

SUPSI

Quaderni di ricerca

Prove standardizzate ticinesi

Italiano nella classe III Scuola Elementare

Alberto Crescentini



Proposta di citazione:

Crescentini, A. (2017). *Prove standardizzate ticinesi. Italiano nella classe III Scuola Elementare*.
Locarno: Centro innovazione e ricerca sui sistemi educativi.

Locarno, 2017
CIRSE - Centro innovazione e ricerca sui sistemi educativi
Piazza San Francesco 19, 6600 Locarno
dfa.cirse@supsi.ch

ISBN 978-88-941240-8-8

Responsabilità del progetto: Alberto Crescentini
Ricercatori coinvolti: Loredana Addimando, Manuela Lagattolla, Miriam Salvisberg, Sandra Zampieri,
Giovanna Zanolla
Impaginazione: Selene Dioli

Ringraziamenti

Questo lavoro si è svolto nel passaggio di responsabilità dell'Ufficio delle Scuole Comunali e ha coinvolto numerose persone. Senza la partecipazione di tutti non sarebbe stato possibile portare a compimento la costruzione delle prove, la ricerca ad esse connessa, l'analisi delle informazioni e la redazione del rapporto stesso.

Innanzitutto è doveroso ringraziare tutti i docenti di Scuola Elementare che hanno permesso che la prova fosse prima testata nella fase di sviluppo e quindi svolta una volta definitivamente approntata. La lista sarebbe eccessivamente lunga ma li ricordiamo tutti assieme al direttore dell'USCO durante il periodo di svolgimento delle prove Mirko Guzzi che ha contribuito alla impostazione e gestione del progetto..

Molte persone sono state coinvolte per la costruzione delle prove. In particolare alla definizione dei settori hanno contribuito: Dario Corno e Alessandra Moretti. Allo sviluppo degli item: Antonella Barrera, Roberto Bassi, Dario Corno, Alessandra Moretti, Lucija Pacak e Pamela Tami-Fontana. Per le analisi finalizzate alla selezione degli item migliori e alla determinazione dei punteggi nei diversi settori dell'italiano: Miriam Salvisberg e Giovanna Zanolla con la collaborazione di Stéphanie Berger e Urs Moser. A Loredana Adimando per la partecipazione alla ricerca e i gli indispensabili contributi.

La gestione e il coordinamento con gli istituti scolastici sarebbero state inoltre impossibili senza la fattiva collaborazione di tutti gli ispettori e i direttori delle scuole e senza il coordinamento di Manuela Lagattolla e Sandra Zampieri.

La rilettura e il confronto su temi e contenuti si è giovata del contributo di Michela Crespi, Claudio Biffi e Rezio Sisini.

Sommario

1	Il progetto	7
2	Analisi dei risultati delle prove	11
2.1	Informazioni generali	11
2.2	Il contesto e la scuola	13
2.2.1	I circondari	13
2.2.2	Il grado di urbanizzazione del comune	16
2.3	La classe	17
2.3.1	dimensione della classe o numero di allievi	17
2.3.2	La pluriclasse	18
2.4	L'allievo	21
2.4.1	Caratteristiche socio demografiche	21
2.4.2	La Nazionalità e la lingua madre	22
2.4.3	Prestazione alle prove e gender	23
2.4.4	Numero di assenze	24
2.4.5	Origine sociale	24
2.4.6	Età dell'allievo	26
2.5	Docente e allievo	28
2.5.1	Il genere del docente	28
2.5.2	Nota di condotta	30
2.6	Risultati al test e note	31
2.6.1	Relazione nota di italiano di fine anno – prestazione alla prova	31
2.6.2	Relazione con le note di fine anno dell'anno successivo	34
3	Sintesi	37
	Bibliografia	39
	Appendici	43
	Le analisi tramite i modelli di Rasch	43
	Competenze fondamentali e prove	50

1 Il progetto

Il progetto di valutazione delle competenze in Italiano si inserisce all'interno di un progetto pluriennale che dall'anno 2012 ad oggi ha testato le competenze in matematica (Cirse, 2014) e italiano di due coorti¹ di scuola elementare. Le classi che sono state oggetto della rilevazione sono state la V elementare, per quanto riguarda la Matematica, e la III elementare per quanto riguarda l'Italiano. Per quanto riguarda il progetto sull'italiano nel suo insieme si è svolto negli anni scolastici dal 2012 al 2015.

L'insieme delle valutazioni tramite prove standardizzate ha il duplice obiettivo di fornire delle informazioni di monitoraggio del sistema educativo (oggetto specifico di questo rapporto) e di fornire ai docenti, direttori e ispettori delle informazioni di dettaglio relative all'andamento delle specifiche classi.

Lo svolgimento del progetto di Italiano e la struttura del rapporto riprendono quelli del progetto sulla matematica. La presentazione e le dimensioni teoriche sono molto simili e quindi si rimanda per gli approfondimenti del caso al documento già esistente; alcune parti che si riteneva potessero giovare alla lettura ancorché già esistenti sono state messe in appendice.

I primi mesi del progetto sono stati impiegati per il consolidamento della rete di collaborazione. Questa è stata composta dai referenti statistici esperti di Item Response Theory (IRT), identificati nell'Istituto per la valutazione della formazione (IBE) di Zurigo che hanno fornito una collaborazione per lo svolgimento del processo di valutazione degli item, e dal gruppo di esperti del territorio (in questo caso il termine non fa riferimento esclusivo agli esperti disciplinari della Scuola Media bensì a persone che fossero portatori di un sapere e di una conoscenza utile alla riflessione su questo oggetto), individuati grazie alla collaborazione con l'Ufficio delle Scuole Comunali (USC) e con il Centro di Didattica dell'italiano e delle lingue nella scuola (DILS). Quando si parla di IRT si fa riferimento a un insieme di procedure statistiche, con degli assunti teorici ben definiti, mirate a individuare il valore di difficoltà di ogni item utilizzato per valutare una specifica caratteristica di una popolazione di riferimento.

Grafico 1: Calendario di svolgimento del progetto

FASE	Attività	gen-ago12	set. 12 - mag. 13	giu. 13	lug. 13	ago. 13	set. 13	ott. 13	nov. 13	dic. 13	gen. 14	feb. 14	mar. 14	apr. 14	mag. 14	giu. 14	lug. 14	ago. 14	set. 14	ott. 14	nov. 14 - ago. 15
1	Fase preliminare																				
2	Sviluppo item inclusa codifica																				
3	Definizione design pretest																				
4	Procedure somministrazione																				
5	Creazione fascicoli e guide di codifica																				
6	Creazione campione pretest e comunicazione scuole																				
7	Organizzazione e somministrazione pretest																				
8	Organizzazione codifica e codifica																				
9	Analisi psicometrica; costruzione scala e livelli competenza																				
10	Creazione fascicoli, organizzazione somministrazione e comunicazione scuole																				
11	Somministrazione																				
12	Codifica																				
13	Produzione rapporti																				
14	Analisi complessive																				

¹ "Gruppo di individui portatori di una caratteristica comune (per es., c. di nascita, di nati in uno stesso anno; c. rurale, di abitanti di un comune rurale; c. lavorativa, di soggetti che lavorano nello stesso ente) o anche di soggetti arruolati in uno studio longitudinale o in una sperimentazione clinica di variabile durata." (Treccani, 2016).

La prima fase del progetto, sino ad agosto 2012, è stata dedicata alla definizione dei contenuti da testare e all'identificazione delle persone che avrebbero potuto sviluppare gli item.

La scelta dei settori da investigare è stata fondamentale in quanto sarebbe estremamente difficoltoso valutare contemporaneamente tutte le competenze presenti in una disciplina in modo esatto. Il modello di analisi che fa riferimento alla IRT è, infatti, costruito al fine di avere delle misure precise di costrutti ben definiti e quanto più possibili unitari, in appendice è presente una spiegazione maggiormente dettagliata (vedi Appendice 2 "Competenze fondamentali e prove").

Nell'elaborazione della prova sono stati presi in considerazione, attingendoli dal modello HarmoS, solo quattro settori di conoscenza: lessico, punteggiatura, ortografia morfologica e ortografia fonologica. L'insieme delle risposte alle domande ha costituito una misura generale chiamata Italiano Generale. La scelta dell'ortografia è stata dettata dalla considerazione secondo la quale HarmoS prescrive che essa sia interamente acquisita dagli allievi entro la fine della terza elementare (in particolare l'ortografia fonologica: corrispondenza fonema-grafema e la punteggiatura: uso dei principali segni di punteggiatura negli impieghi standard). L'acquisizione del lessico, invece, copre tutto il percorso della scuola dell'obbligo.

Nel caso della lingua di insegnamento si deve inoltre considerare la complessità aggiuntiva di valutare un contenuto utilizzando come strumento di valutazione il contenuto stesso che si intende valutare. Gli item, infatti, sono comunque sviluppati nella lingua di insegnamento. La definizione delle competenze da valutare ha riguardato una dimensione di contenuto ed una di prospettiva, quest'ultima proiettata sul profilo di competenze che dovrà essere presente dopo l'entrata a regime degli accordi di armonizzazione.

Successivamente, sono stati sviluppati gli item per testare ognuna delle parti stabilite. Nel costruire gli item si è dovuto tener conto di alcune indicazioni. In primo luogo ogni item doveva essere quanto più possibile mono dimensionale. La mancanza di una esperienza precedente sulla produzione di item per l'italiano ha reso la produzione degli item particolarmente complessa e ricca di regolazioni in corso d'opera collegate agli stimoli che progressivamente si rendevano disponibili da parte del gruppo di creazione. Per poter misurare la capacità di discriminazione dell'item (si parla di capacità discriminativa relativamente al fatto che l'item riceva risposte corrette dagli allievi più abili e scorrette da quelli meno abili) e anche la sua coerenza con la dimensione che si desiderava valutare è infatti necessario che ogni item sia attinente a una e una sola dimensione. Questa caratteristica rende questi item in sé differenti da quelli che normalmente sono utilizzati dai docenti durante la loro attività professionale. In secondo luogo si è dovuto creare un numero di item sovrabbondante rispetto all'uso finale. Si è dovuto infatti prevedere che successivamente alla prova campione sarebbe stato eliminato almeno il 30% degli item prodotti.

Una volta prodotti gli item e sottoposti a verifica di contenuto con l'assistenza degli esperti di italiano, si è proceduto alla preparazione della prova campione. Questa prova aveva lo scopo di valutare la pertinenza degli item e di individuare quelli più efficaci a misurare e a discriminare. Ordinandoli per difficoltà crescente si dovrebbe trovare un numero inizialmente molto elevato di allievi che risponderà correttamente e questo numero dovrebbe ridursi al crescere della difficoltà. Se un item ad esempio, riceverà, un numero di risposte corrette elevato ma solo dagli allievi meno abili questo sarà scartato, parimenti saranno eliminati gli item non discriminanti, quelli cioè ai quali tutti o nessuno avranno risposto. Queste procedure hanno infatti lo scopo di costruire delle scale valide non in termini astratti ma all'interno delle popolazioni reali. Gli item costruiti sono infatti coerenti con i contenuti presentati nella scuola ticinese e la loro difficoltà è valutata rispetto agli allievi della stessa scuola.

Concretamente, sono stati realizzati 4 fascicoli differenti, ciascuno di essi richiedeva un tempo di risposta di 45 minuti. Questa distribuzione apparentemente complessa era necessaria per garantire che ogni item fosse testato su almeno 300 allievi e ogni allievo venisse confrontato con 2 fascicoli, l'uno a distanza di una settimana dall'altro. La distanza di una settimana è stata ritenuta quella minima per poter ritenere l'effetto di apprendimento residuale. Ogni allievo ha quindi ricevuto due fascicoli diversi assegnati casualmente e le classi sono state estratte in modo da essere rappresentative della popolazione degli studenti ticinesi. Si è scelto di utilizzare un campionamento basato sulle classi scolastiche, in quanto meno invasivo e per altro accettato all'interno delle principali ricerche nazionali e internazionali. Sono stati testati circa 1600 allievi pari al 50% della popolazione complessiva.

La somministrazione è stata curata da personale esterno alla scuola che si è occupato di portare le prove nelle singole classi, far eseguire il lavoro agli allievi e recuperare poi i materiali distribuiti. Quest'ultima fase è estremamente rilevante in quanto gli esercizi proposti in questa fase dei lavori non hanno ancora subito alcun processo di validazione e non possono essere ritenuti efficaci alla valutazione delle competenze specifiche, non vi sono infatti valori che ne indichino l'efficacia o la difficoltà in alcun modo. Si deve anche sottolineare come un esercizio diffuso in maniera non corretta (ad esempio tramite fotocopie del materiale) potrebbe rendere l'attività di valutazione non valida in quanto introdurrebbe una condizione di

non equità di fronte alla prova. Un problema più ampio legato alla disponibilità degli esercizi è quello definito in letteratura *“teaching for testing”* del quale si discuterà più oltre.

Tutti i questionari sono stati quindi raccolti e le risposte sono state inserite in un archivio al fine di poter valutare la bontà metrica degli esercizi. I singoli esercizi, le scale e l'insieme degli esercizi sono stati quindi valutati utilizzando il modello di Rasch, al fine di capire come costruire le successive prove e quali esercizi conservare. Queste analisi hanno permesso di identificare 107 esercizi con buone capacità metriche divisi nei 4 settori. Di seguito vedremo un sunto delle analisi effettuate.

Dagli esercizi è stato possibile realizzare un fascicolo, la prova complessivamente comprendeva 107 item dei quali: 28 attinenti il lessico, 19 attinenti la punteggiatura, 34 attinenti l'ortografia morfologica e 26 l'ortografia fonologica. La finestra di tempo nella quale questa prova doveva avvenire è stata di due settimane.

Per evitare inoltre gli effetti visti nelle medesime ricerche si è stabilito di far svolgere le attività di somministrazione esclusivamente a personale formato specificatamente, ciò anche per non portare un aggravio di lavoro al personale insegnante.

Una volta raccolte le prove gli item sono stati nuovamente sottoposti ad una analisi relativa alla identificazione della difficoltà rispetto alla popolazione degli allievi, sulla base di questi valori sono state successivamente svolte le analisi. Nel periodo estivo sono stati elaborati tre tipi di rapporti. Il primo tipo è stato consegnato ad ogni docente che avesse avuto una classe testata, il secondo ad ogni direttore nel cui istituto fosse stata testata una classe e il terzo ad ogni ispettore. Ogni docente ha ricevuto un rapporto relativo alla sua classe nel quale si mostrava per la classe e per ogni allievo relativamente ad ogni singolo settore il punteggio medio² rispetto al circondario e all'insieme della popolazione testata. Analogamente i direttori e gli ispettori sono stati inviati rapporti, relativamente, all'istituto e al circondario di competenza. Il direttore dell'Ufficio Scuole Comunali ha ricevuto l'insieme dei rapporti assieme ad una sintesi generale.

Nel capitolo relativo alle analisi verranno letti i dati raccolti secondo tre punti di osservazione differenti: il territorio, le classi e gli allievi.

² Il punteggio medio era dato dal numero di item eseguiti correttamente ciascuno pesato in rapporto al coefficiente di difficoltà ottenuto con le analisi svolte nella prova campione tramite la IRT.

2 Analisi dei risultati delle prove

Per organizzare i contenuti del rapporto si procederà con una logica a imbuto. Inizialmente saranno presentate le analisi più generali poi si scenderà progressivamente nel dettaglio. Da un punto di vista di informazioni rispetto al sistema dopo quelle generali saranno considerati i circondari, quindi alcune dimensioni territoriali e poi le classi. Nel capitolo successivo saranno approfondite le dimensioni relative alle caratteristiche degli allievi quindi alla relazione tra la prestazione alla prova e le note di fine anno.

Il livello di circondario è stato scelto anche per la funzione di vigilanza e di organizzazione che viene svolta dagli ispettorati e dal collegio degli ispettori. Il Cantone infatti ha compiti di istituzione e direzione della scuola svolti in collaborazione con i comuni. Il collegio degli ispettori e l'ufficio scuole comunali (USC) sono, per le scuole comunali, gli organi di raccordo tra le scuole del territorio e il DECS (Legge della scuola, 1990).

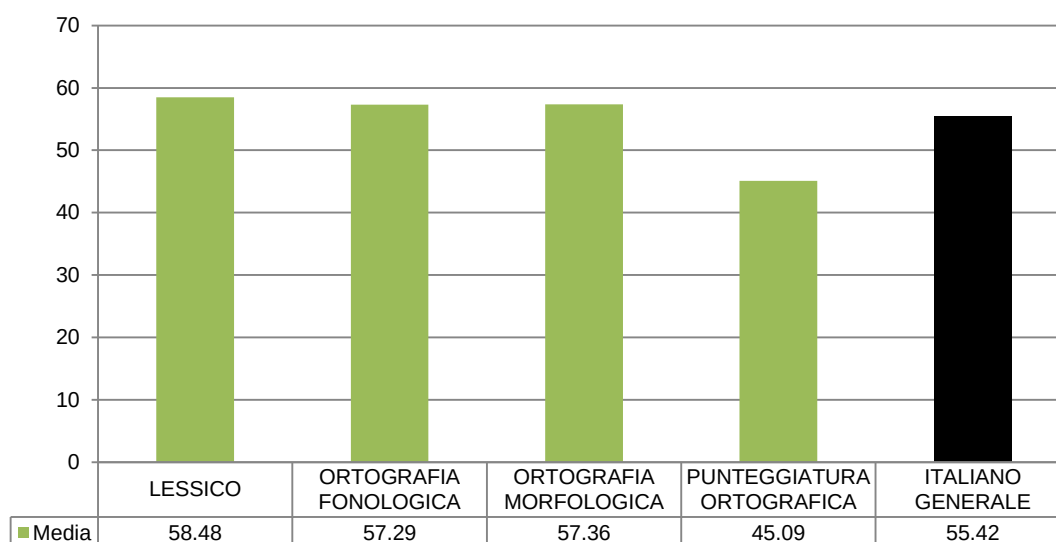
2.1 Informazioni generali

Nel Grafico 2 è possibile vedere le medie dei risultati degli allievi nei quattro settori dell'italiano contemplati dalle prove.

Il valore "Italiano generale" è quella variabile costituita dalla media dei punteggi di tutti i settori. Questa media è stata calcolata per ogni allievo e può essere considerata come un indicatore sintetico della sua prestazione (Cirse, 2014) alla prova.

La dimensione lessico rivela il punteggio medio più elevato (58.31) mentre punteggiatura sembra essere la dimensione affrontata con maggiori difficoltà (45.01) dagli allievi.

Grafico 2: Punteggi medi degli allievi nelle diverse dimensioni



Tutti i punteggi sono stati normalizzati³ in modo da assumere valori compresi tra 0 e 100. I punteggi non equivalgono però a percentuali corrispondenti al numero di esercizi svolti correttamente: ottenere 50 in un

³ L'operazione di normalizzazione consiste nel far sì che i punteggi si distribuiscano su scale che permettano il confronto. Nel nostro caso abbiamo fatto sì che tutte le scale partissero da un punteggio pari a 0 a avessero valore massimo pari a 100.

certo settore non significa infatti aver svolto correttamente il 50% degli item di quel settore. In ciascun settore gli item sono stati infatti ponderati per il rispettivo coefficiente di difficoltà; questo è stato calcolato ponderando la quantità di risposte corrette allo specifico item. In pratica, chi ha svolto correttamente gli esercizi con elevato coefficiente di difficoltà ottiene un punteggio superiore a chi ha svolto un uguale numero di esercizi con coefficiente di difficoltà inferiore. All'interno dei fascicoli gli item utilizzati sono stati proposti in ordine crescente di difficoltà ed è quindi probabile che chi ha risposto correttamente agli item più difficili abbia risposto correttamente anche a quelli precedenti.

Tabella 1: Correlazioni tra le dimensioni valutate con la prova

	LESSICO	ORTOGRAFIA FONOLOGICA	ORTOGRAFIA MORFOLOGICA	PUNTEGGIATURA ORTOGRAFICA	ITALIANO GENERALE
LESSICO	1				
ORTOGRAFIA FONOLOGICA	.557**	1			
ORTOGRAFIA MORFOLOGICA	.454**	.689**	1		
PUNTEGGIATURA ORTOGRAFICA	.379**	.483**	.552**	1	
ITALIANO GENERALE	.736**	.850**	.859**	.729**	1

La tabella precedente mostra come i risultati nelle diverse dimensioni siano tra loro correlati. Questo risultato non è inatteso in quanto si tratta di dimensioni parte del costrutto più generale valutato che corrisponde all'italiano in generale. Si è quindi costruita una dimensione che aggregasse le quattro dimensioni e che rappresentasse il punteggio globale del singolo allievo alla prova.

In statistica per correlazione si intende una relazione tra due variabili tale che a ciascun valore della prima variabile corrisponda con una "certa regolarità" un valore della seconda.

Attenzione deve essere prestata al non considerarla una relazione di causa-effetto. Si tratta infatti della tendenza di una variabile a variare in funzione di un'altra. Il valore della correlazione può essere compreso tra -1 e +1. Se il valore si avvicina a -1 al crescere del valore di una variabile vi sarà il decrescere del valore dell'altra e viceversa, se il valore si avvicina a +1 le due variabili cresceranno o decresceranno coerentemente l'una all'altra.

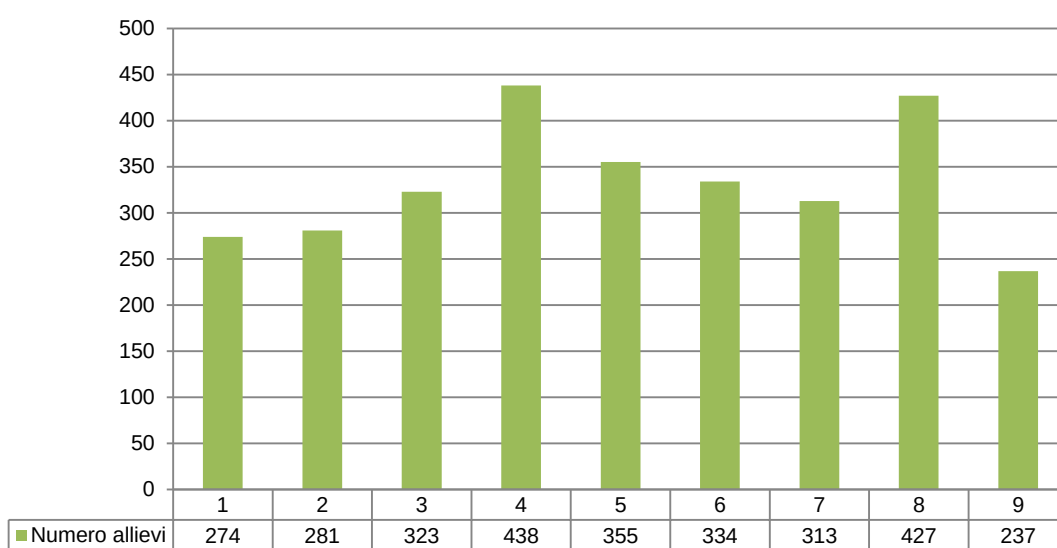
Nel considerare le correlazioni si valuta la loro significatività. Questa è legata al calcolo della probabilità che la correlazione sia dovuta o meno al caso, convenzionalmente si considerano come significative quelle correlazioni che hanno il 99% o il 95% di probabilità di essere non casuali. Nelle tabelle si trovano indicate con un asterisco (nel caso di 95% di probabilità) o con due asterischi (99%).

2.2 Il contesto e la scuola

2.2.1 I circondari

I circondari sono le unità organizzative territoriali più ampie per quanto riguarda le scuole comunali. Il territorio ticinese è diviso in nove aree individuate da un numero crescente (attualmente i circondari sono passati da 9 a 7 a causa di una riorganizzazione del sistema, si è preferito però conservare l'organizzazione precedente per coerenza con la raccolta delle informazioni) procedendo da sud verso nord. La popolazione degli allievi che hanno partecipato al test è composta da 2982 bambini (50.8% maschi e 49.2% femmine). La ricerca ha coinvolto 205 classi in 109 scuole, tra pubbliche (95%) e private (5%), di cui 62 pluriclassi. Tutti i circondari sono stati testati nel mese di maggio 2014. La distribuzione degli allievi per circondario è rappresentata nel Grafico 3.

Grafico 3: Numero di allievi che hanno svolto la prova nei nove circondari



La tabella riepiloga la suddivisione degli allievi delle classi terze nei 9 circondari in cui si è tenuto il test. Di seguito (Tabella 2) i punteggi medi nei 9 circondari per la variabile "italiano generale" e per le diverse dimensioni analizzate.

Le differenze tra i circondari risultano significative in tutte le dimensioni analizzando l'insieme dei dati tramite l'analisi della varianza. Osservando la Tabella 3 è possibile osservare come le differenze tra le medie siano, in termini assoluti, spesso limitate. Si è quindi proceduto a una analisi più di dettaglio per verificare quali fossero le differenze significative in ogni dimensione. Utilizzando la correzione di Bonferroni per Lessico risultano significative le differenze tra 1 e 6, 2 e 4, 4 e 6, 6 e 9; per Ortografia Morfologica risultano significative le differenze tra 1 e 6, 6 e 7; per Punteggiatura risultano significative le differenze tra 1 e 6, 2 e 4, 2 e 5, 2 e 9, 3 e 6, 6 e 7, 6 e 8.

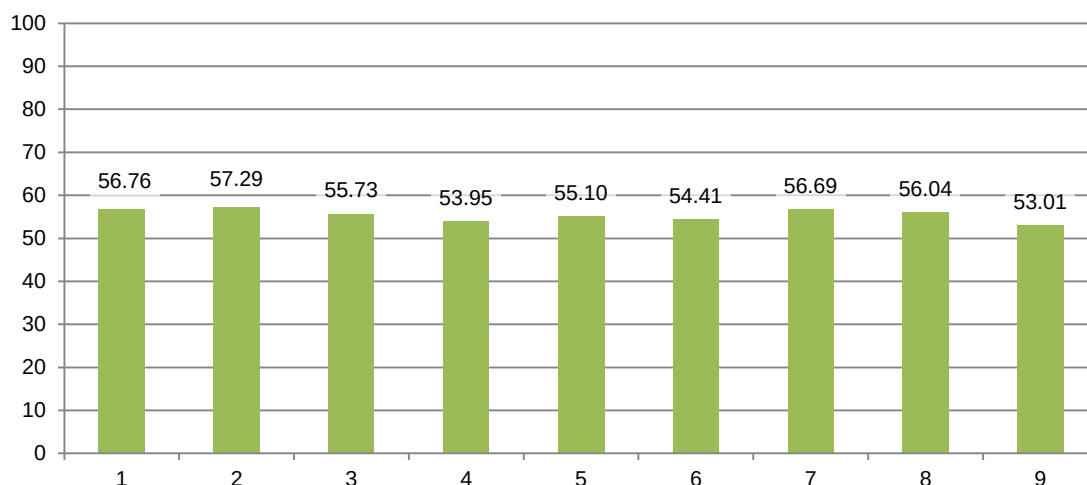
Non è possibile, con le informazioni a nostra disposizione, attribuire delle cause specifiche a queste differenze. Vale la pena notare come vi siano delle differenze nelle prestazioni nelle singole dimensioni non omogenee (un singolo circondario può avere una prestazione complessivamente elevata in una dimensione e modesta in un'altra). Per maggiore chiarezza si riportano i grafici delle singole dimensioni.

Tabella 2: Punteggi medi delle dimensioni nei diversi circondari

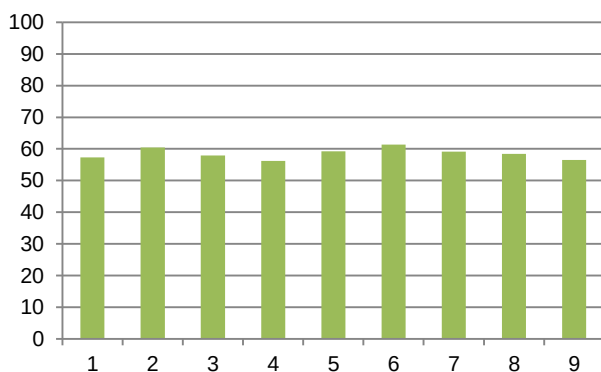
Circondario		LESSICO	ORTOGRAFIA FONOLOGICA	ORTOGRAFIA MORFOLOGICA	PUNTEGGIATURA ORTOGRAFICA	ITALIANO GENERALE
1	Media	57.34	59.44	59.80	46.53	56.76
2	Media	60.45	57.12	58.98	49.81	57.29
3	Media	57.96	57.50	57.68	47.23	55.73
4	Media	56.18	56.29	57.03	43.86	53.95
5	Media	59.21	56.79	55.77	43.95	55.10
6	Media	61.34	56.12	54.50	40.30	54.41
7	Media	59.15	58.14	59.49	46.27	56.69
8	Media	58.45	58.29	58.48	45.79	56.04
9	Media	56.46	56.01	54.35	42.82	53.01

I grafici che seguono mostrano la distribuzione delle medie dei punteggi ottenuti in “italiano generale” e nelle singole dimensioni suddivisi per circondario.

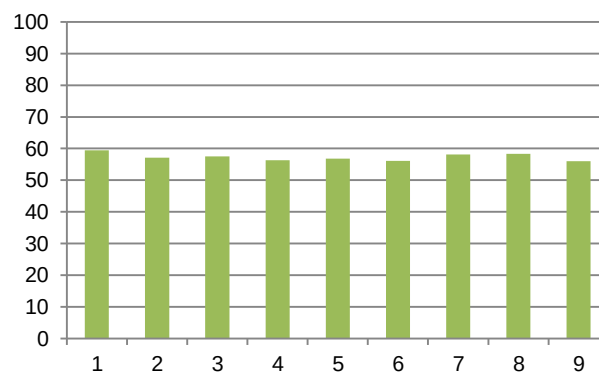
ITALIANO GENERALE



LESSICO



ORTOGRAFIA FONOLOGICA



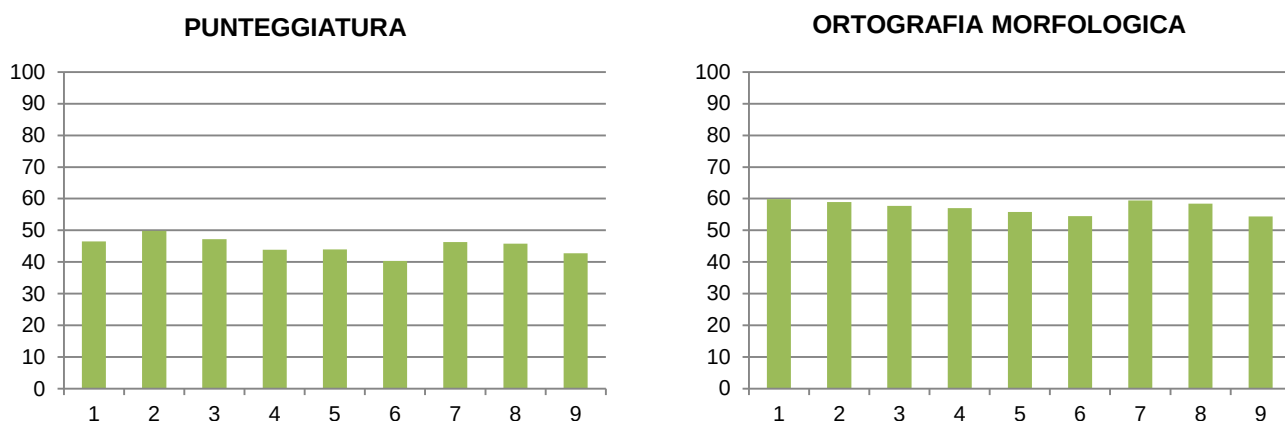
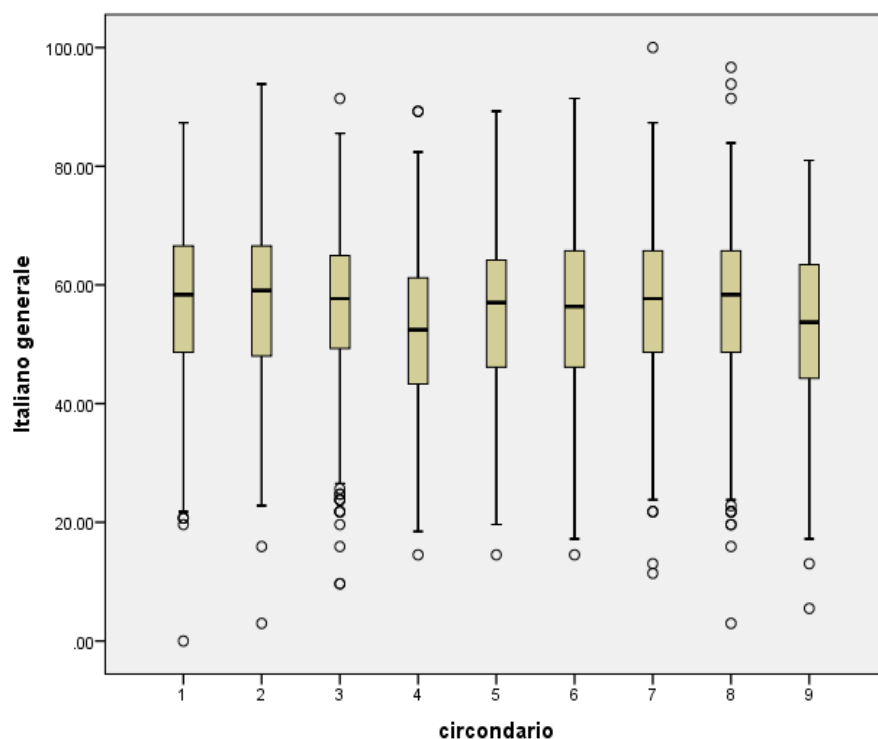


Grafico 4: Confronto delle mediane⁴ di Italiano generale nei circondari



Nel Grafico 4 è possibile vedere come le distribuzioni delle risposte dei diversi circondari siano tra loro molto simili oltre ad esservi una mediana analoga. In questa tipologia di grafico detta "Box-plot" è possibile vedere come vi sia un rettangolo per ogni circondario che rappresenta il 50% della popolazione, si tratta del 50% contiguo al valore mediano ovvero il 25% superiore e il 25% inferiore al valore stesso. La barra contenuta all'interno del rettangolo rappresenta il valore mediano. La barra verticale sopra al rettangolo rappresenta l'intervallo dei risultati del quarto di allievi che hanno ottenuto i punteggi più elevati mentre la barra verticale sotto al rettangolo quello del quarto di allievi che hanno ottenuto i punteggi più bassi.

⁴ Si definisce mediana (o valore mediano) il valore, in una determinata variabile, corrispondente al soggetto che si trovi a metà di un insieme di soggetti ordinati in modo crescente secondo i valori della variabile stessa. Metà dei soggetti ottiene quindi un punteggio superiore, e metà inferiore, rispetto al valore mediano.

2.2.2 Il grado di urbanizzazione del comune

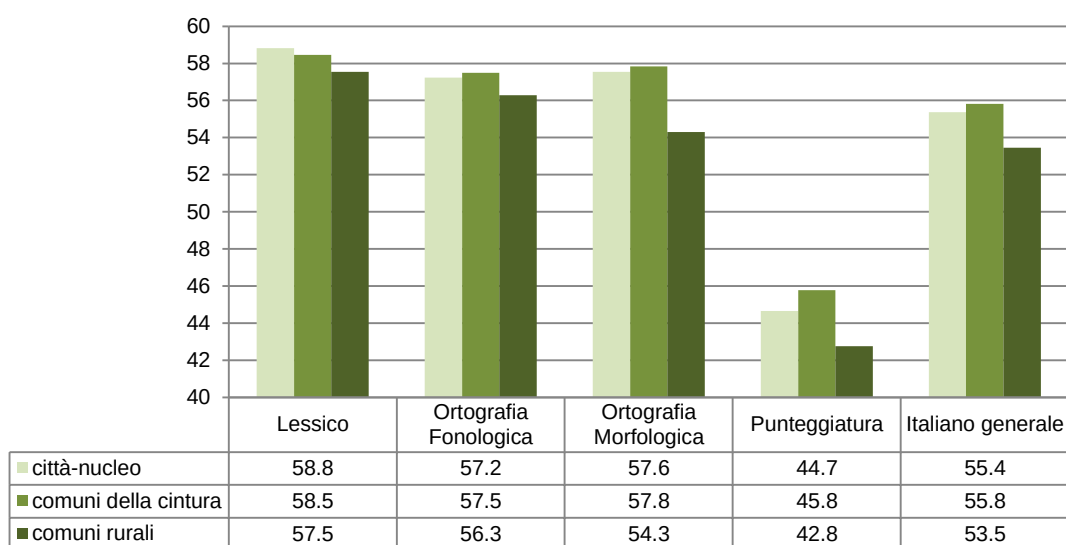
Il livello organizzativo più direttamente visibile delle scuole elementari e dell'infanzia è il comune. Per lo studio dei comuni viene proposto dall'Ufficio Federale di Statistica un indice relativo alla dimensione del singolo comune. Per il ticino vi sono tre gruppi: città nucleo; comuni della cintura; comuni rurali. L'effetto del livello di urbanizzazione sugli apprendimenti scolastici è stato oggetto di studi rivelandosi una delle caratteristiche da considerare (Zanolla, 2014). L'influenza sulla lingua di apprendimento potrebbe anche essere legata alla presenza di nuclei forti di lingue locali differenti dalla lingua di apprendimento.

I diversi comuni in cui sono localizzate le scuole elementari ticinesi sono stati raggruppati per tipologia urbana⁵ e classificati pertanto in città-nucleo (964 allievi), comuni della cintura (1684 allievi) e comuni rurali (333 allievi). Il livello di urbanizzazione sembra influenzare la prestazione in tutte le dimensioni soprattutto per quanto riguarda i comuni rurali che hanno le prestazioni meno elevate. La differenza fra i comuni della cintura (gruppo con prestazione più elevata) e i comuni rurali (gruppo con prestazione più limitata) è significativa (con un errore potenziale dell'1%) esclusivamente per "Ortografia Morfologica" e "Italiano generale".

In statistica quando si dice che i risultati di due gruppi hanno una differenza significativa si intende che la cosiddetta ipotesi nulla, secondo la quale la differenza fra due gruppi è legata a variazioni casuali, può essere scartata.

Ogni decisione in tal senso è legata a una probabilità di errore, nel presente rapporto si è scelto di indicare come significative le differenze non casuali con un errore potenziale dell'1%.

Grafico 5: Prestazioni alla prova e grado di urbanizzazione



Nella prova di matematica le differenze nelle dimensioni erano risultate significate solo nel caso "Geometria; Eseguire e Applicare" nella quale il vantaggio era per i comuni rurali.

⁵ Per la definizione della tipologia urbanistica si veda il sito Internet dell'Ufficio Federale di Statistica: <http://www.bfs.admin.ch/bfs/portal/fr/index/infothek/nomenklaturen/blank/blank/agglom/01.html>

2.3 La classe

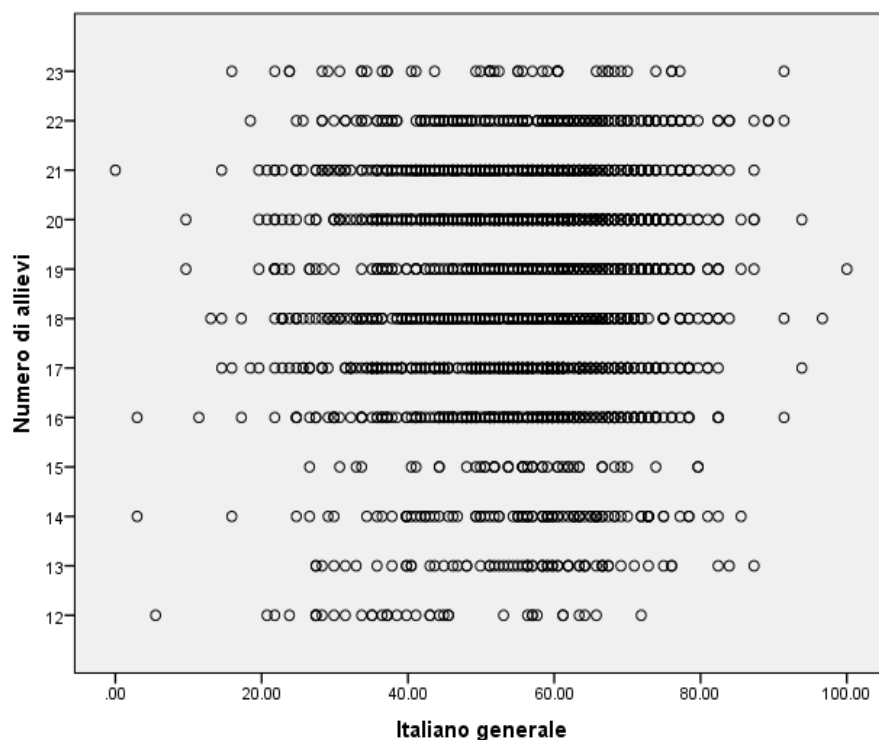
Nel valutare le prestazioni degli allievi le caratteristiche della classe sono ritenute estremamente rilevanti sia in termini di qualità sia in termini di quantità degli apprendimenti stessi (Dumay e Dupriez, 2009). In coerenza con quanto considerato nel precedente rapporto sono state considerate: numero di allievi per classe e la presenza di mono o pluriclasse.

2.3.1 dimensione della classe o numero di allievi

Il dibattito sull'effetto del numero di allievi per classe è, a livello internazionale, estremamente ricco e riguarda prevalentemente l'interazione che può avvenire nella classe tra docenti e allievi al ridurre gli effettivi (Crescentini e Galetta, 2014).

In Ticino le classi possono configurarsi diversamente a seconda del luogo e della popolazione interessata. Nell'anno considerato la media degli allievi di terza nelle classi è pari a 15. Questo numero diventa 18 nelle monoclasse e 8 nelle pluriclassi.

Grafico 6: Diagramma di dispersione Numero di allievi – Punteggio alla prova



Per quanto riguarda i dati da noi analizzati, una breve analisi della relazione tra il numero di allievi per classe e il punteggio ottenuto nella prova standardizzata indica come non sia possibile osservare nessuna tendenza esplicita (si veda Grafico 6). Numerosità della classe e performance in italiano non appaiono correlati tra loro.

Questo risultato è coerente con quanto osservato nel rapporto sulla prova standardizzata di matematica nel quale non si era rilevata alcuna differenza significativa collegata alla numerosità delle classi.

2.3.2 La pluriclasse

La pluriclasse è una classe formata da bambini che frequentano anni di corso differenti. Si tratta di una particolare organizzazione di classe della scuola primaria diffusa soprattutto nelle zone a bassa densità abitativa, laddove non si raggiunge il numero minimo di iscritti necessari per la formazione di classi omogenee. Gli studi sull'efficacia della pluriclasse e sugli effetti sugli allievi hanno iniziato a essere condotti negli anni '70 collegandosi alle riflessioni sulle scuole delle aree discoste negli Stati Uniti (Miller, 1988). Il tema è sfaccettato e vi sono ricerche che identificano sia vantaggi sia svantaggi.

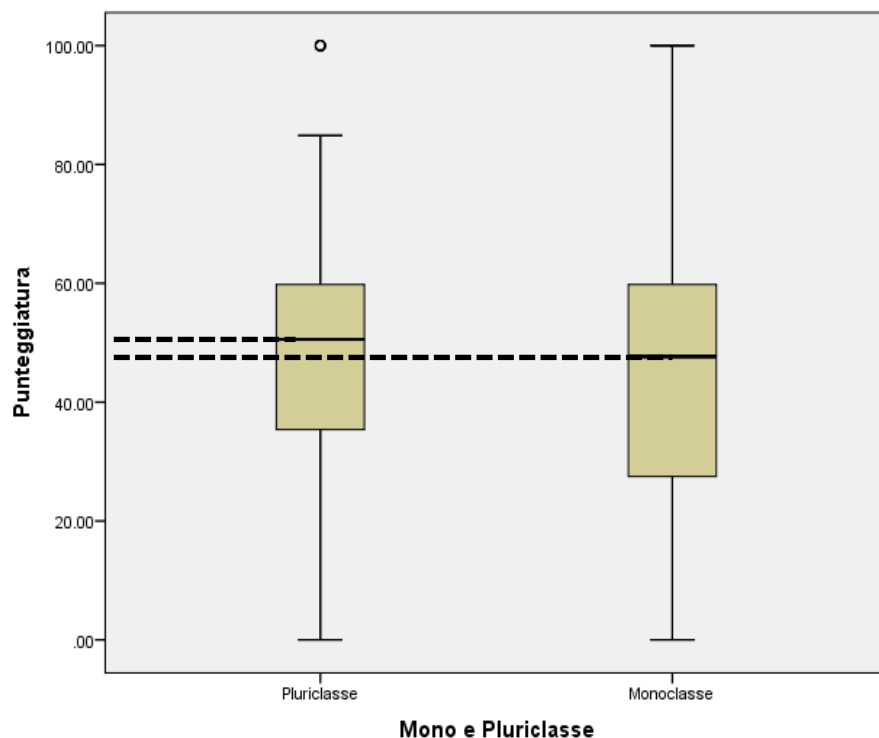
Un elenco non esaustivo può prevedere tra i vantaggi: riduzione della disobbedienza, maggiore sviluppo linguistico (Logue, 2006); migliori apprendimenti nella lettura e nelle capacità linguistiche (Skapski, 1960); migliore e più rapido sviluppo cognitivo (Fosco, Schleser, & Andal, 2003); migliori e più approfondite relazioni tra insegnanti e allievi e loro famiglie (Miller, 1994); maggiore tranquillità sin dal primo giorno (Fu et al., 1999); maggiore possibilità di un apprendimento centrato sulla velocità del singolo studente (Aina, 2001); il rapporto maggiormente individualizzato può portare a un atteggiamento degli studenti migliore nei confronti della scuola e delle relazioni con i pari (Veenman, 1995); le interazioni sociali e il lavoro collaborativo appaiono più sviluppati (Hoffman, 2002); la differenziazione applicata aiuta gli allievi più dotati (Lloyd, 1999); vengono sviluppate strategie di insegnamento maggiormente basate sulla collaborazione tra gli allievi (Broome, 2009); lo sviluppo di strategie di assistenza tra gli allievi promuove la percezione che la differenza sia normale e che si sia parte di una comunità di apprendimento (Kolstad & McFadden, 1998); la presenza di allievi più maturi può essere di stimolo (Leuven & Rønning, 1996).

Tra gli svantaggi emergono invece: difficoltà maggiore per creare i materiali e i gruppi riportata dai docenti e mancanza di una formazione specifica (Farkas & Duffett, 2008); difficoltà gestionali nell'organizzare le classi (Kolstad & McFadden, 1998); l'insegnamento è meno efficace e l'apprendimento più complesso (Mason & Burns, 1996); la presenza di allievi più giovani può avere un effetto negativo sull'apprendimento (Leuven & Rønning, 1996).

Gli allievi frequentanti pluriclassi sono pari a circa il 17% della popolazione mentre il restante frequenta monoclasse. Nel nostro caso le differenze nei punteggi medi delle dimensioni tra monoclasse e pluriclassi non sono significative, salvo in "Punteggiatura" nella quale si mostrano più abili gli allievi delle pluriclassi. Questi ultimi, come è possibile vedere dalla tabella, hanno punteggi maggiori in tutte le dimensioni anche se queste differenze non sono statisticamente significative.

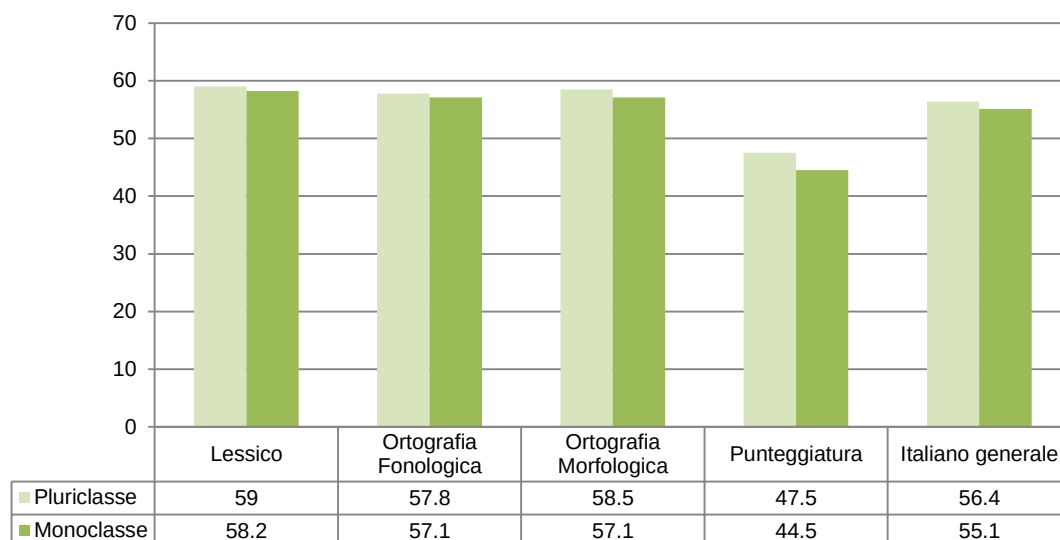
Osservando il grafico relativo ai valori nella dimensione Punteggiatura (Grafico 7) costruito mettendo a confronto le mono e le pluriclassi è possibile rilevare come la differenza sia nella distribuzione dei soggetti più che nel semplice valore mediano. Nelle monoclasse infatti il quartile⁶ inferiore alla mediana ha dei valori sensibilmente più bassi.

Grafico 7: Confronto distribuzione mono e pluriclasse



⁶ Si utilizza il termine quartili quando la distribuzione dei soggetti è ordinata dal valore inferiore sino al superiore e viene divisa in quattro gruppi di pari numerosità. Il primo quartile raggruppa quindi il 25% dei soggetti con valore più basso il secondo il successivo 25% e così via.

Grafico 8: Media valore nelle dimensioni nelle mono e pluri classi



Coerentemente con la letteratura le prestazioni non appaiono avere grandi differenze rispetto a pluri e mono classi (Veenman, 1995). Attualmente la prassi è in uso in diverse nazioni con programmi specifici (50% delle classi in Olanda, Finlandia, Portogallo; un settimo in West Australia; presenza sfaccettata negli Stati Uniti; le Antille Olandesi hanno intrapreso un cambiamento di sistema dal 2004 verso un approccio centrato sugli allievi e con un modello basato sulle pluriclassi). Negli Stati Uniti ha ricevuto una spinta rilevante in relazione al progetto NCLB (No Child Left Behind Act), in Australia e Canada esistono progetti specifici di diffusione. In tutti i contesti viene sottolineato come, al fine di ottenere i maggiori benefici da questo modello di classe, sia rilevante una formazione specifica del personale insegnante.

Nella prova di matematica le differenze tra gli allievi erano risultate significative esclusivamente in una delle dimensioni, anche in quel caso la differenza era a favore degli allievi delle pluriclassi. Rispetto a quanto rilevato in questa prova si può ricordare come la possibilità di confronto con allievi di altre età potrebbe funzionare da stimolo per l'allievo nello sviluppare delle competenze maggiori nella lingua di comunicazione.

2.4 L'allievo

Nel valutare le prestazioni in una prova è rilevante considerare le caratteristiche degli allievi e verificare se non possano esservi caratteristiche dell'allievo che hanno un impatto sulla prestazione stessa. Questo oltre ad avere una rilevanza dal punto di vista dell'equità della prova potrebbe anche essere rilevante rispetto alla organizzazione in generale. Sono state quindi considerate diverse dimensioni delle quali verificare l'eventuale impatto: caratteristiche socio demografiche; nazionalità e lingua materna; gender; origine sociale; età; il numero delle assenze nell'anno della prova (in termini di giorni di assenza).

2.4.1 Caratteristiche socio demografiche

I grafici di seguito forniscono alcune informazioni sulla configurazione sociodemografica della popolazione degli allievi.

Grafico 9: Percentuale di allievi divisi per nazionalità

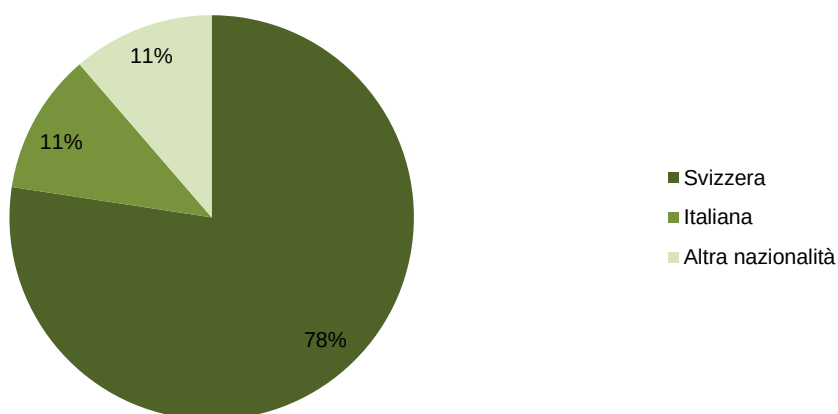
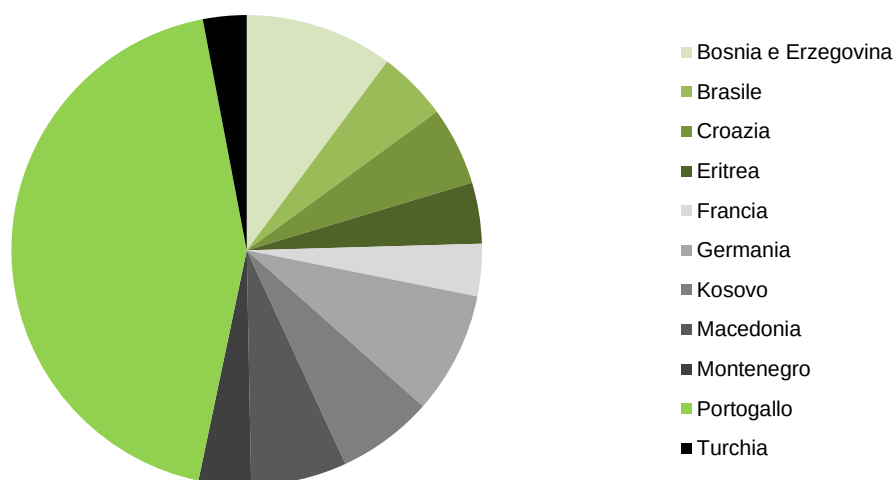


Grafico 10: Percentuale di allievi con nazionalità diversa da Svizzera o Italiana



La maggior parte degli allievi è di nazionalità svizzera (Grafico 9), con una percentuale consistente di allievi provenienti dall'Italia o da altri Paesi (i Paesi più frequentemente rappresentati sono mostrati nel Grafico 10). In questo caso sono stati esclusi i Paesi con meno di 5 allievi.

Considerando il gender degli allievi il 49,2% è femmina e il 50,8 maschio.

2.4.2 La Nazionalità e la lingua madre

Il Grafico 11 mostra la distribuzione delle medie dei punteggi nelle diverse dimensioni suddivise per nazionalità dell'allievo e il Grafico 12 suddivise per lingua madre. Nel primo grafico si rileva come i punteggi migliori siano sempre degli allievi di nazionalità Svizzera. Se invece il confronto è fatto relativamente alla lingua i punteggi peggiori sono ottenuti da allievi non madrelingua italiana. Sebbene sia logico supporre che in una prova di italiano una maggiore padronanza della lingua avvantaggi i gruppi di allievi madrelingua si rende necessario approfondire l'analisi sulle differenze tra le medie dei tre gruppi organizzati per nazionalità anche per capire se le migliori performance degli allievi svizzeri rispetto agli altri risultano statisticamente significative o al contrario non rilevanti dal punto di vista statistico. In questo caso tramite analisi della varianza è stato possibile, ad esempio, verificare che la significatività statistica è rintracciabile quando si guarda al gruppo degli allievi di altra nazionalità, mentre non sussistono differenze significative tra allievi svizzeri e italiani. Rispetto agli allievi di nazionalità svizzera quelli madrelingua italiana hanno ottenuto punteggi superiori con valori statisticamente significativi (con un errore potenziale dell'1%) in tutte le dimensioni rispetto a quelli di madrelingua diversa.

Grafico 11. Medie dei punteggi degli allievi suddivisi per nazionalità

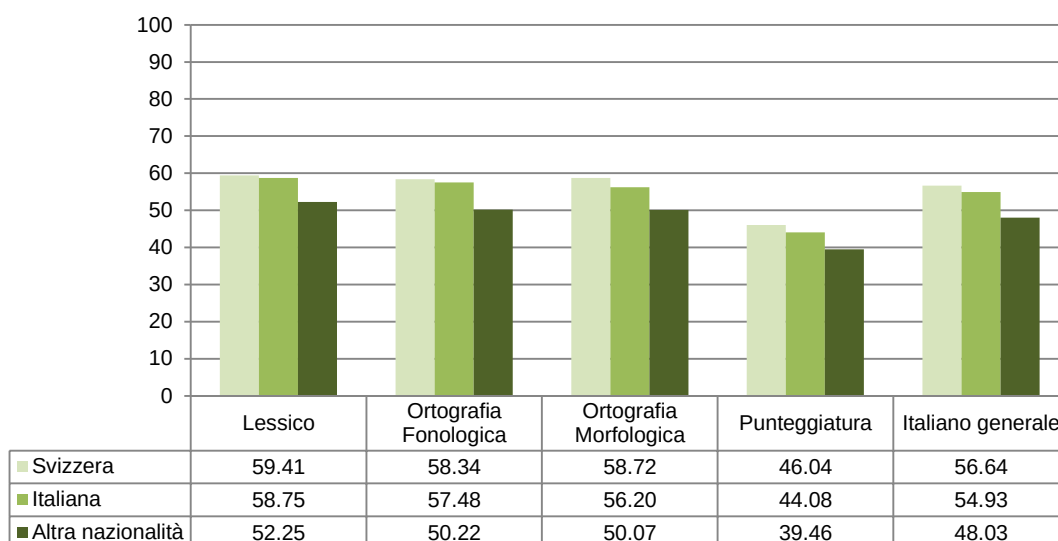
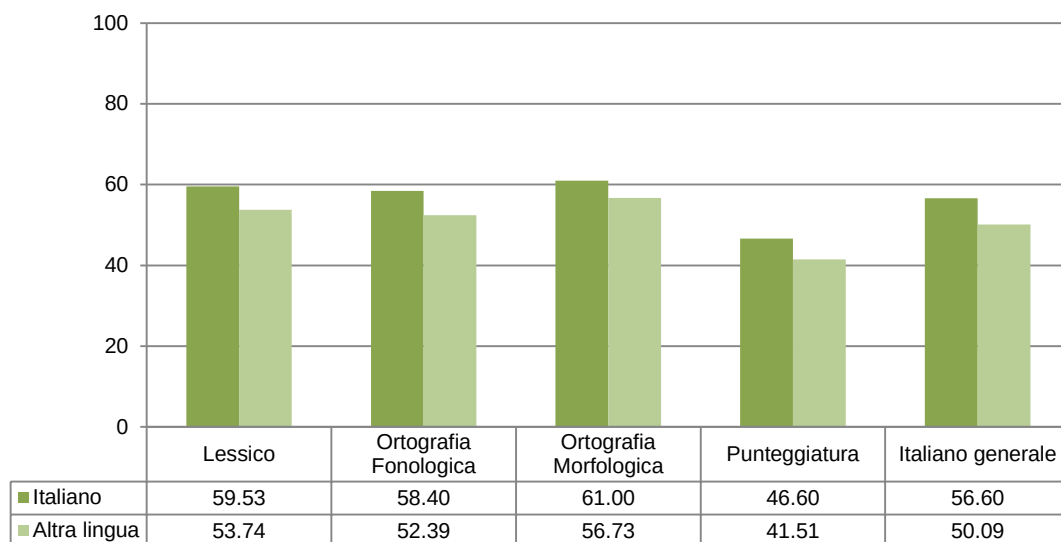


Grafico 12: Medie dei punteggi suddivisi per lingua madre



2.4.3 Prestazione alle prove e gender

Come rilevato precedentemente (Zanolla, 2014) le differenze di gender possono avere un impatto sulla prestazione in considerazione sia della socializzazione primaria (in famiglia) sia di quella esterna alla scuola, sia in riferimento a eventuali effetti presenti a scuola. Riuscire a distinguere questi effetti va oltre le possibilità del presente lavoro ma è importante rendere attento il lettore in proposito.

Nei risultati alla prova non emergono comunque differenze significative tra maschi e femmine in nessuna delle dimensioni.

Tabella 3: Risultati alle prove e gender

	Lessico	Ortografia Fonologica	Ortografia Morfologica	Punteggiatura	Italiano generale
Femminile	58.1	57.9	57.4	45.0	55.2
Maschile	58.9	57.6	57.4	45.2	55.6

E anche confrontando la popolazione suddividendola in quartili (Tabella 4) rispetto alle prestazioni non è possibile identificare differenze rilevanti. Questa suddivisione permette di osservare come vi sia equità di distribuzione anche nei livelli di prestazione.

Tabella 4: Risultati alle prove in quartili e gender

		Lessico	Ortografia Fonologica	Ortografia Morfologica	Punteggiatura	Italiano generale
Femmine	25	50.28	49.15	49.03	27.49	46.15
	50	58.43	57.53	60.64	47.68	56.34
	75	65.74	64.87	68.71	59.78	64.95
Maschi	25	50.28	49.15	50.63	27.49	46.78
	50	58.43	59.82	60.64	47.68	57.01
	75	65.74	67.75	68.71	59.79	65.35

È possibile inoltre rilevare come la distribuzione sui livelli di difficoltà sia omogenea in tutte e quattro le scale iniziali. Rispetto alla prova di matematica erano invece emerse delle differenze e una presenza più marcata di allievi maschi con prestazioni molto alte. Da considerare è comunque la differenza di età (prova di italiano svolta in terza rispetto alla quinta) e come la differenza possa non essersi ancora rivelata.

2.4.4 Numero di assenze

Un'altra variabile presa in esame è il numero di assenze degli allievi nell'anno scolastico 2013-2014, l'idea era di verificare se le assenze degli allievi fossero in relazione con i punteggi ottenuti nelle diverse dimensioni. Le assenze sono state valutate considerando il numero di giorni di assenza giustificati dagli allievi.

Come mostrato in Tabella 5, anche in questo caso non si evidenziano correlazioni significative.

Tabella 8: Correlazioni tra le dimensioni valutate dalla prova, le assenze e la nota in condotta conseguita a fine anno

	LESSICO	ORTOGRAFIA FONOLOGICA	ORTOGRAFIA MORFOLOGICA	PUNTEGGIATURA ORTOGRAFICA	ITALIANO GENERALE
ASSENZE	-.002	-.035	-.009	.004	-.015

2.4.5 Origine sociale

L'origine sociale degli allievi è stata spesso indicata come uno dei fattori rilevanti da considerare nella valutazione delle prestazioni scolastiche. Le ragioni sono molteplici e fanno riferimento principalmente alle aspettative della famiglia e alle risorse che sono messe a disposizione degli allievi sia in termini materiali sia in termini culturali nel senso più ampio del termine.

L'analisi condotta conferma che, in tutte le dimensioni, quanto più elevato è il retroterra socio-culturale⁷ di un individuo, migliore sarà la sua prestazione. I figli di genitori appartenenti alla categoria "colletti bianchi altamente qualificati" riportano risultati significativamente migliori. Sia nelle prove standardizzate, sia nella valutazione da parte degli insegnanti, i figli dei gruppi socialmente più svantaggiati si posizionano peggio. Vale la pena osservare come tra la fascia sociale di origine più modesta e la fascia sociale di origine più elevata vi sia, nella valutazione di fine anno, una differenza media di mezzo punto. Come avviene nella

⁷ Nonostante termini quali "colletti blu" o "colletti bianchi" per definire coloro che svolgono rispettivamente un lavoro manuale o tecnico come ad es. gli operai e gli artigiani o mansioni più o meno fisiche e coloro che invece svolgono un lavoro di tipo impiegatizio, i funzionari, i venditori, ecc., ci sembrassero alquanto desueti, si è scelto di non alterare la nomenclatura delle categorie dell'OECD della variabile HseCateg. Questa è inoltre la scelta che è stata fatta a livello nazionale nel recente Rapporto sul sistema educativo svizzero (CSRE, 2014). Il retroterra socioculturale familiare corrisponde alla categoria occupazionale più elevata tra i due genitori. Come si fa generalmente in questi casi, è stato dunque applicato il criterio di dominanza (Erikson, 1988; Schizzerotto e Schadee, 1990).

maggior parte dei lavori sul tema anche nel presente documento saranno utilizzate le professioni dei genitori per categorizzare i figli con la consapevolezza che ci si riferisce al retroterra e alle opportunità.

Tabella 13: Relazione tra punteggi medi e origine sociale degli allievi

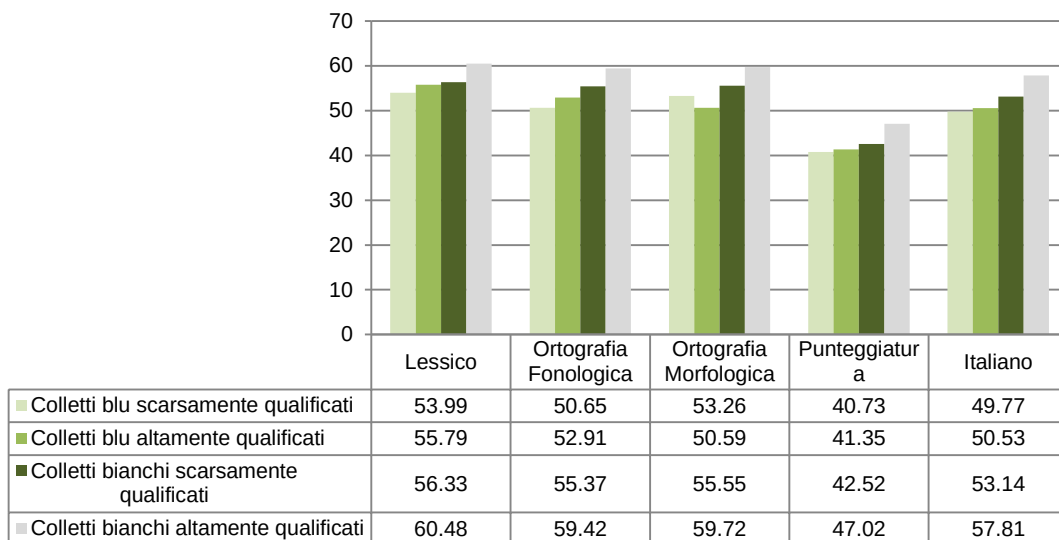
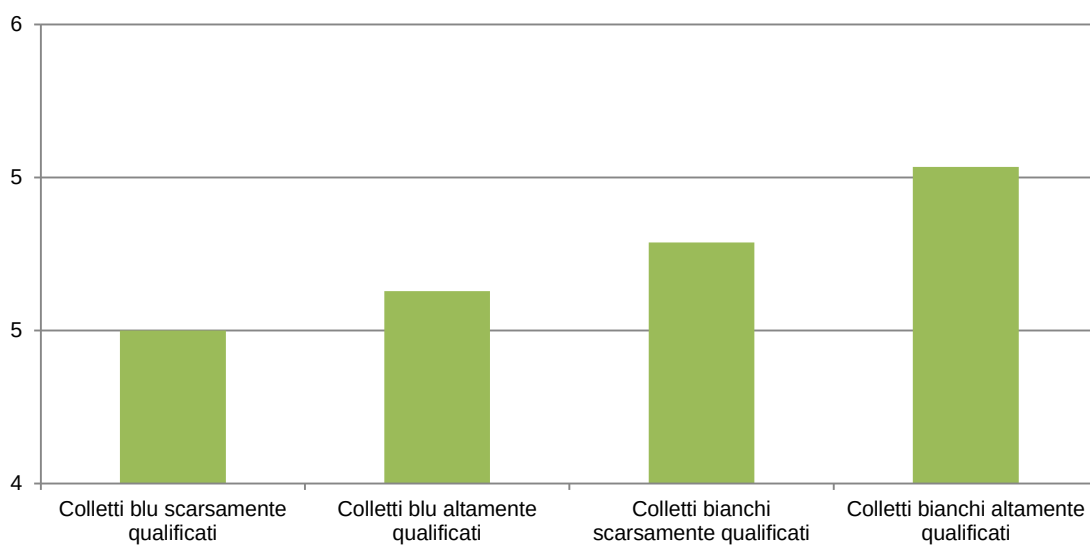
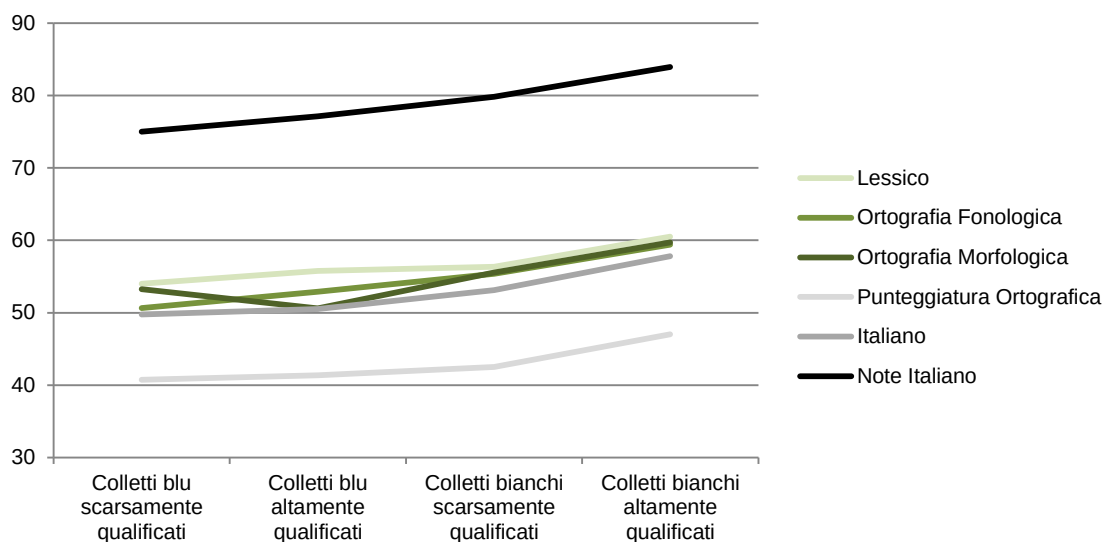


Grafico 14: Media note di italiano



È interessante notare come sia il voto finale sia i punteggi nella prova abbiano valori che crescono coerentemente uno all'altro. Per confrontare i valori, la nota finale è stata normalizzata ovvero i punteggi che in origine possono essere da 0 a 6 sono stati ricalcolati in modo da essere da 0 a 100. Sono messi maggiormente in evidenza il punteggio finale e quello del valore aggregato.

Grafico 15: Punteggi rispetto ai gruppi socioeconomici di provenienza



2.4.6 Età dell'allievo

Nelle classi della scuola primaria sono poco frequenti i casi di ripetizione (Rigoni, 2013) ma come ricordato nel rapporto di matematica sono relative principalmente ad allievi provenienti da gruppi sociali svantaggiati (Zanolla, 2014).

E' stata calcolata la correlazione tra la prestazione in ogni settore e l'età dell'allievo espressa in mesi e in ogni settore essa è risultata negativa e significativa. Con l'aumentare dell'età (cui verosimilmente si associano trascorsi di trasferimenti di residenza e scuola, migrazioni da altri paesi o insuccessi scolastici) gli allievi ottengono dunque punteggi significativamente inferiori in tutti i settori.

Tabella 5: Correlazioni tra punteggi medi alla prova e età degli allievi

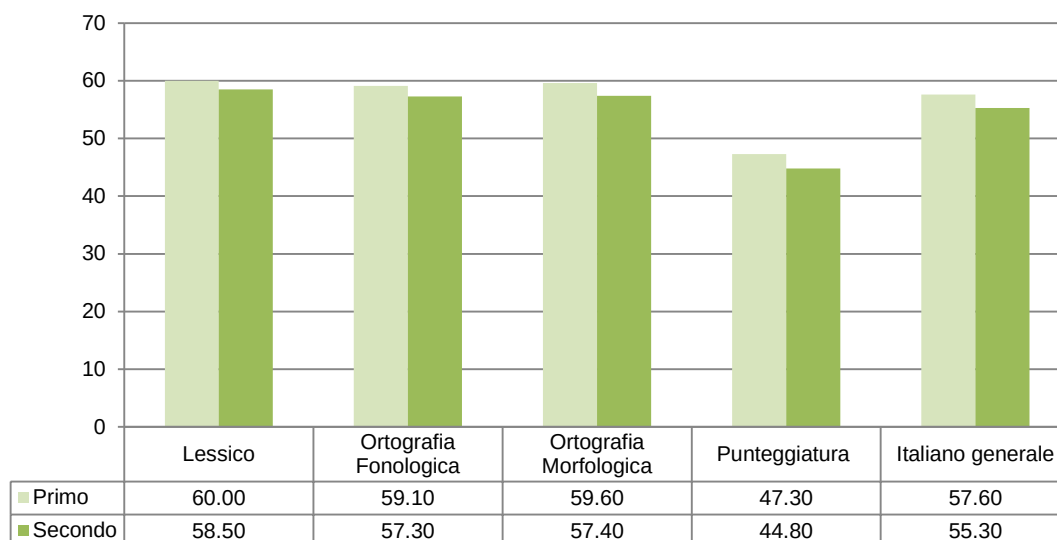
		Lessico	Ortografia Fonologica	Ortografia Morfologica	Punteggiatura Ortografica	Italiano
Età	Correlazione di Pearson	-.106**	-.105**	-.072**	-.058**	-.104**

** La correlazione è significativa a livello 0,01 (a due code).

L'86.9% degli allievi della nostra popolazione è nata nel corso del 2005 e il 12.1% nel 2004.

Si è quindi verificato se tra gli allievi nati nel 2005 sussistessero differenze tra i bambini nati nei primi sei mesi dell'anno (N 1309) e quelli nati a partire da luglio (N 1185): dall'analisi della varianza emerge che i nati nel primo semestre ottengono risultati significativamente migliori.

Grafico 16: Risultati alle prove degli allievi nati nel primo e nel secondo semestre 2005



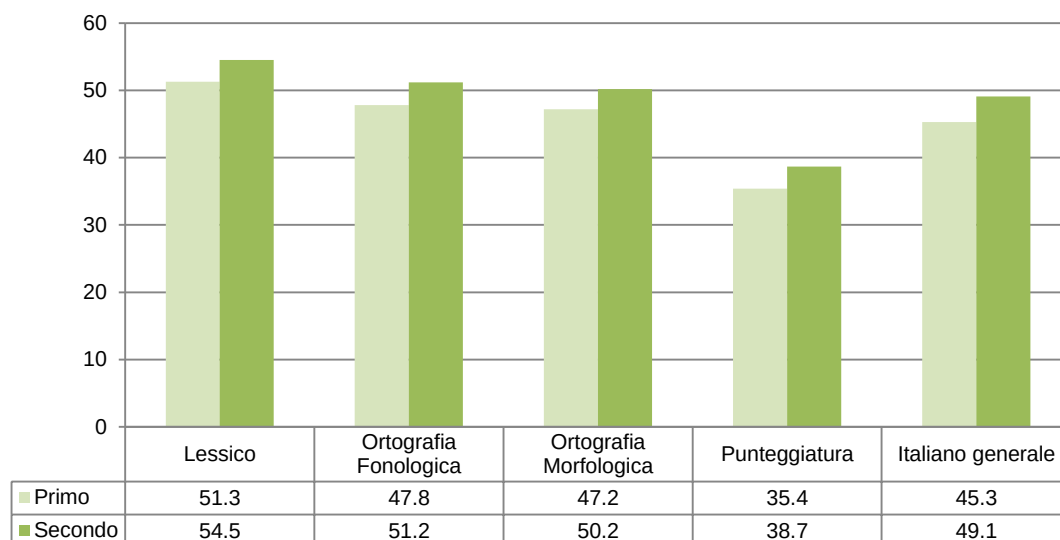
Anche considerando lo specifico mese di nascita nell'anno 2005 si constatano differenze significative in tutte le dimensioni. I valori appaiono decrescenti in modo quasi continuo rispetto al mese di nascita. Questi risultati si differenziano rispetto a quanto individuato nelle analisi relative alle prove standardizzate di matematica. Una possibile spiegazione si può trovare nella classe e nel processo di apprendimento della lingua rispetto alla matematica ma non abbiamo possibilità di sostenere questa ipotesi tramite altre informazioni.

Tabella 6: Punteggi medi nelle diverse dimensioni per mese di nascita (2005)

Mese di nascita	Lessico	Ortografia Fonologica	Ortografia Morfologica	Punteggiatura Ortografica	Italiano
Gennaio	61.9	60.7	62.8	49.6	60.0
Febbraio	59.4	60.0	60.8	47.0	58.1
Marzo	60.6	59.7	59.5	46.8	57.7
Aprile	59.9	59.3	59.3	48.2	57.8
Maggio	59.1	57.1	57.9	47.3	56.1
Giugno	58.9	58.0	56.9	44.7	55.7
Luglio	58.9	57.0	57.7	46.1	55.7
Agosto	59.1	58.1	58.3	44.2	55.7
Settembre	59.8	59.5	59.6	45.6	57.3
Ottobre	58.5	56.9	57.6	45.6	55.3
Novembre	57.1	56.4	55.3	44.2	53.9
Dicembre	57.2	55.5	54.5	41.9	52.9

Replicando la medesima analisi tra i nati del 2004 si constata invece che sono i nati nel secondo semestre a riportare le medie più elevate.

Grafico 17: Risultati alle prove degli allievi nati nel primo e nel secondo semestre 2004



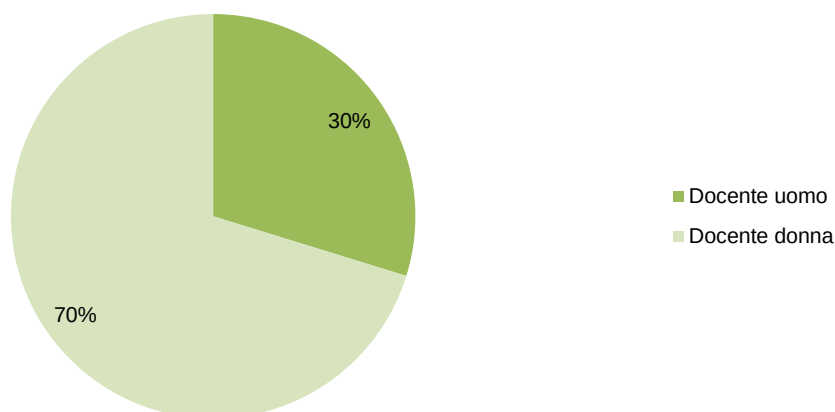
2.5 Docente e allievo

La relazione che si instaura tra docente allievo è una delle caratteristiche che definiscono il contesto scolastico. Questa relazione è particolarmente significativa all'interno delle scuole elementari. Nel presente rapporto siamo andati ad indagare due dimensioni che potrebbero essere influenzate da questa relazione. La prima è relativa al gender e al fatto che vi sia una differenza in relazione alla differenza o omologia tra docente e allievo e la seconda alla valutazione di condotta che identifica la qualità della relazione tra docente e allievo.

2.5.1 Il genere del docente

I docenti titolari coinvolti nell'indagine sono stati 205, per la maggioranza donne (Grafico 18). Ciò non sorprende ed è in linea con i recenti studi che introducono il concetto di "femminilizzazione della professione docente" (Marchesi, 2012; Bombelli, 2009), affermando come questa professione veda sempre più la presenza di donne nei ruoli di docenza.

Grafico 18: Percentuale di docenti titolari uomini e docenti titolari donne



Rispetto all'apprendimento non sono indicati in letteratura effetti specifici del genere del docente. Come indicato nel rapporto sulle prove di matematica (CIRSE, 2014) vi sono però dati generali sugli effetti che il

genere del docente potrebbe avere sull'apprendimento degli allievi. Dall'analisi dei nostri dati risulta che in nessuna delle dimensioni considerate sussistono differenze statisticamente significative tra le classi con docente titolare maschio e quelle con docente titolare donna.

In altre parole il genere del docente non sembra essere di per sé rilevante per la performance degli allievi. Anche quando si mette in relazione il genere del docente con i punteggi medi ottenuti dagli allievi, distinti anch'essi a loro volta per genere (allievi e allieve), non si constatano differenze significative (Grafico 20). Nella analisi dei risultati alla prova di matematica era risultato che le allieve ottenevano risultati migliori nel caso avessero un docente maschio.

Grafico 19: Medie dei punteggi conseguiti dagli allievi con docente titolare donna o uomo.

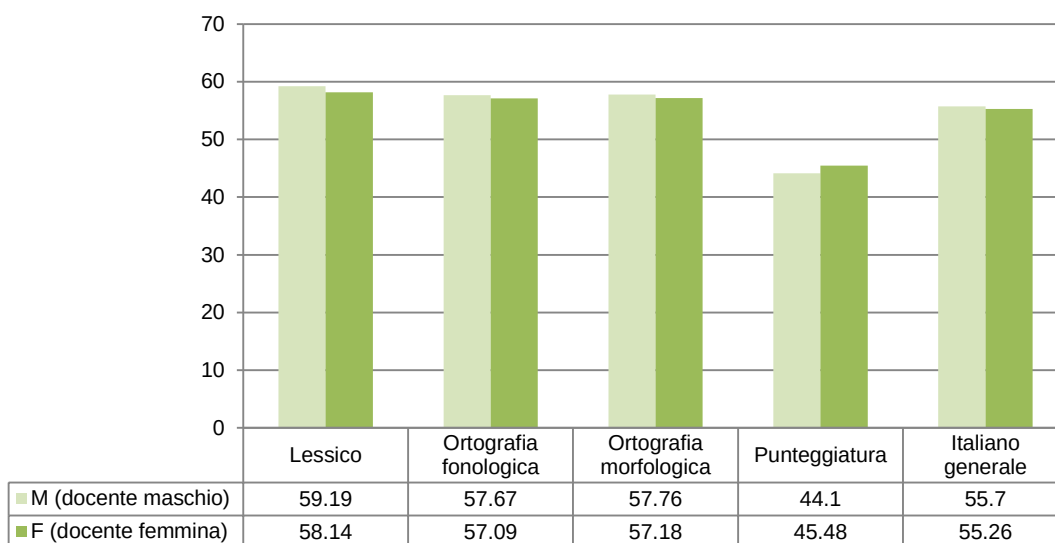
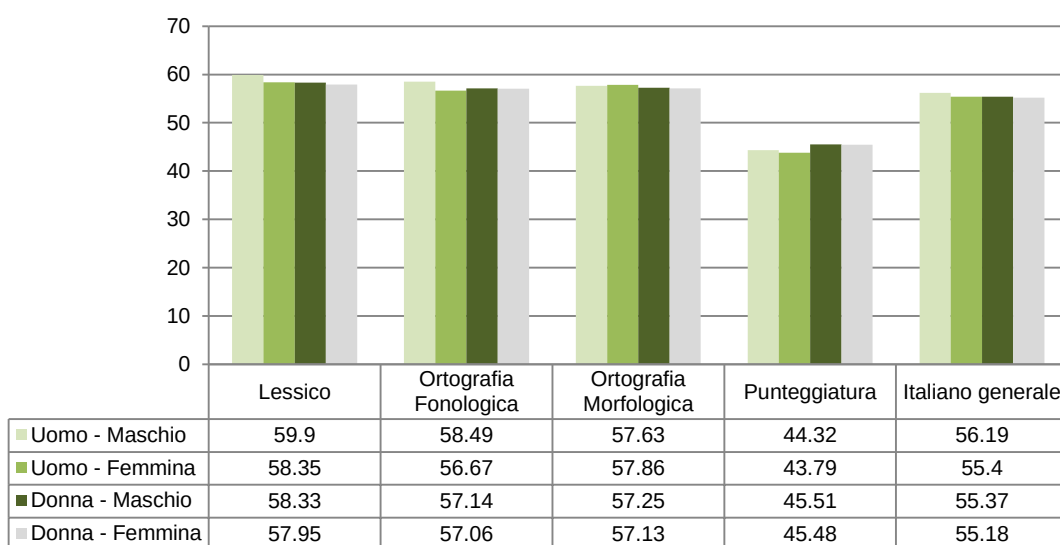


Grafico 20: Medie dei punteggi conseguiti dagli allievi maschi e femmine con docente titolare donna o uomo



2.5.2 Nota di condotta

Le correlazioni diventano significative quando si mettono in relazione le dimensioni testate con la condotta degli allievi, in questo caso tutte le dimensioni risultano statisticamente associate (con probabilità di errore pari all'1%, in considerazione deve essere tenuto il valore basso, in termini assoluti, del valore della correlazione) agli aspetti di condotta valutati attraverso la nota in condotta.

È importante ricordare come una analisi delle correlazioni non individui delle causalità quanto piuttosto dei cambiamenti coerenti tra loro. In questo caso è possibile vedere come la nota di condotta e i valori nelle dimensioni crescano o calino coerentemente.

Tabella 7: Correlazioni tra le dimensioni valutate dalla prova e la nota in condotta conseguita a fine anno

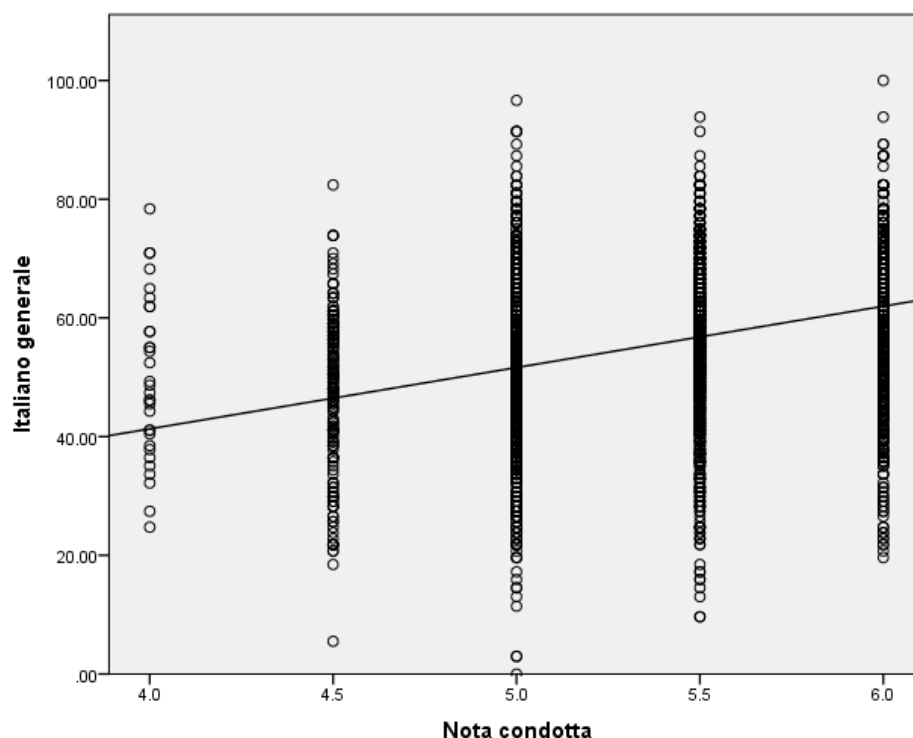
	LESSICO	ORTOGRAFIA FONOLOGICA	ORTOGRAFIA MORFOLOGICA	PUNTEGGIATURA ORTOGRAFICA	ITALIANO GENERALE
CONDOTTA	.142**	.203**	.216**	.146**	.231**

Nella Tabella 8 si riportano i valori medi in Italiano generale raggruppati per i diversi valori della nota di condotta. La relazione è resa più evidente nel grafico (Grafico 21) successivo nel quale sono riportati i singoli allievi rappresentati mettendo in rapporto i due valori e disegnando la retta di regressione che meglio rappresenta questa relazione.

Tabella 8: Media Italiano generale e nota di condotta

Nota condotta	Media Italiano generale
4.0	49.65
4.5	47.77
5.0	53.20
5.5	56.29
6.0	59.64

Grafico 21: Dispersione prestazioni nella prova rispetto alla nota di condotta



Vale la pena ricordare come la nota di condotta rilevi una dimensione di relazione tra l'allievo e il docente e come nella scuola primaria questa relazione sia particolarmente forte e presente in tutte le materie.

2.6 Risultati al test e note

2.6.1 Relazione nota di italiano di fine anno – prestazione alla prova

Sebbene sia evidente la differenza di impostazione e di obiettivo delle valutazioni di fine anno e delle prove standardizzate abbiamo ritenuto opportuno utilizzare comunque i valori di entrambe per fornire delle informazioni generali. Confrontare valutazioni che sono nate con logiche e finalità diverse è molto delicato. Come richiamato da Salvisberg e Zampieri (2015) “Sin dagli anni ‘30, i lavori docimologici svolti sulla validità delle notazioni degli insegnanti sui test degli allievi, sollevano critiche rispetto alla difficoltà di essere oggettivi” (19) porle quindi a confronto con prove che hanno lo scopo di essere oggettive pone in essere un problema molto forte al quale si aggiunge la differenza di scopo (Motier Lopez & Crahay, 2009)”. Le note scolastiche rilevano infatti solo in parte le competenze disciplinari riportando anche valori maggiormente soggettivi ma comunque rilevanti per la valutazione completa dell'allievo. Motier, Lopez e Crahay (2009) sostengono inoltre che le due tipologie di valutazione dovrebbero essere tra loro integrate per fornire una fotografia maggiormente strutturata nella quale le dimensioni di soggettività e oggettività avrebbero la possibilità di integrarsi.

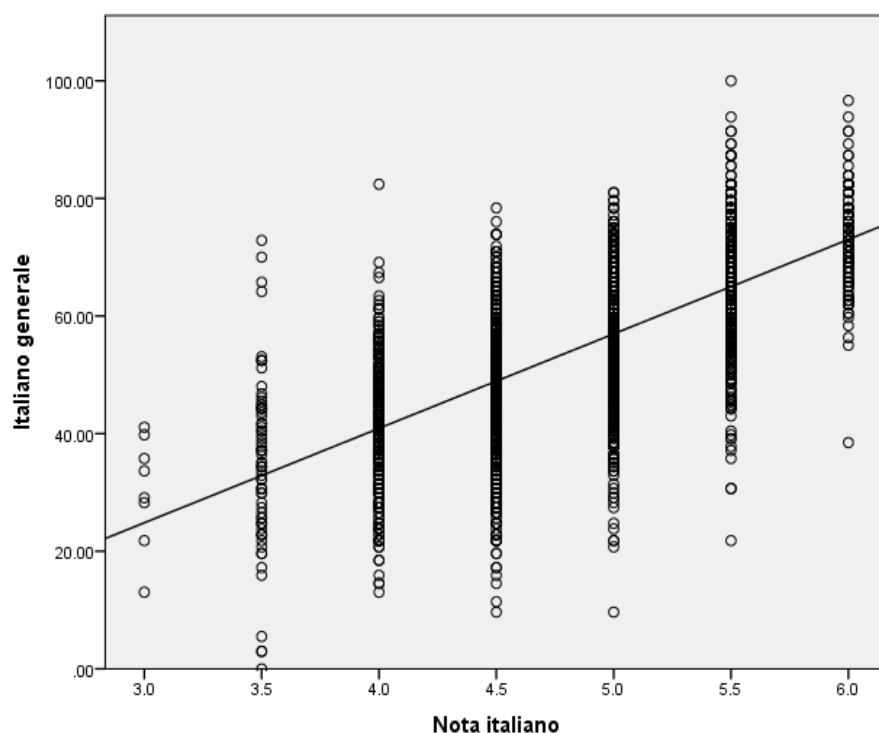
La correlazione tra nota in italiano e test è statisticamente significativa. Nella Tabella 9 sono riportati i valori di correlazione ($p < .01$). Ciò indica una generale conferma della coerenza tra valutazioni dei docenti e performance degli allievi alla prova standardizzata.

Esplorando la relazione tra nota di Italiano e punteggio nella dimensione di Italiano generale è possibile osservare una progressiva crescita delle due valutazioni. Questo valore contribuisce a rafforzare la validità della prova fornendo una validazione concorrente delle dimensioni testate.

Tabella 9. Correlazioni tra le dimensioni valutate con la prova e la nota conseguita dagli allievi in italiano

	LESSICO	ORTOGRAFIA FONOLOGICA	ORTOGRAFIA MORFOLOGICA	PUNTEGGIATURA ORTOGRAFICA	ITALIANO GENERALE
Nota italiano	.498**	.601**	.540**	.454**	.667**

Grafico 12: Dispersione prestazioni nella prova rispetto alla nota di Italiano



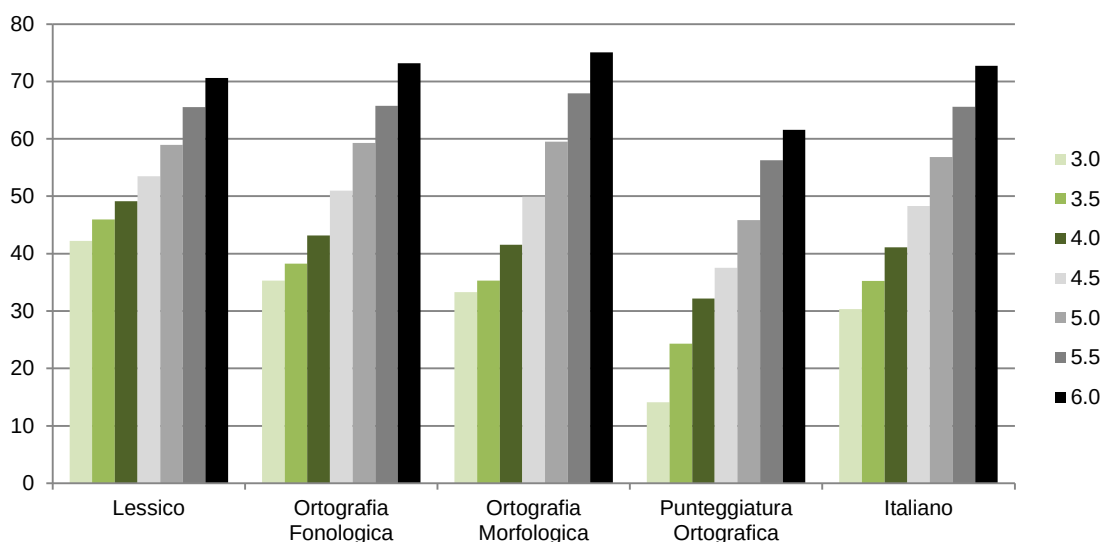
A livello delle singole dimensioni testate, la Tabella 9 mostra una relazione più debole tra lessico e punteggiatura. In altre parole sembra che gli allievi bravi in Ortografia Fonologica diano buone performance anche in Ortografia Morfologica, mentre non è ugualmente detto che chi è bravo in lessico sia bravo anche in Punteggiatura.

Dai dati GAGI risulta che gli allievi che hanno sostenuto le prove a fine anno scolastico hanno ricevuto una valutazione di 3 o 3.5 nel 2.7% dei casi, mentre quasi un terzo ha riportato una nota superiore a 5. Ponendo in relazione la nota ottenuta in Italiano e i risultati delle prove standardizzate si constata che, in tutte le dimensioni, all'aumentare della nota aumenta significativamente il punteggio riportato nella prova. Nonostante il test copra solo una parte del programma di materia dell'anno precedente, c'è dunque coerenza tra la valutazione del docente e i risultati dei test.

Tabella 10: Voto finale in Italiano

Nota italiano	
	Percentuale
3.0	.3
3.5	2.6
4.0	10.1
4.5	24.7
5.0	33.6
5.5	24.0
6.0	4.8
Totale	100.0

Grafico 23: Relazione punteggi nelle dimensioni e nota di fine anno



Anche la prova di matematica mostrava dei valori che correlavano con le note di fine anno di matematica. Il fatto che le prove siano state sviluppate da docenti ed esperti del territorio facendo riferimento ai contenuti insegnati nelle scuole del Cantone ha certamente contribuito a questa coerenza di valutazione.

I risultati alle prove risultano inoltre significativamente (con una probabilità di errore pari all'1%) correlati con le note in tutte le materie e anche i risultati in ogni materia sono correlati con quelle in tutte le altre. Utilizzando come indicatore della prova esclusivamente il fattore aggregato (questo effetto è comunque presente anche nelle singole dimensioni) è possibile osservare che la correlazione tra la nota di Italiano e le altre note è più forte che non tra il risultato alla prova e le altre note. Questi risultati potrebbero essere interpretabili considerando alcune caratteristiche delle quali è difficile valutare l'effetto in modo esatto. In primo luogo occorre ricordare come l'Italiano sia la lingua di insegnamento e come sia strumento di mediazione dei contenuti; una competenza in Italiano potrebbe portare a una facilitazione anche nelle altre discipline. In secondo luogo si deve considerare il cosiddetto "effetto docente", per la maggior parte degli allievi il docente è unico per tutte le discipline e la valutazione dell'allievo può essere influenzata da fattori relazionali. La nota scolastica è infatti prodotta da una serie di fattori concomitanti tra i quali ci sono le competenze disciplinari, un fattore allievo (che può comprendere a sua volta impegno, abilità, motivazione e altri fattori individuali) e anche la relazione docente allievo.

Tabella 11: Correlazioni note di fine anno in tutte le discipline e la nota di fine anno di italiano e il punteggio nella dimensione Italiano generale

	Italiano generale	Nota italiano	Nota ed. fisica	Nota studio dell'ambiente	Nota att. grafico pittoriche	Nota religione	Nota applicazione	Nota francese	Nota condotta	Nota ed. musicale	Nota matematica	Nota att. creative
Italiano generale	1	.667**	.197**	.485**	.327**	.221**	.381**	.472**	.231**	.290**	.549**	.300**
Nota italiano	.667**	1	.262**	.670**	.502**	.240**	.583**	.643**	.352**	.339**	.716**	.388**

** La correlazione è significativa ($p < .01$).

2.6.2 Relazione con le note di fine anno dell'anno successivo

La lingua di apprendimento è usata come mezzo di comunicazione e trasmissione di tutte le discipline, è stato quindi valutato se il risultato alla prova e nella nota di fine anno avessero una relazione con le note di fine anno della classe quarta in tutte le discipline. Sia il risultato alla prova sia la nota di fine anno di Italiano sono correlati significativamente con le note di fine anno dell'anno successivo.

Tabella 12: Correlazioni note di fine anno in tutte le discipline nell'anno successivo alla prova e la nota di fine anno (anno della prova) di italiano e il punteggio nella dimensione Italiano generale

	Nota condotta IV	Nota applicazione IV	Nota italiano IV	Nota matematica IV	Nota studio dell'ambiente IV	Nota francese IV	Nota educazione fisica IV	Nota educazione musicale IV	Nota attività grafiche e pittoriche IV	Nota attività creative IV	Nota educazione religiosa IV
Italiano generale	.197**	.358**	.655**	.530**	.503**	.486**	.185**	.296**	.330**	.273**	.203**
Nota italiano	.291**	.525**	.866**	.659**	.620**	.641**	.250**	.343**	.472**	.372**	.250**

** La correlazione è significativa ($p < .01$).

Ma una correlazione ancora più forte emerge tra i risultati nelle medesime discipline nell'anno della prova e nell'anno successivo. Le valutazioni da un anno con l'altro appaiono quindi abbastanza coerenti questo potrebbe essere sia collegato alla relazione di apprendimento sviluppata tra docente e allievo sia alla presenza di un sapere progressivamente costruito nel quale delle basi solide costruite in un anno possono avere delle influenze con i risultati anche nell'anno successivo.

Tabella 13: Correlazioni note di fine anno (IV) – Note di fine anno (III). In grigio le correlazioni tra la medesima disciplina nei due anni

	Nota condotta IV	Nota applicazione IV	Nota matematica IV	Nota studio dell'ambiente IV	Nota francese IV	Nota educazione fisica IV	Nota educazione musicale IV	Nota attività grafiche e pittoriche IV	Nota attività creative IV	Nota educazione religiosa IV
Nota condotta	.790**	.599**	.269**	.271**	.308**	.231**	.353**	.430**	.433**	.275**
Nota applicazione	.590**	.764**	.533**	.487**	.488**	.309**	.376**	.515**	.487**	.296**
Nota matematica	.214**	.468**	.840**	.622**	.554**	.279**	.242**	.329**	.321**	.204**
Nota studio dell'ambiente	.243**	.467**	.610**	.724**	.558**	.246**	.268**	.354**	.346**	.212**
Nota francese	.244**	.439**	.523**	.539**	.749**	.223**	.291**	.400**	.338**	.191**
Nota ed. fisica	.265**	.308**	.304**	.290**	.257**	.668**	.227**	.260**	.280**	.200**
Nota ed. musicale	.343**	.365**	.236**	.222**	.313**	.210**	.691**	.309**	.355**	.269**
Nota att. grafico pittoriche	.414**	.512**	.358**	.364**	.422**	.214**	.313**	.743**	.541**	.231**
Nota att. creative	.421**	.496**	.367**	.364**	.360**	.306**	.353**	.551**	.774**	.247**
Nota religione	.281**	.282**	.201**	.208**	.194**	.141**	.298**	.207**	.251**	.519**

** La correlazione è significativa ($p < .01$).

3 Sintesi

L'uso delle prove standardizzate in ambito educativo è un uso ormai consolidato sia in ambito nazionale sia in ambito internazionale. Nelle nostre scuole elementari è già stato pubblicato un rapporto rispetto alle prove di matematica in quarta ed è in pubblicazione un ulteriore rapporto sulla matematica in quinta.

A differenza delle prove internazionali quelle utilizzate in questa ricerca sono state create all'interno del sistema educativo ticinese da persone coinvolte nell'insegnamento, nella gestione del sistema e nella formazione dei docenti.

I risultati presentati sono stati organizzati in modo da poter essere confrontabili tra di loro. Tutti i valori nelle quattro dimensioni sono stati calcolati su di una scala da 0 a 100 e così è stato anche per la scala denominata Italiano generale che è stata calcolata come media tra le altre quattro e rappresenta l'insieme dei valori ottenuti da ogni allievo nella prova nel suo insieme.

I risultati sono stati analizzati prima considerando i contesti organizzativi (Circondari e variabili geografiche) poi le variabili relative alla classe quindi le caratteristiche degli allievi e la relazione docente allievo per terminare con la relazione tra le prestazioni alla prova e le note scolastiche.

Osservando in generale le scale è possibile vedere come la scala Punteggiatura ortografica sia stata in media per gli allievi la più difficile e come Lessico sia stata la più facile. I risultati non sono stati omogenei tra i circondari anche se con distribuzioni molto simili tra loro. Non è possibile con le informazioni presenti attribuire delle cause a queste differenze. Entrando nel dettaglio dei comuni, utilizzando la classificazione proposta dall'Ufficio Federale di Statistica di tipologia urbana, è stato possibile rilevare come gli allievi delle scuole localizzate nei comuni rurali abbiano avuto prestazioni complessivamente più modeste.

La dimensione della classe non appare essere una variabile rilevante per la prestazione degli allievi mentre in una delle dimensioni (Punteggiatura) gli allievi delle pluriclasse appaiono ottenere punteggi significativamente migliori. Si deve ricordare come questa modalità di insegnamento stia ottenendo nuova attenzione a livello internazionale in relazione alle possibilità di favorire un insegnamento maggiormente inclusivo.

Considerando le caratteristiche degli allievi è possibile osservare come gli allievi di nazionalità svizzera ottengono punteggi migliori rispetto agli altri allievi e gli allievi di madrelingua italiana ottengono punteggi migliori rispetto agli altri. Tra gli allievi di nazionalità svizzera quelli di madrelingua italiana hanno ottenuto punteggi più elevati rispetto agli altri. Contrariamente a quanto rilevato in altre ricerche non sono state individuate differenze rispetto al gender o al numero di assenze.

Le prestazioni degli allievi appaiono differenziate rispetto all'origine socio economica degli allievi in tutte le dimensioni con un vantaggio per gli allievi di origine socio economica più avvantaggiata. La presenza di ritardo scolastico è stata rilevata come correlata con dei punteggi più bassi in tutte le dimensioni.

Il genere del docente non appare influenzare le prestazioni degli allievi indipendentemente da quello degli allievi stessi. Esiste invece una relazione tra la nota di condotta e i punteggi ottenuti alla prova in tutte le dimensioni.

Le note di fine anno in italiano e i risultati alla prova sono correlati. Questo risultato mostra una stabilità nella valutazione e una coerenza tra la prova e le valutazioni degli insegnanti. I risultati alla prova sono correlati anche con le note in tutte le altre materie sia nell'anno della prova sia nell'anno successivo.

La relazione tra le diverse valutazioni non stupisce in considerazione anche del ciclo di studio nel quale è stata fatta la prova. Nella scuola elementare le materie sono insegnate in classe da un numero limitato di docenti. Questo aumenta l'impatto della dimensione relazionale ovvero di quanto la relazione docente – allievo possa influenzare sia la prestazione sia la valutazione. Meriterebbe un approfondimento di merito il valutare come si componga la nota di fine anno e quanto la nota in ogni materia sia effettivamente generata dalle prestazioni dell'allievo nella materia stessa. Una prima prova svolta sui dati mostra come esistano almeno tre grandi fattori che contribuiscono alla valutazione di fine anno. Da una parte vi è una dimensione trasversale di competenza scolastica che ritroviamo in ogni disciplina, e maggiormente nelle note delle materie ritenute più sfidanti (matematica, francese e italiano), da un'altra vi è l'abilità o competenza specifica alla disciplina (in questo caso valutabile con la prova standardizzata che si caratterizza per essere una prova a-relazionale rispetto al docente) e il terzo fattore è rappresentato dalla relazione docente allievo che contribuisce alla motivazione dell'allievo ma influisce anche sull'atteggiamento valutativo del docente. La nota scolastica è infatti un indice sintetico che racchiude un anno di vita condivisa e

un processo di apprendimento nel quale docenti e allievi sono chiamati, con ruoli naturalmente diversi, a confrontarsi per un bene comune.

Bibliografia

- Achilles, C. A. (1999). *Let's Put Kids First, Finally: Getting Class Size Right*. Thousand Oaks, California: Corwin Press.
- Agbo, S. A. (2007). Addressing school-community relations in a cross-cultural context: A collaborative action to bridge the gap between first nations and the school. *Journal of Research in Rural Education*, 22, 1-14.
- Aina, O. E. (2001). Maximizing Learning in Early Childhood Multiage Classrooms: Child, Teacher, and Parent Perceptions. *Early Childhood Education Journal*, 28(4), 219-224.
- Aina, O. E. (2001). Maximizing Learning in Early Childhood Multiage Classrooms: Child, Teacher, and Parent Perceptions. *Early Childhood Education Journal*, 28(4), 219-224.
<http://doi.org/10.1023/A:1009590724987>
- Ammermüller, A., & Dolton, P. (2006). *Pupil-teacher gender interaction effects on scholastic outcomes in England and the USA*. ZEW Discussion Paper 06-600.
- Blatchford, P., Bassett, P., Goldstein, H., & Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the Institute of Education Class Size Study of children aged 5-7 Years. *British Educational Research Journal*, 29(5), 709-730.
- Blatchford, P., Kutnick, P., Baines, E., & Galton, M. (2003). Toward a social pedagogy of classroom group work. In Blatchford, P., and Kutnick, P. (Ed.) *Special Edition of International Journal of Educational Research*, 39, 153-172.
- Blatchford, P., Moriarty, V., Edmonds, S., & Martin, C. (2002). Relationships between class size and teaching: a multi-method analysis of English infant schools. *American Educational Research Journal*, 39(1), 101-132.
- Bombelli, M.C. (2009). *Alice in business land. Diventare leader rimanendo donne*. Milano: Guerini e Associati.
- Boudon, R. (1974). *Education, Opportunity, and Social Inequality*. New York: Wiley.
- Bourdieu, P., & Passeron, J. C. (1964). *Les Héritiers*, Les éditions de Minuit, Paris (trad. it. 1971 I delfini. Gli studenti e la cultura, Guaraldi, Bologna-Firenze).
- Bourdieu, P., & Passeron, J. C. (1972). *La riproduzione. Sistemi di insegnamento e ordine culturale*. Firenze: Guaraldi Editore. (Original work published 1971).
- Breen, R., Luijkx, R., Müller, W., & Pollak, R. (2009). Non-persistent inequality in educational attainment. Evidence from eight European Countries. *American Journal of Sociology*, 114(5), 1475-1521.
- Broome, J. L. (2009). A Descriptive Study of Multi-Age Art Education in Florida. *Studies in Art Education: A Journal of Issues and Research in Art Education*, 50(2), 167-183. Retrieved from <http://eric.ed.gov/?id=EJ867990>
- Brown, R. (1995). *Prejudice. Its Social Psychology*. Oxford: Blackwell Publishers.
- Chevalier, A., Denny, K., & McMahon, D. (2003). *A multi-country study of inter-generational educational mobility*. ISSC Discussion Paper Series 2003/06. Dublin: University College. Disponibile in: http://irserver.ucd.ie/bitstream/handle/10197/1105/dennyk_workpap_053.pdf;jsessionid=7D3CEC8D789157FDF5BE710C7DFF8212?sequence=1 [15 febbraio 2013].
- CIRSE (2014). *Prove standardizzate di Matematica per la SE*. Locarno: Centro Innovazione e Ricerca sui Sistemi Educativi
- Cooper, H. M. (1989). Does reducing student-to-teacher ratios affect achievement? *Educational Psychologist*, 24(1), 79-98.
- De Graaf, N. D, De Graaf, P., & Kraaykamp, G. (2000). Parental Cultural Capital and Educational Attainment in the Netherlands : A Refinement of the Cultural Capital Perspective. *Sociology of Education*, 73, 92-111.
- Dee Thomas, S. (2007). Teachers and the Gender Gaps in Student Achievement. *Journal of Human Resources*, 42(3), 528-554.

- Driessen, G., Smit, F., & Slegers, P. (2005). Parental Involvement and Educational Achievement. *British Educational Research Journal*, 31(4), 509-532.
- Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Willms, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest*, 2(1).
- Esping-Andersen, G., & Mestres, J. (2003). Ineguaglianza delle opportunità ed eredità sociale. *Stato e Mercato*, 1, 123-152.
- Farkas, S., & Duffett, A. (2008). Results from a National Teacher Survey. In *High-achieving students in the era of NCLB*, 49-82.
- Fosco, A. M., Schleser, R., & Andal, J. (2003). Multiage Programming Effects on Cognitive Developmental Level and Reading Achievement in Early Elementary School Children. *Reading Psychology an International Quarterly*, 25(1), 1-17. Retrieved from <http://eric.ed.gov/?id=EJ682436>
- Fu, D., Hartle, L., Lamme, L. L., Copenhaver, J., Adams, D., Harmon, C., & Reneke, S. (1999). A Comfortable Start for Everyone: The First Week of School in Three Multi-Age (K-2) Classrooms. *Early Childhood Education Journal*, 27(2), 73-80. <http://doi.org/10.1023/A:1026092009745>
- Galton, M. (1998). Class size: a critical comment on the research. *International Journal of Educational Research*, 29, 809-818.
- Glass, G., Cahen, L., Smith, M. L., & Filby, N. (1982). *School Class Size*. Beverley Hills, C.A.: Sage.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). Interpreting Research on School Resources and Student Achievement: A Rejoinder to Hanushek. *Review of Educational Research*, 66(3), 411-416.
- Grolnick, W. S., & Slowiaczek, M. L. (1994). Parents' involvement in children's schooling: A multidimensional conceptualization and motivational model. *Child Development*, 65, 237-252.
- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The Role of Parents and Teachers. *Development of Gender-Related Math Attitudes. Sex Roles*, 66, 153-166.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbache, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8, 1-51.
- Hanushek, E. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *Journal of Human Resources*, 14, 351-388.
- Hanushek, E. (1986). The Economic of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, 24(3), 1141-1177.
- Hanushek, E. (1997). Assessing the Effects of school Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*, 19, 141-164.
- Hanushek, E. (2003). The Failure of Input-Based Schooling Policies. *Economic Journal*, 113, 64-89.
- Hoffman, J. (2002). Flexible Grouping Strategies in the Multiage Classroom. *Theory Into Practice*, 41(1), 47-52. Retrieved from <http://www.jstor.org/stable/1477537>
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, 57, 469-480.
- Kasten, W. C. (1998). Why does multiage make sense? Compelling arguments for educational change. *Primary Voices K-6*, 6(2), 2-9.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H., & Vollmer, H. (2004). *Le développement de standards nationaux de formation. Une expertise*. Bonn : ministère fédéral de l'éducation et de la recherche.
- Kolstad, R., & McFadden, A. (1998). Multiage Classrooms: An Age-Old Educational Strategy Revisited. *Journal of Instructional Psychology*, 25(1), 14-18. Retrieved from <http://eric.ed.gov/?id=EJ565426>
- Krueger, A. B., & Whitmore, D. (2002). Would Smaller Classes Help Close the Black-White Achievement Gap? In J. E. Chubb & T. Loveless (Eds.). *Bridging the Achievement Gap*. Washington: Brookings Institution Press.
- Leuven, E., & Rønning, M. (2011). *Classroom Grade Composition and Pupil Achievement*. IZA Discussion Paper No. 5922. Disponibile in: <http://ftp.iza.org/dp5922.pdf>.

- Leuven, E., & Rønning, M. (1996). Classroom Grade Composition and Pupil Achievement. Retrieved from <http://papers.ssrn.com/abstract=1921973>
- Lloyd, L. (1999). Multi-Age Classes and High Ability Students. *Review of Educational Research*, 69(2), 187–212. <http://doi.org/10.3102/00346543069002187>
- Logue, M. E. (2006). Teachers Observe to Learn: Differences in Social Behavior of Toddlers and Pre-schoolers in Same-Age and Multiage Groupings. *Young Children*, 61(3), 70–76. Retrieved from <http://eric.ed.gov/?id=EJ768997>
- Marc, V., Wirthner, M. (2013). *Développement d'un modèle d'évaluation adapté au PER. Rapport scientifique du projet d'épreuves romandes communes*. Neuchâtel: IRDP.
- Marchesi, A. (2012). Qui ci vuole (ancora) il maschio. Dialogo tra generazioni di educatori. In B. Mapelli, & S. Olivieri Stiozzi (a cura di). *Uomini in educazione*. Rho: Stripes Edizioni (153-177).
- Mason, D. A., & Burns, R. B. (1996). "Simply No Worse and Simply No Better" May Simply Be Wrong: A Critique of Veenman's Conclusion About Multigrade Classes. *Review of Educational Research*, 66(3), 307–322. <http://doi.org/10.3102/00346543066003307>
- Mason, D. A., & Burns, R. B. (1996). Simply no worse, and simply no better may simply be wrong: A critique of Veenman's conclusion about multigrade classes. *Review of Educational Research*, 66, 307-322.
- McIntire, W., Marion, S., & Quaglia, R. (1990). Rural school counselors: Their communities and schools. *The School Counselor*, 37(3), 166-172.
- Miedel, W. T., & Reynolds, A. J. (1999). Parent involvement in early intervention for disadvantaged children: Does it matter? *Journal of School Psychology*, 37, 370-402.
- Miller, B. A. (1988). Teacher Preparation for Rural Schools. Retrieved from <http://eric.ed.gov/?id=ED295772>
- Miller, B. A. (1991). A review of the qualitative research on multigrade instruction. *Journal of Research in Rural Education*, 7(2), 3-12.
- Miller, B. A. (1994). CHILDREN AT THE CENTER. Implementing the Multiage Classroom. (N. R. E. Laboratory, Ed.). Portland: ERIC Clearinghouse on Educational Management. Retrieved from <https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/3237/children.pdf?sequence=1>
- Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, 21(2), 165-177.
- Mons, Nathalie. (2009). *Les effets théoriques et réels de l'évaluation standardisée*. Disponible in: http://eacea.ec.europa.eu/education/eurydice/documents/thematic_reports/111FR.pdf [15 febbraio 2013].
- Mottier Lopez, L. & Crahay, M. (Ed.) (2009). *Evaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*. Bruxelles : De Boek.
- Neugebauer, M., Helbig, M., & Landmann, A. (2011). Unmasking the Myth of the Same-Sex Teacher Advantage. *European Sociological Review*, 27(5), 669-689.
- Ong, W., Allison, J., & Haladyna, T. M. (2000). Student achievement of 3rd-graders in comparable single-age and multiage classrooms. *Journal of Research in Childhood Education*, 14(2), 205-214.
- Owens, J. S., Richerson, L., Murphy, C. E., Jagelewski, A., & Rossi, L. (2007). The parent perspective: Informing the cultural sensitivity of parenting programs in rural communities. *Child Youth Care Forum*, 36, 179-194.
- Prater, D. L., Bermudez, A. B., & Owens, E. (1997). Examining parental involvement in rural, urban, and suburban schools. *Journal of Research in Rural Education*, 13, 72-75.
- Raccolta delle leggi (1990). Legge della scuola 5.1.1.1. 1 febbraio 1990
- Rasch, G. (1961). On General Laws and the Meaning of Measurement in Psychology. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Medicine*, 321--333, University of California Press, Berkeley, Calif.

- Rigoni, B. (2013). *Statistica degli allievi fine anno 2011/12*. Bellinzona: Ufficio del monitoraggio e dello sviluppo scolastico (DECS).
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectatins and student intellectual development*. New York: Holt, Rinehart & Winston.
- Salvisberg, M., & Zampieri, S.(2014). Valutazioni sotto esame. Piste esplorative per un confronto tra PISA e note scolastiche 2009 e 2012. Locarno: Centro Innovazione e Ricerca sui Sistemi Educativi
- Shapson, S. M., Wright, E. N., Eason, G. & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, 17, 144-52.
- Shavit, Y. & Blossfeld, H. P. (1993). *Persistent Inequality: Changing Educational Attainment in Thirteen Countries*. Boulder: Westview Press.
- Skapski, M. K. (1960). Ungraded Primary Reading Program: An Objective Evaluation. *The Elementary School Journal*, 61(1), 41–45. Retrieved from <http://www.jstor.org/stable/999752>
- Slavin, R. E. (1989). Class size and student achievement: small effects of small classes. *Educational Psychologist*, 24(1), 99-110.
- Spelke, E. S. (2005). Sex differences in intrinsic aptitudes for math and science? A critical review. *American Psychologist*, 60, 950-968.
- Sullivan, A. (2007). Cultural Capital, Cultural Knowledge and Ability. *Sociological Research Online*, 12(6). Disponibile in: <http://www.socresonline.org.uk/12/6/1.html>.
- Thélot, C. (2002). Evaluer l'Ecole. *Études*. Tome 397, octobre 2002, 323 – 334.
- Tomasetto, C., Galdi, S., & Cadinu, M. (2012). Quando l'implicito precede l'esplicito: gli stereotipi di genere sulla matematica in bambine e bambini di 6 anni. *Psicologia Sociale*, 7, 169-185.
- Treccani (2016). Enciclopedia. Consultato il 20.06.2016: <http://www.treccani.it/enciclopedia/>
- UNESCO (2006). *EFA Global Monitoring Report 2007: Strong Foundations*. Parigi: UNESCO.
- UNESCO (2007). *EFA Global Monitoring Report 2008. Education for All by 2015: Will we make it?*. Parigi: UNESCO.
- Veenman, S. (1995). Cognitive and Noncognitive Effects of Multigrade and Multi-Age Classes: A Best-Evidence Synthesis. *Review of Educational Research*, 65(4), 319–381.
<http://doi.org/10.3102/00346543065004319>
- Woolley e Grogan-Kaylor, 2006; Woolley e Bowen, 2007
- Zanolla, G. (2014). Why do children differ in mathematical competencies? The experience of a standardized test in the primary school of Canton Ticino, Switzerland. *International Journal of Teaching and Education*, 2 (3), 155-167.

Appendici

Le analisi tramite i modelli di Rasch

Nel 1961, Rasch affermò che la risposta ad un item dipende da due fattori indipendenti: l'abilità del soggetto e la difficoltà intrinseca dell'item. Sulla base di questa riflessione propose un modello di analisi delle risposte che permettesse di misurare queste due dimensioni (Rasch, 1961).

Le caratteristiche del modello lo rendono uno strumento ideale per testare la validità delle scale ordinali (Wright e Linacre, 1989) e, in particolare, ne hanno fatto un riferimento nella validazione delle scale di competenze. Perché il modello possa dare dei riferimenti corretti è necessario che nelle prove gli item siano quanto più possibili monodimensionali, ovvero che venga valutata una e una sola dimensione, e che siano ordinati per difficoltà dal più semplice al più difficile.

Le analisi sul campione della popolazione chiamato a svolgere il pre-test sono state condotte su 1175 allievi della quarta classe di SE. Ogni studente si è confrontato con due fascicoli amministrati a distanza di una settimana.

Ogni studente ha risposto a due fascicoli in modo da essere confrontato con item di tutte le dimensioni previste. Lo stesso fascicolo è stato somministrato a due gruppi ognuno dei quali era composto da un numero variabile di allievi. Il numero esatto è stato determinato dal numero di allievi presenti nella singola classe. Le indicazioni per l'applicazione dei modelli di Rasch prevedono la presenza di un minimo di 200 soggetti per la valutazione della qualità degli item. Le dimensioni testate erano le seguenti:

1. lessico;
2. punteggiatura;
3. ortografia morfologica;
4. ortografia fonologica.

La scelta dell'ortografia è stata dettata dalla considerazione secondo cui HarmoS prescrive che tale competenza sia interamente acquisita dagli allievi entro la fine della terza elementare (in particolare l'ortografia fonologica: corrispondenza fonema-grafema e la punteggiatura: uso dei principali segni di punteggiatura negli impieghi standard). L'acquisizione del lessico, invece, copre tutto il percorso della scuola dell'obbligo.

Per la somministrazione è stato utilizzato un design che è detto *a matrice multipla* in modo che nella singola classe vi fosse una distribuzione equa di allievi che si confrontassero con i diversi fascicoli. Questo si rende necessario per rendere maggiormente aleatorio un campione che, di per sé non lo è, essendo estratto sulla base di un criterio, cioè la classe. Ogni item è stato quindi sottoposto ad un minimo di 270 allievi.

Dopo aver raccolto i fascicoli, le risposte sono state inserite in un database unitario. Ricordiamo come per la valutazione della bontà degli item, fosse rilevante la loro capacità discriminativa e quindi le risposte sono state codificate in corrette o sbagliate (1 e 0), in modo da rendere dicotomica la scala di risposta anche nei casi in cui questa fosse politomica o aperta. La valutazione era infatti relativa alla capacità di fornire la risposta corretta.

Le analisi descrittive per individuare eventuali errori di codifica o inserimento sono state condotte con il software SPSS, mentre le analisi relative alla coerenza delle osservazioni con il modello previsto (il cosiddetto model FIT) sono state realizzate utilizzando il software ConQuest. Il FIT è stato valutato rispetto alle seguenti ipotesi:

- che vi fosse un modello multidimensionale (ovvero che tutti i settori fossero perfettamente indipendenti);
- che vi fosse un modello monodimensionale;
- che vi fosse un modello monodimensionale con sottodomini.

Per ognuno dei modelli si è verificato quanto riuscisse a spiegare la complessità delle informazioni raccolte, i tre modelli sono stati posti a confronto reciproco per individuare il migliore da un punto di vista metrico. Il terzo modello ha mostrato di avere le caratteristiche metriche migliori e quindi è stato scelto di proseguire le analisi in questa direzione.

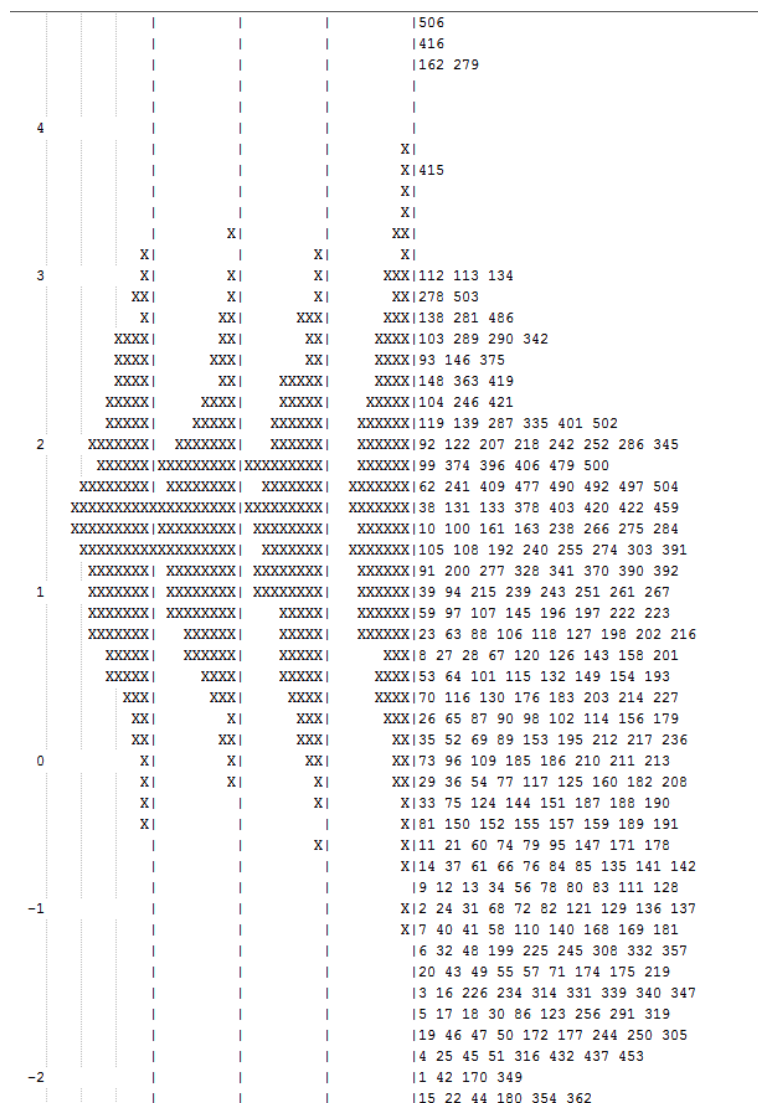
Il secondo passaggio è consistito nella valutazione di ogni singolo item rispetto alle sue capacità metriche e alla sua coerenza con il modello generale. Nella pagina seguente è riportato lo schema nel quale sono ordinati tutti gli item in ordine di difficoltà (rappresentati dai numeri presenti sulla destra) e tutti i soggetti in ordine di abilità rispetto a ogni singola dimensione (ogni X rappresenta 9.1 soggetti), ogni dimensione rappresenta uno dei settori misurati.

Dal grafico è possibile vedere come molti degli item testati fossero facili o molto facili mentre ve ne fossero relativamente pochi difficili o molto difficili.

È stata quindi prodotta una tabella contenente tutti gli item, per ognuno dei quali sono state riportate: la percentuale di allievi che hanno risposto correttamente, la capacità discriminativa dell'item e, nel caso vi fossero, la capacità discriminativa di ogni risposta alternativa e, infine, la coerenza con il modello⁸. Per ogni item è stata inoltre prodotta la curva caratteristica di risposta e la possibilità che vi fossero funzionamenti molto differenziati in relazione al livello degli allievi.

⁸ Per questo si utilizza un parametro detto MNSQ che fornisce informazioni riguardo l'errore nella stima del singolo item, questo errore può essere causato da molti fattori, tra gli altri formulazioni ambigue o presenza di ridondanze nelle domande; il valore atteso per MNSQ è pari ad 1 con una soglia di accettabilità di più o meno 0,3.

Grafico 2: Rapporto difficoltà degli item – numero di soggetti



Esempio di uno dei tipici grafici prodotti dalle analisi fatte sulla base del modello di Rasch. Con le X sono indicati dei raggruppamenti di soggetti ordinati dal più abile al meno abile rispetto ai singoli settori. Sulla destra sono indicati tutti gli item ordinati per difficoltà. Questi grafici permettono una visualizzazione immediata degli item troppo facili o troppo difficili. L'item 362, ad esempio, è eccessivamente facile; in pratica tutti i soggetti rispondono correttamente. L'item 506, al contrario, risulta troppo complesso.

Adeguatezza del modello

L'analisi di dettaglio dei modelli di descrizione delle risposte ha portato a considerare, come riportato precedentemente, il terzo come quello con il maggiore potenziale esplicativo. L'insieme delle risposte si distribuisce su quattro dimensioni con un costrutto latente che potrebbe essere definito "italiano" il quale aggrega i quattro fattori.

A titolo di esempio si riporta il confronto tra il modello unidimensionale con quello multidimensionale dove è possibile vedere come il secondo sia significativamente migliore.

Tabella 1: Confronto modello unidimensionale senza e con sottodomini

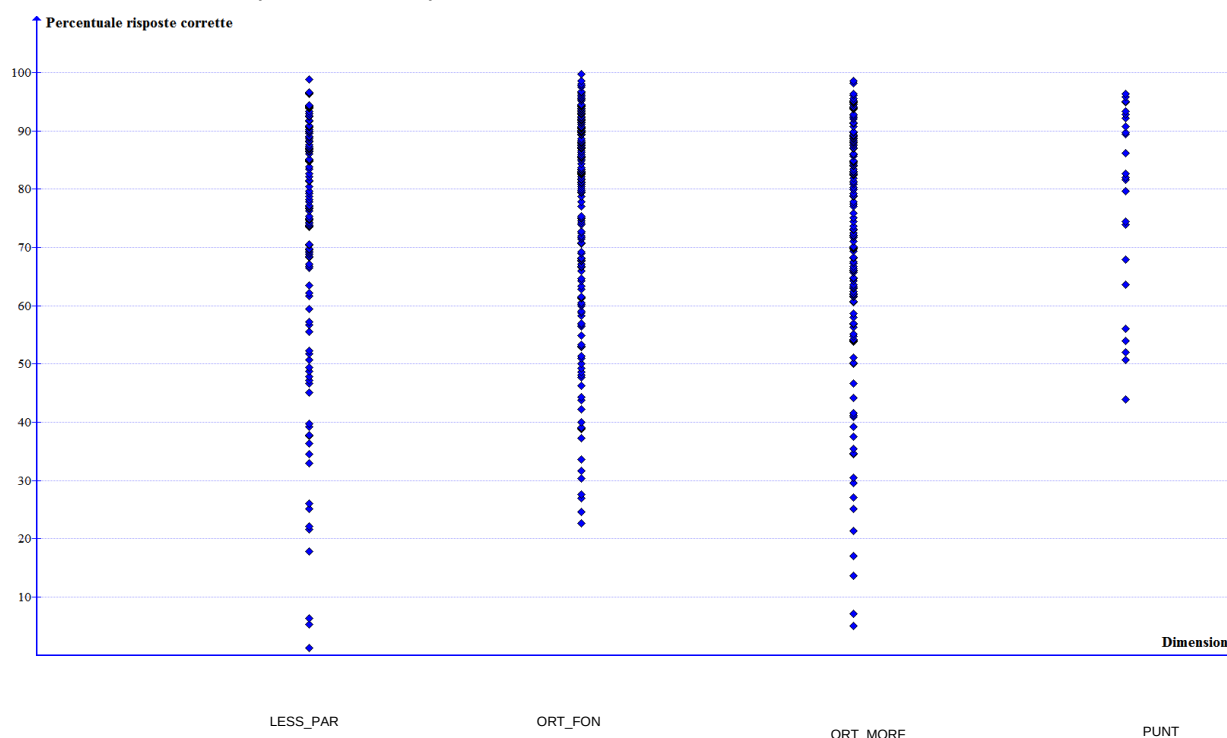
Modelli	Devianza	Numero di parametri/gl	Significatività
Modello unidimensionale	147666.96998	510	
Modello unidimensionale con sotto domini	146423.49965	15	
Differenza/Confronto	1243.47	495	P < 0.001

Criteri di scelta degli item

Per ogni item sono stati considerati sia il numero di non risposte sia quello di risposte corrette o sbagliate. Questo perché nel caso di non risposte è necessario interrogarsi sulle ragioni di questo fenomeno (per esempio: problemi di tempo, di consegna, di eccessiva difficoltà), nel caso di risposte errate si deve invece approfondire rispetto al livello di difficoltà dell'item.

Osservando nel grafico la distribuzione delle percentuali di risposte corrette per item in ogni dimensione è possibile vedere come fossero presenti pochi item difficili o molto difficili, questo soprattutto per la punteggiatura. Le altre dimensioni sono invece distribuite in modo più equilibrato. La presenza di item che coprono tutti i livelli di difficoltà permette di identificare gli allievi per tutti i livelli di competenza, mancando item di difficoltà molto elevata diventa invece difficile identificare gli allievi molto competenti. La presenza di item distribuiti su tutti i livelli di difficoltà ha permesso di costruire una prova equilibrata.

Grafico 3: Percentuale di risposte corrette per item



Un fattore da considerare nello scegliere se tenere o scartare un item è la capacità di discriminare dello stesso. Questa si calcola valutando la correlazione tra l'item e il punteggio grezzo (numero di risposte corrette) del soggetto agli item della dimensione di riferimento. Un coefficiente tra 0,3 e 1 indica che l'item è in grado di differenziare bene tra soggetti più e meno competenti, un coefficiente negativo indicherebbe, invece, che ottengono punteggi più positivi gli allievi deboli che non quelli forti. Questa ultima condizione è sovente indicativa di un item mal formulato nella consegna o nelle opzioni di risposta.

Per valutare la coerenza con il modello di Rasch sono stati calcolati, per ogni item, anche i punteggi di coerenza, detti *infit* (Weighted Mean Square MNSQ). Questi si ottengono considerando il numero di risposte inattese, ovvero incoerenti, rispetto a quelle previste dal modello. Un valore di 1 corrisponde a una coerenza perfetta, un valore minore di 1 indica che l'item discrimina maggiormente del previsto (*over fit*) e un valore maggiore di 1 indica che l'item discrimina meno del previsto (*under-fit*).

I valori consigliati in letteratura si situano tra 0,7 e 1,3. Gli item prodotti hanno tutti valori che si situano in questi parametri tranne 44 che hanno valori *over-fit*. Di ogni item è stata valutata la differenza rispetto al valore previsto allo scopo di esplorarne la significatività.

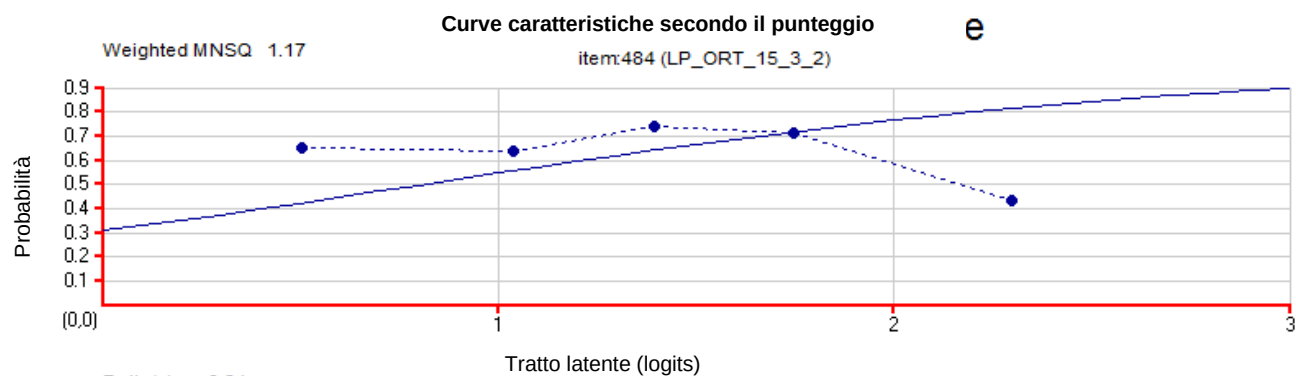
In aggiunta ai valori riportati per ogni item, è stata disegnata la curva caratteristica (*item characteristic curve*, ICC) che permette di osservare direttamente la coerenza tra la distribuzione reale e quella prevista dal modello. Di seguito due esempi (Grafico 4 e 5) nei quali la curva teorica è indicata dalla linea continua e quella realmente osservata è rappresentata invece da quella punteggiata.

Pur non esistendo in letteratura indicazioni completamente convergenti per escludere item sulla base esclusiva delle curve caratteristiche, si ritiene che una forte incoerenza con la distribuzione prevista possa essere un criterio di eliminazione.

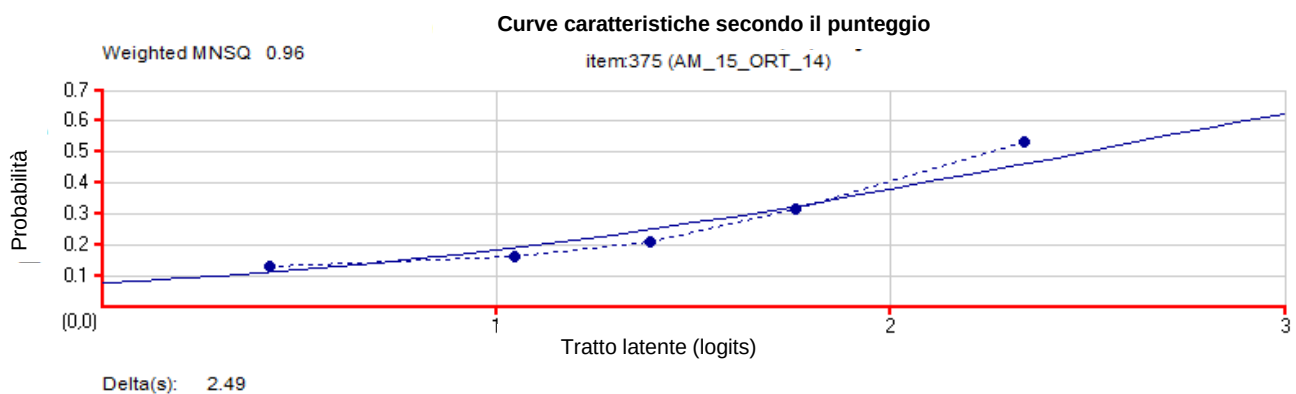
Un ulteriore parametro di scelta è stata la identificazione di differenze eccessive tra la prima e la seconda somministrazione. Come detto, ogni item è stato somministrato a un campione pari di allievi nella prima somministrazione e nella seconda. Tendenzialmente i punteggi sono stati lievemente migliori nella seconda, gli item con punteggi eccessivamente differenziati sono stati ritenuti problematici e quindi esclusi.

Gli item che mostravano capacità metriche adeguate, rispondendo a tutti i parametri e permettendo di mappare in modo completo i livelli di competenza, alla fine del processo di selezione erano quindi 173, di questi sono stati scelti per comporre i fascicoli esclusivamente i 107 che mostravano le capacità migliori e che si distribuivano più coerentemente sulla scala di difficoltà. Gli item complessivamente prodotti e testati inizialmente sono stati 509.

Grafico 4 e 5: Curve caratteristiche di risposta all'item



Tramite questi grafici è possibile osservare se le risposte ad un item (linea punteggiata) sono coerenti con quelle attese (linea continua) in considerazione della sua difficoltà. Ad esempio l'item 484 non si distribuisce in modo adeguato mentre il 375 si distribuisce coerentemente al modello



Le medesime analisi sono state condotte al termine della somministrazione alla popolazione in modo da avere il valore esatto di difficoltà dell'item e usare questo valore per costruire il punteggio finale. Questa procedura ha permesso di avere un punteggio maggiormente preciso che poi è stato riportato su di una scala da 0 a 100 per facilitare la lettura.

Di seguito si riporta la nuova distribuzione degli allievi rispetto alla difficoltà degli item e il rapporto tra difficoltà degli item e prestazione degli allievi. È possibile osservare come item e allievi si trovino distribuiti in modo quasi parallelo.

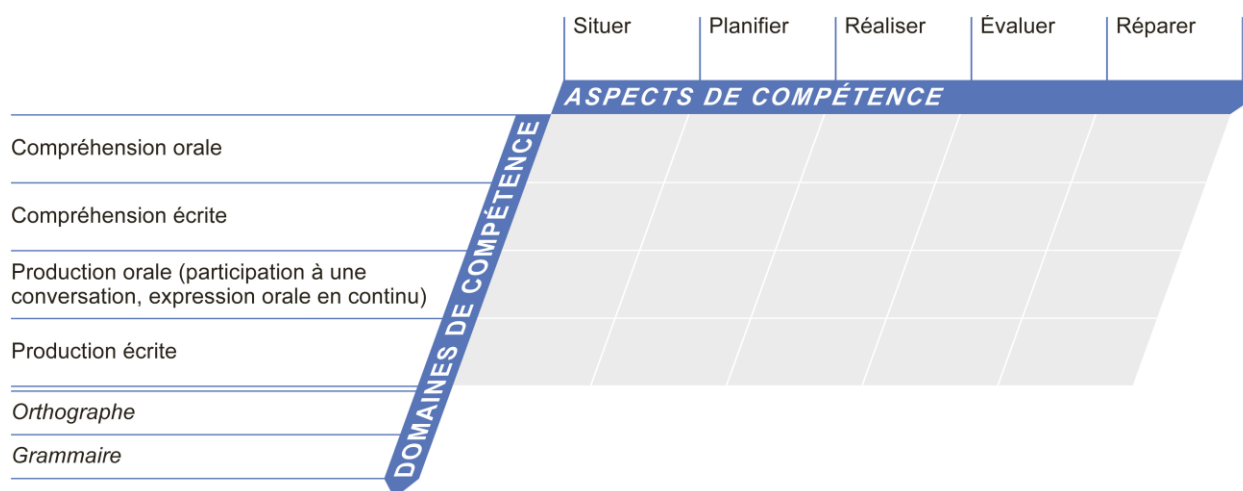
Una conferma ulteriore sulla coerenza interna delle scale è venuta dalla analisi di affidabilità che ha mostrato dei coefficienti (usando l'indice di Kuder-Richardson) per le singole scale elevati

Competenze fondamentali e prove

Le competenze fondamentali per la lingua di scolarizzazione

Il modello di competenza HarmoS per la lingua di scolarizzazione si fonda su una concezione linguistica e didattica della lingua focalizzata sulla capacità di utilizzare la lingua in modo appropriato alla situazione di riferimento (CDPE, 2011), mettendo così al centro dell'attenzione l'azione e la comunicazione. Il modello di competenza (figura 1), si costruisce su due assi principali: gli ambiti di competenza e gli aspetti costitutivi della competenza. I paragrafi che seguono riprendono in modo sintetico questi due livelli.

Grafico 7: Modello di competenza



Ambiti di competenza

Il modello HarmoS relativo alla lingua di scolarizzazione si compone di sei ambiti di competenza: comprensione orale, comprensione scritta, produzione orale, produzione scritta, ortografia, grammatica.

La *comprensione orale* e *scritta* definiscono l'attitudine a comprendere il senso di testi orali/scritti vari, di percepire il senso come espressione di determinati contesti culturali e di farsi una rappresentazione personale del testo.

La *produzione orale* si divide in due grandi campi, da una parte vi è la capacità di partecipare ad una conversazione, dall'altra la capacità di produrre un testo orale di tipo monologo. La partecipazione a una conversazione si descrive attraverso l'attitudine a esprimersi in un linguaggio appropriato, che tenga conto della situazione di riferimento e dell'interlocutore. La produzione di un monologo mette invece in evidenza la capacità di pianificare un discorso, di anticipare le attese di chi ascolta e di saperne cogliere le sensazioni.

La *produzione scritta* descrive l'attitudine a redigere testi di vario tipo tenendo conto delle norme formali, del pubblico di riferimento e del contesto nel quale si scrive. La produzione scritta include anche la dimensione estetica.

Dal punto di vista dell'*ortografia*, il modello di competenza mira due obiettivi principali: fare in modo che lo scolaro scriva un testo con il minor numero possibile di errori; fare in modo che lo scolaro sia in grado di individuare in un proprio testo gli errori commessi. La competenza in *grammatica* si esprime invece attraverso un uso della lingua che ne rispetti le regole costitutive (fonologia, morfologia, sintassi, semantica), ma anche attraverso l'identificazione e la comprensione di queste regole che molte volte sono unicamente utilizzate in modo implicito.

Una dimensione propria all'uso della lingua è assente dagli ambiti di competenza del modello HarmoS: la *percezione dell'estetica della lingua*. La CDPE ne evoca comunque la presenza affermando che, anche se non è considerata una competenza fondamentale, la percezione dell'estetica della lingua definisce la

capacità di percepire una lingua come espressione della sensibilità e della creatività umane. In questo senso l'uso della lingua va oltre agli aspetti meramente comunicativi.

Aspetti costitutivi dell'uso della lingua e elaborazione della prova scritta

Il modello identifica, oltre agli ambiti di competenza, cinque aspetti costitutivi dell'uso della lingua (aspetti di competenza). Essi sono in sintesi:

Situare: ogni attività linguistica si orienta rispetto ad un contesto e a degli obiettivi che sono più o meno coscienti. Data la finalità comunicativa della lingua, risulta fondamentale capire il contesto e gli obiettivi nei quali si situa l'atto linguistico.

Pianificare: procedura che consiste nello scegliere i diversi elementi di contenuto finalizzati alla produzione del testo orale o scritto.

Realizzare: procedura che si descrive normalmente con i termini "ascoltare", "leggere", "parlare", "scrivere".

Valutare: procedura attraverso la quale si prendono le distanze per rapporto al produzione effettuata in modo da poterne misurare la qualità e il raggiungimento degli obiettivi prefissati.

Riparare: procedura che consiste in una seconda o terza realizzazione che interviene dopo aver messo in evidenza i problemi e gli errori che caratterizzavano il primo prodotto.

Il gruppo di lavoro composto da docenti della disciplina coinvolto nell'elaborazione della prova scritta ha proceduto tenendo in considerazione quattro settori di conoscenza che sono trasversali agli ambiti di competenza sopra elencati. Si tratta precisamente di: lessico, punteggiatura, ortografia morfologica e ortografia fonologica. Così facendo si è quindi messo l'accento sulle competenze proprie agli ambiti dell'ortografia e della grammatica. La prova somministrata agli allievi comprende in totale 107 item di cui: 28 esercizi attinenti il lessico, 19 attinenti la punteggiatura, 34 attinenti l'ortografia morfologica e 26 l'ortografia fonologica. L'insieme delle risposte alle domande ha costituito una misura generale chiamata Italiano Generale. Per ciascun settore gli esercizi sono stati scelti in modo da garantire la completezza dell'intera scala di difficoltà.

Repubblica e Cantone Ticino
Dipartimento dell'educazione, della cultura
e dello sport

Scuola universitaria professionale
della Svizzera italiana

Prove standardizzate ticinesi
Italiano nella classe III Scuola Elementare

Quaderni di ricerca – n. 20

Scuola universitaria professionale della Svizzera italiana
Dipartimento formazione e apprendimento
Centro innovazione e ricerca sui sistemi educativi

Piazza San Francesco 19
6600 Locarno
www.supsi.ch/dfa

ISBN
978-88-941240-8-8

