrazioni audio acquisite in ambiente controllato. A questo scopo abbiamo utilizzato delle basi di dati audio disponibili nei nostri laboratori contenenti le voci di: bambini inglesi che parlano in inglese (madrelingua), bambini tedeschi che parlano sia tedesco che inglese, bambini italiani che parlano in italiano, inglese e tedesco. Sono stati effettuati dei test di riconoscimento automatico della voce e, in questa maniera, si è configurato il sistema ASR sviluppato in FBK in maniera che possa fornire prestazioni ottimali sulla voce di bambini non nativi (i risultati ottenuti sono riportati in un articolo scientifico: Non-Native Children Speech Recognition Through Transfer Learning, Marco Matassoni, Roberto Gretter, Daniele Falavigna, Diego Giuliani) sottomesso alla conferenza ICASSP 2018. Si è quindi applicato tale sistema, facendo gli opportuni adattamenti, ai dati audio forniti da IPRASE.

Come sopra menzionato è stato controllato e trascritto manualmente un insieme di dati estratto dalle registrazioni del try-out 2017. A questo scopo ci si è avvalsi della collaborazione di una classe del Liceo linguistico "Marie Curie" di Pergine (TN).

L'intero corpus audio try-out 2017 è stato acquisito a novembre 2017 e contiene:

1. registrazioni di 542 ragazzi (ognuno pronuncia da 1 a 21 frasi);
2. 7851 frasi di lunghezza variabile (in totale circa 32 ore di parlato);
3. parlato in inglese: 4112 frasi, 511 ragazzi;
4. parlato in tedesco: 3739 frasi, 478 ragazzi;
5. in totale le risposte sono state ottenute a partire da 47 domande diverse.

Dall'insieme sopra descritto sono state selezionate, controllate e trascritte manualmente diverse ore di parlato: i dati sono riportati in tabella A6.2. Parte di tale materiale compone dei test set che sono stati usati per valutare le prestazioni dei vari sistemi ASR sviluppati nell'ambito del progetto.

Risultati di classificazione

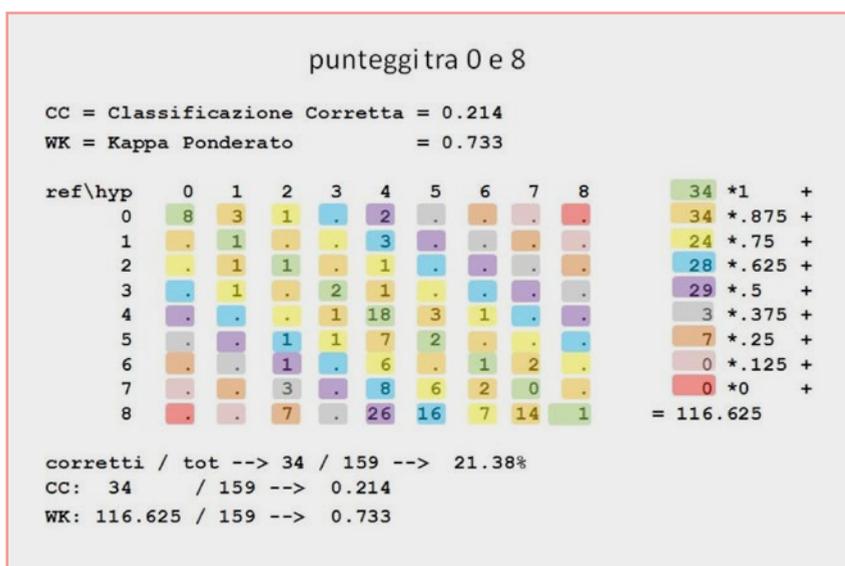
Per misurare le prestazioni di classificazione, consideriamo tre parametri: Classificazione corretta (CC), Kappa ponderato (WK), Correlazione (Corr) tra i punteggi dell'esperto e quelli ipotizzati dal sistema. I primi due vengono calcolati su una matrice di confusione, che riporta nella maniera più completa possibile il comportamento del classificatore. L'elemento i, j della matrice contiene il numero di volte in cui una frase ha ricevuto punteggio i da parte dell'esperto e j da parte del sistema automatico. Le metriche usate cercano di esprimere con un numero tutta la matrice, per permettere il confronto tra versioni successive del sistema automatico. Un esempio di calcolo di CC e WK è riportato in figura A6.2.

CC è il numero di elementi sulla diagonale (numero di classificazioni corrette) sul numero totale di elementi da classificare;

WK assegna a ogni elemento della matrice di confusione un peso che vale 1 sulla diagonale, 0 sugli angoli opposti, e un valore scalato in maniera lineare mano a mano che ci si allontana dalla diagonale;

Corr è la usuale misura di correlazione e viene calcolata sulla sequenza di coppie <punteggio esperto, punteggio sistema>.

Figura A6.2 – Matrice di confusione ed esempio di calcolo delle metriche di classificazione CC e WK, per un caso in cui il punteggio complessivo varia tra 0 e 8 (4 indicatori individuali). CC considera positivamente solo i valori sulla diagonale, mentre WK assegna punteggi via via scalati mano a mano che ci si allontana dalla diagonale.



Gli indici CC e WK variano tra 0 e 1, mentre Corr varia tra -1 e +1. Per tutti e tre gli indici, 0 indica classificazione completamente sbagliata, mentre 1 indica classificazione perfetta.

Per ciascun indicatore, abbiamo utilizzato i dati di addestramento per addestrare una rete neuronale di tipo *feedforward*, raggruppando tutte le frasi che condividono lingua, livello di competenza e sessione. I risultati della classificazione media sono riportati in tabella A6.3 per training e test 2018, mentre la tabella A6.4 riporta risultati in termini di CC e WK su tutti i dataset di test.

Tabella A6.3 - Risultati della classificazione media sui dati 2018, parlato, riportati sia su dati di addestramento che di test, in termini di classificazione corretta (CC), Kappa ponderato (WK), correlazione (Corr).

DATASET	INGLESE			TEDESCO		
	CC	WK	CORR	CC	WK	CORR
2018 train	0.712	0.840	0.684	0.763	0.866	0.763
2018 test	0.596	0.775	0.532	0.667	0.822	0.613

Tabella A6.4 - Risultati della classificazione sui dati di test per inglese, tedesco; scritto, parlato; campagne 2016, 2017, 2018. Le colonne 2 e 3 (ind) riportano i risultati medi di classificazione per i singoli indicatori (con punteggio 0, 1, 2) mentre le colonne 4 e 5 riportano i risultati di classificazione rispetto ai punteggi complessivi (da 0 a 8/10/12 a seconda del dataset).

DATI DI TEST, SCRITTO	IND CC	IND WK	TOT CC	TOT WK
inglese 2016	0.57	0.77	0.24	0.80
inglese 2017	0.55	0.74	0.17	0.75
inglese 2018	0.59	0.78	0.22	0.80
tedesco 2016	0.62	0.80	0.29	0.81
tedesco 2017	0.58	0.77	0.21	0.81
tedesco 2018	0.63	0.81	0.26	.082

DATI DI TEST, PARLATO	IND CC	IND WK	TOT CC	TOT WK
inglese 2016	0.60	0.65	0.38	0.60
inglese 2017	0.56	0.72	0.31	0.73
inglese 2018	0.59	0.77	0.31	0.78
tedesco 2016	0.76	0.82	0.48	0.75
tedesco 2017	0.58	0.71	0.38	0.72
tedesco 2018	0.66	0.81	0.40	0.84

Guardando i risultati nelle tabelle A6.3 e A6.4, le prestazioni in termini di tutte le metriche riportate (CC, WK e Corr) sono buone, mostrando che i punteggi assegnati automaticamente non sono lontani da quelli manuali assegnati da esperti umani. La bassa differenza tra prestazioni in training e test in tabella A6.3 indica che i modelli non hanno imparato a memoria i dati di addestramento. Ancora più importante, i valori dei coefficienti di correlazione raggiunti sono simili a quelli riportati in Ramana 2017, relativi alla correlazione tra punteggi assegnati da esperti umani su un task simile, in termini di difficoltà per gli studenti a quanto presentato in questo lavoro.

Si è cercato inoltre di verificare se alcuni parametri (features), specificatamente progettati per dare un contributo significativo ad alcuni indicatori, permettessero di migliorare le prestazioni di classificazione proprio per tali indicatori. Ad esempio:

- indicatori di fluenza: il riconoscimento vocale permette di conoscere istante di inizio e di fine di ogni parola riconosciuta, quindi di estrarre parametri come “numero di parole pronunciate”, “numero di sillabe al secondo”, “durata di tutte le pause tra parole”, “durata complessiva della frase pronunciata (inclusi/esclusi silenzi)”. Ci si aspetta che questi indicatori possano migliorare le prestazioni per gli indicatori di fluenza ma non, ad esempio, per gli indicatori di contenuto semantico.
- indicatori di contenuto: dalla parte di addestramento si possono estrarre le parole di contenuto più frequenti (bag of words), per le frasi che hanno ricevuto punteggi alti dagli esperti. Per una frase da valutare si possono calcolare parametri “numero di parole nella bag of words”, “percentuale di parole nella bag of words”, che indicano se la risposta contiene alcune delle parole che ci si aspetta.

Non si è tuttavia riusciti a trovare evidenza di tale fenomeno, peraltro atteso. I motivi sono probabilmente da ricercare nei seguenti fattori:

- il numero di parametri (features) utilizzati è relativamente alto (attorno al centinaio) quindi l'aggiunta di qualche parametro specifico non riesce a dare un contributo significativo alla classificazione;
- la dimensionalità dei dati su cui vengono addestrati i sistemi di classificazione potrebbe essere troppo bassa;
- la correlazione tra punteggi degli esperti relativa ai vari indicatori è risultata piuttosto alta (vedi tabella A6.5), quindi un miglioramento su uno specifico indicatore tende comunque a riflettersi anche su altri indicatori.

Tabella A6.5 - Correlazione media tra punteggi degli esperti assegnati ai singoli indicatori, nettamente più alta nel caso parlato rispetto al caso scritto.

SOTTOINSIEMI	TEDESCO		INGLESE	
	CORRELAZIONE MEDIA TRA INDICATORI	CORRELAZIONE MEDIA TRA INDICATORI E PUNTEGGIO TOTALE	CORRELAZIONE MEDIA TRA INDICATORI	CORRELAZIONE MEDIA TRA INDICATORI E PUNTEGGIO TOTALE
scritto				
A1 S2	0.68	0.86	0.68	0.85
A2 S1	0.66	0.85	0.68	0.85
A2 S2	0.68	0.85	0.67	0.85
B1 S1	0.72	0.88	0.64	0.83
B1 S2	0.71	0.87	0.68	0.85
parlato				
A1 P1	0.87	0.94	0.85	0.93
A1 P2	0.82	0.92	0.82	0.92
A2 P1	0.66	0.85	0.71	0.87
A2 P2	0.55	0.68	0.76	0.89
B1 P1	0.77	0.89	0.74	0.88
B1 P2	0.68	0.86	0.68	0.86

Finito di stampare
nel mese di marzo 2019
presso **la grafica** srl, Mori (TN)

Dopo una prima edizione svolta nel 2016, si presentano in questo volume gli esiti della seconda rilevazione delle competenze linguistiche degli studenti trentini al grado quinto, ottavo, decimo per la secondaria di secondo grado e undicesimo per l'istruzione e formazione professionale.

La rilevazione si propone in particolare di restituire in modo dettagliato gli esiti delle prove standardizzate nelle abilità ricettive (ascolto e lettura) e nelle abilità produttive (produzione orale e scritta) nonché di presentare un'analisi puntuale delle determinanti dei risultati ottenuti dagli studenti utilizzando informazioni raccolte con un questionario sociolinguistico predisposto ad hoc. Il quadro viene completato dalla presentazione del campione di studenti oggetto di analisi, da un approfondimento psicometrico sui risultati della rilevazione insieme a un lavoro sperimentale e innovativo di valutazione automatica delle abilità linguistiche del parlato e della scrittura.

Il presente studio rappresenta quindi un tassello unico nel panorama trentino e nazionale nella direzione, avviata nel 2016, di rafforzare know how, competenze e strumenti di testing degli apprendimenti linguistici a livello provinciale. Ciò al fine di mettere tutte le istituzioni scolastiche trentine in condizione di migliorare la didattica e le abilità linguistiche delle nuove generazioni di studenti, individuandone punti di forza e criticità alla luce di evidenze empiriche fondate e mirate.