



UNIVERSITÀ DEGLI STUDI
DI GENOVA



Dipartimento Di Scienze Della Formazione (DISFOR)
CORSO DI DOTTORATO DI RICERCA IN SCIENZE SOCIALI
XXXIV CICLO- CURRICULUM SOCIOLOGIA

**Relazione annuale sul progetto di ricerca e sulle attività svolte
nell'anno accademico 2019/2020**

Tutor: Prof. Mauro Palumbo
Co-Tutor: Dott. Roberto Ricci (INVALSI)

Dottoranda DANIELA TORTI matricola 4622202

**E-mail: daniela.torti@edu.unige.it
daniela.torti_ext@invalsi.it**

CONTENUTI:

- A) ATTIVITÀ FORMATIVA, SEMINARI E CONVEGNI**
- B) ATTIVITÀ DI RICERCA SVOLTA PRESSO INVALSI**
- C) ATTIVITA' DIDATTICA**
- D) PUBBLICAZIONI**
- E) MISSIONI**
- F) PROGETTO DI RICERCA**
- G) BIBLIOGRAFIA**

Io sottoscritta Daniela Torti, nata a Sora (FR) il 09-09-1989 C.F. TRTDNL89P49I838C, iscritta al secondo anno del Dottorato di ricerca in scienze sociali, curriculum sociologia (34° ciclo) con borsa di dottorato INVALSI, dichiaro di aver partecipato alle attività di formazione e ricerca esposte di seguito.

A) ATTIVITÀ FORMATIVA, SEMINARI E CONVEGNI

Ho partecipato ai seguenti convegni e seminari in qualità di **relatore** presentando i seguenti contributi:

-**Torti, D.** (2019) "Le opinioni degli insegnanti sulla valutazione e le sue implicazioni con il concetto di Apprendere a Apprendere: un'analisi preliminare di un'intervista in alcune scuole italiane". The teachers' opinions about assessment and its implications with Learning to learn concept: a preliminary analysis of an interview in Italian schools". IV Seminario "I dati INVALSI: uno strumento per la ricerca e la didattica" tenutosi il 29-30 novembre e 1° dicembre 2019 a Roma.

-**Torti, D.** (2019) "Il punto di vista degli insegnanti sulla valutazione" XXIII congresso nazionale AIV, associazione italiana di Valutazione previsto a Bari 02-04 aprile 2020 ma rinviato a data da destinarsi causa emergenza COVID-19 e delle disposizioni dettate dal D.P.C.M. 4 marzo 2020.

Ho partecipato ai seguenti convegni e seminari in qualità di **discente**:

- "Risultati della sperimentazione del RAV per le scuole dell'infanzia" INVALSI, 15 luglio 2020.
- "La scuola non si ferma: costruire il Piano per la didattica digitale integrata", Fondazione per la scuola, Riconessioni, 07 settembre 2020.
- Seminario riservato a personale interno INVALSI su formative testing il 12 maggio 2020.
- "FLIP+ ITEM LIBRARY WORKSHOP" tenuto a Roma presso INVALSI il 28 ottobre 2019.
- Iscrizione al "SEMINARIO INTERNAZIONALE ADI" previsto a Bologna 28-29 febbraio 2020 ma rinviato al 26-27 febbraio 2021 causa emergenza COVID-19.

B) ATTIVITÀ DI RICERCA SVOLTA PRESSO INVALSI

Nell'ambito del dottorato ho continuato a svolgere attività di ricerca nel progetto Apprendere ad apprendere in Italia, Europa e America Latina (L2L) e nel progetto Sperimentazione RAV Infanzia. Nello specifico ho svolto le seguenti attività per il progetto L2L:

- supporto al coordinamento dello studio qualitativo e messa a punto di strumenti d'indagine qualitativi;

- costruzione e pilotaggio strumento d'indagine qualitativa (traccia intervista semi strutturata);
- realizzazione di interviste semi-strutturate ad insegnanti di scuola dell'infanzia, primaria, secondaria di primo e secondo grado;
- analisi dati qualitativi;
- analisi dati qualitativi con programma di analisi MAXQDA;
- partecipazione a convegni e pubblicazioni.

Nello specifico, ho svolto le seguenti attività per il progetto Sperimentazione RAV Infanzia:

- Assistenza alle scuole dell'infanzia, del gruppo di riferimento e auto-candidate, durante la fase di sperimentazione;
- gestione contatti con le scuole e casella di posta elettronica;
- redazione dei cap. 3, 4 e 12 del rapporto sulla sperimentazione del RAV per la scuola dell'Infanzia (https://www.invalsi.it/infanzia/docs/Rapporto_RAV_Infanzia_def.pdf);
- supporto per l'organizzazione della videoconferenza "RAV Infanzia: i risultati della sperimentazione" realizzato il 15 luglio 2020.

C) ATTIVITA' DIDATTICA

Come da autorizzazione del Tutor Prof. Mauro Palumbo e dal Coordinatore del Corso di Dottorato in Scienze Sociali Prof. Luca Andrighetto, svolgo attività di tutor didattico presso l'università telematica Unitelma Sapienza di Roma per la cattedra della Prof.ssa Alby per il corso di psicologia sociale dei gruppi e per la cattedra dei Proff. Ajello Annamaria e Giampaolo Nicolais per il corso di Psicologia dello sviluppo e dell'educazione. L'attività consiste nell'allestire il corso in piattaforma, caricare video-lezioni, materiali di approfondimento ed altro, supporto al docente per la gestione del corso in generale e delle attività richieste, assistenza agli studenti tramite casella e-mail personale ed assistenza al docente per lo svolgimento dell'esame.

Da marzo 2020 Collaborazione all'attività didattica in qualità di cultore della materia per l'insegnamento di psicologia sociale della Professoressa Francesca Alby presso la facoltà di medicina e psicologia, università La Sapienza di Roma.

Da marzo 2020 Collaborazione all'attività didattica in qualità di cultore della materia per l'insegnamento di psicologia dello sviluppo e dell'educazione della Professoressa Anna Maria Ajello presso la facoltà di medicina e psicologia, università La Sapienza di Roma.

D) PUBBLICAZIONI

- Brito-Rivera, H.A., **Torti, D.**, Malheiro Ferraz de Carvalho,A., (2020 in press) "Diseño, traducción y adaptación de una entrevista sociocultural a docentes sobre Aprender a Aprender". Aula Abierta.
- Stringher, C., Brito Rivera, H.A., Patera, S.,Silva Silva, I., Castro Zubizarreta, A., Davis Leme, C.,**Torti, D.**, Huerta, M.d.C., Scrocca, F. (2020 in press). Learning to learn and assessment: complementary concepts or different worlds?. Educational Research.
- **Torti, D.** (2020 in press). Le opinioni degli insegnanti sulla valutazione e le sue implicazioni con il concetto di "Apprendere ad Apprendere": un'analisi preliminare su interviste svolte in alcune scuole italiane. INVALSI per la ricerca, Franco Angeli.
- Redazione dei cap. 3, 4 e 12 del rapporto sulla sperimentazione del RAV per la scuola dell'Infanzia (https://www.invalsi.it/infanzia/docs/Rapporto_RAV_Infanzia_def.pdf).

E) MISSIONI

Per quanto riguarda il FONDO SPESA TORTI 10% dichiaro di non aver usufruito del budget disponibile nell'anno accademico 2019/2020. Avevo inizialmente richiesto di effettuare una missione formativa per la partecipazione al “seminario internazionale ADI” previsto a Bologna 28-29 febbraio 2020, ma a causa dell'emergenza COVID-19 l'evento è stato posticipato e quindi al momento non risulta nessuna richiesta di rimborso inoltrata per mio conto. Considerato lo slittamento della ricerca sul campo, conto di utilizzare il budget per le missioni, le attività di formazione e l'acquisto di eventuale strumentazione necessaria nel nuovo anno accademico.

F) PROGETTO DI RICERCA

A causa dell'emergenza sanitaria Covid-19, lo stato di avanzamento della ricerca ha subito un arresto. Considerando che il progetto si svolge nell'ambito dell'attività di ricerca di INVALSI e più specificatamente delle prove nazionali, le quali per l'anno 2020 sono state sospese, e che prevede il coinvolgimento delle organizzazioni scolastiche, è in atto una profonda ridefinizione del disegno della ricerca con campione, fasi e strumenti.

La prima fase di ricerca quantitativa sul campo sarebbe dovuta iniziare nel mese di maggio 2020 in corrispondenza delle prove nazionali. Con il tutor Prof. Mauro Palumbo e co-tutor Dott. Roberto Ricci, si è convenuto di investire il periodo di fermo nell'analisi della letteratura al fine di orientare al meglio la ristrutturazione del progetto. Si sta valutando la discesa sul campo per il periodo di gennaio 2021 con una fase di ricerca qualitativa.

Il progetto di ricerca, dal titolo provvisorio "*Carta, computer e tablet: uno studio quali-quantitativo per esplorare il mode effect sulla comprensione del testo in una prova INVALSI nel grado 6*", è volto ad indagare il *mode effect* sulla comprensione del testo in una prova INVALSI di italiano in studenti di classe prima di scuola secondaria di primo grado.

Dall'anno accademico 2017/2018 l'INVALSI ha introdotto le prove CBT nelle rilevazioni nazionali per le classi III secondaria di primo grado e la II secondaria di secondo grado, sulla base di una specifica prescrizione di legge (D. Lgs. n. 62 del 13 aprile 2017).

Per l'implementazione di questa nuova procedura è stato di fondamentale importanza considerare alcuni parametri tra cui la finalità della valutazione, la comparabilità fra il test cartaceo e il test computer based, la costruzione di prove equivalenti, l'identificazione dell'oggetto della misurazione e, molto importante, la familiarità nell'uso dello strumento digitale da parte dell'utenza. Il passaggio da prove basate su carta e penna a CBT, quindi, non è comunque un processo scontato, bensì un terreno su cui si possono individuare diverse caratteristiche che vanno approfondite, come ad esempio la variazione dei contenuti, il formato dei compiti oppure ancora le operazioni disponibili al risolutore in relazione alla natura dello strumento utilizzato.

Su questi temi sono state condotte diverse ricerche che hanno messo in evidenza le differenze nei risultati degli studenti nei tradizionali test basati su carta e matita (PBT) rispetto ai test computer based (CBT) (Noyes & Garland, 2008). Tale fenomeno, chiamato "test mode effect", genericamente fa riferimento alla probabilità di ottenere prestazioni degli studenti differenziate a causa delle diverse modalità di somministrazione della stessa prova (Clariana & Wallace, 2002; Wang et al., 2007; Parshall & Kromrey, 1993;)

Sono stati identificati molti fattori che potrebbero essere responsabili di tale differenza nella prestazione degli studenti (Leeson, 2006). Tuttavia, la questione non appare del tutto risolta, con risultati ancora contrastanti e per certi versi ambigui e non conclusivi (Way, Davis, Keng, & Strain-Seymour, 2016). Sembra però che il verificarsi del "mode effect" sia correlato a fattori specifici, come ad esempio le caratteristiche degli studenti, la familiarità nell'uso del dispositivo, le caratteristiche del test, ecc (Dadey, Lyons e DePascale, 2018; Jeong, 2014; Jerrim, 2016). La maggior parte degli studi condotti inoltre, si è concentrata particolarmente sull'uso dei computer e meno sull'uso specifico dei tablet per la valutazione degli studenti.

In considerazione dei diversi vantaggi che una valutazione digitalizzata potrebbe garantire (riduzione sostanziale del fenomeno del *cheating*, riduzione dei costi, maggiore coinvolgimento degli studenti, riduzione degli errori in fase di codifica dei dati, ecc), nei prossimi anni questa innovazione potrebbe

essere estesa anche alla scuola primaria. Pertanto, si ritiene necessario implementare uno studio qualitativo volto ad approfondire l'effetto della modalità di somministrazione su una prova INVALSI di Italiano da intendersi come operazione preliminare per immaginare e prevedere quali problemi si potrebbero incontrare implementando questa ulteriore innovazione.

Tale studio ha l'obiettivo di esplorare *il test mode effect* in una prova di comprensione del testo in base a diversi dispositivi di somministrazione in studenti di grado 6, al fine di considerare le possibili implicazioni dell'introduzione di una valutazione digitalizzata nella scuola primaria; vuole inoltre rappresentare un'azione valutativa ben distante dall'intento esclusivo di fornire un "giudizio" degli esiti, bensì quello di promuovere un'attività di riflessione sul senso stesso dell'azione innovativa (Lipari, 1995). In altre parole, come suggerito anche da Mauro Palumbo e Claudio Bezzi, (1998) "una buona valutazione è quella che viene pensata e progettata assieme al sorgere del processo decisionale, e che viene realizzata assieme al realizzarsi delle successive fasi di implementazione-gestione" (Bezzi, Palumbo, 1998, p.101).

Come è ampiamente noto, quando una innovazione riguarda il mondo della scuola, gli effetti si generano come una sorta di onde in espansione perché si riverberano al presente su studenti, docenti e personale della scuola in generale, oltre che sulle famiglie e, nel futuro, sulla diversa formazione che viene proposta alle nuove generazioni. Quindi, in linea con la funzione "learning" della valutazione (Palumbo, 2001), sviluppare uno studio volto ad esplorare il *test mode effect*, può concorrere ad offrire indicazioni utili ad INVALSI al fine di orientare il successivo processo decisionale, per implementare al meglio l'introduzione di tale innovazione nella scuola primaria.

Stato dell'arte

***Test mode effect* nella transazione da PBT a CBT: letteratura di riferimento**

Con il termine CBT (computer based testing) si fa riferimento all'insieme generalmente inteso dei test somministrati in ambiente informatizzato, indipendentemente dal tipo di piattaforma utilizzata e al tipo di codifica dei risultati richiesta (manuale, elettronica). Con l'acronimo CAT (computer adaptive testing) si indica una tipologia particolare di CBT nella quale la struttura del test viene modificata in fase di risoluzione dell'utente; sono due gli approcci principali che studiano i problemi legati al passaggio da test paper based a test computer based: migratory e transformative (Ripley, 2009; Horkay et al., 2006). Con l'approccio migratory si fa riferimento all'utilizzo del supporto tecnologico come semplice mezzo di somministrazione. In altri termini, si tratta di un trasferimento degli item di un test, nati in formato cartaceo, in formato digitale. L'approccio transformative, al contrario, prevede la trasformazione del test originale sulla base dell'integrazione di nuovi strumenti informatizzati che producano altri stimoli di tipo interattivo. La valutazione condotta mediante questa modalità si discosta in modo netto dalla valutazione tradizionale condotta in ambiente carta e penna, aprendo nuove prospettive per la valutazione. A partire dall'anno scolastico 2017/2018, l'INVALSI ha utilizzato delle prove migrate da carta e penna a supporto digitale, conservando sostanzialmente le stesse caratteristiche che si sono avute nella modalità carta e penna degli anni precedenti; nonostante questo, però, il passaggio al CBT non è un processo del tutto analogo al precedente. In altre parole, la transizione dalla valutazione basata su carta a quella basata su computer, solleva un numero di questioni epistemologiche e metodologiche importanti; in particolare, si tratta di comprendere come tale procedura possa influenzare le prestazioni e le strategie di risposta degli

alunni. Diverse sono le esperienze internazionali condotte in tal senso nell'ultimo ventennio. I risultati di questi studi però non forniscono risposte definitive, poiché in alcuni studi i risultati delle performance degli studenti sono migliori nella modalità carta e matita, in altri studi invece i risultati degli studenti sono più alti nella prova su computer, in altri studi ancora, non vi è nessuna differenza nelle prestazioni degli studenti tra CBT e PBT (Bennett et al., 2008; Clariana & Wallace, 2002; Jerrim, 2016; Nardi & Ranieri, 2018; Noyes & Garland, 2008; Russell, Goldberg, & O'Connor, 2003; Wang et al., 2007). Le ricerche condotte in tale ambito suggeriscono quindi che il verificarsi del "*test mode effect*" dipende dalle caratteristiche degli studenti che svolgono il test (ad es. il genere, il livello socioeconomico e la familiarità con il dispositivo) nonché dalle caratteristiche stesse del test come la disciplina oggetto del test, la facilità nell'uso del dispositivo e la modalità di presentazione degli item.

Genere e familiarità

Per quanto riguarda il genere, alcune ricerche hanno dimostrato che le studentesse ottengono risultati leggermente inferiori su CBT rispetto a PBT (Gallagher et al., 2002). Secondo Jeong (2014), le discrepanze tra i punteggi CBT e PBT dipendono dal genere e dalla disciplina/argomento in esame. Nel suo studio nell'ambito dell'istruzione primaria in Corea, l'autore infatti rileva punteggi CBT nelle ragazze significativamente inferiori rispetto ai punteggi PBT, specialmente nei test di matematica e scienze (Jeong, 2014). Le alunne e gli alunni mostrano punteggi PBT comparabili in scienza e matematica, ma per quanto riguarda i punteggi CBT quelli delle ragazze sono inferiori a quelli dei ragazzi. Un'esplorazione dei dati PISA 2012 ha mostrato che nella maggioranza dei paesi partecipanti (28 su 32) il punteggio medio dei ragazzi in matematica è superiore alla media delle ragazze in entrambe le modalità di valutazione. Inoltre, il divario di genere tende ad essere leggermente più ampio nella valutazione basata su computer (Jerrim, 2016). Generalmente le prestazioni tecnologiche delle ragazze sono inferiori a quelle dei ragazzi perché sono utenti meno interessati ai computer e mostrano maggiore ansia nell'uso dei computer rispetto ai maschi (Cooper, 2006; Meelissen & Drent, 2008). Dai recenti studi di Hamhuis, Glas e Meelisse (2020) su studenti di 10 anni nella scuola primaria olandese, emerge che nelle prove TIMSS 2015 le studentesse ottengono punteggi leggermente superiori rispetto agli studenti nella prova su tablet mentre non appaiono differenze significative nelle PBT. Questo conferma che il tradizionale divario di genere nelle competenze informatiche negli ultimi anni si sta via via assottigliando (Fraillon et al., 2014; Punter et al., 2017). Ragazzi e ragazze si comportano in modo simile sull'applicazione delle funzionalità tecniche; le ragazze sembrano avere performance migliori nella condivisione e nella comunicazione di informazioni e i ragazzi sembrano avere un punteggio più alto nell'autoefficacia per le abilità ICT (Fraillon et al., 2014; Punter et al., 2017).

Una variabile che è stata spesso citata in letteratura come uno dei principali fattori che contribuiscono alle differenze nelle prestazioni degli studenti è la familiarità con il dispositivo con cui si effettua la prova. Secondo Davis et al. (2016), se gli studenti non hanno familiarità con il dispositivo digitale, è meno probabile che avranno una performance di pari livello in una prova CBT rispetto ad un PBT equivalente (Davis et al., 2016). I lavori di Russell (1999) e Russell e Plati (2001) effettuati nell'ambito del sistema di valutazione NAEP (principale sistema di valutazione scolastico nazionale americano), mettono in luce come le performance degli studenti che prendono parte a una valutazione computer-based siano strettamente connesse alla loro familiarità con lo strumento utilizzato per la somministrazione. Ad esempio, nel caso di item a risposta multipla occorre che gli

studenti sappiano maneggiare il mouse, mentre nelle caselle di testo è fondamentale che essi abbiano una conoscenza e un corretto utilizzo della tastiera, oppure nei casi più complessi in cui si presenti un'interfaccia interattiva, è necessario avere familiarità con gli strumenti digitali forniti dall'ambiente. Come indicano alcuni studi svolti sulle capacità di utilizzo della tastiera del computer (Sandene et al., 2005), avere familiarità con questa può avere effetti positivi sui risultati di un test. Russell e Haney (1997) per esempio, registrano delle differenze significative in termini di tempo impiegato nelle domande a risposta aperta, mentre non c'è nessuna differenza significativa nelle domande a risposta multipla. Per gli autori infatti, gli strumenti computerizzati richiedono delle tempistiche maggiori per la risposta nel caso in cui gli studenti abbiano una scarsa familiarità nella scrittura attraverso la tastiera. In questa prospettiva, è possibile ipotizzare che nella migrazione da prove su carta a CBT, cambiando il formato della domanda ci si possono aspettare delle differenze almeno dal punto di vista delle performance degli studenti. Per quanto riguarda il nostro Paese, la maggior parte dei nativi digitali in Italia è in grado di utilizzare i social oppure di utilizzare la tecnologia per passatempo ma continuano a mostrare carenze nell'utilizzo nel computer per scopi didattici. Il rapporto studenti, computer e apprendimento redatto dal MIUR nel 2016, che riprende alcuni dei risultati più significativi che emergono dallo studio dell'OCSE sui dati PISA 2012, analizza in particolare i risultati degli studenti nelle prove CBA di matematica e di lettura in digitale. Lo studio entra nei dettagli soprattutto per quanto riguarda la digital reading, ovvero l'accertamento delle capacità degli studenti di raggiungere, decifrare e valutare i contenuti degli ipertesti proposti nelle prove. Da questo lavoro, emerge che i nostri studenti hanno buone capacità di navigazione generica sul web (addirittura sopra la media OCSE), ma quando si tratta di fare ricerche per la scuola, solo un quarto degli studenti è in grado di "navigare" in modo "orientato" e "critico". Inoltre, c'è un gap tra l'uso della tecnologia che i ragazzi fanno a casa e a scuola: gli studenti usano molto il pc nell'ambiente domestico ma ancora troppo poco a scuola.

L'effetto della familiarità nell'uso del tablet sulla comprensione della lettura è stato studiato con studenti universitari del secondo anno (Chen et al., 2014). Il gruppo di studenti con un alto livello di familiarità con i tablet ha ottenuto risultati significativamente migliori rispetto al gruppo con un basso livello di familiarità con i tablet, suggerendo che la familiarità con il dispositivo è una variabile importante nel test digitale. Davis e Strain-Seymour (2013) hanno scoperto che la familiarità con i dispositivi varia in base all'età e che le preferenze nell'uso del dispositivo possono variare in base al contenuto del test. Jeong (2014), nel suo studio su alunni Coreani di 10-12 anni, evidenzia la necessità di un'ulteriore distinzione tra la familiarità con il dispositivo e la capacità di adattamento alle prove computerizzate degli studenti. Sembrerebbe infatti che, oltre alla familiarità, un fattore che incide sulle differenze di performance degli studenti sia l'adattamento e la disposizione dello studente alla specifica modalità della prova (Jeong, 2014).

Funzionalità del dispositivo

Altri studi presenti in letteratura evidenziano che le differenze nelle prestazioni degli studenti sono dovute a determinate funzionalità del dispositivo di somministrazione della prova (Dadey, 2018). Per quanto riguarda il layout della prova cioè la disposizione degli elementi sulla pagina, le dimensioni e l'orientamento della pagina o dello schermo differiscono tra le modalità di somministrazione, il che ha un effetto sul numero di pagine di testo, nonché sulla dimensione e sul posizionamento del testo come, ad esempio, interruzioni di colonna e di riga. La presentazione di più elementi su una pagina, come è comunemente fatto su carta, rispetto alla presentazione di un elemento

alla volta come accade generalmente sullo schermo di un computer, può causare effetti sulle prestazioni degli studenti (Schroeders & Wilhelm, 2010).

Un'altra funzionalità legata al dispositivo digitale è lo scrolling e cioè lo scorrimento in senso orizzontale o verticale di un testo o di altre tipologie di dati sullo schermo. Se la quantità di informazioni su una pagina è maggiore rispetto alle dimensioni dello schermo, scorrere o sfogliare con il mouse/touchpad per leggere un testo e gli elementi associati è un'altra potenziale fonte di test mode effect. La funzione di scrolling si è mostrata più difficile rispetto alla funzione di paging (Higgins, Russell, & Hoffmann, 2005; Bridgeman et al., 2003; Schwarz, Rich, & Podrabsky, 2003). Tuttavia, gli studi condotti da Yamamoto (2012) nell'ambito del *Programme for the International Assessment of Adult Competencies* (PIAAC), suggeriscono che la funzione di scorrimento non ha avuto un impatto significativo sulla difficoltà nella risoluzione degli item nella versione computer based (Yamamoto, 2012).

La possibilità di rivedere gli item e correggere le risposte precedentemente date è una funzione tipica del test cartaceo e spesso vietata sul test computerizzato (Pommerich & Burden, 2000; Vispoel, 2000). Nei test computerizzati, tale mancanza di flessibilità può comportare un effetto sulla performance degli esaminati (Bodmann e Robinson, 2004). Tuttavia, Vispoel (2000) in uno studio con studenti universitari, ha evidenziato che impedire la revisione degli item e delle risposte non producono differenze statisticamente significative nei punteggi medi e nelle stime di capacità poiché solo una percentuale molto piccola di coloro che hanno effettuato il test ha sfruttato l'opportunità per modificare le proprie risposte. In tali casi, il vantaggio individuale derivante dall'uso della revisione degli oggetti è aumentato con il livello di abilità dei partecipanti al test. Vispoel ha anche mostrato che i test takers hanno espresso un forte desiderio di revisione degli item, in particolare quelli che mostravano ansia dovuta al test (Vispoel, 2000).

Alla luce di tali premesse, è evidente l'impossibilità a generalizzare sulla comparabilità di test PB e CB e la presenza di test mode effect; pertanto si rende evidente la necessità di implementare uno studio specifico volto ad esplorare il *test mode effect* in una specifica prova di comprensione della lettura INVALSI per studenti di scuola primaria;

Vantaggi di una valutazione Computer Based

Sebbene l'efficacia del computer-based testing (CBT) rispetto a Paper-Based Testing (PBT) sia ancora da verificare a fondo e i risultati degli studi inerenti al test mode effect siano misti (Way, Davis, Keng & Strain-Seymour, 2016), non sembrano esserci dubbi sui numerosi punti di forza che accompagnano tale modalità di somministrazione (Ripley, 2009; Bridgeman, 2009; Raikes & Harding, 2003; Wang et al., 2006; Lemmo, 2017). Un importante vantaggio delle prove CB risiede nella loro possibilità di fornire un feedback formativo continuo, creando un ambiente di apprendimento in cui gli studenti possono rivedere in modo efficace le risposte errate ed i contenuti connessi (Martin 2008). L'apprendimento efficace prevede che gli studenti esaminino attivamente il feedback dei risultati dei loro test in tempo reale. I test basati su computer consentono agli studenti di interiorizzare il feedback e di utilizzare queste informazioni per esprimere giudizi sul proprio operato (Gardner 2006). Inoltre, i test computerizzati sono particolarmente appropriati per la valutazione formativa e l'autovalutazione e consentono agli studenti di sviluppare competenze di auto-orientamento (Nicol & Macfarlane-Dick, 2006). La modalità computerizzata risulta avere riscontri positivi anche per gli insegnanti. Un esempio è il caso dell'Olanda (Nusche et. al, 2014), in cui i dati analizzati relativi alle rilevazioni nazionali vengono restituiti il giorno seguente alla somministrazione del test. In questo modo gli insegnanti hanno la possibilità di utilizzare

tempestivamente le informazioni provenienti dai dati al fine di promuovere una valutazione formativa implementando attività volte a colmare le carenze individuate o ad intervenire sulle difficoltà riscontrate dagli studenti.

Da un punto di vista logistico e funzionale, l'uso di test in modalità informatizzata abbassa i costi di somministrazione riducendo ad esempio i costi legati alle operazioni di stampa e distribuzione delle prove. Inoltre, la possibilità di misurare il tempo di somministrazione da un lato consente una maggiore flessibilità di utilizzo in quanto lo studente può iniziare il test quando si sente pronto ed il conteggio del tempo ha inizio con l'avvio della prova sullo schermo; dall'altro, consente informazioni accurate sui tempi di completamento della prova per ciascuno studente e informazioni specifiche sui tempi impiegati da ogni studente in ciascun item della prova (Liu et al. 2001).

Una valutazione CBT può contare su una riduzione dello stress e delle pressioni sulle organizzazioni scolastiche. Pensando all'esperienza condotta da INVALSI negli ultimi anni ad esempio, il passaggio a CBT ha permesso di allargare le finestre di somministrazione consentendo alle scuole una maggiore flessibilità nell'organizzare lo svolgimento delle prove.

Rispetto a un PBT, un CBT dovrebbe essere più accurato poiché è possibile utilizzare il punteggio automatizzato eliminando, ad esempio, gli errori di codifica umani (Segall et al. 2005).

Grazie all'utilizzo di test computer based adattivi inoltre, è possibile proporre una valutazione "personalizzata" per ogni studente e dei risultati individuali più accurati.

Dal punto di vista della somministrazione inoltre, una valutazione informatizzata garantisce un rapporto diretto con gli studenti permettendo una maggiore autenticità dei risultati garantita dalla riduzione del *cheating* (INVALSI, 2018;2019; Dermo 2009; Dawson 2016; Sindre and Vegendla, 2015). Con il termine *cheating*, che in italiano significa "imbrogliare", si fa riferimento all'insieme di anomalie che possono alterare l'esito della prova, per esempio ai tentativi degli studenti di copiare o degli insegnanti di facilitare gli studenti nello svolgimento della prova o nel modificarne gli esiti durante la correzione di prove cartacee. Infine, oltre a questi dati di per sé positivi, la somministrazione digitale può contare su un'alta motivazione degli studenti che hanno maggiore disponibilità verso l'utilizzo di questi strumenti (Banister, 2010). Esse offrono anche un potenziale significativo per l'uso di funzionalità dinamiche come video, animazioni e simulazioni. pertanto, le CBT rendono possibile l'uso dell'audio o di altri stimoli sensoriali (Thelwall, 2000, huang et al. 2009).

Valutazione su Tablet (TBT) per la comprensione del testo: alcuni studi comparativi

L'utilizzo di strumenti digitali innovativi rappresenta un elemento centrale nella vita della cosiddetta "Generazione App" (Davis & Gardner, 2014) sia all'interno che all'esterno dei contesti scolastici. Tra questi dispositivi digitali, i tablet sono diventati una tecnologia ampiamente utilizzata in ambito educativo in quanto questo tipo di tecnologia unisce le funzionalità di computer, laptop, smartphone e precedenti versioni del tablet con connessione internet sempre disponibile e migliaia di applicazioni con cui personalizzarne l'utilizzo (Johnson et al., 2013). Con schermi significativamente più grandi e interfacce grafiche più ricche rispetto alle altre tecnologie mobili, il tablet rappresenta uno strumento ideale per l'apprendimento. Sembra infatti che, viste le sue ampie funzionalità grazie all'installazione di App, al suo costo contenuto, alla connessione internet continua e alla sua mobilità, tale dispositivo possa considerarsi una via maestra per migliorare i risultati di apprendimento degli studenti (Banister, 2010). Diversi studi sono stati sviluppati per approfondire l'uso del tablet nelle pratiche valutative e la comparabilità di prove somministrate su tablet con altre modalità (Davis, Kong, McBride, & Morrison, 2016; Davis, Orr, Kong, & Lin, 2015; Keng, Davis, McBride, & Glaze,

2015; Ling, 2016; Olsen, 2014; Kong et al., 2018). Per quanto riguarda le caratteristiche legate alla funzionalità del dispositivo come per esempio l'uso del touchscreen, generalmente la precisione nella funzione di input può essere compromessa quando si utilizza il dito su un touchscreen rispetto a quando si utilizza un mouse (Way et al., 2016). Tale riduzione di precisione potrebbe aumentare i tempi impiegati dagli studenti per rispondere alle domande di una prova somministrata su tablet (circa 3-4 secondi in più) ma non comprometterebbe la corretta esecuzione del compito (Kong et al., 2018). Gli studi che si sono occupati di esaminare le prestazioni degli studenti con le tastiere su schermo e quelle esterne, hanno messo in luce che nelle domande che richiedono brevi risposte scritte entrambe le tipologie funzionano bene. Generalmente, quando vengono utilizzate le tastiere sullo schermo, le risposte degli studenti sono ridotte in termini di lunghezza, probabilmente a causa di un maggiore sforzo legato all'uso di tale strumento (Davis, Orr et al., 2015; Strain-Seymour et al., 2013; Pisacreta, 2013). Nello studio di Pisacreta (2013), la maggior parte dei partecipanti afferma di avere preferenza nell'uso di una tastiera esterna per rispondere alle domande su un testo scritto. Tuttavia, questa preferenza potrebbe non essere generalizzabile a tutti gli studenti. Infatti, Strain-Seymour et al. (2013) hanno scoperto che gli studenti più giovani, non abili nella dattilografia, non mostrano preferenze per l'uso della tastiera esterna. La chiave sembra essere dunque la padronanza degli studenti nell'uso della tastiera touchscreen: è importante quindi, che gli studenti sappiano utilizzare le funzionalità della tastiera su schermo come ad esempio la funzione di alternare lettere e numeri, evidenziare e spostare parti del testo ed inoltre, poiché la tastiera utilizza spazio sullo schermo che può impedire la corretta visualizzazione della prova, l'attivazione e la disabilitazione della tastiera (Davis, Strain-Seymour e Gay, 2013; Pisacreta, 2013; Strain-Seymour et al., 2013). Davis, Janiszewska et al. (2016) nell'ambito dei test NAPLAN, hanno notato che gli studenti che per lo svolgimento della prova hanno usato un tablet con l'aggiunta di una tastiera esterna hanno ottenuto risultati peggiori rispetto a quelli che hanno effettuato la prova con tablet con tastiera su schermo. Per quanto riguarda la dimensione dello schermo ed il contenuto visualizzato, ci sono due problemi separati ma correlati da considerare quando si valuta l'effetto sull'esecuzione della prova: la dimensione del display e la quantità di contenuto mostrato sul display. Dei due, il secondo sembra essere più rilevante ai fini dell'effetto sui risultati della performance degli studenti. Gli studi condotti in tale ambito, infatti, suggeriscono che quando le informazioni mostrate sullo schermo sono costanti, schermi di 10 pollici o più grandi, sono adatti per la visualizzazione e per la possibilità di operare sulla prova di valutazione, con poca evidenza di differenze nelle prestazioni complessive del test o differenze a livello dei singoli item (Davis et al., 2013; Keng, Kong e Bleil, 2011). Schermi di dimensioni inferiori a 10 pollici possono comportare differenze nelle prestazioni degli studenti, con punteggi decisamente più bassi nei casi in cui si utilizza un dispositivo con uno schermo più piccolo (Davis et al., 2013; Schroeders & Wilhelm, 2010).

Ancora pochi studi si sono occupati di indagare l'effetto della familiarità con i tablet sulla performance degli studenti. Tra questi, uno studio di Cheng et al. (2014) nell'ambito universitario, ha esplorato gli effetti della comprensione della lettura in prove su carta, tablet e computer. Gli autori, inoltre, si sono posti il problema di capire l'impatto della familiarità con i media sulla comprensione della lettura. I risultati hanno indicato che il gruppo di studenti che ha svolto la prova cartacea ha ottenuto risultati significativamente migliori nella lettura rispetto a chi ha svolto la prova su computer manifestando un livello di comprensione più basso. Il gruppo di studenti con alto livello di familiarità con i tablet ha ottenuto risultati significativamente migliori nella comprensione del testo rispetto al gruppo con basso livello di familiarità. Tale studio conferma che il tablet, può essere un buon dispositivo di apprendimento dal punto di vista educativo utilizzabile anche per la valutazione ma è

fondamentale che gli insegnanti abbiano un'adeguata formazione ed inoltre è importante che gli insegnanti e gli studenti abbiano abbastanza tempo per adattarsi a questi dispositivi digitali di apprendimento e valutazione innovativi (Cheng et al., 2014). Per quanto riguarda la motivazione degli studenti, diversi studi confermano che gli studenti più giovani sembrano preferire il tablet come dispositivo per svolgere la prova di valutazione in quanto rappresenta il dispositivo con cui hanno maggiore esperienza d'uso (Davis et al., 2017; Davis et al., 2013, 2015; Strain-Seymour et al., 2013). Nello studio sopracitato di Hamhuis e collaboratori (Hamhuis et al., 2020) condotto nel 2017 su studenti di quarta elementare nei Paesi Bassi, gli autori hanno esplorato, tra gli altri aspetti, l'influenza del genere sulle differenze di performance degli studenti tra il PBT e TBT (Tablet Based Assessment) e l'influenza del compito di lettura richiesta agli studenti suddivisi in alto, medio e basso (Punter et al., 2018). sulle differenze di performance degli studenti tra il PBT e TBT. Tra i principali risultati dello studio, è emerso che le studentesse hanno leggermente superato i ragazzi nel test su tablet e che invece non vi sono differenze significative tra maschi e femmine nelle prove cartacee. Per quanto concerne invece il verificarsi del mode effect su item matematici con compito di lettura con diverso livello di complessità, a differenza di quanto previsto in letteratura - secondo cui la complessità e la lunghezza della lettura di testi sullo schermo di un computer o di un tablet sono due dei principali svantaggi legati all'uso di CBT che possono avere influenze negative sulle performance degli studenti (Choi & Tinkler, 2002; Kingston, 2009; Ling, 2016; Mangen et al., 2013; Nardi & Ranieri, 2018) - tale studio mette in evidenza che, indipendentemente dalla modalità di test utilizzata, gli studenti hanno mostrato lo stesso comportamento nei diversi item. Probabilmente tale risultato è dovuto al fatto che nella versione somministrata su tablet del test TIMSS, la necessità di scorrimento all'interno del test degli item con un'elevata richiesta di lettura è stato ridotto al minimo.

La letteratura sintetizzata in questa rassegna non è da intendersi come fondamento per considerare il tablet uno strumento utile solo come dispositivo alternativo al computer finalizzato ad espandere il set di dispositivi a disposizione il giorno della prova; Obiettivo della rassegna è invece quello di fornire un quadro di insieme in cui inserire le opportune considerazioni per l'implementazione di una valutazione tablet based.

Le ragioni di questo studio

Alla luce di quanto emerso detto finora dall'analisi della letteratura, si ritiene necessario implementare uno studio volto ad approfondire tali tematiche nel nostro contesto scolastico per diverse ragioni. Di seguito le principali motivazioni per cui si ritiene opportuno condurre tale ricerca:

- dall'analisi della letteratura è evidente che molti autori hanno approfondito le differenze nella modalità test tra CBT e PBT ma i risultati dei loro studi sono spesso divergenti ed incompleti;
- molte ricerche si sono soffermate nell'indagare le modalità CBT e poche ancora si sono concentrate in modo specifico sull'utilizzo dei tablet per la valutazione degli studenti;
- dalla letteratura emerge che pochi studi si sono concentrati sulla scuola primaria rispetto agli altri livelli scolastici (Tonelli et al., 2018);

- l'utilizzo del tablet è ampiamente diffuso tra i "nativi digitali" (Prensky, 2001) in tutti i contesti e non solo a scuola, al punto che negli ultimi anni tale etichetta si è modificata in "Generazione app" (Davis & Gardner, 2014);
- approfondire il *mode effect* nella comprensione del testo è molto importante anche per le altre discipline scolastiche. Infatti, la comprensione del testo si può considerare un'abilità trasversale indispensabile oltre che per l'italiano anche per altri ambiti disciplinari (ad es. matematica, logica ecc) (Franchini et al., 2017; D'Amore, 1996; Zan, 2007, 2016);
- la "generazione app" molto probabilmente avrà una maggiore predisposizione all'uso del tablet per svolgere una prova di comprensione del testo in quanto strumento ampiamente utilizzato per leggere e comprendere testi anche nella vita quotidiana (dentro e fuori la scuola). A tal proposito, conoscere la loro performance sulla comprensione del testo in un test tablet based è importante per approfondire la performance degli studenti nell'ambito di una "valutazione situata" caratterizzata da maggiore autenticità rispetto ad una prova cartacea;
- in considerazione dei risultati provenienti dagli studi condotti in letteratura, è evidente l'impossibilità a generalizzare e la necessità di avviare uno studio specifico per esplorare il *test mode effect* in una prova di comprensione della lettura per studenti italiani di grado 6;
- In linea con la finalità *learning* della valutazione (Palumbo, 2001) i risultati della presente ricerca potrebbero offrire spunti di riflessione interessanti al fine di orientare l'introduzione di prove INVALSI digitalizzate nella scuola primaria;

Fasi ed Implementazione sul campo

Considerata l'attuale situazione pandemica che ha improvvisamente rivoluzionato il sistema scolastico e non solo, considerato il periodo ancora caratterizzato da incertezza ed emergenza sanitaria si stanno ripensando le fasi di ricerca e le modalità di discesa sul campo al fine di adeguarsi all'attuale situazione.

BIBLIOGRAFIA

- Banister, S. (2010). Integrating the iPod Touch in K–12 education: Visions and vices. *Computers in the Schools*, 27(2), 121-131.
- Bezzi, C., & Palumbo, M. (Eds.). (1998). *Strategie di valutazione: materiali di lavoro*. Gramma Linee di Ricerca sociale, Perugia.
- Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, 33(5), 593–602.
- Cooper, J. (2006). The digital divide: The special case of gender. *Journal of Computer Assisted Learning*, 22, 320–334. <https://doi.org/10.1111/j.1365-2729.2006.00185.x>.
- D'Amore, B. (1996a). Difficoltà nella lettura e nella interpretazione del testo di un problema. *Bollettino degli insegnanti di matematica del Canton Ticino*, 32, 57-64.
- Dadey, N., Lyons, S., & DePascale, C. (2018). The comparability of scores from different digital devices: A literature review and synthesis with recommendations for practice. *Applied Measurement in Education*, 31(1), 30-50.
- Davis, L. L., & Strain-Seymour, E. (2013a, June). Digital devices research. Paper presented at the National Conference on Student Assessment, National Harbor, Maryland.
- Davis, L. L., Janiszewska, I., Schwartz, R., & Holland, L. (2016). NAPLAN device effects study. Melbourne, VIC: Pearson.
- Davis, L. L., Orr, A., Kong, X., & Lin, C. H. (2015). Assessing student writing on tablets. *Educational Assessment*, 20(3), 180-198.
- Decreto Legislativo 13 aprile 2017, n. 62 Norme in materia di valutazione e certificazione delle competenze nel primo ciclo ed esami di Stato, a norma dell'articolo 1, commi 180 e 181, lettera i), della legge 13 luglio 2015, n. 107. (17G00070).
- Dermo, J. (2009). e-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203-214.
- Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). Preparing for life in a digital age. The IEA International and Information Literacy Study international Report. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1009&context=ict_literacy.
- Franchini, E., Lemmo, A., & Sbaragli, S. (2017). Il ruolo della comprensione del testo nel processo di matematizzazione e modellizzazione The role of text comprehension in the process of mathematization and modeling.
- Gardner, J., ed. 2006. *Assessment and learning*. London: Sage Publications.

- Hamhuis, E., Glas, C., & Meelissen, M. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS'paper-and-pencil test and tablet test among Dutch grade-four students?. *British journal of educational technology*.
- Jeong, H. (2014). A comparative study of scores on computer-based tests and paper-based tests. *Behaviour and Information Technology*, 33(4),410–422. <https://doi.org/10.1080/0144929X.2012.710647>.
- Jerrim, J. (2016). PISA 2012: How do results for the paper and computer tests compare? *Assessment in Education: Principles, Policy and Practice*, 23(4), 495–518. <https://doi.org/10.1080/0969594X.2016.1147420>.
- Kong, X., Davis, L. L., McBride, Y., & Morrison, K. (2018). Response time differences between computers and tablets. *Applied Measurement in Education*, 31(1), 17-29.
- Leeson, H. V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing*, 6(1), 1-24.
- Lipari D. (1995), *Progettazione e valutazione nei processi formativi*, Edizioni Lavoro, Roma.
- MIUR (2016). *Studenti, computer e apprendimento: dati e riflessioni. Uno sguardo agli esiti delle prove in Lettura in Digitale dell'indagine OCSE PISA 2012 e alla situazione in Italia*.
- Noyes, J. M., & Garland, K. J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, 51(9), 1352–1375.
- Parshall C G and Kromrey J D (1993) *Computer testing versus paper-and-pencil: an analysis of examinee characteristics associated with mode effect* A paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA, April (Educational Resources Document Reproduction Service (ERIC) # ED363272).
- Palumbo, M. (2001). *Il processo di valutazione*. Milano: Franco Angeli.
- Pisacreta, D. (2013, June). Comparison of a test delivered using an iPad versus a laptop computer: Usability study results. In Council of Chief State School Officers, National Conference on Student Assessment, National Harbor, MD.
- Pommerich, M., & Burden, T. (2000, April). From Simulation to Application: Examinees React to Computerized Testing. Paper presented at the annual meeting of the National Council of Measurement in Education, New Orleans.
- Pommerich, M. (2004). Developing Computerized Versions of Paper-and-Pencil Tests: Mode Effects for Passage-Based Tests. *The Journal of Technology, Learning, and Assessment*, 2 (6), 3-44.
- Punter, R. A., Meelissen, M. R., & Glas, C. A. (2017). Gender differences in computer and information literacy: An exploration of the performances of girls and boys in ICILS 2013. *European educational research journal*, 16(6), 762-780.

Ripley, M. (2009). Transformational computer-based testing. In F. Scheuermann, & J. Sandene, B., Horkay, N., Bennett, R. E., Allen, N., Braswell, J., Kaplan, B., et al. (2005).

Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil.

Russell, M., & Plati, T. (2001). Effects of computer versus paper administration of a statemandated writing assessment. *The Teachers College Record*.

Strain-Seymour, E., Craft, J., Davis, L. L., & Elbom, J. (2013). Testing on tablets: Part I of a series of usability studies on the use of tablets for K-12 assessment programs. White paper in Pearson Always Learning series, July, 1-35.

Thelwall, M. (2000). Computer-based assessment: a versatile educational tool. *Computers & Education*, 34(1), 37-49.

Tonelli, D., Grion, V., & Serbati, A. (2018). L'efficace interazione fra valutazione e tecnologie: evidenze da una rassegna sistematica della letteratura. *Italian Journal of Educational Technology*, 26(3), 6-23.

Wang, S., Jiao, H., Young, M. J., Brooks, T. E., & Olson, J. (2007). A meta-analysis of testing mode effects in Grade K-12 mathematics tests. *Educational and Psychological Measurement*, 67, 219–238.

Way, W. D., Davis, L. L., Keng, L., & Strain-Seymour, E. (2016). From standardization to personalization: The comparability of scores based on different testing conditions, modes, and devices. *Technology and testing: Improving educational and psychological measurement*, 2, 260-284.

Yamamoto, K. (2012). Outgrowing the Mode Effect Study of Paper and Computer Based Testing. Retrieved from http://www.umdcipe.org/conferences/EducationEvaluationItaly/COMPLETE_PAPERS/Yamamoto/YAMAMOTO.pdf

Zan, R. (2007). La comprensione del problema scolastico da parte degli allievi: alcune riflessioni. *L'insegnamento della matematica e delle scienze integrate*, 30 A-B (6), 741-762.

Zan, R. (2016). *I problemi di matematica: difficoltà di comprensione e formulazione del testo*. Roma: Carocci.

Genova,

FIRMA

A handwritten signature in black ink, appearing to read "Dowda torti", is written over a horizontal line.

