## La ricerca nel complesso

La ricerca che abbiamo deciso di replicare è stata condotta negli Stati Uniti ed investiga l’impatto di un intervento di alfabetizzazione sugli studenti di prima elementare di lingua spagnola che hanno difficoltà. L’intervento applicato, **Descubriendo La Lectura (DLL)**, fornisce istruzione di alfabetizzazione one-to-one in lingua spagnola agli studenti di prima elementare che parlano spagnolo come lingua madre ed è implementato da insegnanti bilingui per un periodo che va dalle 12 alle 20 settimane a seconda delle esigenze del singolo studente.

I ricercatori, per valutarne l’impatto sulle competenze di alfabetizzazione in spagnolo e in inglese, hanno predisposto un randomized control trial multisito e multicohort a livello di studente.

Il **Randomized Control Trial (RCT**) è un disegno di studio in cui i partecipanti vengono assegnati in modo casuale a uno dei due gruppi: un gruppo di trattamento che riceve l'intervento o il trattamento in studio, e un gruppo di controllo che non riceve l'intervento. L'assegnazione casuale ai gruppi aiuta a ridurre il bias e a rendere i gruppi confrontabili.

**Multisito** in quanto lo studio coinvolge più siti o location, come diverse scuole o istituti, invece di essere condotto in un unico luogo. Ciò permette di generalizzare i risultati su una più ampia popolazione o contesto. Nel nostro caso, diverse scuole di diversi stati.

**Multicohort** perchè lo studio coinvolge più gruppi di studenti anche in diversi periodi temporali.

Scopo finale della ricerca è trovare una metodologia efficace che colmi il gap tra i madrelingua inglese e gli studenti di lingua spagnola, i quali costituiscono già una cospicua minoranza nel tessuto sociale americano, minoranza che è destinata a crescere nei prossimi anni.

Lo studio inoltre può essere utile a tutti quelli Stati che hanno dei tassi di immigrazione elevati, come ad esempio gli Stati europei che ogni anno accolgono centinaia di migranti provenienti dall’Africa e dall’Asia in fuga da guerra e crisi climatica.

Le ricerche accademiche suggeriscono che l'uso estensivo dell'istruzione nella lingua madre supporta lo sviluppo linguistico e della lettura degli **ELL (English Language Learners)**. Inoltre, imparare le competenze di base nella lingua madre tende a trasferirsi allo sviluppo delle competenze nella seconda lingua. Ci sono prove infatti che la capacità di lettura precoce nella lingua madre è predittiva della successiva competenza in inglese. Altre ricerche, invece, suggeriscono che il tutoring uno a uno si è dimostrato efficace per gli studenti che richiedono il maggior livello di supporto, soprattutto per migliorare il rendimento nella lettura.

## Metodo e campionamento

Dopo aver identificato gli studenti di prima elementare che parlano spagnolo con difficoltà di lettura, questi sono stati divisi in due gruppi: coloro che avrebbero ricevuto i servizi di DDL all’inizio dell’anno scolastico (gruppo immediato) e coloro i quali li avrebbero ricevuti successivamente (gruppo ritardato), il quale è servito da gruppo di controllo. Le scuole sono state scelte tra quelle che già avevano implementato i servizi di DDL per almeno un anno prima dell’inizio dell’esperimento e gli studenti tra quelli che erano di madrelingua spagnola e facevano parte del 25% degli studenti con punteggio più basso nelle loro scuole. Il campione del primo gruppo include 152 studenti provenienti da 22 scuole di tre Stati (Texas, Illinois e Arizona).

L’assegnazione ai due gruppi è stata effettuata mediante l’uso di un generatore di numero casuale e quelli con numeri più bassi sono stati serviti per primi.

Tutti gli studenti sono stati testati su 3 misure: **IdO**, **Logramos** e **ITBS** prima e dopo il trattamento del gruppo “immediato”.

Riguardo alle scale di valutazione sono 2 in spagnolo e 1 in inglese rispettivamente IdO e Logramos per lo spagnolo e ITBS per l’inglese. Lo scopo delle diverse somministrazioni è dovuto al voler effettuare analisi diverse tra loro. Le due valutazioni in lingua spagnola per non confondere la conoscenza della lingua inglese con le abilità di alfabetizzazione generali, mentre il test ITBS per misurare in che misura gli studenti possono aver ottenuto miglioramenti nelle loro abilità di alfabetizzazione in inglese.

Purtroppo non era presente il dataset che raccoglieva i dati riguardanti l’IdO, pertanto abbiamo potuto effettuare soltanto una replica parziale dell’esperimento.

In particolare, i dataset contengono l'ID dello studente, l'ID della scuola di provenienza, il gruppo di appartenenza ovvero se hanno ricevuto il trattamento o se non lo hanno ricevuto, i punteggi pre e post trattamento in reading, language, vocabulary che poi vengono combinati in un punteggio complessivo (ELA.TOTAL), oltre che la variabile booleana T\_assignment che è uguale a 0 per gli studenti non trattati e 1 per gli studenti trattati. I dataset per **Logramos** e **ITBS** sono strutturati in maniera identica.

## Elaborazione dati

Dopo aver caricato i due dataset disponibili e importato le librerie di cui abbiamo fatto uso, abbiamo dapprima verificato le **attriction** presenti nel paper e abbiamo creato le due tabelle **attrition\_df** e **diff\_attrition\_df** che ne riassumono i risultati confermando quanto espresso nel paper.

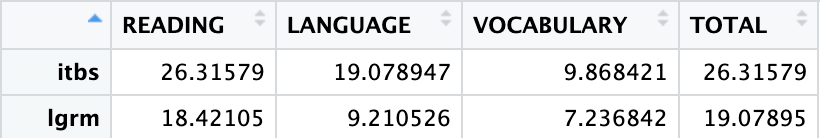


Figura 1. Attrition\_df

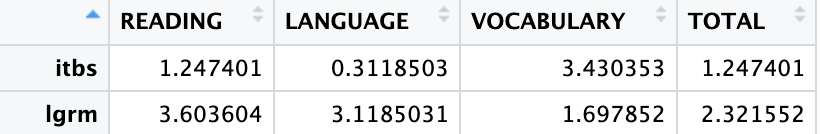


Figura 2. Diff\_attrition\_df

Successivamente, abbiamo creato la funzione “*cleaning\_itbs\_lgrm”* che prende in input un dataset (**DATA\_a**) contenente i dati relativi ai test ITBS e Logramos, insieme ai nomi delle colonne dei punteggi post-test (**POSTTESTM**) e pre-test (**PRETESTM**). La funzione innanzitutto standardizza i punteggi dei test: ovvero vengono trasformati in z-score. La funzione *"Make.Z"* viene utilizzata per calcolare gli z-score dei punteggi dei test specificati nel dataset DATA\_a. L’operazione si rende necessaria in quanto i due test hanno scale diverse che devono essere standardizzate per poter essere rese equivalenti tra loro. In particolare, tutti i punteggi di pre e post-test sono stati standardizzati con una media di 0 e una deviazione standard di 1. Successivamente la funzione rinomina gli z-score. Ad esempio, i nomi delle colonne degli z-score per i punteggi di lettura post-test e pre-test vengono rinominati rispettivamente come **"Posttest\_READING\_z"** e **"Pretest\_READING\_z"**. Poi, il dataset originale viene unito con gli z-score creando il nuovo dataset chiamato **DATA\_b** utilizzando la funzione *"cbind".* Infine, elimina tutti i casi in cui mancano i dati del pre-test o del post-test: cioè vengono rimossi i casi dal dataset **DATA\_b** in cui i valori per il pre-test o il post-test sono mancanti (NA).

Abbiamo poi applicato la funzione alle quattro valutazioni (READING, LANGUAGE, VOCABULARY e TOTAL) sia per la scala ITBS che per quella Logramos.

Ciò ci ha consentito di visualizzare il numero di studenti che hanno ricevuto il trattamento immediatamente oppure in maniera ritardata a seconda dell’assegnazione casuale:

* **table(DLL\_itbs$Group).** Risultato: Delayed: 74 Immediate: 78
* **table(DLL\_lgrm$Group).** Risultato: Delayed: 74 Immediate: 78

Ma soprattutto visualizziamo per ogni categoria di quanti siamo riusciti a tenere traccia:

* **table(itbs\_READING$Group).** Risultato: Delayed: 55 Immediate: 57
* **table(itbs\_LANGUAGE$Group).** Risultato: Delayed: 60 Immediate: 63
* **table(itbs\_VOCABULARY$Group)**. Risultato: Delayed: 68 Immediate: 69
* **table(itbs\_ELA.TOTAL$Group).** Risultato: Delayed: 55 Immediate: 57
* **table(lgrm\_READING$Group).** Risultato: Delayed: 59 Immediate: 65
* **table(lgrm\_LANGUAGE$Group).** Risultato: Delayed: 66 Immediate: 72
* **table(lgrm\_VOCABULARY$Group).** Risultato: Delayed: 68 Immediate: 73
* **table(lgrm\_ELA.TOTAL$Group).** Risultato: Delayed: 59 Immediate: 64

Dopo aver creato il vettore di stringhe *cov.nam\_itbs\_lgrm*, contenente i nomi delle variabili corrispondenti alle misure dei test ITBS (Iowa Tests of Basic Skills) e Logramos. lo utilizziamo per creare una tabella riassuntiva (summary table) utilizzando la funzione “*CreateTableOne”*. La tabella riassuntiva viene costruita per ogni dataset pulito precedentemente costruito. Le variabili vengono suddivise in base alla colonna "Group" presente in tutti i dataset, che rappresenta la variabile di trattamento. L'opzione test = FALSE indica che non verranno eseguiti test statistici sulla differenza tra i gruppi.

Da ognuna di queste tabelle, ricaviamo un nuovo dataframe che contiene la tabella riassuntiva e gli SMD (Standardized Mean Differences) utilizzando l'opzione smd = TRUE. Gli SMD rappresentano una misura della dimensione dell'effetto tra i gruppi di trattamento e controllo, standardizzata rispetto alla deviazione standard dei gruppi.

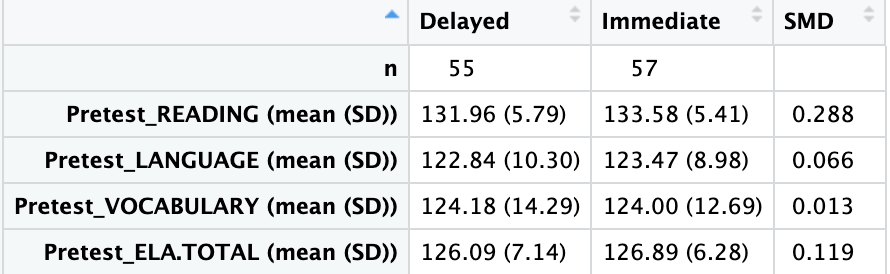


Figura 3.Itbs\_read\_by\_trt\_smd

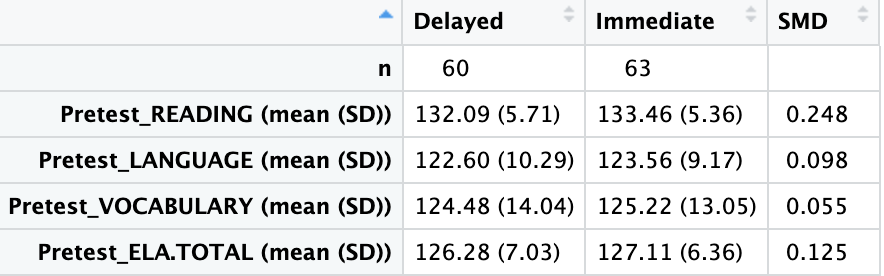


Figura 4.Itbs\_lang\_by\_trt\_smd

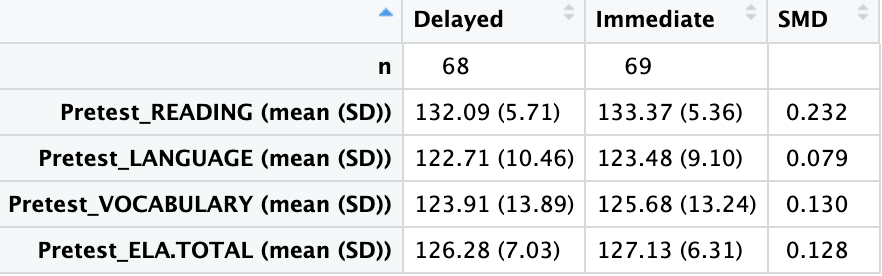


Figura 5.Itbs\_voca\_by\_trt\_smd

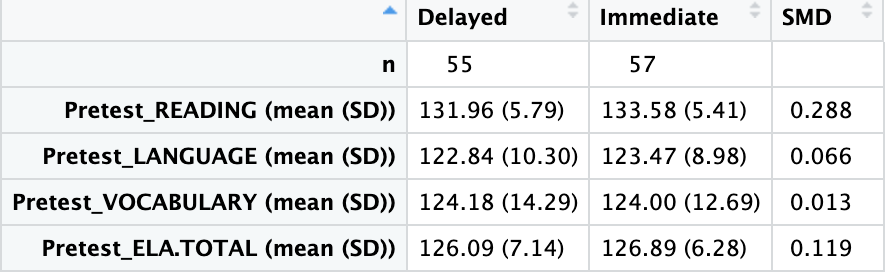


Figura 6.Itbs\_ttl\_by\_trt\_smd

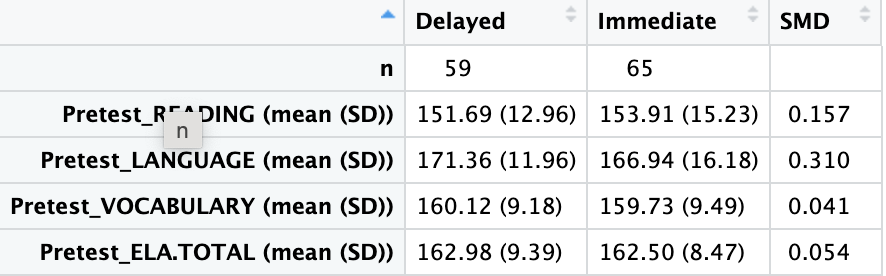


Figura 7.Lgrm\_read\_by\_trt\_smd

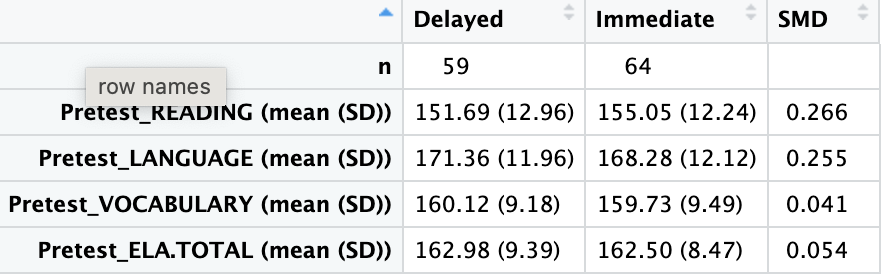


Figura 8. Lgrm\_lang\_by\_trt\_smd

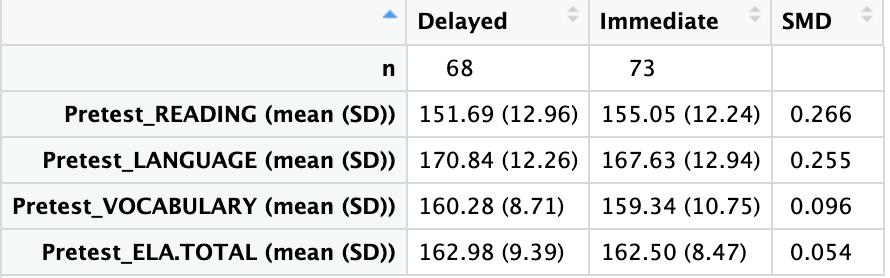


Figura 9.Lgrm\_voca\_by\_trt\_smd

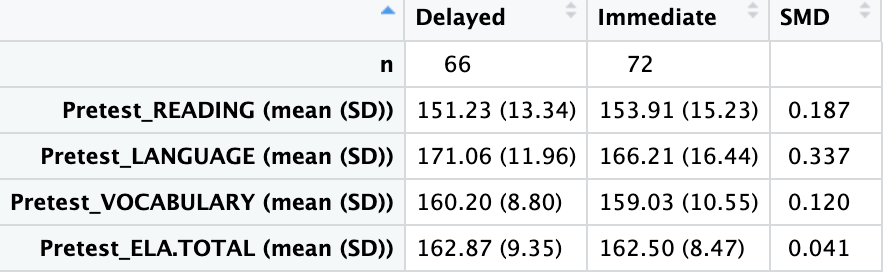


Figura 10.Lgrm\_ttl\_by\_trt\_smd

Una volta fatto ciò, siamo passati allo step successivo: la costruzione di una funzione che crei dei **modelli random intercept** per le quattro viariabili. I modelli *random intercept* utilizzano la funzione *“lme”* del pacchetto *“nlme”.* Ogni modello include le variabili indipendenti T\_assignment (assegnazione del trattamento) e Pretest\_ (valore del pretest) come predittori fissi. Inoltre, viene specificato un termine casuale **~ 1 | SCHOOLID** per catturare le variazioni casuali tra le scuole.

I modelli vengono memorizzati nella lista *“modellist”*, e successivamente viene utilizzata la funzione *“screenreg”* del pacchetto *“texreg”* per generare una tabella dei risultati dei modelli. La funzione *“screenreg”* restituisce la tabella dei risultati dei modelli con i nomi personalizzati delle variabili specificati in “*custom.model.names”.*

La funzione restituisce la tabella dei risultati dei modelli. “*DLL\_RI\_itbs\_lgrm”* è una funzione che sarà applicata indifferentemente alle variabili di tipo itbs che logramos in quanto simili e aventi lo stesso numero di dataset.

Di seguito i risultati relativi all’ITBS:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VARIABILI** | **READING** | **LANGUAGE** | **VOCABULARY** | **TOTAL** |
| (Intercept) | -0.120 | -0.035 | -0.099 | -0.075 |
|  | (0.141) | (0.138) | (0.144) | (0.140) |
| T\_assignment | 0.189 | 0.122 | 0.292 | 0.167 |
|  | (0.178) | (0.176) | (0.160) | (0.178) |
| Pretest\_READING\_z | 0.152 |  |  |  |
|  | (0.096) |  |  |  |
| Pretest\_LANGUAGE\_z |  | 0.118 |  |  |
|  |  | (0.091) |  |  |
| Pretest\_VOCABULARY\_z |  |  | 0.062 |  |
|  |  |  | (0.089) |  |
| Pretest\_ELA.TOTAL\_z |  |  |  | 0.204\* |
|  |  |  |  | (0.094) |
|  |  |  |  |  |
| AIC | 321.980 | 362.866 | 398.256 | 323.563 |
| BIC | 335.436 | 376.804 | 412.746 | 337.019 |
| Log Likelihood | -155.990 | -176.433 | -194.128 | -156.781 |
| Num. obs. | 112 | 123 | 137 | 112 |
| Num. groups: SCHOOLID | 18 | 20 | 22 | 18 |
|  |  |  |  |  |
| \*\*\* p < 0.001 | \*\* p < 0.01 | \* p < 0.05 |  |  |

Di seguito i risultati relativi al Logramos:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **VARIABILI** | **READING** | **LANGUAGE** | **VOCABULARY** | **TOTAL** |
| (Intercept) | \*-0.278 | -0.204 | -0.240 | \*-0.292 |
|  | (0.115) | (0.132) | (0.122) | (0.123) |
| T\_assignment | \*\*\*0.542 | \*0.407 | \*\*0.473 | \*\*\*0.548 |
|  | (0.159) | (0.163) | (0.157) | (0.159) |
| Pretest\_READING\_z | \*\*\*0.417 |  |  |  |
|  | (0.080) |  |  |  |
| Pretest\_LANGUAGE\_z |  | 0.153 |  |  |
|  |  | (0.083) |  |  |
| Pretest\_VOCABULARY\_z |  |  | \*\*0.239 |  |
|  |  |  | (0.082) |  |
| Pretest\_ELA.TOTAL\_z |  |  |  | \*\*\*0.425 |
|  |  |  |  | (0.081) |
| AIC | 336.137 | 399.337 | 400.444 | 337.141 |
| BIC | 350.116 | 413.863 | 415.080 | 351.078 |
| Log Likelihood | -163.068 | -194.668 | -195.222 | -163.571 |
| Num. obs. | 124 | 138 | 141 | 123 |
| Num. groups: SCHOOLID | 20 | 21 | 22 | 20 |
| \*\*\* p < 0.001 | \*\* p < 0.01 | \* p < 0.05 |  |  |