

SUPSI

Quaderni di ricerca

Prove standardizzate ticinesi

Matematica nella classe V Scuola Elementare

Alberto Crescentini



Proposta di citazione:

Crescentini, A. (2017). *Prove standardizzate ticinesi. Matematica nella classe V 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 978-88-941240-7-1

Responsabilità del progetto: Alberto Crescentini

Ricercatori coinvolti: Loredana Addimando, Francesco Galetta, Manuela Lagattolla, Miriam Salvisberg e
Giovanna Zanolla

Impaginazione: Selene Dioli

Ringraziamenti

Il lavoro presente ha richiesto il contributo di numerose persone con ruoli e responsabilità differenti. Elen-carli tutti sarebbe difficile ma senza la partecipazione di tutti si sarebbe compiere il lavoro.

Innanzitutto è doveroso ringraziare tutti i docenti di Scuola Media che hanno permesso che la prova fosse testata nella fase di sviluppo e quindi quelli di Scuola Elementare nelle classi dei quali è stata svolta una volta definitivamente approntata. La lista sarebbe eccessivamente lunga ma li ricordiamo tutti e ad essi aggiungiamo I somministratori che hanno gestito la somministrazione della prova sul territorio e i collaboratori che hanno inserito i dati di tutti i questionari.

Molte persone sono state coinvolte per la costruzione delle prove. In particolare alla scelta dei traguardi hanno contribuito: Silvia Sbaragli e Aldo Frapolli. Allo sviluppo degli item: Alberto Battaini, Giovanna Cor-rent Lepori, Francesco Dagani, Rossana Falcade, Guido Gottardi, Michela Rapp, Rodolfo Roncoroni. Il coordinamento delle attività del gruppo di costruzione degli item è stato garantito da Loredana Addiman-do e Miriam Salvisberg mentre la supervisione disciplinare è stata curata da: Tito Franchi, Aldo Frapolli, Silvia Sbaragli, Oliver Villa. Per le analisi finalizzate alla selezione degli item migliori e alla determinazio-ne dei punteggi nei diversi settori della matematica: Miriam Salvisberg e Giovanna Zanolla. A Francesco Galetta per il contributo che ha fornito al gruppo di lavoro durante molta parte del processo.

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	18
2.3	La classe	19
2.3.1	Dimensione della classe o numero di allievi	19
2.3.2	La pluriclasse	20
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	26
2.4.4	Origine sociale	29
2.4.5	Età dell'allievo	31
2.5	Docente e allievo	34
2.5.1	Il genere del docente	34
2.5.2	Nota di condotta	37
2.6	Risultati al test e note	38
2.6.1	Relazione nota di Matematica di fine anno – prestazione alla prova	38
3	Sintesi	43
	Appendici	45
	Le analisi tramite i modelli di Rasch	45
	Competenze fondamentali e prove	51

1 Il progetto

Il presente rapporto fa seguito a due rapporti analoghi prodotti il primo sulla base dei risultati alle prove di matematica in IV elementare (Cirse, 2014) e il secondo alle prove di italiano in III elementare (Cirse, 2016).

Per quanto riguarda questo progetto nel suo insieme si è svolto negli anni scolastici dal 2013 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 Matematica e la struttura del rapporto riprendono quelli degli altri rapporti. La presentazione e le dimensioni teoriche sono molto simili e si rimanda per gli approfondimenti del caso ai documenti già esistenti; 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 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 Scuole Comunali (USC) e con il Dipartimento Formazione e Apprendimento (DFA).

Grafico 1: Calendario di svolgimento del progetto

ATTIVITA'	feb. - mag. 13	giu. - ago. 13	set. 13 - mag. 14	giu. - ott. 14	nov. 14	dic. 14	gen. 15	feb. 15	mar. 15	apr. 15	mag. 15	giu. 15	lug. 15	ago. 15	set. 15	ott. 15	nov. 15	dic. 15	
Definizione traguardi da testare	■																		
Identificazione gruppo di sviluppo item e gruppo di supervisione		■																	
Creazione item inclusa codifica			■																
Organizzazione e preparazione materiali				■															
Somministrazione pretest					■														
Codifica						■	■												
Ripresa dati						■	■												
Analisi psicometriche							■	■											
Costruzione fascicoli e stampa								■	■										
Organizzazione e somministrazione											■								
Ripresa dati												■	■	■					
Produzione reportini													■	■	■	■			
Stesura rapporto																	■	■	■

La prima fase del progetto, sino ad agosto 2013, 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 fondamentale in quanto sarebbe estremamente difficoltoso valutare contemporaneamente tutte le competenze presenti in una disciplina in modo esatto. Il modello di analisi che è stato utilizzato è legato alla Item Response Theory (IRT) ed è 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 al Piano di studio della scuola dell'obbligo (Divisione della scuola, 2015), sei Traguardi di competenza: Grandezze e misure – Eseguire e applicare, Grandezze e misure – Sapere, riconoscere e descrivere, Numeri e calcolo – Matematizzare e modellizzare, Grandezze e misure – Matematizzare e modellizzare, Numeri e calcolo – Sapere, riconoscere e descrivere, Numeri e calcolo – Eseguire e applicare. L'insieme delle risposte alle domande ha costituito una misura generale chiamata Matematica Generale.

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. Il lavoro in questa fase è stato svolto in stretta collaborazione con gli esperti di Didattica della Matematica operanti nel DECS e nel DFA. 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, 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 10 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*" (Flukiger, 2004) 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 90 esercizi con buone capacità metriche divisi nei 6 settori. Di seguito vedremo un sunto delle analisi effettuate.

Dagli esercizi è stato possibile realizzare due fascicoli, la prova complessivamente comprendeva 90 item, 15 per ogni traguardo di competenza. 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 isti-

tuto 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 medio¹ rispetto al circondario e all'insieme della popolazione testata. Analogamente ai direttori e agli 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.

¹ 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

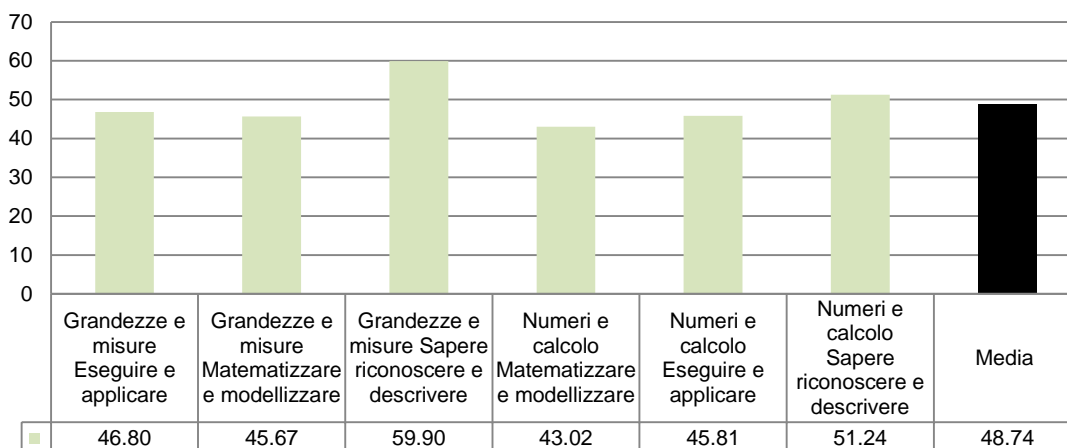
Per organizzare i contenuti del rapporto si procederà con una logica a imbuto. Inizialmente saranno presentate le analisi più generali poi si scenderà progressivamente nel dettaglio. Da un punto di vista di informazioni rispetto al sistema dopo quelle generali saranno considerati i circondari, quindi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

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 (USCO) 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

Per avere una visione di insieme sono stati inizialmente analizzati i punteggi ottenuti nei singoli traguardi e la loro media. Questo permette un rapido raffronto tra le differenti parti che compongono la prova. Il punteggio nel traguardo è stato calcolato definendo per ogni item la difficoltà sul traguardo specifico e quindi sommandone i valori. Al fine di ottenere un valore che tenesse conto della difficoltà dei singoli item rispetto alla prova nel suo insieme è stato quindi fatto il medesimo calcolo tenendo però conto non più solo dei singoli traguardi ma della difficoltà complessiva questo valore è stato definito "Matematica generale".

Grafico 2: Punteggi medi degli allievi nelle diverse dimensioni



Il traguardo che ha raggiunto il valore medio più elevato è "Grandezze e misure Sapere, riconoscere e descrivere" mentre quello con il valore più basso è "Numeri e calcolo Matematizzare e modellizzare". Si può notare come, all'interno dei due Ambiti di competenza scelti (Grandezze e misure e Numeri e calcolo) gli Aspetti di competenza relativi a Matematizzare e modellizzare risultino quelli appartenemente più difficili e quelli relativi a Sapere riconoscere e descrivere risultino i due con i risultati più elevati.

Tutti i punteggi sono stati normalizzati² 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 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.

Il valore "Matematica generale" è quindi quella variabile costituita dalla somma dei valori di difficoltà degli item ai quali l'allievo ha risposto poi normalizzata rispetto a una scala da 0 a 100. Questo valore è stato calcolato per ogni allievo e può essere considerato come un indicatore sintetico della sua prestazione (Crescentini, Salvisberg, & Zanolla, 2014) alla prova. Il sistema con il quale è stato calcolato il valore di Matematica Generale è, rispetto ai precedenti rapporti, più raffinato e permette di valutare meglio la competenza generale. Non si tratta infatti della media dei valori degli altri traguardi ma è stato ricalcolato rispetto alla difficoltà di tutti gli item, il valore medio per tutta la popolazione è pari a 62.52.

Tabella 1: Correlazioni tra le dimensioni valutate con la prova

	Grandezze e misure Eseguire e applicare	Grandezze e misure Matematizzare e modellizzare	Grandezze e misure Sapere, riconoscere e descrivere	Numeri e calcolo Matematizzare e modellizzare	Numeri e calcolo Eseguire e applicare	Numeri e calcolo Sapere, riconoscere e descrivere	Matematica generale
Grandezze e misure Eseguire e applicare	1						
Grandezze e misure Matematizzare e modellizzare	.676**	1					
Grandezze e misure Sapere, riconoscere e descrivere	.637**	.566**	1				
Numeri e calcolo Matematizzare e modellizzare	.726**	.685**	.643**	1			
Numeri e calcolo Eseguire e applicare	.655**	.689**	.557**	.682**	1		
Numeri e calcolo Sapere, riconoscere e descrivere	.672**	.679**	.610**	.681**	.707**	1	
Matematica generale	.861**	.847**	.781**	.871**	.849**	.858**	1

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

² 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.

La tabella precedente mostra come i risultati nelle diverse dimensioni siano tra loro correlati. Le correlazioni tra loro esistenti sono forti. Questo risultato non è inatteso in quanto si tratta di dimensioni parte del costruito più generale valutato che corrisponde alla matematica in generale. Questa è una giustificazione aggiuntiva alla costruzione della dimensione Matematica generale che rappresentasse il punteggio globale del singolo allievo alla prova.

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%.

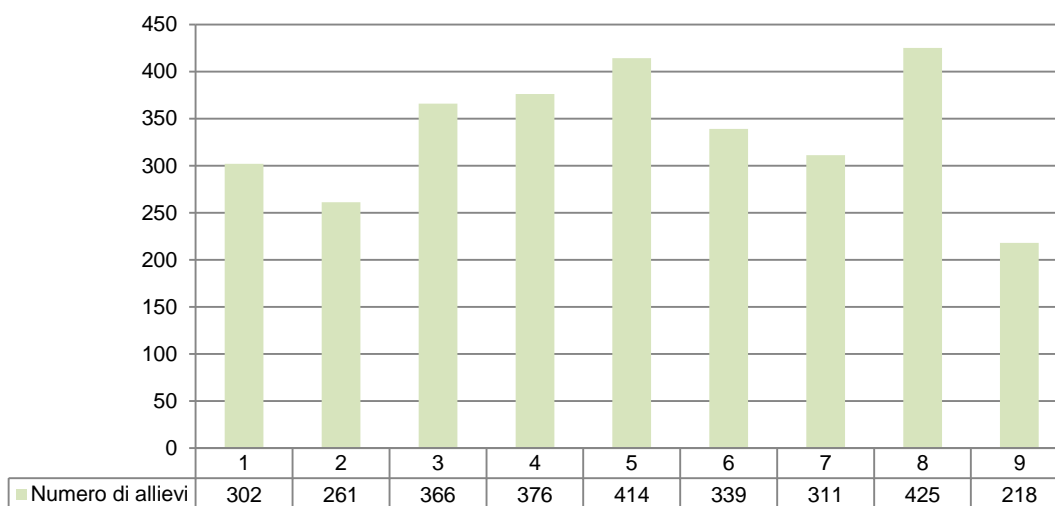
Il quadro generale delle analisi di seguito riportate rispecchia in gran parte il percorso di indagine prodotto nei rapporti precedenti (Crescentini et al., 2014) che ha riguardato le prove standardizzate di matematica e di italiano. Analogamente a quanto fatto allora, in questa sede verranno presentate le analisi descrittive generali dei punteggi ottenuti dagli allievi e le analisi di approfondimento riguardanti i circondari, le classi e gli allievi.

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 nostro territorio per quanto riguarda le scuole comunali 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 193 classi, tra pubbliche (95%) e private (5%), di queste 54 sono 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

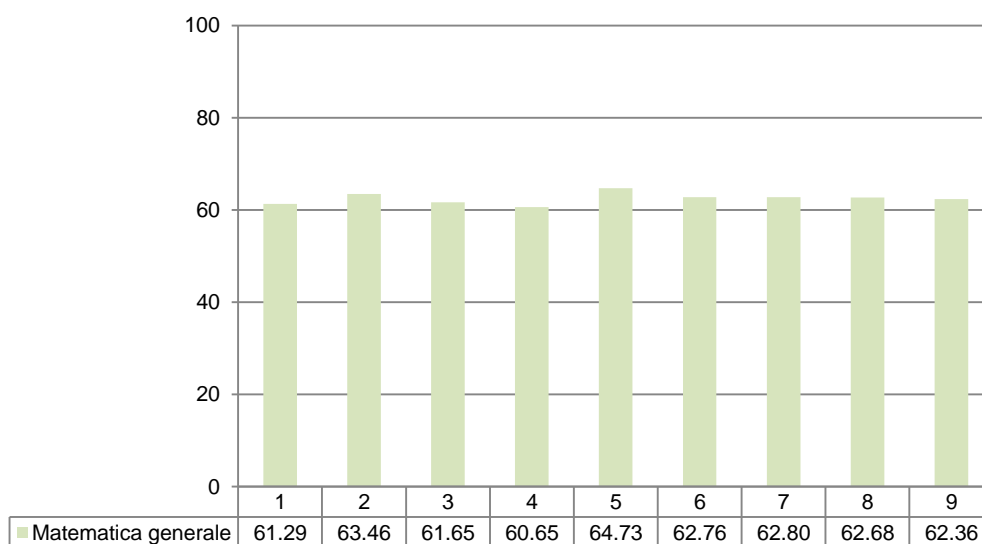


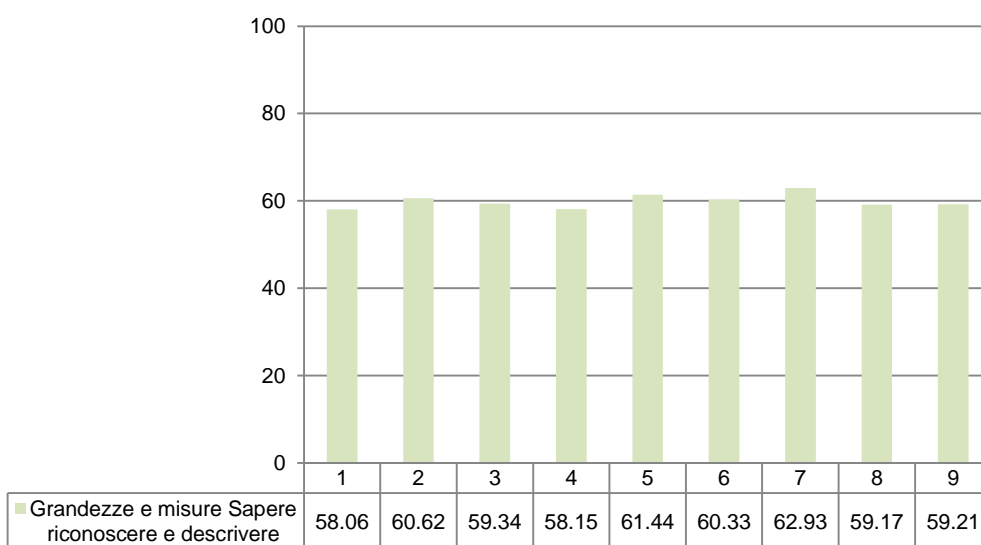
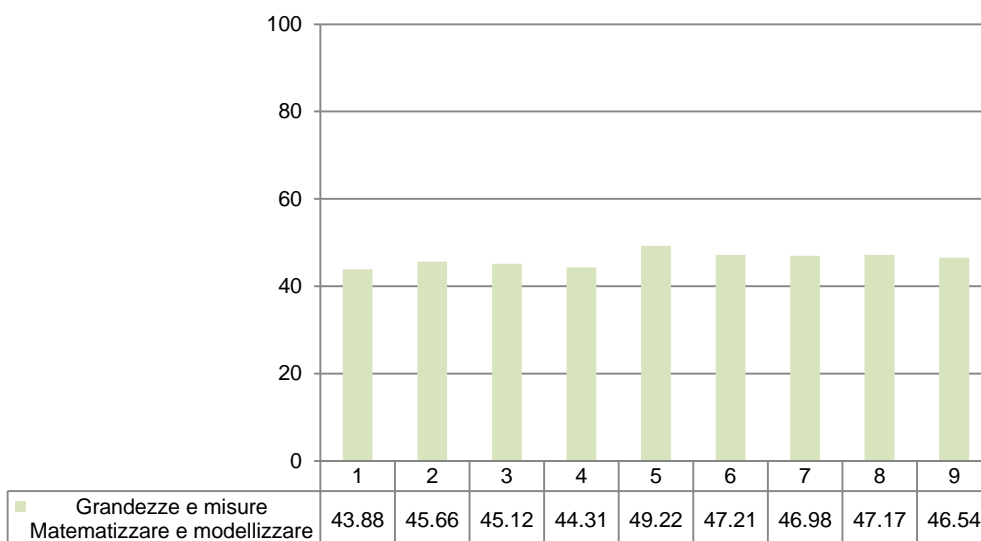
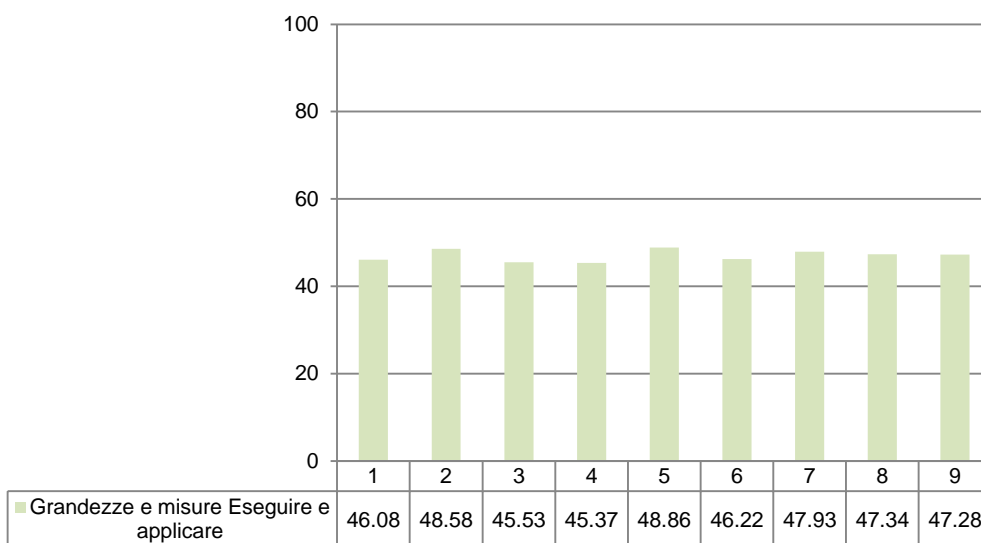
Di seguito (Tabella 2) i punteggi medi nei 9 circondari per la variabile per le diverse sottodimensioni analizzate. Le differenze tra i circondari risultano significative in tutte le dimensioni analizzando l'insieme dei dati tramite l'analisi della varianza. Non è possibile attribuire delle cause specifiche 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).

Tabella 2: Punteggi medi delle dimensioni nei diversi circondari

	Grandezze e misure Eseguire e applicare	Grandezze e misure Matematizzare e modellizzare	Grandezze e misure Sapere, riconoscere e descrivere	Numeri e calcolo Matematizzare e modellizzare	Numeri e calcolo Eseguire e applicare	Numeri e calcolo Sapere, riconoscere e descrivere	Matematica generale
1	46.08	43.88	58.06	42.31	44.40	50.12	61.29
2	48.58	45.30	60.62	43.81	46.27	52.57	63.46
3	45.27	43.86	59.34	42.67	44.93	49.63	61.65
4	44.63	43.32	58.15	40.35	43.08	49.79	60.65
5	48.74	48.48	61.44	46.52	47.63	54.03	64.73
6	45.94	46.78	60.33	42.28	47.19	51.41	62.76
7	47.77	46.05	62.73	43.46	46.12	51.71	62.80
8	47.22	46.72	59.17	42.87	46.64	50.81	62.68
9	47.28	45.89	59.21	42.43	45.72	51.01	62.36

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





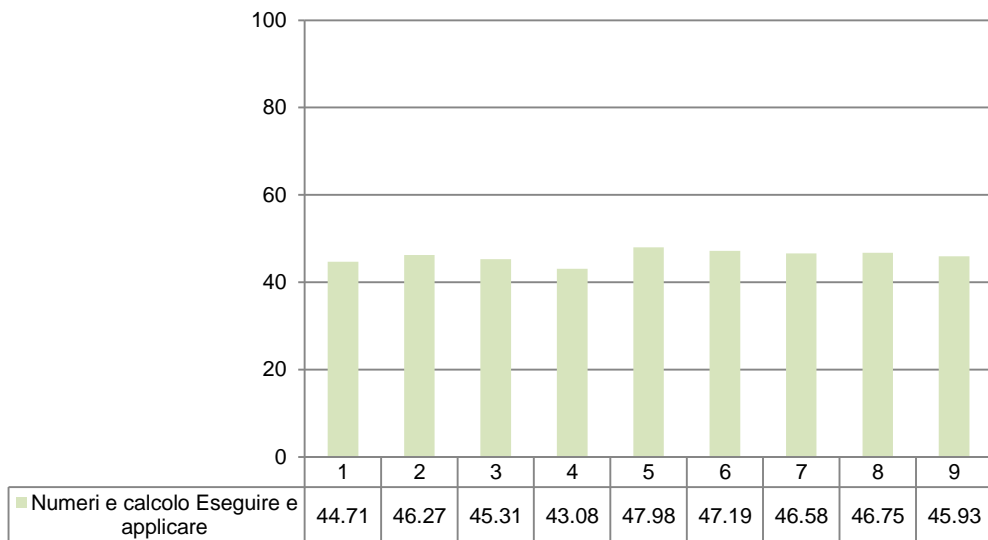
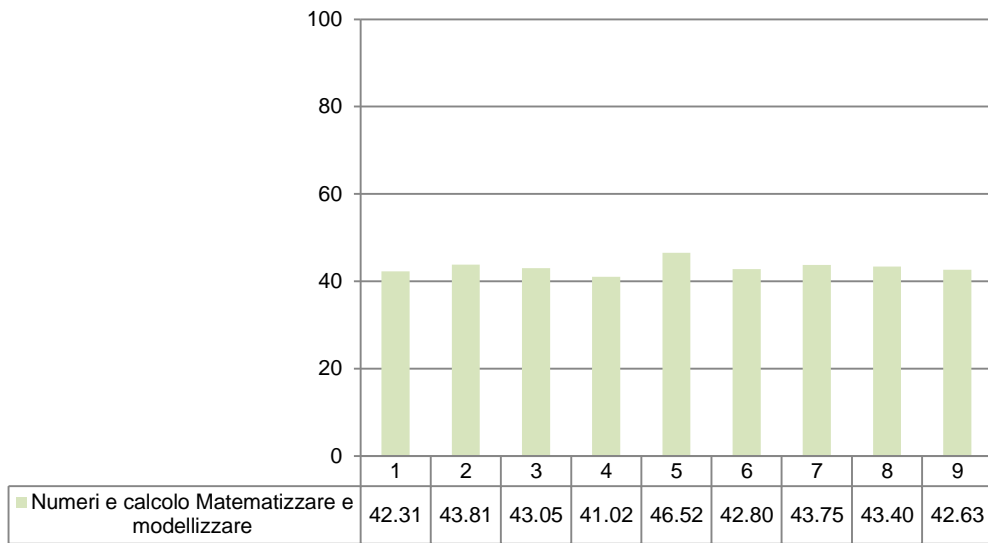
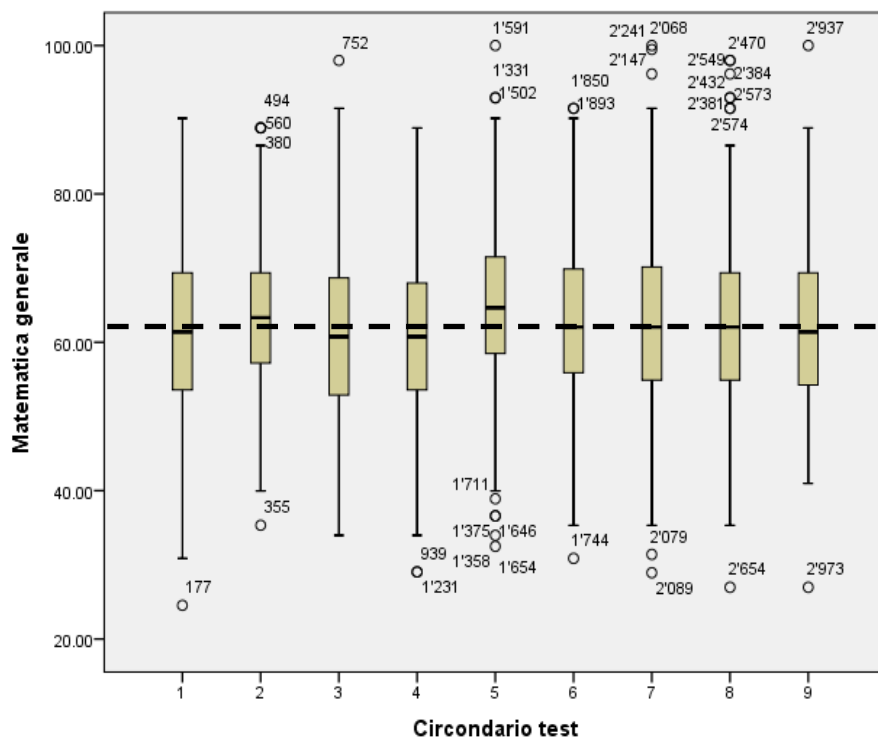


Grafico 4: Confronto delle mediane di Matematica generale nei circondari



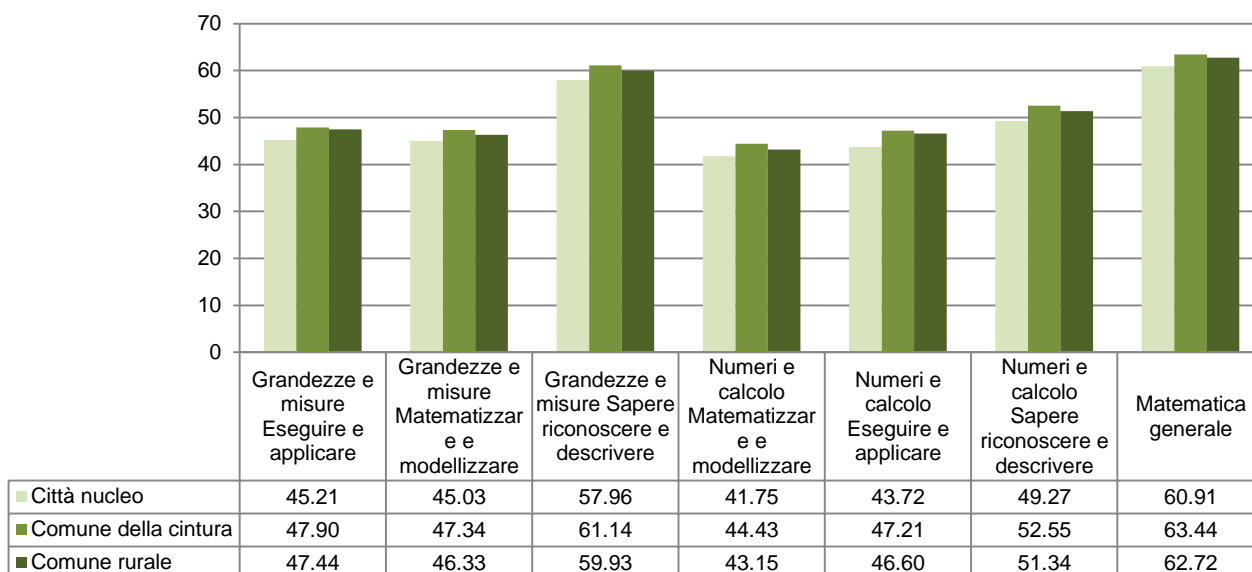
Nel Grafico 4 è possibile vedere come le distribuzioni delle risposte dei diversi circondari siano tra loro molto simili oltre ad esservi una mediana analoga. In questa tipologia di grafico detta “Box-plot” la mediana è indicata dalle barre orizzontali, il 25% degli allievi è nel rettangolo superiore e il 25% in quello inferiore, le due barre verticali rappresentano gli altri allievi e i cerchi in alto e in basso rappresentano gli allievi con punteggi molto più elevati o molto più bassi rispetto alla popolazione di riferimento.

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).

I diversi comuni in cui sono localizzate le scuole elementari ticinesi sono stati raggruppati per tipologia urbana³ e classificati pertanto in città-nucleo (27,7% degli allievi), comuni della cintura (61,8% degli allievi) e comuni rurali (10,5% degli allievi). Il livello di urbanizzazione e la prestazione appaiono correlati in tutte le dimensioni soprattutto per quanto riguarda le città nucleo che hanno le prestazioni meno elevate. La differenza fra i comuni della cintura (gruppo con prestazione più elevata) e le città nucleo (gruppo con prestazione più limitata) è significativa (con un errore potenziale dell'1%) per tutte le dimensioni testate.

Grafico 5: Prestazioni alla prova e grado di urbanizzazione



³ 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

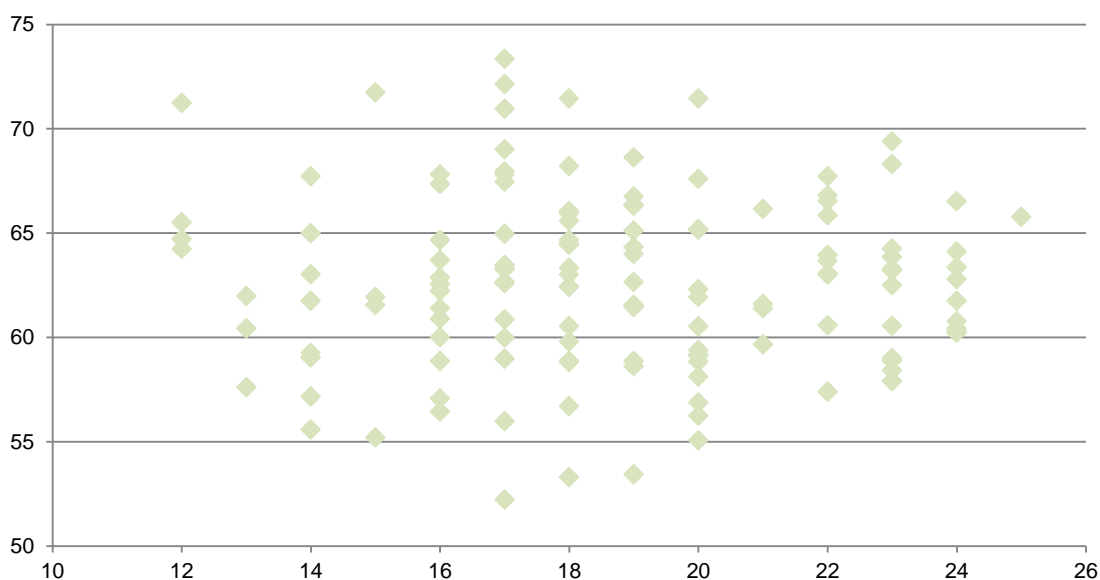
Nel valutare le prestazioni degli allievi le caratteristiche della classe sono ritenute estremamente rilevanti (Dumay e Dupriez, 2009). In coerenza con quanto considerato nei precedenti rapporti sono state considerate: numero di allievi per classe e la presenza di mono o pluriclassi.

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 & 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 quinta 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 Matematica non appaiono quindi collegati 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 così come anche nella prova di italiano.

2.3.2 La pluriclasse

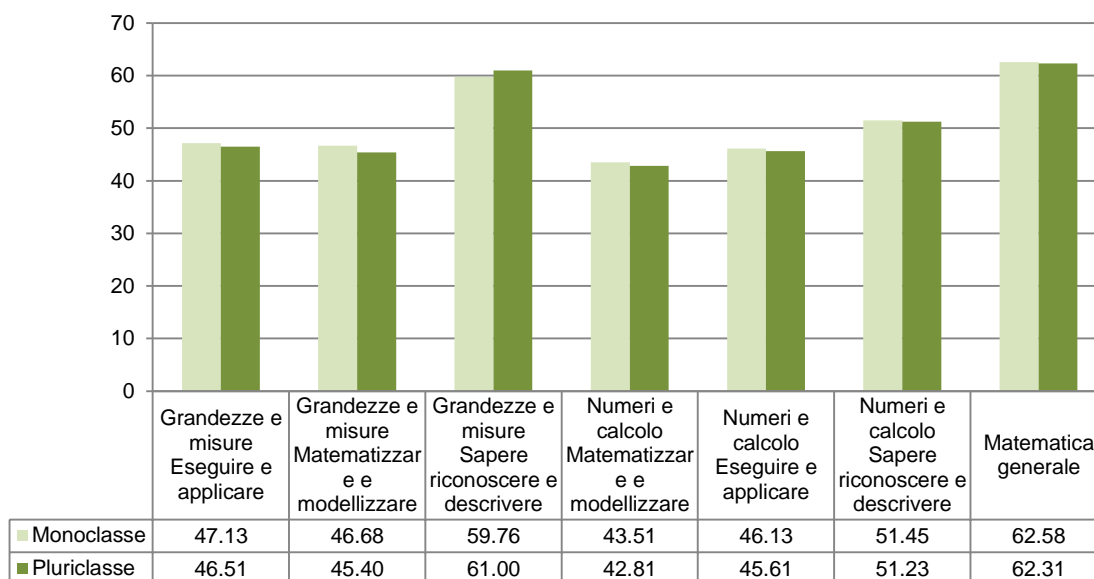
La pluriclasse è una classe formata da bambini che frequentano anni di corso differenti. Si tratta di una particolare organizzazione di classe della scuola primaria diffusa soprattutto nelle zone a bassa densità abitativa, laddove non si raggiunge il numero minimo di iscritti necessari per la formazione 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 20% della popolazione mentre il restante frequenta monoclasse. Nel nostro caso le differenze nei punteggi medi delle dimensioni tra monoclasse e pluriclassi non sono significative in nessuna dimensione.

Grafico 7: Media valore nelle dimensioni nelle mono e pluri classi



Coerentemente con la letteratura le prestazioni non appaiono avere grandi differenze confrontando 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 precedente e anche in quelle di italiano erano risultate delle differenze significative a favore degli allievi delle pluriclassi.

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 abbiano 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 8: Percentuale di allievi divisi per nazionalità

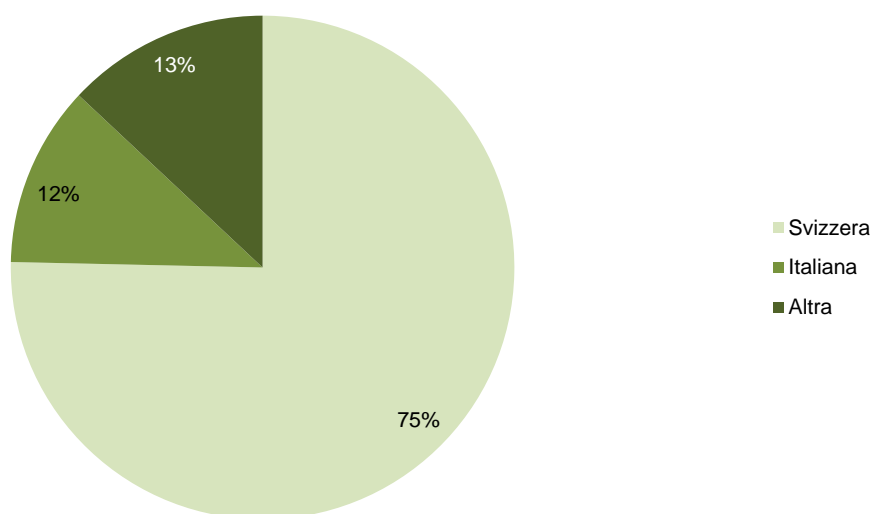
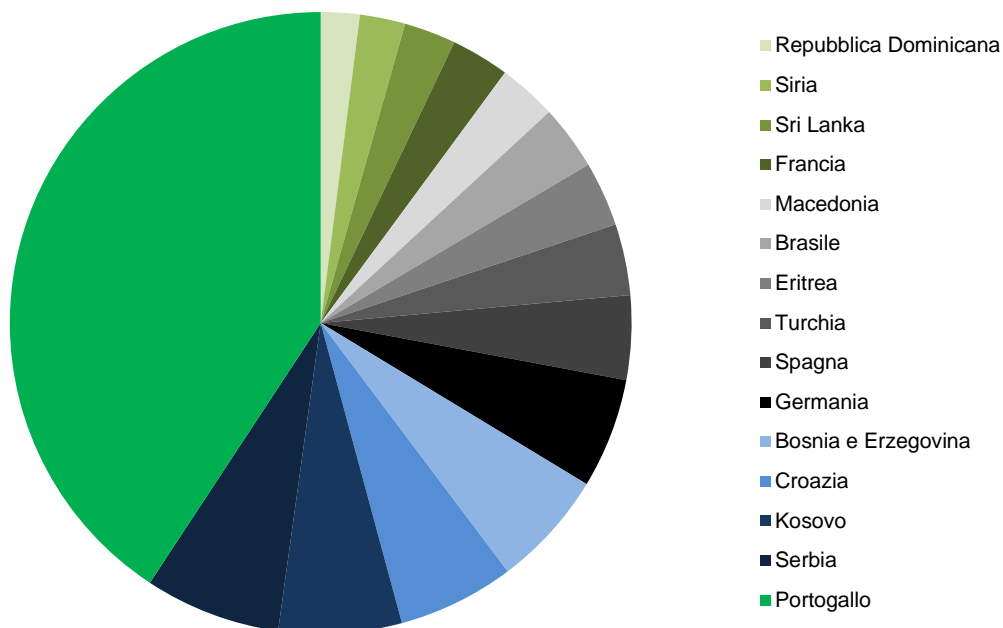


Grafico 9: Percentuale di allievi con nazionalità diversa da Svizzera o Italiana



La maggior parte degli allievi è di nazionalità svizzera (Grafico 13), con una percentuale consistente di allievi provenienti dall'Italia (12%) o da altri Paesi (13%). I Paesi più frequentemente rappresentati sono mostrati in dettaglio nel Grafico 9. Sono stati esclusi i Paesi con meno di 5 allievi.

2.4.2 La Nazionalità e la lingua madre

Il grafico 10 mostra la distribuzione delle medie dei punteggi nelle diverse dimensioni suddivise per nazionalità dell'allievo e il grafico 11 suddivise per lingua madre. Le differenze sono significative (al 99% di probabilità) sempre per quanto riguarda gli allievi di nazionalità Svizzera rispetto agli altri mentre si tratta di differenze significative tra gli allievi di nazionalità Italiana e gli allievi di altra nazionalità nei casi di "Grandezze e misure Sapere, riconoscere e descrivere" e "Numeri e calcolo Matematizzare e modellizzare".

Per i due traguardi che hanno differenze significative rispetto ai tre gruppi si riportano i grafici a scatola che permettono di osservare come vi siano differenze rilevanti nelle modalità con le quali si distribuiscono i punteggi nei tre gruppi. L'insieme dei punteggi degli allievi di nazionalità non Svizzera e non Italiana si caratterizza infatti in un caso per un punteggio mediano che si scosta in maniera evidente dagli altri e nel secondo per uno schiacciamento dei punteggi verso valori più bassi.

Grafico 10: Medie dei punteggi degli allievi suddivisi per nazionalità

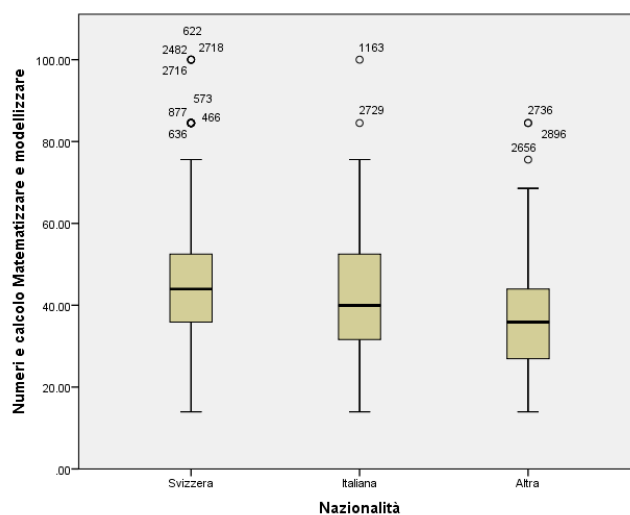
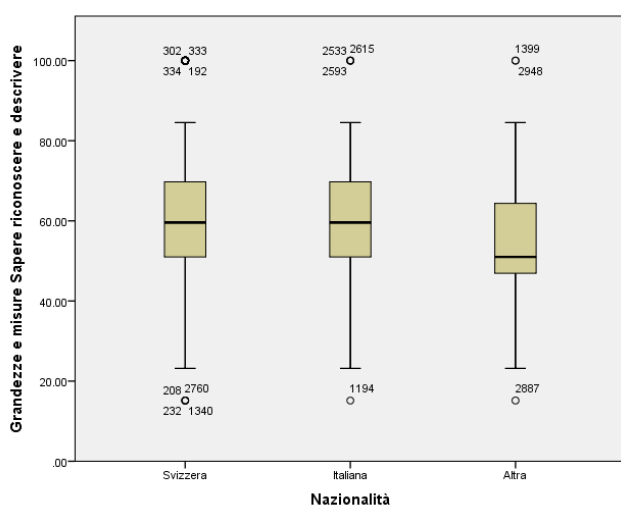
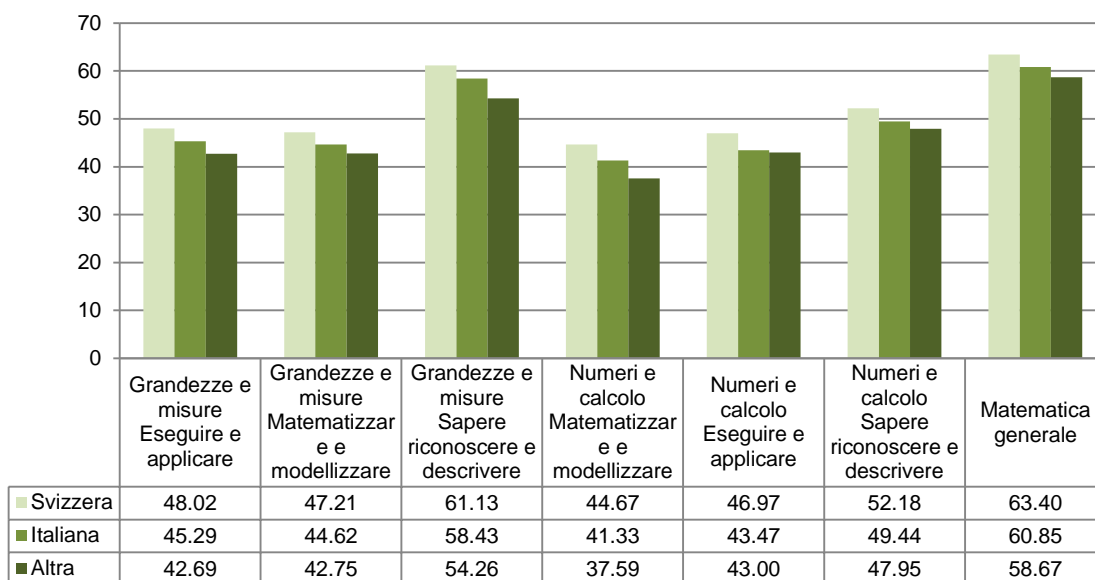
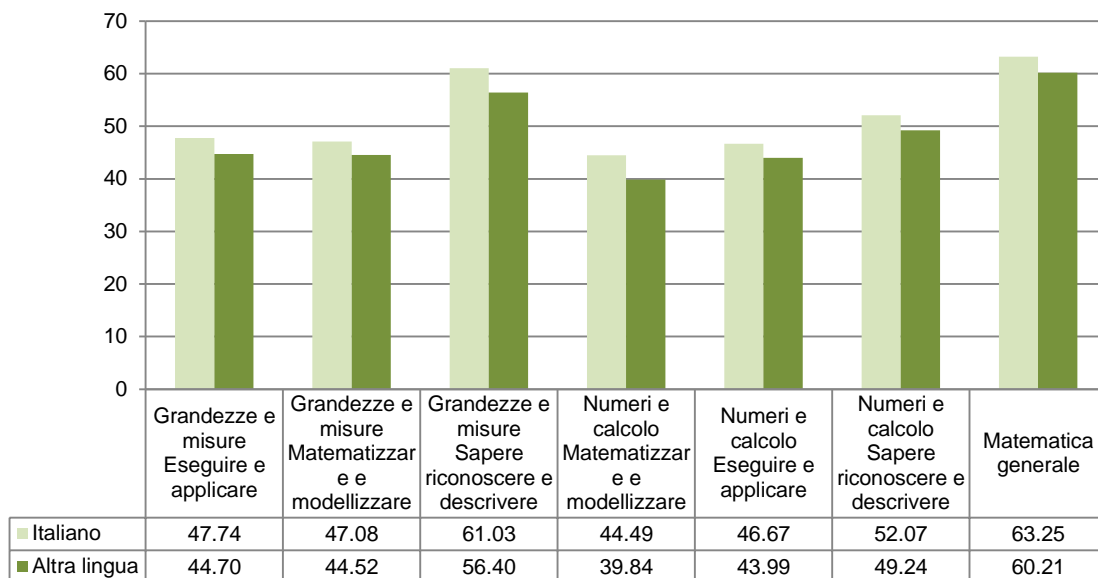
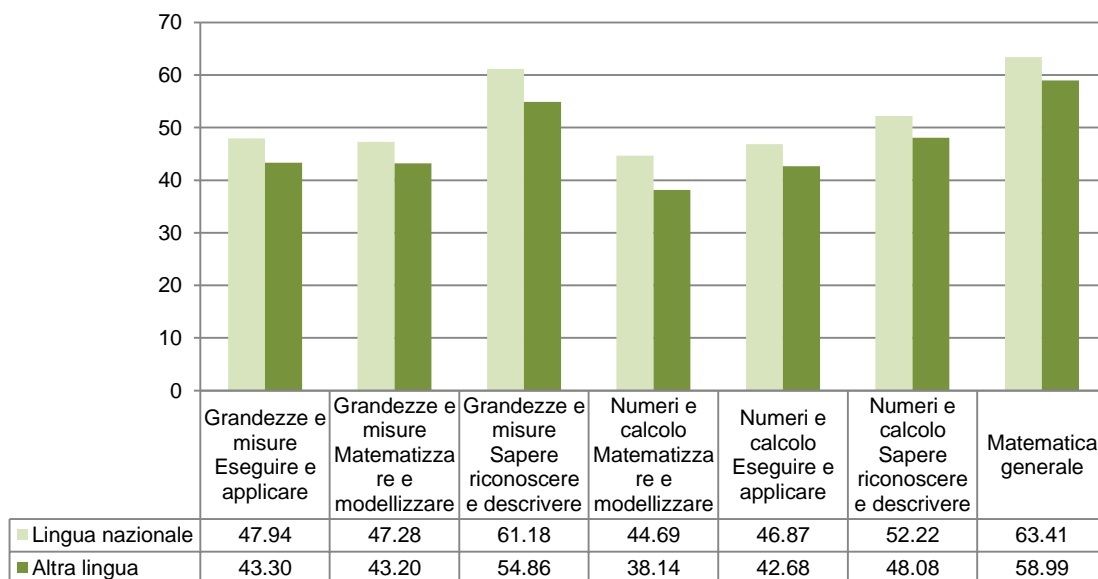


Grafico 11: Medie dei punteggi degli allievi suddivisi per lingua madre



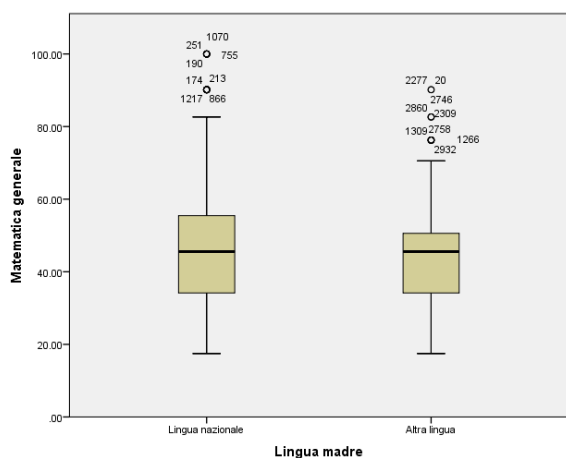
Nel caso della lingua madre le differenze sono sempre significative (con una probabilità del 99%) con punteggi maggiori per gli allievi di lingua madre italiana

Grafico 12: Medie dei punteggi degli allievi suddivisi per lingua madre considerando le lingue nazionali



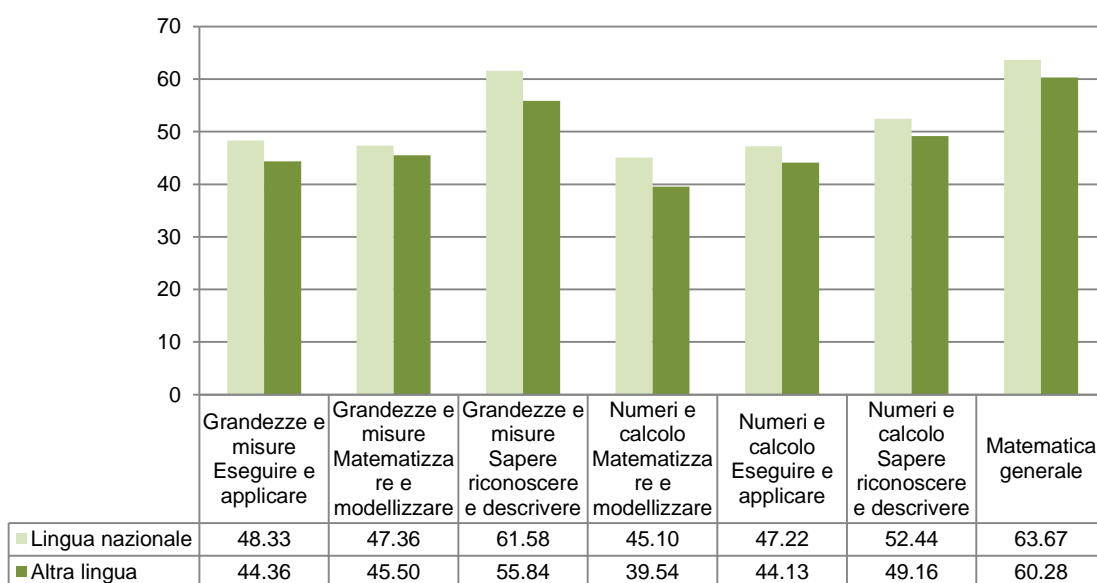
Considerando gli allievi che parlano una lingua madre nazionale rispetto agli allievi di lingua madre diversa risultano anche delle differenze elevate e significative (al 99% di probabilità) in tutte le dimensioni considerate. Le distribuzioni in questo caso si differenziano per la presenza di uno schiacciamento dei punteggi degli allievi di lingua madre non nazionale verso i punteggi più bassi. A titolo esemplificativo si riporta il grafico a scatola di "matematica generale" dal quale si può vedere come sia la parte corrispondente ai punteggi più elevati ad essere schiacciata verso il basso e ad essere meno estesa.

Grafico 13: Distribuzione delle medie dei punteggi degli allievi suddivisi per lingua madre aggregando le lingue nazionali



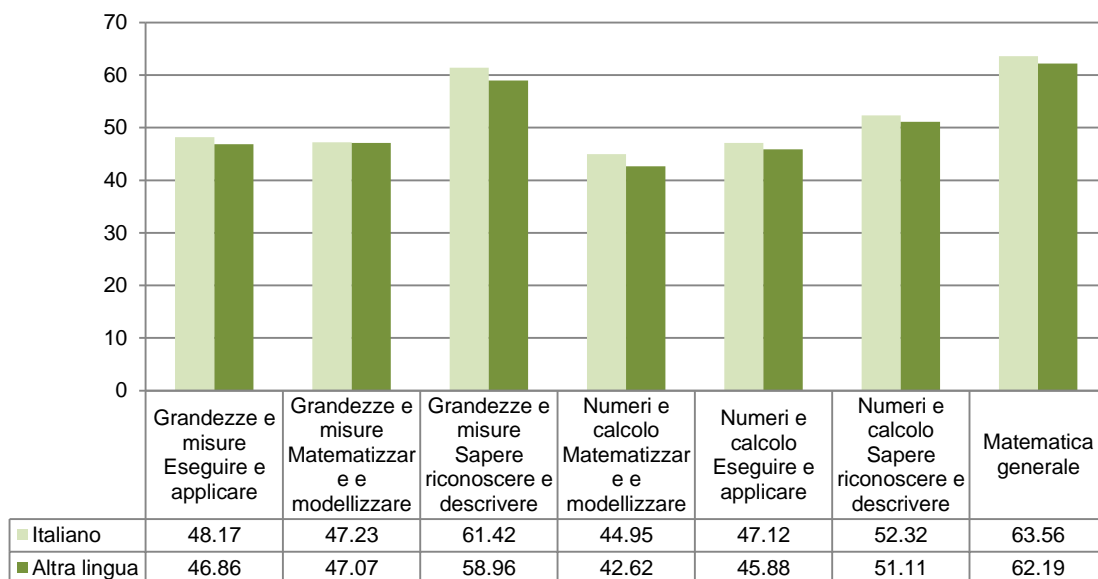
Considerando i soli allievi di nazionalità Svizzera confrontando quelli di lingua madre nazionale rispetto a quelli di lingua madre diversa risultano delle differenze significative (al 99% di probabilità) in tutte le dimensioni valutate).

Grafico 14: Medie dei punteggi degli allievi di nazionalità svizzera suddivisi per lingua madre considerando le lingue nazionali



Considerando gli allievi di nazionalità Svizzera divisi tra quelli madrelingua italiana e quelli di madrelingua differente è possibile rilevare delle differenze in positivo per gli allievi madrelingua italiani in tutti i traguardi testati, risultano significative (al 99% di probabilità) solamente “Grandezze e misure Sapere, riconoscere e descrivere” e “Numeri e calcolo Matematizzare e modellizzare”.

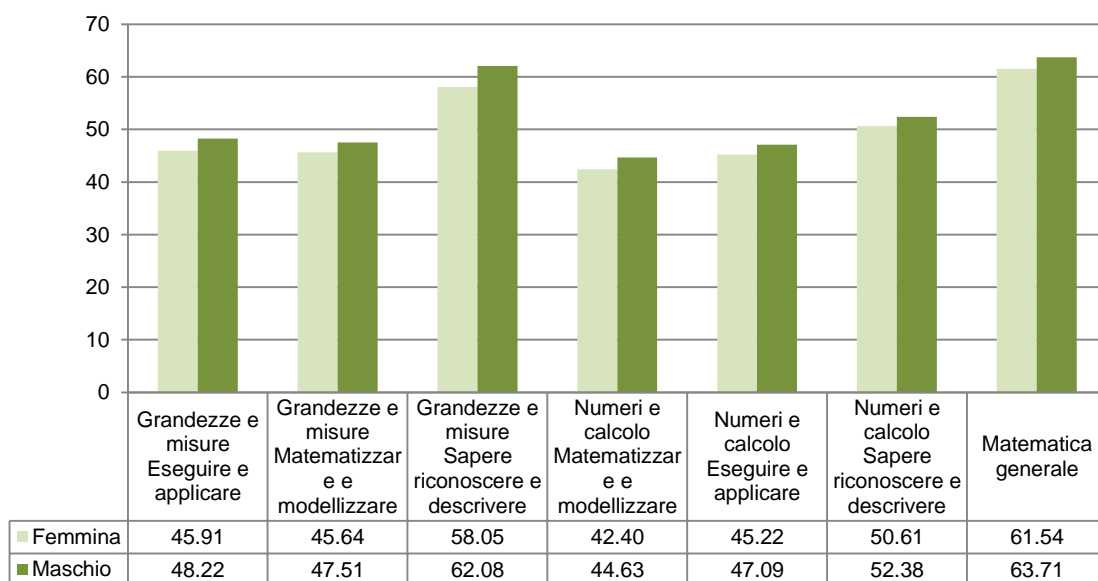
Grafico 15: Medie dei punteggi degli allievi di nazionalità svizzera suddivisi per lingua madre divisi tra allievi di lingua madre italiana e di altre lingue



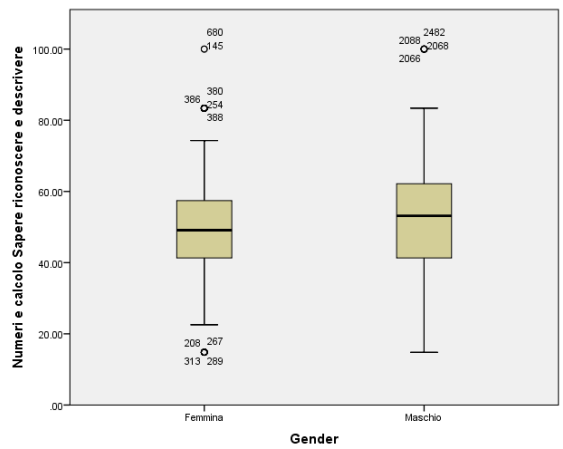
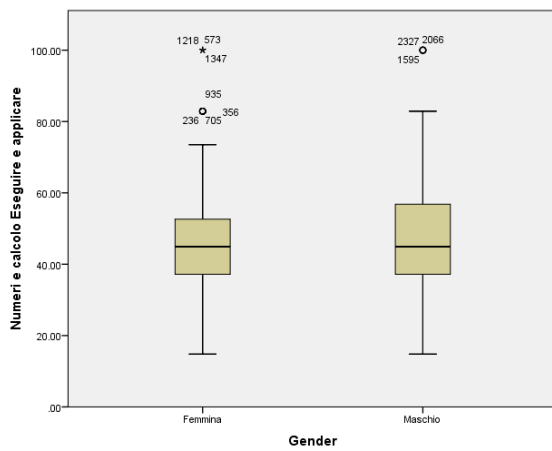
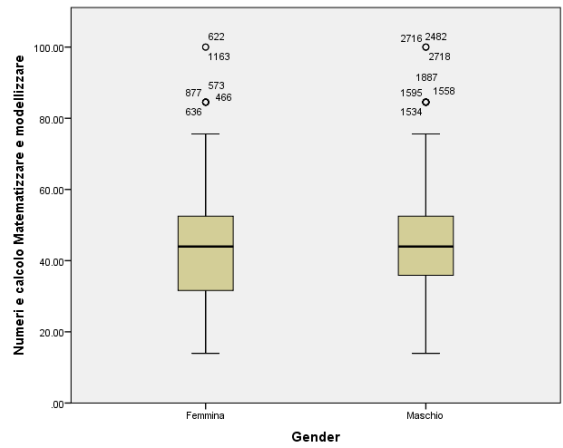
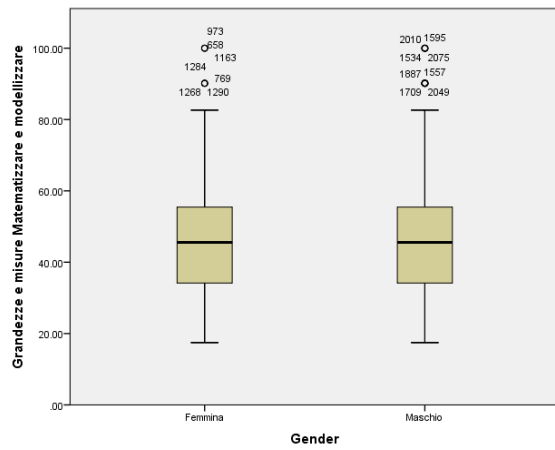
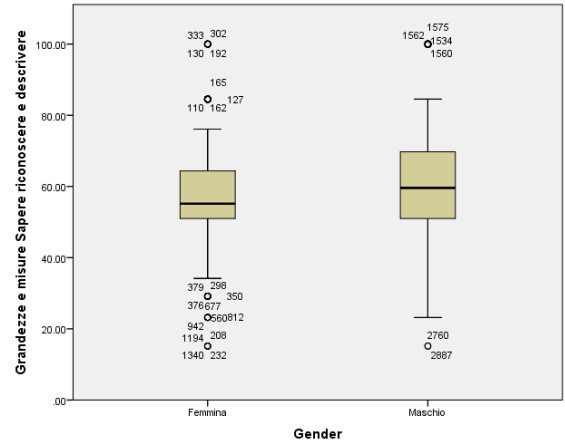
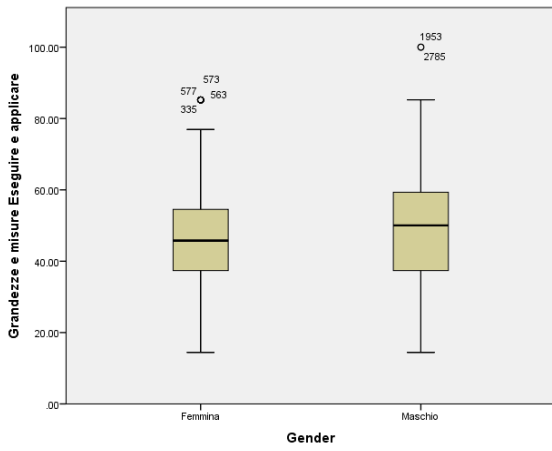
2.4.3 Prestazione alle prove e gender

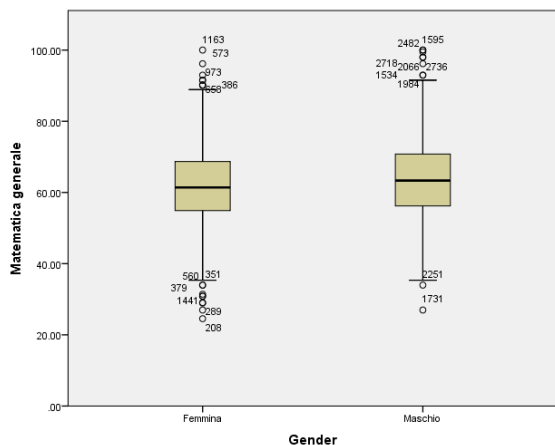
Come rilevato precedentemente (Zanolla, 2014) le differenze di gender possono avere un impatto sulla prestazione in considerazione sia della socializzazione primaria (in famiglia) sia di quella esterna alla scuola, sia in riferimento a eventuali effetti presenti a scuola. Riuscire a distinguere questi effetti va oltre le possibilità del presente lavoro ma è importante rendere attento il lettore in proposito. La differenza nei risultati ottenuti dagli allievi maschi e femmine risulta significativa (al 99% di probabilità) in tutte le dimensioni testate.

Grafico 16: Risultati alle prove e gender



Osservando i grafici a scatola dei punteggi rispetto ai diversi traguardi è possibile osservare come in alcuni casi vi siano distribuzioni differenti che permettano di vedere meglio le differenze nelle prestazioni.





Alcuni traguardi hanno infatti punteggi differenti in modo molto rilevante. Se infatti guardiamo le due distribuzioni relative a “Sapere, riconoscere e descrivere” è possibile osservare come la distribuzione dei punteggi delle femmine sia più compatta rispetto a quella dei punteggi dei maschi. Questa differenza è presente sia relativamente ai punteggi alti sia relativamente ai punteggi bassi.

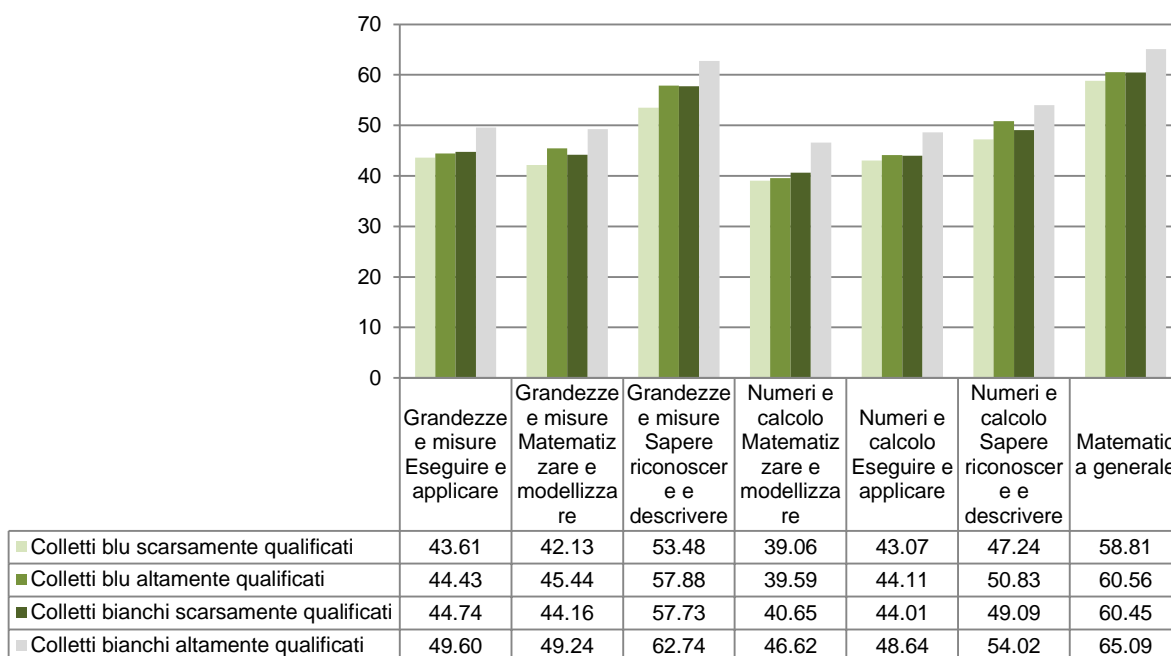
È possibile inoltre rilevare come la distribuzione sui livelli di difficoltà non sia omogenea. Nella prova di italiano non esistevano delle differenze rilevanti relativamente al sesso degli allievi mentre nella prova di matematica erano emerse delle differenze relativamente alla presenza più marcata di allievi maschi con prestazioni molto alte. Questa differenza si è ritrovata anche in questa prova. Come era stato rilevato nella prova di italiano non è possibile dire, sulla base dei nostri risultati, se si tratti di un fattore legato alla crescita.

2.4.4 Origine sociale

L'origine sociale degli allievi è stata spesso indicata come uno dei fattori rilevanti 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.

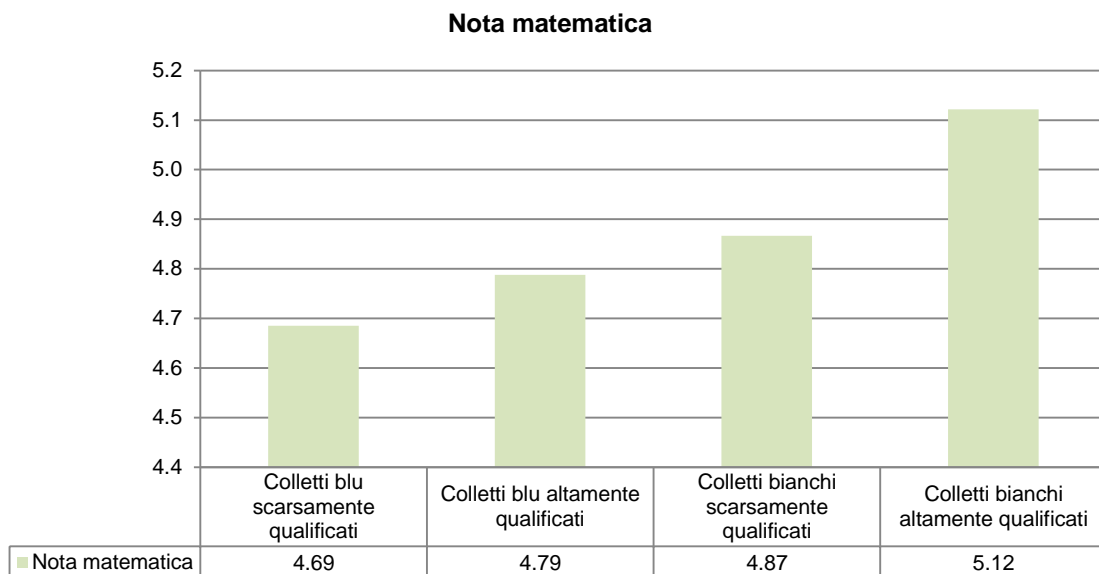
L'analisi condotta conferma che, in tutte le dimensioni, quanto più elevato è il retroterra socio-culturale⁴ di un individuo, migliore sarà la sua prestazione. I figli di genitori appartenenti alla categoria “colletti bianchi altamente qualificati” riportano risultati significativamente migliori. Sia 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 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.

Grafico 17: Relazione tra punteggi medi e origine sociale degli allievi



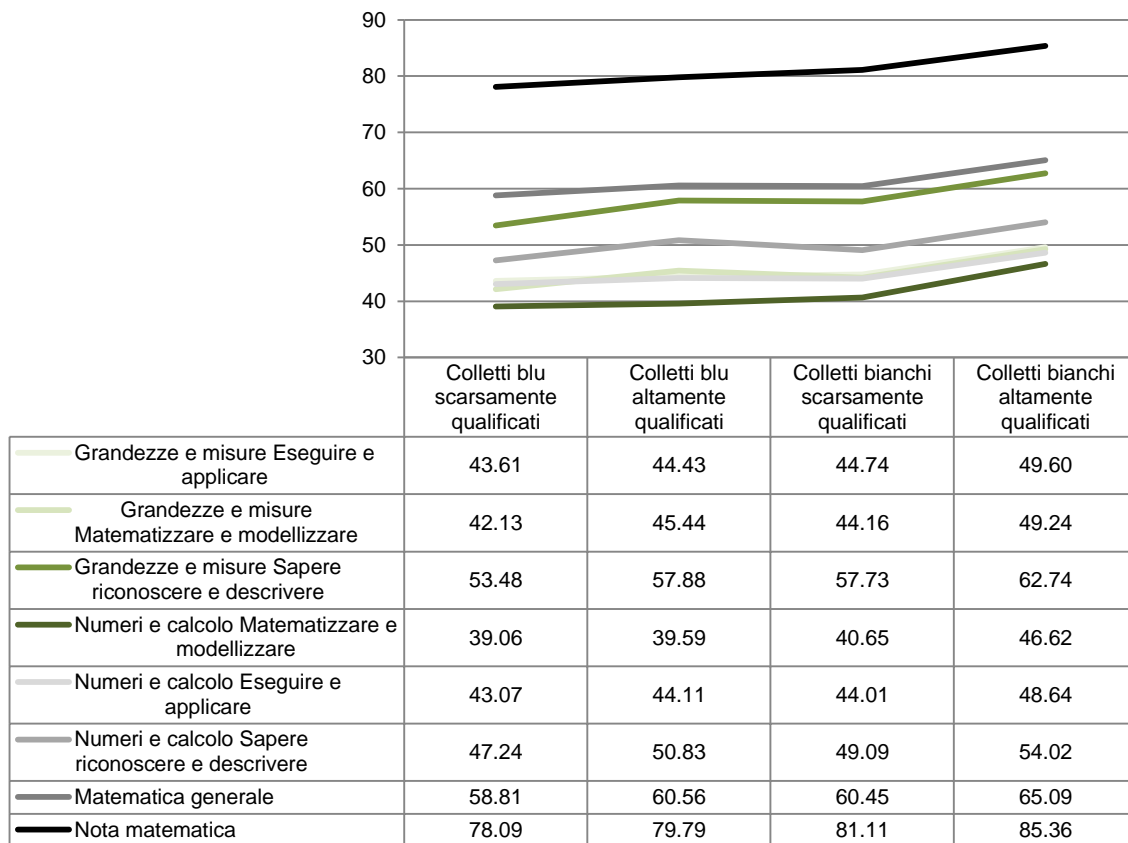
⁴ Nonostante termini quali “colletti blu” o “colletti bianchi” per definire coloro che svolgono rispettivamente un lavoro manuale o tecnico come ad es. gli operai e gli artigiani o mansioni più o meno fisiche e coloro che invece svolgono un lavoro di tipo impiegatizio, i funzionari, i venditori, ecc., 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).

Grafico 18: Media nota di matematica



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.

Grafico 19: Punteggi rispetto ai gruppi socioeconomici di provenienza



2.4.5 Età dell'allievo

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

Coerentemente con quanto detto la maggior parte degli allievi sono nati nel 2004 (85.3%) e una percentuale limitata nel 2003 (13.3%).

Tabella 3: Anno di nascita degli allievi

	Frequenza	Percentuale valida
2001	1	.0
2002	32	1.1
2003	385	13.3
2004	2472	85.3
2005	9	.3

È stata calcolata la correlazione tra la prestazione in ogni settore matematico e l'età dell'allievo 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 più bassi in tutti i traguardi che sono stati misurati.

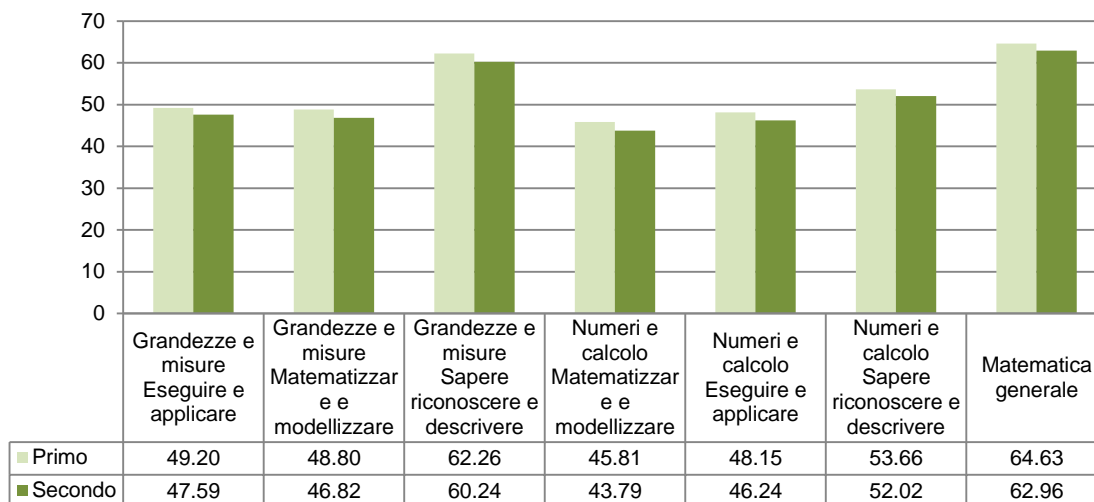
Tabella 4: Correlazioni tra punteggi medi alla prova e età degli allievi

		Grandezze e misure Eeguire e applicare	Grandezze e misure Matematizzare e modellizzare	Grandezze e misure Sapere, riconoscere e descrivere	Numeri e calcolo Matematizzare e modellizzare	Numeri e calcolo Eeguire e applicare	Numeri e calcolo Sapere, riconoscere e descrivere	Matematica generale
Età	Correlazione di Pearson	-.116**	-.101**	-.102**	-.099**	-.086**	-.132**	-.133**

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

Si è quindi verificato se tra gli allievi nati nel 2004 sussistessero differenze tra i bambini nati nei primi sei mesi dell'anno (N 1305, 53%) e quelli nati a partire da luglio (N 1167, 47%).

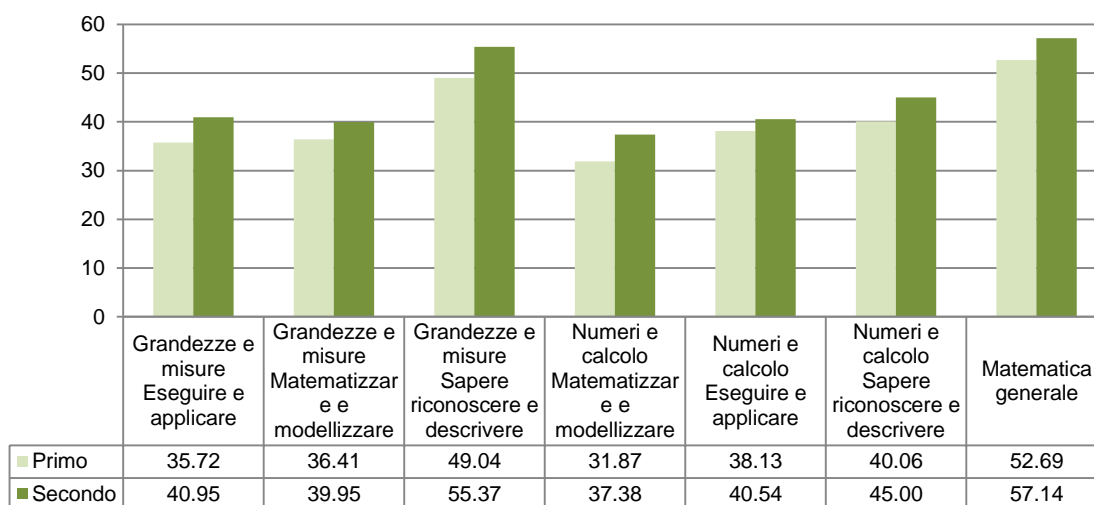
Grafico 20: Risultati alle prove degli allievi nati nel primo e nel secondo semestre 2004



Le differenze risultano essere sempre significative (con una probabilità del 99%) con un punteggio maggiore per gli allievi nati nel primo semestre.

Replicando la medesima analisi tra i nati del 2003 si constata invece che sono i nati nel secondo semestre a riportare le medie più elevate.

Grafico 21: Risultati alle prove degli allievi nati nel primo e nel secondo semestre 2003



Le differenze risultano essere sempre significative (con una probabilità del 99%) con un punteggio maggiore per gli allievi nati nel secondo semestre.

I valori appaiono decrescenti in modo quasi continuo rispetto al mese di nascita. Questi risultati sono coerenti con quelli delle prove di italiano ma non con quelle precedenti di matematica. La spiegazione che era stata proposta della possibile relazione con la classe e il processo di apprendimento della lingua rispetto alla matematica non pare più essere sostenibile.

Tabella 5: Punteggi medi nelle diverse dimensioni per mese di nascita (2004)

mese di nascita	Grandezze e misure Eeguire e applicare	Grandezze e misure Matematizzare e modellizzare	Grandezze e misure Sapere, riconoscere e descrivere	Numeri e calcolo Matematizzare e modellizzare	Numeri e calcolo Eeguire e applicare	Numeri e calcolo Sapere, riconoscere e descrivere	Matematica generale
gennaio	49.33	47.25	63.60	45.37	47.94	53.09	64.43
febbraio	49.46	47.83	62.49	46.89	47.47	53.54	64.54
marzo	48.27	48.29	59.64	45.64	47.78	53.52	64.12
aprile	49.55	48.11	62.48	45.80	48.28	53.31	64.34
maggio	49.25	51.01	61.99	46.17	48.87	53.83	65.12
giugno	49.22	49.97	63.00	45.11	48.40	54.58	65.08
luglio	48.65	47.99	61.97	44.15	47.44	52.50	64.05
agosto	48.92	47.87	60.85	44.68	45.67	53.39	63.67
settembre	47.52	46.77	59.77	44.32	46.75	51.29	62.97
ottobre	47.13	45.94	60.13	43.33	46.45	52.61	62.68
novembre	47.56	47.12	59.48	43.90	45.66	51.59	62.47
dicembre	44.41	44.17	58.06	41.49	44.65	49.70	60.91

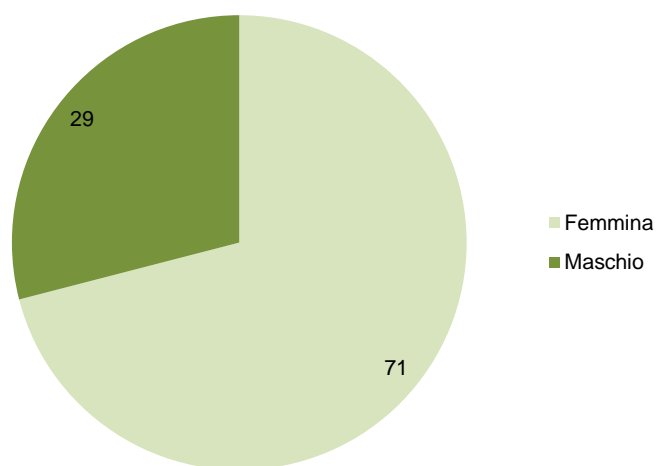
2.5 Docente e allievo

La relazione che si instaura tra docente allievo è una delle caratteristiche che definiscono la qualità del contesto scolastico. Questa relazione è particolarmente significativa all'interno delle scuole elementari. La significatività deriva sia dalle caratteristiche dell'età sia dalla quantità di ore trascorse assieme. 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 185, per la maggioranza donne (Grafico 22). 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 22: 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.

Dalle analisi dei risultati alle prove emerge come vi siano delle prestazioni più elevate degli allievi che hanno docenti maschi rispetto a quelli che hanno docenti femmine. Questo risultato è visibile in tutti i traguardi testati ed è statisticamente significativo (con una probabilità del 99%), non emergono però differenze significative per quanto riguarda le note di fine anno nelle diverse discipline.

Grafico 23: Punteggi medi ottenute dagli allievi secondo il genere del docente

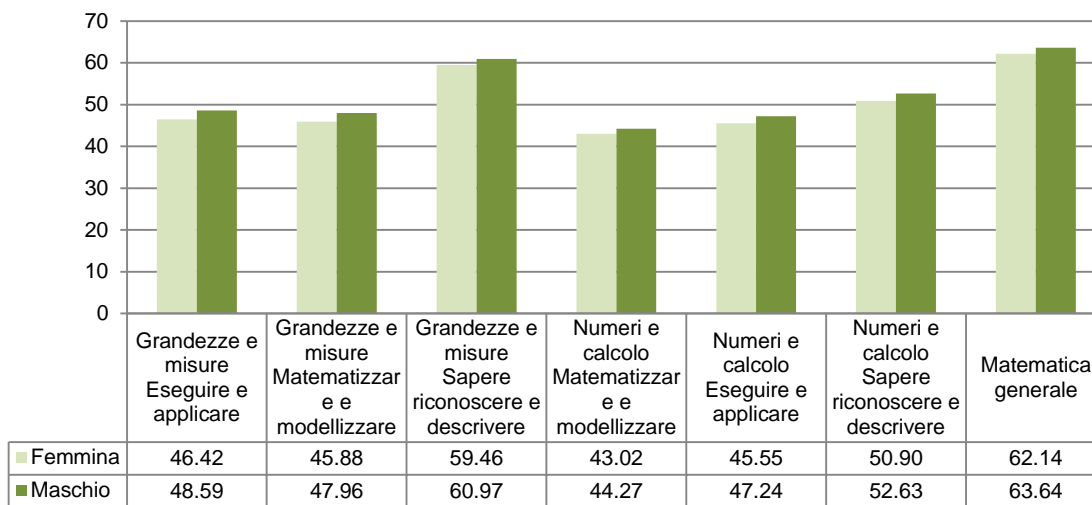
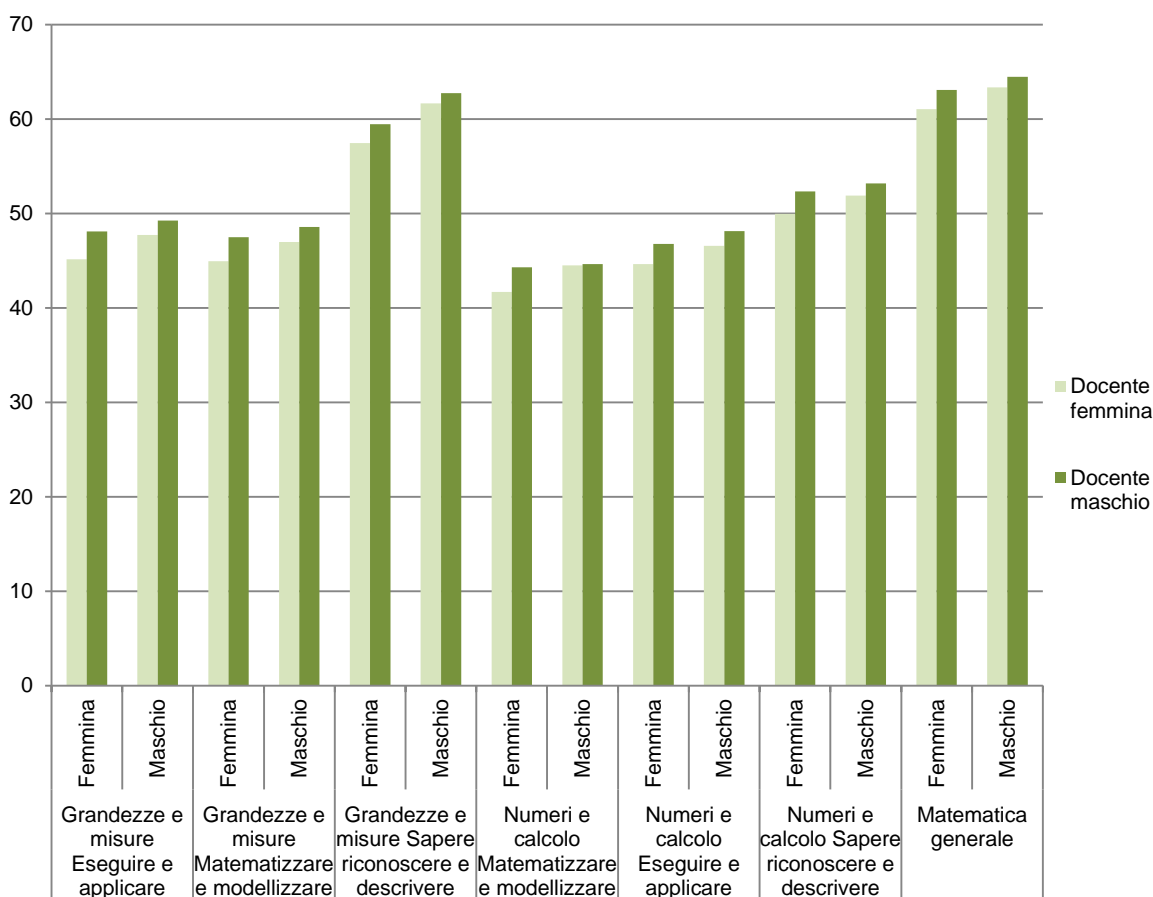


Grafico 24: Medie dei risultati degli allievi divisi per sesso degli allievi e dei docenti



Abbiamo poi osservato l'incrocio tra la differenza di gender degli allievi e la possibile relazione con la differenza di gender dei docenti.

In tutte le dimensioni gli allievi maschi hanno risultati più elevati degli allievi femmine. Nel caso di docenti femmine queste differenze sono sempre significative mentre nei casi di docenti maschi queste differenze sono significative solo nella dimensione "Grandezze e misure Sapere, riconoscere e descrivere". Osservando la Tabella 5 è possibile vedere come le allieve con docenti femmine abbiano costantemente punteggi più bassi rispetto agli allievi con docenti maschi.

Il genere del docente risulta quindi una caratteristica rilevante da considerare. Nelle precedenti prove vi erano stati risultati differenti. Nella prova di Italiano in terza elementare era risultato non rilevante, nella prova di Matematica in quarta era risultato debolmente rilevante mentre nella presente risulta rilevante. Una possibile ipotesi è che sia un effetto che si struttura nel tempo tramite confronto con la realtà sociale nella quale si è immersi e adeguamento ai prototipi che vengono prospettati.

Tabella 5: Media prestazioni per maschi e femmine rispetto al genere del docente

	Gender Allievi	Docente femmina	Docente maschio
Grandezze e misure Eseguire e applicare	Femmina	45.14	48.10
	Maschio	47.73	49.27
Grandezze e misure Matematizzare e modellizzare	Femmina	44.96	47.50
	Maschio	46.99	48.59
Grandezze e misure Sapere, riconoscere e descrivere	Femmina	57.48	59.46
	Maschio	61.68	62.77
Numeri e calcolo Matematizzare e modellizzare	Femmina	41.69	44.32
	Maschio	44.50	44.65
Numeri e calcolo Eseguire e applicare	Femmina	44.65	46.78
	Maschio	46.57	48.16
Numeri e calcolo Sapere, riconoscere e descrivere	Femmina	49.99	52.36
	Maschio	51.92	53.21
Matematica generale	Femmina	61.05	63.11
	Maschio	63.36	64.49

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%) 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 6: Correlazioni tra i traguardi valutati dalla prova e la nota in condotta conseguita a fine anno

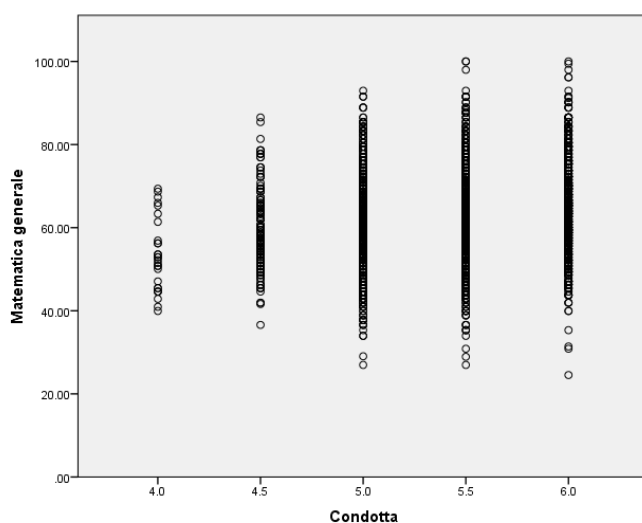
	Grandezze e misure Eseguire e applicare	Grandezze e misure Matematizzare e modellizzare	Grandezze e misure Sapere, riconoscere e descrivere	Numeri e calcolo Matematizzare e modellizzare	Numeri e calcolo Eseguire e applicare	Numeri e calcolo Sapere, riconoscere e descrivere	Matematica generale
Condotta	.161**	.133**	.082**	.141**	.161**	.161**	.166**

Nella tabella 7 si riportano i valori medi in Matematica generale raggruppati per i diversi valori della nota di condotta. La relazione è resa più evidente nel grafico (Grafico 15) successivo nel quale sono riportati i singoli allievi rappresentati mettendo in rapporto i due valori.

Tabella 7: Media Matematica generale e nota di condotta

	Matematica generale
4.0	53.37
4.5	59.25
5.0	61.19
5.5	62.90
6.0	64.77

Grafico 25: Dispersione prestazioni nella prova rispetto alla nota di condotta



2.6 Risultati al test e note

2.6.1 Relazione nota di Matematica 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 matematica e test è statisticamente significativa. Nella Tabella 8 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. La correlazione⁵ è l’indice di associazione tra le due variabili. Indici prossimi all’unità mostrano la forza dell’associazione tra le dimensioni testate dalla prova standardizzata nel mese di maggio 2014 e le valutazioni date dai docenti agli allievi alla fine dell’anno scolastico.

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.

È necessario però rilevare come vi siano numerosi allievi che ottengono risultati sensibilmente differenti nelle due valutazioni. Nel grafico 26 sono stati infatti raggruppati gli allievi per nota di fine mettendo a confronto i punteggi ottenuti alla prova nella dimensione Matematica generale. È stato evidenziato graficamente come la valutazione mediana degli allievi che hanno ottenuto una nota pari a 3 sia analoga a quanti hanno ottenuto una nota pari a 4.

Tabella 8. Correlazioni tra le dimensioni valutate con la prova e la nota conseguita dagli allievi in italiano

		Grandezze e misure Eseguire e applicare	Grandezze e misure Matematizzare e modellizzare	Grandezze e misure Sapere, riconoscere e descrivere	Numeri e calcolo Matematizzare e modellizzare	Numeri e calcolo Eseguire e applicare	Numeri e calcolo Sapere, riconoscere e descrivere	Matematica generale
Nota Matematica	Correlazione di Pearson	.595**	.571**	.501**	.616**	.587**	.594**	.686**

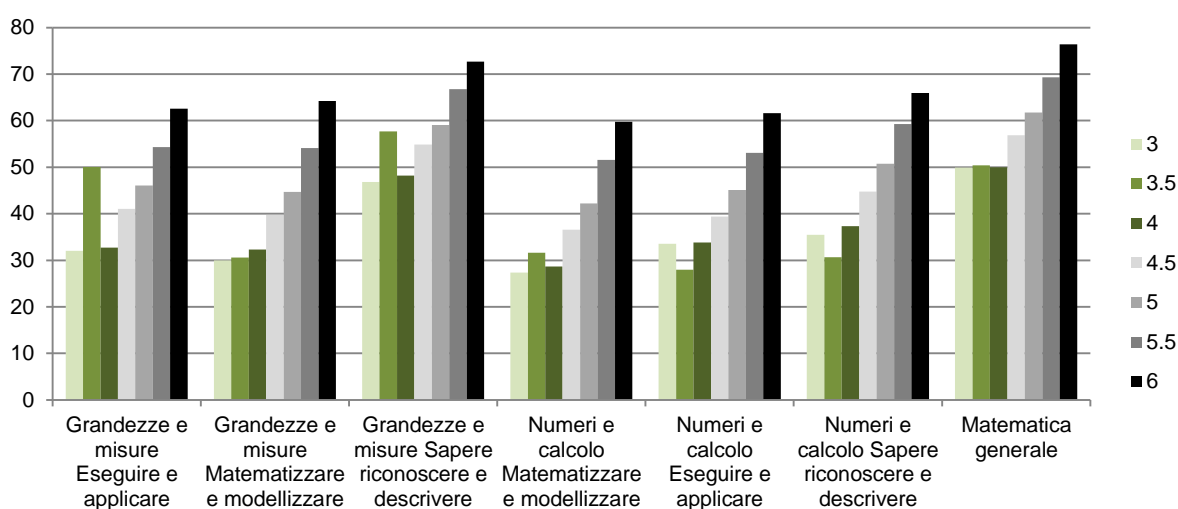
** La correlazione è significativa a livello 0,01 (a due code).

⁵ Il coefficiente di correlazione r può assumere valori compresi fra -1 e 1. I valori positivi indicano l’esistenza di una relazione lineare positiva; i valori negativi indicano una relazione lineare negativa; il valore 0 indica assenza di correlazione.

Tabella 9: Voto finale in Matematica

Matematica	Percentuale
3.0	.6
3.5	.1
4.0	12.3
4.5	20.2
5.0	30.9
5.5	26.9
6.0	9.0
Totale	100.0

Grafico 28: Relazione punteggi nelle dimensioni e nota di fine anno



Il risultato fondamentale delle valutazioni nella prova e nella nota di fine anno è allineato con quanto era stato rilevato nelle precedenti prove. 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 nei singoli traguardi) è possibile osservare che la correlazione tra la nota di Matematica e le altre note è più forte che non tra il risultato alla prova e le altre note. Questa osservazione era stata fatta anche per l'Italiano ipotizzando che il fatto che fosse lingua di insegnamento potesse esserne l'origine. Naturalmente questa non può essere considerata una valida spiegazione nel caso della Matematica. Rimane valida l'ipotesi dell'effetto docente formulata anche per l'italiano. Secondo la quale 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 10: Correlazioni note di fine anno – note di matematica e risultato alla prova

	Educazione fisica	Studio dell'ambiente	Attività grafiche e pittoriche	Educazione religiosa	Applicazioni	Francese	Italiano	Condotta	Educazione musicale	Attività creative
Matematica	.310**	.704**	.405**	.239**	.561**	.603**	.700**	.291**	.315**	.369**
Matematica generale	.238**	.532**	.224**	.205**	.351**	.471**	.543**	.166**	.266**	.279**

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

Un'analisi specifica è stata condotta sull'uso dell'approccio DiMat all'interno della classe, i risultati di questa analisi saranno presentati in un rapporto in fase di pubblicazione.

3 Sintesi

Il presente rapporto esce in parallelo al rapporto sulle prove di italiano, secondo rapporto sulle prove standardizzate nella scuola elementare. Il primo è stato sempre sulla matematica ma nella classe terza.

Giova ricordare che queste prove sono costruite come interne al sistema educativo ticinese da persone che vi lavorano all'interno. Questo implica una maggiore adesione ai programmi scolastici rispetto ad analoghe prove internazionali o nazionali.

Nel rapporto 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.

Tramite la prova sono stati misurati sei traguardi di competenza identificabili negli incroci tra ambiti (Numeri e calcolo e Grandezze e misure) e aspetti di competenza (Eeguire e applicare, Misurare e modellizzare e Sapere, riconoscere e descrivere). La scala più facile per gli allievi è risultata essere quella relativa al traguardo Grandezze e misure – Sapere riconoscere e descrivere mentre quella risultata più difficile è quella relativa a Numeri e calcolo – Matematizzare e modellizzare. Le distribuzioni della abilità nei diversi circondari è risultata differente anche se in modo non eclatante e con un andamento delle distribuzioni analogo. Analizzando le differenze rispetto al grado di urbanizzazione le prestazioni degli allievi delle città nucleo sono risultate meno elevate mentre quelle dei comuni della cintura sono risultate le più elevate, questo in tutti i traguardi testati.

La dimensione della classe non appare essere una variabile rilevante per la prestazione degli allievi così come anche il fatto che l'allievo sia in una mono o in una pluriclasse mostrando così come non vi sia un effetto penalizzante sulle prestazioni degli allievi dovuto a questa variabile organizzativa.

Gli allievi di nazionalità svizzera ottengono punteggi migliori rispetto agli altri allievi e gli allievi di madrelingua italiana ottengono punteggi migliori rispetto agli altri. Gli allievi di madrelingua italiana hanno ottenuto punteggi più alti in tutte le scale. Tra gli allievi di nazionalità svizzera quelli di madrelingua italiana hanno ottenuto punteggi più elevati rispetto agli altri. Analizzando le differenze rispetto alla variabile di genere sessuale è stato possibile rilevare come le allieve abbiano ottenuto risultati meno elevati in tutte le dimensioni, questa caratteristica presente anche nella prova precedente pone degli interrogativi su quale ne sia l'origine. La tipologia di dati raccolti non rende possibile rispondere a questa domanda che richiederebbe un approfondimento specifico.

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 probabile ritardo scolastico è stata rilevata come correlata con dei punteggi più bassi in tutte le dimensioni.

Gli allievi con un docente maschio hanno ottenuto punteggi più elevati in tutte le scale. Nel caso in cui gli allievi abbiano una docente femmina la differenza tra maschi e femmine è sempre significativa mentre nel caso abbiano un docente maschio la differenza è significativa solo per una dimensione. Questo effetto meriterebbe un approfondimento anche in relazione alla progressiva femminilizzazione del corpo docente.

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. I risultati alla prova sono risultati correlati con la nota di condotta. Le note di fine anno in Matematica 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. Si deve altresì rilevare come esistano allievi con note di fine anno molto diverse che però hanno ottenuto punteggi molto vicini alla prova. I risultati alla prova sono correlati anche con le note in tutte le altre materie.

La relazioni 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 meri-

to 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.

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 quinta 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. I traguardi testati erano i seguenti:

1. Grandezze e misure – Eseguire e applicare,
2. Grandezze e misure – Sapere, riconoscere e descrivere,
3. Numeri e calcolo – Matematizzare e modellizzare,
4. Grandezze e misure – Matematizzare e modellizzare,
5. Numeri e calcolo – Sapere, riconoscere e descrivere,
6. Numeri e calcolo – Eseguire e applicare.

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 vi fosse un buon allineamento tra la difficoltà degli item e le competenze degli allievi.

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 modello⁶. 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.

⁶ 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.

Criteria 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.

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.

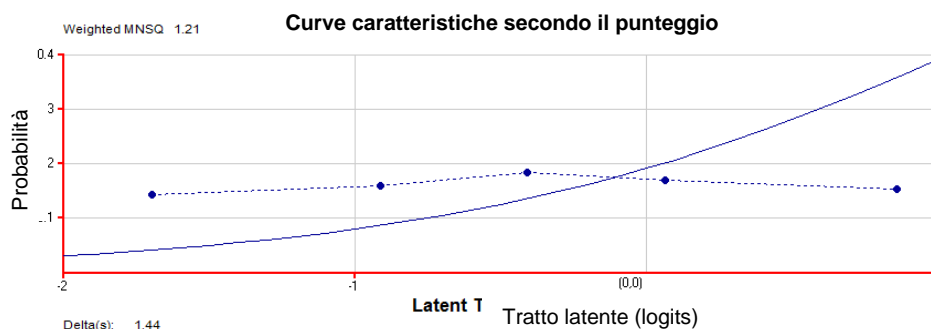
In aggiunta ai valori riportati per ogni item, è stata disegnata la curva caratteristica (*item characteristic curve*, ICC) che permette di osservare direttamente la coerenza tra la distribuzione reale e quella prevista dal modello. Di seguito due esempi (Grafico 4 e 5) nei quali la curva teorica è indicata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

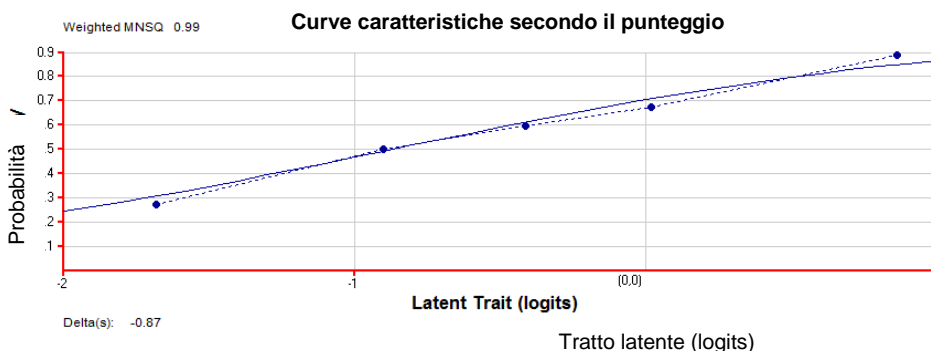
Un ulteriore parametro di scelta è stata la identificazione di differenze eccessive tra la prima e la seconda somministrazione. Come detto, ogni item è stato somministrato a un campione pari di allievi nella prima somministrazione e nella seconda. Tendenzialmente i punteggi sono stati lievemente migliori 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 3 e 4: 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 116 non si distribuisce in modo ottimale mentre il 140 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 al quale si è fatto riferimento è quello presente nel Piano di studio della scuola dell'obbligo (Decs, 2015). Per la matematica il modello di competenza (Figura 1), si costruisce su due assi principali: gli ambiti di competenza e gli aspetti di competenza.

Figura 1: Modello di competenza

		1°- 2°- 3° ciclo				
		Ambiti di competenza				
		Numeri e calcolo	Geometria	Grandezze e misure	Funzioni	Probabilità e statistica
Aspetti di competenza	Risorse cognitive	Sapere e riconoscere	Manifestazioni di competenza			
		Eeguire e applicare				
	Processi cognitivi	Esplorare e provare				
		Matematizzare e modellizzare				
		Interpretare e riflettere sui risultati				
		Comunicare e argomentare				

Il gruppo di lavoro composto da docenti della disciplina coinvolto nell'elaborazione della prova scritta ha proceduto tenendo in considerazione tre aspetti di competenza in due degli ambiti definendo quindi complessivamente 6 dimensioni. Si tratta precisamente di: Grandezze e misure – Eeguire e applicare, Grandezze e misure – Sapere e riconoscere, Numeri e calcolo – Matematizzare e modellizzare. La seconda prova comprende i traguardi di competenza: Grandezze e misure – Matematizzare e modellizzare, Numeri e calcolo– Sapere e riconoscere, Numeri e calcolo – Eeguire e applicare L'insieme delle risposte alle domande ha costituito una misura generale chiamata Matematica Generale. Per ciascuna dimensione 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
Matematica nella classe V Scuola Elementare

Quaderni di ricerca – n. 19

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
978-88-941240-7-1

