

LE PAROLE DEL RAV

L'analisi testuale per la metavalutazione

Analisi dei punti di forza e di debolezza:
area dei Risultati nelle Prove Standardizzate
Nazionali





INVALSI - Via Ippolito Nievo 35, Via Marcora 18/20, 00153 Roma.

Le parole del Rav. L'analisi testuale per la metavalutazione

A cura di:

Donatella Poliandri, Ughetta Favazzi, Monica Perazzolo, Isabella Quadrelli, Emanuela Vinci

Ughetta Favazzi ha elaborato i dati e redatto il presente capitolo.

Alla riflessione sull'autovalutazione e valutazione delle scuole e al progetto PON Valu.E, hanno contribuito in questi anni: Fabio Alivernini, Mattia Baglieri (consulente), Paola Bianco, Roberta Cristallo (consulente), Nicoletta Di Bello, Graziana Epifani, Stefano Famiglietti (responsabile settore Web), Ughetta Favazzi, Brunella Fiore (assegnista), Francesca Fortini, Michela Freddano, Letizia Giampietro, Filippo Gomez Paloma (consulente), Angela Litteri, Beba Molinari (consulente), Lorenzo Mancini, Sara Manganelli, Daniela Marinelli, Flora Morelli, Enrico Nerli Ballati, Monica Perazzolo, Donatella Poliandri (responsabile dell'Area di Ricerca INVALSI – Innovazione e Sviluppo e del progetto PON Valu.E), Elisabetta Prantera, Isabella Quadrelli (consulente), Maria Ranieri (consulente), Sara Romiti, Simone Russo, Stefania Sette, Consuelo Torelli (assegnista), Emanuela Vinci.

Le prove standardizzate nazionali rilevano alcune fra le competenze linguistiche e matematiche degli studenti. L'analisi dei punteggi alle prove standardizzate nazionali consente di riflettere sugli esiti ottenuti da una scuola in relazione a quelli ottenuti dalle scuole presenti su un determinato territorio, al valore medio nazionale e alle scuole con background socio-economico simile. L'analisi degli esiti alle prove consente, altresì, di rilevare la capacità della scuola di assicurare a tutti gli studenti il raggiungimento di alcuni livelli essenziali di apprendimento. L'azione della scuola dovrebbe essere volta a ridurre l'incidenza degli studenti con livelli di apprendimento sotto una determinata soglia, considerando la variabilità di risultati interna alla scuola (tra le classi, tra le sedi, tra gli indirizzi), così come la distribuzione degli studenti sui diversi livelli di apprendimento. Il criterio di qualità della rubrica per l'area Prove standardizzate nazionali presuppone che la scuola assicuri l'acquisizione dei livelli di apprendimento per tutti gli studenti.

3.1 Introduzione¹⁻²

In questo contributo presentiamo l'analisi dei testi prodotti dalle scuole e inseriti nei campi aperti per la descrizione dei punti di forza e di debolezza per l'area Risultati nelle Prove standardizzate nazionali così come redatti dalle scuole nel Rapporto di Autovalutazione (RAV). Nello specifico, le scuole sono chiamate a descrivere la propria situazione in merito ai risultati che gli studenti hanno conseguito nelle prove standardizzate nazionali, comparando i risultati ottenuti con i riferimenti territoriali disponibili e prestando attenzione a eventuali differenze tra classi, indirizzi e plessi.

Per ciascuna partizione del testo sono state condotte una serie di analisi preliminari del profilo lessicale, che hanno consentito di esaminare le dimensioni del corpus testuale, ossia di individuare le fasce di frequenza delle parole di cui si compone il corpus. È stata poi effettuata una analisi di specificità, finalizzata all'individuazione di eventuali parole caratteristiche dell'area oggetto di analisi.

¹ I livelli di Italiano e Matematica sono stabiliti su una base empirica e corrispondono a fasce di punteggio su una scala quantitativa (Rasch) sulla quale le risposte degli studenti e il loro livello di competenza sono ordinati su uno stesso continuo. I livelli di Italiano e Matematica sono cinque, in ordine crescente dal livello 1 al livello 5. Cfr. Rapporto Prove Invalsi 2018

Nella seconda fase, attraverso la procedura di *Topic extraction* di Wordstat³ (finalizzata a individuare la struttura tematica nascosta di una raccolta di testo), si è proceduto all'individuazione di nuclei tematici principali. La procedura di estrazione delle componenti principali è stata effettuata utilizzando la rotazione degli assi (Varimax) e un peso fattoriale minimo pari a 0,40.

Infine, a partire dai nuclei semantici individuati con la procedura di *Topic extraction* e effettuando approfondimenti di tipo qualitativo attraverso strumenti per l'esplorazione dei testi (Keyword Retrieval, Keyword in Context⁴), abbiamo elaborato un vocabolario tematico che ha previsto la costruzione di categorie che rappresentano, sinteticamente, i principali temi contenuti nei testi.

3.2 L'analisi del profilo lessicale

La partizione, che comprende i testi prodotti da 725 scuole per la descrizione dei punti di forza per l'area "Prove standardizzate nazionali", si presenta di

https://www.invalsi.it/invalsi/doc_evidenza/2018/Rapporto_prove_INVALSI_2018.pdf

² I percorsi valutativi delle scuole - Inquadramento teorico del RAV, 2014. https://www.invalsi.it/snv/docs/271114/Inquadramento_teorico_RAV.pdf

³ Cfr. Glossario

⁴ Cfr. Glossario

medie dimensioni (Bolasco, 1999): risulta composta da 66.738 *token*⁵ e da 3.528 *types*⁶ e ciascun frammento è composto mediamente da 92 parole. Il *type/token ratio* risulta pari a 5,3% e la percentuale di *hapax* presenti nel testo (parole citate solo una volta) risulta inferiore al 50% (47,0%), di conseguenza la partizione si presenta adeguata per poter essere sottoposta ad analisi statistiche (Bolasco, 1999, Giuliano e La Rocca, 2008).

Anche la partizione che raccoglie i testi prodotti per la descrizione dei punti di debolezza si presenta di medie dimensioni; tuttavia, dal confronto, risulta composta da un numero più ridotto di occorrenze (56.171) e un numero leggermente più elevato di *types* (3.848), con frammenti formati mediamente da 77 parole. Inoltre, il *type/token ratio* risulta pari a 6,9% e la percentuale di *hapax* pari a 46,8%. Si tratta di dati da cui si evince una tendenza delle scuole a scrivere di più quando si parla di aspetti positivi relativi agli esiti degli studenti, anche se rispetto al numero di parole diverse utilizzate (*type*) la differenza tra punti di forza e di debolezza si riduce. Questa tendenza si rileva in tutti i testi di questo tipo considerati.

Per ciascuna delle due partizioni (punti di forza e punti di debolezza) abbiamo condotto un'analisi della frequenza delle parole che ha portato all'individuazione delle fasce di frequenza⁷.

Partendo dall'analisi dei campi aperti relativi ai **punti di forza**, la parola più

frequentemente utilizzata è la preposizione semplice "di", citata 2818 volte e, a seguire, la congiunzione "e" (1897).

Circoscrivendo l'analisi alle sole parole piene⁸ (Bolasco, 2013, p.93) (fig.1), le parole a elevata frequenza sono 7, di cui 6 sono anche caratteristiche dell'area analizzata, vale a dire si tratta di parole sovra-utilizzate in questa area dal confronto con la partizione relativa ai punti di forza nel suo complesso.

Sono caratteristiche di questa partizione (con valori di Z superiori a 60) le parole a elevata frequenza "classi" (72,7%), "Italiano" (69,2%), "risultati" (68,5%), "Matematica" (67,3%), "media" (53,9%) e "prove" (68,7%) (fig. 3.1).

Considerando le parole che ricorrono con una frequenza media (114 in valore assoluto), circa l'86% sono anche caratteristiche dell'area analizzata. Ne sono esempio le parole "nazionale" (47,3%), "livello" (24,9%), "superiore" (24,9%), "punteggio" (18,9%), "cheating" (19,6%), "variabilità" (12,7%). Le parole a elevata e media frequenza utilizzate nei campi aperti dell'area "Prove standardizzate" sono, quindi, in larga parte anche caratteristiche dell'area oggetto di analisi, proprio perché si tratta di termini, in parte anche tecnici, utilizzati dalle scuole per descrivere la propria situazione in merito agli esiti delle prove (fig. 3.2).

⁵ cfr. Glossario

⁶ cfr. Glossario

⁷ Per la partizione punti di forza rientrano nella fascia di frequenza elevata le parole che ricorrono con una frequenza compresa tra 789 e 2818; nella fascia di frequenza media le parole con una frequenza compresa tra 57 e 718; nella fascia di frequenza bassa le parole con una frequenza uguale e inferiore a 55.

Per la partizione punti di debolezza rientrano nella fascia di frequenza elevata le parole che ricorrono con una frequenza compresa tra 147 e 2494; nella fascia di frequenza media le parole con una frequenza compresa tra 47 e 144; nella fascia di frequenza bassa le parole con una frequenza pari e inferiore a 44.

⁸ Cfr. Glossario

studenti e si tratta di parole sovra-utilizzate in questa area, con valori di Z molto elevati.

Tra le parole a media frequenza, sono presenti termini tecnici utilizzati per la descrizione dei punteggi come “varianza”, “dato”, “percentuale”, e si tratta di parole caratteristiche dell’area analizzata con valori di Z molto elevati (rispettivamente, per le parole considerate, 40,5, 13,8, 6,6) (fig.4).

Concentrando l’attenzione sull’utilizzo delle parole “Italiano” e “Matematica” nelle due partizioni (forza e debolezza), si osserva che parola “Italiano” ricorre in maniera lievemente più frequente (1027 volte) rispetto alla parola “Matematica” (983) nei frammenti per la descrizione dei punti di forza. Inoltre, se si considera la partizione

relativa ai punti di forza, la parola “Italiano” è presente in maniera più frequente nei testi prodotti dalle scuole del I ciclo di istruzione (71,8%) rispetto ai testi prodotti da scuole del II ciclo di istruzione, (64,0%)⁹.

Al contrario, si rileva una ricorrenza più frequente della parola “Matematica” (741 volte) rispetto alla parola “Italiano” (630) nei frammenti per la descrizione dei punti di debolezza. Inoltre, diversamente da quanto registrato per i punti di forza, si osserva un utilizzo più frequente della parola “Matematica” tra le scuole del II ciclo di istruzione (61,1%, a fronte del 52,5% delle scuole del I ciclo di istruzione).

⁹ È stato applicato il test del Chi-quadrato

3.3 Il vocabolario tematico.

3.3.1 I temi trattati nei campi aperti per la descrizione dei punti di forza

Al fine di individuare i principali temi trattati dalle scuole nei campi aperti per la descrizione dei punti di forza è stata avviata la procedura di *Topic extraction* presente nel software WordStat. Nello specifico, è stato individuato un valore soglia della frequenza delle parole pari a 10 che ha permesso una copertura pari al 90,3% del corpus. Inoltre, al fine di focalizzare l'attenzione sulle parole piene, è stata applicata una *stop list* che ha limitato l'analisi al 49,1% delle parole presenti nella partizione.

Con la procedura di *Topic extraction* abbiamo individuato 13 temi principali, tenendo conto della frequenza con cui una o più parole sono presenti almeno una volta nei frammenti prodotti dalle scuole (tab. 3.1).

La componente a cui risultano associate parole contenute nella quasi totalità dei documenti è stata denominata "Esiti in Italiano e Matematica" e contiene parole utilizzate dalle scuole per descrivere i risultati ottenuti dai propri studenti nelle Prove standardizzate di Italiano e Matematica.

La seconda componente, in ordine di frequenza, denominata "Livelli di abilità" si compone di parole che ricorrono nell'oltre 75% dei casi e che sono utilizzate per descrivere la collocazione degli studenti nei diversi livelli di abilità. La terza e la quarta componente contengono, invece, parole usate per sottolineare, rispettivamente, la capacità della scuola di attenuare le eventuali disparità tra i risultati degli alunni e di assicurare esiti più uniformi tra le classi.

Tra le componenti che ricorrono con minore frequenza, se ne segnalano alcune che, come punti di forza della scuola, fanno riferimento alla mancanza di *cheating* e alla registrazione di punteggi più elevati se confrontati con scuole popolate da studenti provenienti da un contesto socio-economico e culturale simile. Inoltre, tra le componenti estratte ve ne sono alcune contenenti parole che le scuole utilizzano per mettere in luce eventuali differenze nei risultati degli studenti (con l'indicazione di valori numerici) in base al ciclo o all'area geografica.

L'ultima fase di analisi ci ha permesso di elaborare un vocabolario tematico relativo ai punti di forza per l'area 'Prove standardizzate nazionali'. Il vocabolario presenta un buon livello di copertura ed è articolato in 8 categorie (tabb. 3.2, 3.3 e fig. 3.4); ciascuna categoria è popolata da parole e segmenti ripetuti relativi allo specifico tema trattato. Dopo aver applicato la *stop list* risultano coperte l'86,0% delle frasi, il 91,2% dei paragrafi e il 97,7% dei casi (RAV).

A partire dai risultati della *Topic extraction*, e con il supporto degli strumenti di Keyword Retrieval e Keyword in Context¹⁰, sono state costruite le categorie del vocabolario tematico procedendo alla revisione, all'accorpamento di alcuni *topic* e all'inserimento di parole e frasi.

Nello specifico, la categoria a cui sono associate *keyword* e segmenti ripetuti che ricorrono nella percentuale più elevata dei casi è denominata **Esiti in Italiano e Matematica** (94,6% dei casi) e contiene parole e frasi che rimandano, come punti di forza della scuola, ai risultati degli studenti in Italiano e Matematica. In questa categoria, oltre alle parole "Italiano" e "Matematica", sono contenuti termini tecnici utilizzati per commentare i risultati degli studenti come "media", "punteggio medio",

¹⁰ Cfr Glossario

“standardizzate”. Analizzando i dati in base al ciclo di istruzione, si osserva che questa categoria ricorre in maniera più frequente nei testi prodotti da scuole del I ciclo di istruzione; abbiamo, inoltre, osservato una tendenza più diffusa a trattare il tema dei punteggi ottenuti dagli studenti nei campi aperti per la descrizione dei punti di forza.

La scuola nelle prove standardizzate nazionali di Italiano e Matematica per l'a.s. 2013-2014 ha raggiunto risultati significativamente superiori al punteggio della Lombardia, del Nord Ovest e dell'Italia [caso 49].

Ricorrono nel 64,7% dei casi parole e frasi contenute nella categoria denominata **Livelli di apprendimento** e utilizzate per descrivere la situazione della scuola in merito alla distribuzione degli studenti per livelli di abilità, e individuando, come aspetto di forza, una quota inferiore dal confronto con i riferimenti territoriali di studenti collocati sui primi livelli, oppure una quota superiore di studenti collocati sui livelli più elevati.

La quota di studenti che si colloca nel livello 1 è notevolmente inferiore alla media nazionale sia in Italiano che in Matematica [caso 132].

La terza categoria, in ordine di frequenza, è stata denominata **Confronto interno scuola** (42,1%). Si tratta di una categoria presente in quasi tutti i vocabolari tematici delle diverse aree analizzate, proprio perché comprende parole e frasi che le scuole utilizzano per illustrare la propria situazione effettuando dei confronti in base ai diversi ordini, plessi, classi, indirizzi presenti nell'istituzione. Anche in questa area si conferma un utilizzo più frequente di questa categoria da parte delle scuole del I ciclo di istruzione, giustificabile se si pensa che

queste sono prevalentemente Istituti comprensivi.

La quarta categoria è stata denominata **Analisi della varianza** perché è popolata da parole e frasi utilizzate per approfondire il tema della variabilità dei punteggi ottenuti alle prove standardizzate INVALSI tra le classi e dentro le classi. Nel 33% circa dei frammenti sono presenti parole volte a segnalare come punto di forza della scuola la capacità di assicurare esiti uniformi tra le classi. In particolare, si segnala che la varianza dei punteggi tra le classi si presenta più ridotta dal confronto con i riferimenti territoriali. Si segnalano, tuttavia, rari casi¹¹ in cui una più ridotta variabilità dei punteggi dentro le classi, in base al confronto con i riferimenti territoriali, è descritta come un aspetto positivo, mostrando come il fenomeno della variabilità dei risultati non sia stato compreso.

Il basso indice di variabilità (3,4%) tra le classi e l'alto indice di variabilità interna (96,6%) indicano rispettivamente un alto tasso di omogeneità e di equilibrio nella composizione delle classi e la presenza di tutti i livelli di apprendimento in ciascuna classe. La quota di studenti collocata nei livelli 1 e 2 in Matematica è decisamente inferiore alla media nazionale; è invece superiore la percentuale degli studenti nei livelli più alti [caso 103].

Infine, è stata denominata **Differenze per background socio - economico** la categoria contenente parole e frasi utilizzate dalle scuole per illustrare i risultati ottenuti dagli studenti nel confronto con scuole con contesto socio-economico e culturale simile,

¹¹ (11 RAV)

con riferimento all'indice ESCS. Anche in questo caso si pone l'accento sugli esiti positivi considerati come punti di forza.

In generale, l'Istituto, nonostante i livelli più bassi rispetto alla media delle diverse aree geografiche, fa registrare, nel complesso, una differenza positiva nel punteggio rispetto alle scuole con contesto socio-economico culturale simile (ESCS), segno di un maggior livello di competenza a confronto con scuole similari [caso 312].

Nella sesta categoria, che è stata denominata **Disparità tra alunni**, sono presenti parole e segmenti ripetuti usati per segnalare l'assenza o la riduzione nel tempo di disparità tra i risultati degli alunni nelle prove. Anche rispetto a questa categoria si segnalano casi residuali¹² in cui la concentrazione in una o più classi di alunni con i punteggi più elevati è illustrata come un aspetto positivo. Seppur residuali, questi casi esistono; questa difficoltà a leggere e interpretare i dati offre però uno spaccato della necessità di sviluppare tale genere di competenze nelle scuole, al fine di evitare di condurre processi autovalutativi sviati e dunque a non individuare nel concreto i propri punti di debolezza e a porvi correttamente rimedio in un'ottica di miglioramento. Inoltre una maggiore capillarizzazione della valutazione esterna delle scuole, così come prevista dal DPR 80/2013, potrebbe supportare le scuole in tale direzione, garantendo un continuo confronto sui dati e una maggiore corretta interpretazione degli stessi.

La categoria **Cheating** è la settima e in essa sono contenute le parole "cheating" e

"opportunistici" e il segmento ripetuto "comportamenti opportunistici", menzionati dalle scuole in accezione positiva, ossia per descrivere l'assenza, o la scarsa diffusione, al momento delle prove, di comportamenti di cheating nelle classi. Si tratta di parole e segmenti ripetuti contenuti nel 23% dei frammenti.

Nelle prove standardizzate il valore del cheating risulta essere pari a 0 nella quasi totalità delle classi oppure, dove presente, al massimo dell'1,2 (registrata in una sola classe, quindi da considerarsi occasionale). ciò può far evidenziare una correttezza nella somministrazione delle prove e assenza di comportamenti opportunistici [caso 282]

Nell'ultima categoria (**Differenze per macro-area**, 22,1%) sono stati, infine, inseriti i termini e i segmenti ripetuti utilizzati dalle scuole per descrivere i punteggi ottenuti alle prove INVALSI o la distribuzione degli alunni sui diversi livelli facendo dei confronti per regione o macro-area.

¹² (4 RAV)

Tabella 3.1 – Temi estratti con la procedura Topic extraction nell’area Risultati nelle Prove standardizzate nazionali (punti di forza)

TEMI	KEYWORDS	EIGENVALUE	% VAR	FREQ	CASI	% CASI
PUNTEGGI IN ITALIANO E MATEMATICA	MEDIA; NAZIONALE; MATEMATICA; ITALIANO; RISULTATI; STANDARDIZZATE; SCUOLA; SUPERIORI; PRIMARIA; SECONDARIA; PROVE	4,03	1,29	6045	699	96,4%
LIVELLI DI APPRENDIMENTO	1; 2; QUOTA; LIVELLI; COLLOCATA; STUDENTI; INFERIORE; NAZIONALE; 5	4,61	1,19	2123	544	75,0%
DISPARITA' TRA ALUNNI	DOTATI; MENO; REGRESSIONE; PERMANENZA; CORSO; DISPARITÀ; ALUNNI; PIÙ; PIU; TRA; DISPARITA	4,46	1,38	1670	497	68,5%
ESITI UNIFORMI	ASSICURARE; UNIFORMI; RIESCE; VARIE; ESITI; TRA	3,09	0,93	1028	424	58,5%
LIVELLI RITENUTI AFFIDABILI	AFFIDABILE; RAGGIUNTO; RITENUTO; ABITUALE; ANDAMENTO; LIVELLO; CONOSCENDO	3,7	0,97	893	375	51,7%
CHEATING	COMPORTAMENTI; OPPORTUNISTICI; NON	2,47	0,81	334	221	30,5%
DIFFERENZE PER BACKGROUND SOCIO-ECONOMICO	ECONOMICO; SOCIO; SIMILE; CULTURALE; SCUOLE; BACKGROUND; CONTESTO	5,25	1,31	820	210	28,9%
CONFRONTO CON VALORI DI RIFERIMENTO	REGIONALI; VALORI; RIFERIMENTO	2,54	0,91	362	178	24,5%
CONFRONTO INTERNO SCUOLA	USCITA; INGRESSO; PRIMO; CICLO; GRADO	2,93	0,90	295	165	22,7%
ANALISI DELLA VARIANZA	VARIABILITÀ; INTERNO; CRITERI	2,71	0,89	215	149	20,5%
VALORI DI CONFRONTO	61; 9; 57; 54; 62; 7; 63; 6; 65; 0	6,63	1,48	444	119	16,4%
COMPOSIZIONE CLASSI	EQUILIBRIO; TASSO; COMPOSIZIONE; INDICA; OMOGENEITÀ; ALTO	3,5	0,91	143	86	11,8%
DIFFERENZE PER AREA GEOGRAFICA	NORD; OVEST; EST	2,69	0,75	184	64	8,8%

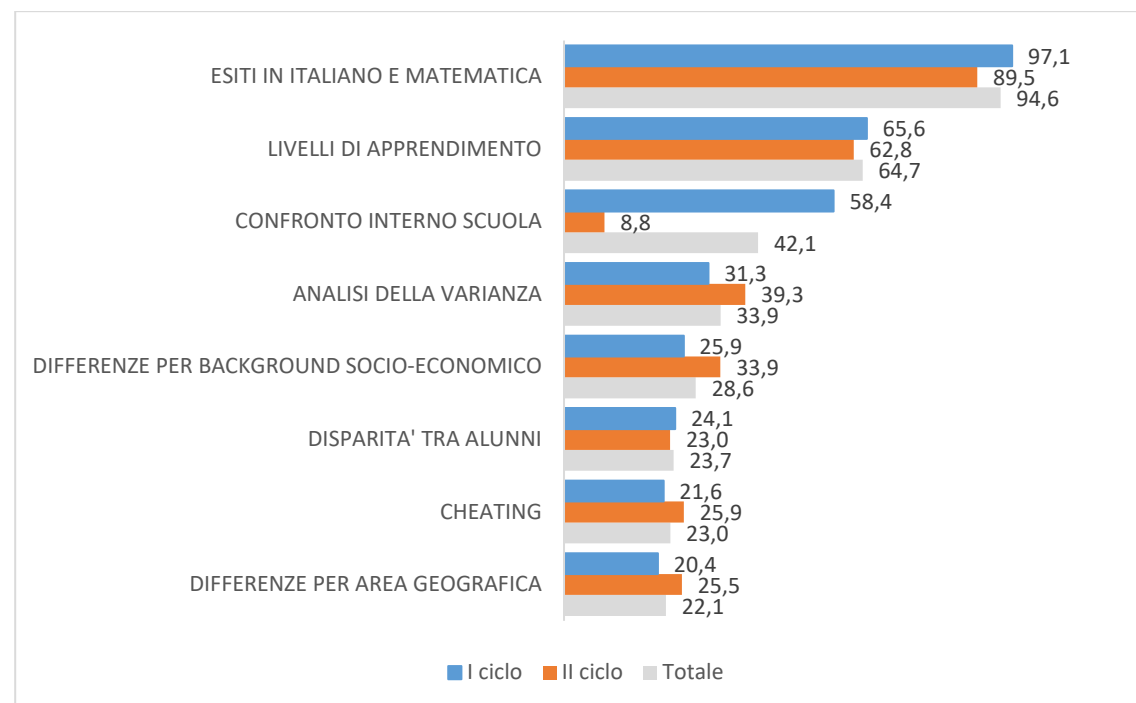
Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella 3.2 – Frequenza e diffusione delle categorie del vocabolario tematico dei punti di forza (Risultati nelle Prove standardizzate nazionali)

	FREQ.	% FREQ.	N. CASI	% CASI	TF • IDF
ESITI IN ITALIANO E MATEMATICA	5647	54,7%	686	94,6%	135,6
LIVELLI DI APPRENDIMENTO'	1564	15,1%	469	64,7%	295,9
CONFRONTO INTERNO SCUOLA	899	8,7%	305	42,1%	338,1
ANALISI DELLAVARIANZA	510	4,9%	246	33,9%	239,4
DIFFERENZE PER BACKGROUND SOCIO-ECONOMICO	754	7,3%	207	28,6%	410,5
DISPARITÀ TRA ALUNNI	339	3,2%	172	23,7%	211,8
CHEATING	206	2,0%	167	23,3%	131,4
DIFFERENZE PER AREA GEOGRAFICA	401	3,8%	160	22,1%	263,1

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Figura. 3.4 – Distribuzione delle categorie del vocabolario tematico (punti di forza), per ciclo di istruzione (val.%)



Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella 3.3 – Frequenza e diffusione delle keyword nelle categorie del vocabolario tematico dei punti di forza (Risultati nelle Prove standardizzate nazionali)

ESITI IN ITALIANO E MATEMATICA		
	FREQUENZA	% CASI
MATEMATICA	741	55,3%
RISULTATI	706	60,6%
ITALIANO	630	49,6%
PROVE	556	49,9%
MEDIA	496	37,2%
NAZIONALE	335	29,5%
STANDARDIZZATE	139	17,2%
PUNTEGGIO	97	10,3%
PUNTEGGIO_MEDIO	41	3,8%
LIVELLI DI ABILITA'		
	FREQUENZA	% CASI
STUDENTI	254	23,3%
LIVELLO	200	19,5%
LIVELLI	159	16,6%
INFERIORE	149	15,0%
LIVELLO_DI_RISULTATI	46	6,3%
LIVELLO_1	43	4,4%
LIVELLI_1	42	5,3%
LIVELLI_DI_APPRENDIMENTO	42	4,9%
LIVELLO_5	35	4,1%
LIVELLO_2	22	2,4%
QUOTA	20	2,4%
LIVELLO_RAGGIUNTO	19	2,2%
LIVELLI_PIÙ_BASSI	16	2,0%
1_E_2	15	1,2%
LIVELLO_1_E_2	14	1,6%
ANALISI DELLA VARIANZA		
	FREQUENZA	% CASI
TRA_LE_CLASSI	254	27,1%
VARIABILITÀ	176	17,9%
DENTRO_LE_CLASSI	52	6,0%
ASSICURARE	39	5,3%
OMOGENEITÀ	24	2,7%
VARIABILITÀ_DEI_PUNTEGGI	18	2,2%
UNIFORMITÀ	16	1,9%
COMPOSIZIONE	14	1,7%
VARIANZA_INTERNA	13	1,7%

CHEATING		
	FREQUENZA	% CASI
CHEATING	93	11,0%
COMPORTAMENTI_OPPORTUNISTICI	20	2,6%
CONFRONTO INTERNO SCUOLA		
	FREQUENZA	% CASI
PRIMARIA	353	31,0%
SECONDARIA	207	21,6%
LICEO	67	5,9%
GRADO	59	6,6%
PRIMO_GRADO	44	5,7%
PROFESSIONALE	37	3,0%
LINGUISTICO	19	2,3%
SCIENZE_UMANE	9	1,1%
CICLO	6	0,6%
DIFFERENZE PER BACKGROUND SOCIO-CULTURALE		
	FREQUENZA	% CASI
SOCIO	98	12,4%
SIMILE	81	10,2%
CULTURALE	76	10,0%
ESCS	88	9,7%
ECONOMICO	69	9,2%
BACKGROUND	70	8,1%
CONTESTO	61	7,5%
DIFFERENZE PER AREA GEOGRAFICA		
	FREQUENZA	% CASI
SUD	61	5,9%
NORD	54	5,1%
AREA	33	3,8%
OVEST	31	2,9%
EST	23	2,3%
GEOGRAFIC*	14	1,6%
DISPARITA' TRA ALUNNI		
	FREQUENZA	% CASI
DISPARITÀ	173	20,5%
PIÙ_DOTATI	42	5,5%
REGRESSIONE	14	1,7%
DISPARITA	6	0,8%
PIU_DOTATI	6	0,8%

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

3.3.2 I temi trattati nei campi aperti per la descrizione dei punti di debolezza

La procedura di *Topic extraction* è stata eseguita anche sui testi prodotti dalle scuole per la descrizione dei punti di debolezza del servizio offerto. Nello specifico, il valore soglia della frequenza delle parole pari a 10 ha permesso una copertura pari all'87,0% del corpus. Inoltre, è stata applicata una *stop list* per limitare l'analisi alle parole piene che ha consentito di concentrare l'analisi sul 46,0% delle parole presenti nella partizione.

L'analisi delle componenti principali ha permesso di individuare 11 temi principali (tab. 3.4.) che, in linea generale, coincidono con i temi principali che le scuole hanno trattato per descrivere i propri punti di forza e che, in questa partizione, sono invece evidenziati come punti di debolezza.

Un primo aspetto da segnalare, confrontando questa partizione con quella relativa ai punti di forza, riguarda la frequenza più ridotta con cui si distribuiscono le parole associate a ciascuna componente sui documenti. Come abbiamo visto in precedenza, le scuole tendono a fornire descrizioni più sintetiche quando sono chiamate a parlare dei propri punti di debolezza.

La componente a cui sono associate parole che ricorrono con la più alta frequenza (77%) è stata denominata **Variabilità tra le classi**, e contiene parole e segmenti ripetuti utilizzati dalle scuole per descrivere come punto di debolezza l'elevata variabilità dei punteggi tra le classi.

Tra le componenti estratte, quelle a cui sono associate parole che ricorrono nell'oltre metà dei documenti, fanno riferimento ai punteggi ottenuti dagli studenti alle prove e confrontati con i valori di riferimento, alla presenza di disparità tra gli studenti in base

ai punteggi, alla mancanza di esiti uniformi tra le classi, a una quota elevata di studenti collocata nei livelli 1 e 2.

A partire dai risultati emersi dalla *Topic extraction*, e attraverso una esplorazione dei testi con gli strumenti di *Keyword Retrieval* e *Keyword in context*, si è proceduto alla costruzione del vocabolario tematico relativo ai punti di debolezza. Il vocabolario presenta un buon livello di copertura ed è articolato in 8 categorie (tabb. 3.5, 3.6 e fig. 3.5). Dopo aver applicato la *stop list* risultano coperte l'85,6% delle frasi, il 90,9% dei paragrafi e il 97,5% dei casi (RAV).

Nel vocabolario dei punti di debolezza dell'area Risultati nelle Prove standardizzate nazionali, così come si registra per il vocabolario dei punti di forza, la categoria a cui sono associate *keyword* e segmenti ripetuti che ricorrono nella percentuale più elevata dei casi è denominata **Esiti in Italiano e Matematica** (88,7% dei casi). Le parole contenute in questa categoria sono utilizzate dalle scuole per descrivere i risultati ottenuti dagli studenti in Italiano e Matematica descritti come "negativi", "non soddisfacenti" dal confronto con i riferimenti territoriali.

I risultati ottenuti in entrambe le prove standardizzate nazionali dagli studenti delle classi quinte della primaria sono mediamente inferiori: in Italiano cinque classi su sei non raggiungono le medie provinciali, regionali e nazionali, in Matematica questo fenomeno si verifica in quattro classi su sei [caso 44].

Ricorrono nel 56% circa dei casi parole e frasi contenute nella categoria denominata **Livelli di apprendimento** e utilizzate per descrivere la situazione della scuola in merito alla distribuzione degli studenti per livelli di apprendimento.

Trattandosi di campi aperti per la descrizione di aspetti di debolezza, in questi frammenti si segnala, come punto di debolezza, una quota di alunni collocata nei primi due livelli superiore ai riferimenti territoriali.

Per le III della s.s. Le differenze tra i livelli si registrano in Italiano. Il livello 1 supera le medie e il 2 supera solo la media regionale, mentre il livello 5 si discosta di poco dalle medie. In Matematica il livello 1 è superiore alla media nazionale e si riscontra una percentuale bassa di alunni di livello 5[caso 4].

Si osserva, inoltre, che analizzando i dati in base al ciclo di istruzione, la percentuale di parole e segmenti ripetuti contenuti in questa categoria si presenta più elevata, in maniera statisticamente significativa, tra le scuole del II ciclo di istruzione (63,2% a fronte del 52,7% rilevato tra le scuole del I ciclo).

La terza categoria è stata denominata **Confronto interno scuola** (46,6%) che, come rilevato in diverse partizioni, comprende parole e segmenti ripetuti che le scuole utilizzano per illustrare la propria situazione effettuando dei confronti in base ai diversi ordini, plessi, indirizzi, classi. Anche in questa partizione si conferma una tendenza più frequente a utilizzare parole e frasi contenute in questa categoria da parte di scuole del I ciclo di istruzione (54,1%).

Fra i vari indirizzi che caratterizzano l'istituto si rilevano apprezzabili disparità :il livello di preparazione degli allievi del liceo linguistico è mediamente più alto rispetto a quello di scienze umane e soprattutto rispetto a quello del liceo economico sociale [caso 150].

La quarta categoria (41,4%), in ordine di frequenza, è stata denominata **Analisi**

della varianza, e in essa sono contenute parole e frasi utilizzate per approfondire il tema della variabilità dei punteggi ottenuti alle prove tra le classi e dentro le classi. Si osserva, inoltre, che analizzando i dati in base al ciclo di istruzione, la percentuale di parole e segmenti ripetuti contenuti in questa categoria si presenta più elevata in maniera statisticamente significativa tra le scuole del I ciclo di istruzione (44,2% a fronte del 35,6% rilevato tra le scuole del II ciclo). Un risultato che sembrerebbe suggerire che il tema della variabilità dei punteggi, declinato come punto di debolezza della scuola (vale a dire una elevata variabilità dei punteggi tra le classi), sia stato affrontato maggiormente dalle scuole del I ciclo.

Un' altra delle categorie del vocabolario è stata denominata **Differenze per background socio - culturale** in quanto è popolata da parole e frasi utilizzate dalle scuole per illustrare i risultati ottenuti dagli studenti nel confronto con scuole con contesto socio-culturale simile, con riferimento all'indice ESCS.

La categoria 6, **Disparità tra alunni**, è popolata da parole e segmenti ripetuti usati per segnalare la presenza, questa volta come aspetto di debolezza, di disparità nei risultati tra le classi o tra plessi e indirizzi diversi presenti nella scuola.

I risultati indicano che nella scuola primaria le prestazioni sono inferiori alle valutazioni quadrimestrali. Nella scuola secondaria di primo grado ancora qualche disparità tra i diversi livelli nelle varie classi[caso 152].

La settima categoria è stata denominata **Differenze per area geografica** (14,9%) e in essa sono stati inseriti termini e segmenti ripetuti utilizzati dalle scuole per descrivere i punteggi ottenuti alle prove o la distribuzione degli alunni sui diversi livelli

facendo dei confronti per regione o macro-area.

L'ultima categoria è composta dalla parola ***Cheating*** e dal segmento "comportamenti opportunistici" che ricorrono nel 12,5% dei documenti. In questo caso, la parola *cheating* è utilizzata per segnalare, come aspetto di debolezza, la diffusione di questo tipo di comportamento durante le prove

Tabella 3.4 - Temi estratti con la procedura Topic extraction nell'area Risultati nelle prove standardizzate nazionali (punti di debolezza)

TEMI	KEYWORDS	EIGENVALUE	% VAR	FREQ	CASI	% CASI
VARIABILITA' TRA LE CLASSI	CLASSI; TRA; VARIABILITÀ; VARIANZA	2,03	0,94	1926	565	77,9%
CONFRONTO CON VALORI DI RIFERIMENTO	MEDIA; NAZIONALE; REGIONALE; ITALIANO	2,65	1,09	1460	452	62,3%
ESITI UNIFORMI	UNIFORMI; RIESCE; ASSICURARE; VARIE; ESITI; NON	3,13	1,22	942	405	55,8%
DISPARITA' TRA ALUNNI	DOTATI; MENO; PERMANENZA; DISPARITÀ; CORSO; AUMENTO; ALUNNI; PIÙ; REGRESSIONE	3,28	1,35	1058	400	55,1%
RISULTATI PROVE INVALSI	NAZIONALI; STANDARDIZZATE; PROVE; REGIONALI	1,97	0,95	855	393	54,2%
CONFRONTO INTERNO SCUOLA	GRADO; SECONDARIA; PRIMO; SCUOLA; TERZE	2,34	0,94	958	387	53,3%
LIVELLI DI ABILITA'	1; 2; LIVELLI; QUOTA; COLLOCATA; STUDENTI; 5	7,38	1,16	981	340	46,9%
DIFFERENZE PER BACKGROUND	SOCIO; CULTURALE; SIMILE; ECONOMICO; SCUOLE; CONTESTO; BACKGROUND	4	1,63	566	178	24,5%
DIFFERENZE PER TIPOLOGIA DI LICEO	SCIENZE; UMANE; LINGUISTICO; LICEO	2	0,81	105	58	8,0%
CHEATING	COMPORTAMENTI; OPPORTUNISTICI; SOSPETTO; ABITUALE	2,48	0,99	76	40	5,5%
DIFFERENZE PER AREA GEOGRAFICA	NORD; OVEST; EST	2,06	0,83	105	37	5,1%

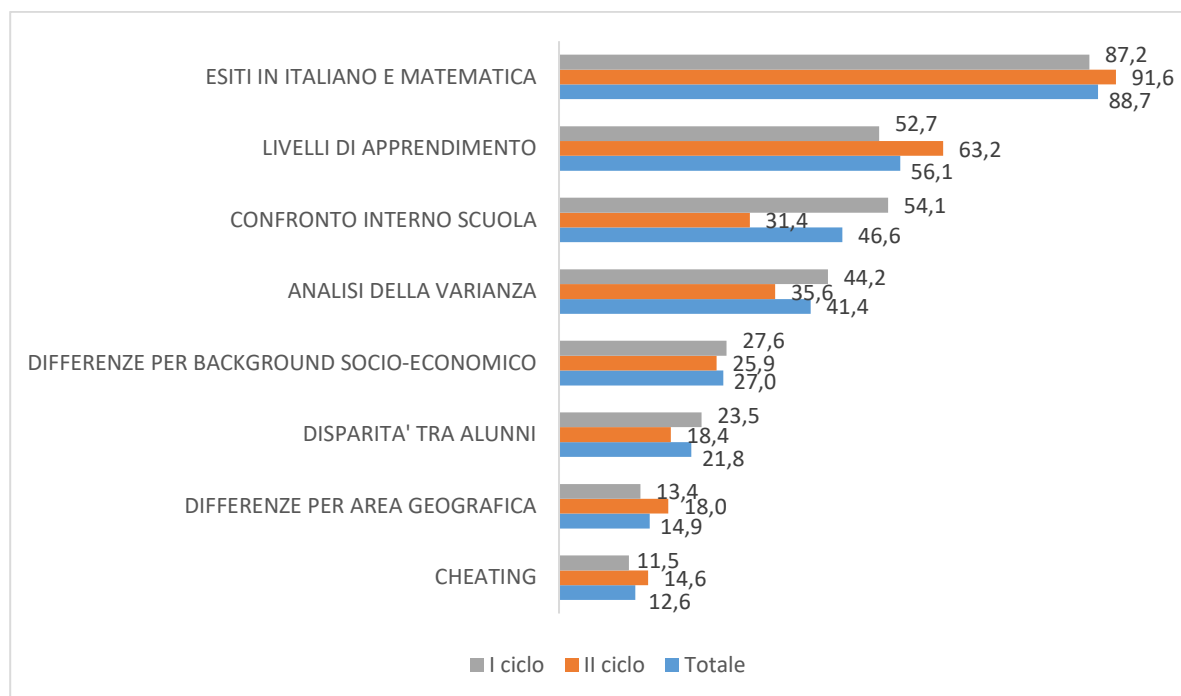
Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella 3. 5 - Frequenza e diffusione delle categorie del vocabolario tematico dei punti di debolezza (Risultati nelle Prove standardizzate nazionali)

	FREQ.	%FREQ.	N. CASI	% CASI	TF • IDF
ESITI IN ITALIANO E MATEMATICA	3695	50,6%	643	88,7%	192,6
LIVELLI DI ABILITA'	1076	14,7%	407	56,1%	269,8
CONFRONTO INTERNO SCUOLA	801	10,9%	338	46,6%	265,5
ANALISI DELLA VARIANZA	606	8,3%	300	41,4%	232,2
DIFFERENZE PER BACKGROUND SOCIO-ECONOMICO	543	7,4%	196	27,0%	308,5
DISPARITA' TRA ALUNNI	241	3,3%	158	21,8%	159,5
DIFFERENZE PER AREA GEOGRAFICA	216	2,9%	108	14,9%	178,6
CHEATING	113	1,5%	91	12,6%	101,8

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Figura 3.5 - Distribuzione delle categorie del vocabolario tematico (punti di debolezza), per ciclo di istruzione (val.%)



Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella 3.6 - Frequenza e diffusione delle keyword nelle categorie del vocabolario tematico dei punti di debolezza (Risultati nelle Prove standardizzate nazionali)

ESITI IN ITALIANO E MATEMATICA		
	FREQUENZA	% CASI
MATEMATICA	741	55,3%
RISULTATI	706	60,6%
ITALIANO	630	49,6%
PROVE	556	49,9%
MEDIA	496	37,2%
NAZIONALE	335	29,5%
STANDARDIZZATE	139	17,2%
PUNTEGGIO	97	10,3%
PUNTEGGIO_MEDIO	41	3,8%
LIVELLI DI ABILITA'		
	FREQUENZA	% CASI
STUDENTI	254	23,3%
LIVELLO	200	19,5%
LIVELLI	159	16,6%
INFERIORE	149	15,0%
LIVELLO_DI_RISULTATI	46	6,3%
LIVELLO_1	43	4,4%
LIVELLI_1	42	5,3%
LIVELLI_DI_APPRENDIMENTO	42	4,9%
LIVELLO_5	35	4,1%
LIVELLO_2	22	2,4%
QUOTA	20	2,4%
LIVELLO_RAGGIUNTO	19	2,2%
LIVELLI_PIÙ_BASSI	16	2,0%
1_E_2	15	1,2%
LIVELLO_1_E_2	14	1,6%
ANALISI DELLA VARIANZA		
	FREQUENZA	% CASI
TRA_LE_CLASSI	254	27,1%
VARIABILITÀ	176	17,9%
DENTRO_LE_CLASSI	52	6,0%
ASSICURARE	39	5,3%
OMOGENEITÀ	24	2,7%
VARIABILITÀ_DEI_PUNTEGGI	18	2,2%
UNIFORMITÀ	16	1,9%
COMPOSIZIONE	14	1,7%
VARIANZA_INTERNA	13	1,7%

CHEATING		
	FREQUENZA	% CASI
CHEATING	93	11,0%
COMPORTAMENTI_OPPORTUNISTICI	20	2,6%
CONFRONTO INTERNO SCUOLA		
	FREQUENZA	% CASI
PRIMARIA	353	31,0%
SECONDARIA	207	21,6%
LICEO	67	5,9%
GRADO	59	6,6%
PRIMO_GRADO	44	5,7%
PROFESSIONALE	37	3,0%
LINGUISTICO	19	2,3%
SCIENZE_UMANE	9	1,1%
CICLO	6	0,6%
DIFFERENZE PER BACKGROUND SOCIO-CULTURALE		
	FREQUENZA	% CASI
SOCIO	98	12,4%
SIMILE	81	10,2%
CULTURALE	76	10,0%
ESCS	88	9,7%
ECONOMICO	69	9,2%
BACKGROUND	70	8,1%
CONTESTO	61	7,5%
DIFFERENZE PER AREA GEOGRAFICA		
	FREQUENZA	% CASI
SUD	61	5,9%
NORD	54	5,1%
AREA	33	3,8%
OVEST	31	2,9%
EST	23	2,3%
GEOGRAFIC*	14	1,6%
DISPARITA' TRA ALUNNI		
	FREQUENZA	% CASI
DISPARITÀ	173	20,5%
PIÙ_DOTATI	42	5,5%
REGRESSIONE	14	1,7%
DISPARITA	6	0,8%
PIU_DOTATI	6	0,8%

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

RISULTATI IN SINTESI

- Uno degli aspetti considerati nell'analisi dei campi aperti per la descrizione dei punti di forza e di debolezza è relativo al modo in cui le scuole si sono approcciate alla redazione del RAV. Si osserva anche in questa area la presenza di frammenti di testo composti da un numero più elevato di parole nei campi aperti dedicati ai punti di forza e la tendenza a descrivere i punti di debolezza utilizzando mediamente un numero più contenuto di parole. Si tratta di un risultato a cui possono essere associate due dimensioni interpretative, la prima riconducibile a un piano comunicativo, e quindi alla tendenza, quando si parla di aspetti positivi a essere più prolissi, ridondanti, viceversa, quando si parla di aspetti negativi, ad essere più concisi; la seconda, relativa a una conoscenza più solida degli aspetti da descrivere come punti di forza e alla –probabile– maggiore inconsapevolezza di quelli che potrebbero essere definiti aspetti di debolezza.
- Quasi tutti i principali temi estratti attraverso la procedura di *Topic extraction* risultano coincidenti nelle due partizioni (punti di forza e punti di debolezza), di conseguenza i due rispettivi vocabolari sono composti da categorie popolate in grande parte dalle stesse parole e segmenti ripetuti. I temi trattati dalle scuole, che risultano coincidenti nelle due partizioni, sono stati, quindi, declinati dalle scuole in punti di forza e di debolezza in base ai risultati ottenuti dagli studenti alle prove standardizzate, tenendo conto delle differenze tra classi, plessi, cicli.
- I temi principalmente trattati sono legati a quelli proposti dalle domande guida alla redazione del RAV (gli esiti in Italiano e Matematica in relazione ai riferimenti territoriali, la distribuzione degli studenti sui diversi livelli, la presenza o assenza di disparità tra alunni nei punteggi ottenuti, la presenza o assenza di *cheating*).
- Il riferimento alla parola “Italiano” ricorre in maniera lievemente più frequente nei frammenti per la descrizione dei punti di forza e a farne riferimento sono maggiormente le scuole del I ciclo di istruzione.
- Il riferimento alla parola “Matematica” ricorre in maniera lievemente più frequente nei frammenti per la descrizione dei punti di debolezza e a farne riferimento sono in particolare le scuole del II ciclo di istruzione
- Tra le scuole del I ciclo di istruzione si rileva una più diffusa tendenza a parlare in linea generale dei risultati ottenuti dagli studenti tra i punti di forza della scuola
- Tra le scuole del II ciclo di istruzione si osserva una più diffusa tendenza a parlare, come punto specifico di debolezza, della distribuzione degli studenti sui livelli.

RIFLESSIONI E PROSPETTIVE

- Si osserva la tendenza da parte delle scuole a descrivere i propri punti di forza utilizzando un numero più elevato di parole, al contrario le descrizioni dei punti di debolezza appaiono più concise. Sarebbe interessante riflettere su possibili strategie da adottare per supportare le scuole nell'individuazione e nella descrizione dei propri punti di debolezza.
- È più diffuso il riferimento ai risultati in Matematica come punto di debolezza e ai risultati in Italiano come punto di forza. Sarebbe interessante verificare se, nell'ambito del campione, vi è coerenza tra i risultati emersi dall'analisi dei campi aperti di questa area e i risultati delle Prove Standardizzate Nazionali.
- Dall'esplorazione dei testi prodotti dalle scuole emergono casi, seppur residuali, di scuole che segnalano come aspetto di forza la presenza di una elevata variabilità dei punteggi tra le classi. Si tratta di una tema complesso e che chiama in causa aspetti relativi alla composizione delle classi e alla capacità della scuola di assicurare esiti il più possibile omogenei tra le classi. Sarebbe interessante approfondire il fenomeno effettuando degli studi di caso per comprendere come i nuclei interni di autovalutazione si avvicinano alla lettura dei dati dei RAV per descrivere la propria situazione.
- In un'ottica strategica, sarebbe utile anche comprendere quanto, una maggiore diffusione della valutazione esterna delle scuole, condotta dai NEV (Nuclei Esterni di Valutazione) possa supportare le scuole nell'interpretazione corretta e coerente dei propri dati. Alcune ricerche hanno, infatti, già messo in luce l'utilità del confronto fra nuclei di autovalutazione interni alle scuole e NEV in tale direzione.

BIBLIOGRAFIA

Bolasco S. (1999), *Analisi multidimensionale dei dati*, Carocci, Roma.

Bolasco S. (2013), *L'analisi automatica dei testi*, Carocci, Roma.

della Ratta-Rinaldi F. (2007), "L'analisi testuale computerizzata", in Cannavò L. e Frudà L., *Ricerca sociale. tecniche speciali di rilevazione, trattamento e analisi*, Carocci, Roma, pp.133-152.

Giuliano L., La Rocca G., (2008), *L'analisi automatica e semi-automatica dei dati testuali. Software e istruzioni per l'uso*, Led, Milano.

I percorsi valutativi delle scuole - Inquadramento teorico del RAV, 2014.
https://www.invalsi.it/invalsi/doc_stampa/27112014/Inquadramento_teorico_RAV.pdf.

APPENDICE DELLE TABELLE

Tabella 3.7 - Parole a frequenza elevata presenti nei punti di forza dell'area Risultati nelle prove standardizzate nazionali

	FREQ.	% FREQ.	N. CASI	% CASI	TF • IDF
DI	2818	4,1%	642	88,6%	148,8
E	2239	3,2%	605	83,5%	175,9
IN	1897	2,7%	549	75,7%	229,1
CLASSI	1393	2,0%	527	72,7%	193,0
LA	1214	1,8%	511	70,5%	184,4
LE	1099	1,6%	481	66,3%	195,8
ITALIANO	1027	1,5%	502	69,2%	163,9
RISULTATI	985	1,4%	497	68,6%	161,5
MATEMATICA	983	1,4%	488	67,3%	169,0
I	977	1,4%	469	64,7%	184,8
MEDIA	903	1,3%	391	53,9%	242,1
DELLA	890	1,3%	412	56,8%	218,4
SCUOLA	875	1,3%	453	62,5%	178,7
È	823	1,2%	388	53,5%	223,4
PER	820	1,2%	351	48,4%	258,3
PROVE	817	1,2%	498	68,7%	133,3
IL	795	1,1%	392	54,1%	212,3
A	789	1,1%	396	54,6%	207,2

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella 3.8 - Parole a frequenza media presenti nei punti di forza dell'area Risultati nelle prove standardizzate nazionali

	FREQ.	% FREQ.	N. CASI	% CASI	TF • IDF
CON	718	1,0%	380	52,4%	201,4
SONO	718	1,0%	403	55,6%	183,1
DELLE	697	1,0%	386	53,2%	190,8
NELLE	689	1,0%	395	54,5%	181,7
SI	689	1,0%	374	51,6%	198,1
CHE	670	1,0%	342	47,2%	218,6
ALLA	666	1,0%	344	47,5%	215,6
NAZIONALE	664	1,0%	343	47,3%	215,8
LIVELLO	633	0,9%	332	45,8%	214,7
DEL	589	0,9%	297	41,0%	228,3
TRA	518	0,7%	337	46,5%	172,3
RISPETTO	508	0,7%	276	38,1%	213,1
AL	495	0,7%	262	36,1%	218,8
NELLA	434	0,6%	237	32,7%	210,7

GLI	429	0,6%	307	42,3%	160,1
STUDENTI	420	0,6%	273	37,7%	178,2
DEI	408	0,6%	263	36,3%	179,7
ALUNNI	400	0,6%	269	37,1%	172,2
LIVELLI	398	0,6%	260	35,9%	177,3
SIA	395	0,6%	225	31,0%	200,7
UN	395	0,6%	242	33,4%	188,2
PRIMARIA	377	0,5%	234	32,3%	185,2
INVALSI	364	0,5%	273	37,7%	154,4
NEL	354	0,5%	243	33,5%	168,1
UNA	354	0,5%	231	31,9%	175,8
SUPERIORI	349	0,5%	251	34,6%	160,8
NAZIONALI	348	0,5%	243	33,5%	165,2
PIÙ	340	0,5%	220	30,3%	176,1
L	338	0,5%	238	32,8%	163,5
SUPERIORE	327	0,5%	181	25,0%	197,1
DEGLI	322	0,5%	228	31,5%	161,8
SECONDARIA	313	0,5%	219	30,2%	162,7
5	299	0,4%	160	22,1%	196,2
2	292	0,4%	178	24,6%	178,1
ESITI	290	0,4%	229	31,6%	145,1
1	281	0,4%	182	25,1%	168,7
DATI	280	0,4%	190	26,2%	162,8
NON	278	0,4%	210	29,0%	149,6
DELL	269	0,4%	191	26,3%	155,8
PUNTEGGIO	267	0,4%	137	18,9%	193,2
LINEA	266	0,4%	190	26,2%	154,7
ALLE	252	0,4%	188	25,9%	147,7
REGIONALE	233	0,3%	150	20,7%	159,4
STANDARDIZZATE	233	0,3%	211	29,1%	124,9
4	230	0,3%	139	19,2%	165,0
SECONDE	223	0,3%	165	22,8%	143,4
ISTITUTO	218	0,3%	161	22,2%	142,5
SCUOLE	213	0,3%	163	22,5%	138,1
NEI	212	0,3%	154	21,2%	142,6
INFERIORE	209	0,3%	144	19,9%	146,7
PROVA	209	0,3%	105	14,5%	175,4
CLASSE	208	0,3%	145	20,0%	145,4
ITALIA	208	0,3%	112	15,5%	168,7
3	205	0,3%	122	16,8%	158,7
QUINTE	204	0,3%	145	20,0%	142,6
PERCENTUALE	202	0,3%	125	17,2%	154,2

MEDIO	199	0,3%	118	16,3%	156,9
AI	197	0,3%	131	18,1%	146,4
QUELLI	197	0,3%	158	21,8%	130,4
DOTATI	190	0,3%	99	13,7%	164,3
ALL	183	0,3%	141	19,5%	130,1
DISPARITÀ	171	0,3%	156	21,5%	114,1
DA	167	0,2%	126	17,4%	126,9
RISULTA	164	0,2%	117	16,1%	129,9
GRADO	163	0,2%	133	18,3%	120,0
CHEATING	162	0,2%	142	19,6%	114,7
ANCHE	160	0,2%	125	17,2%	122,1
PUNTEGGI	155	0,2%	106	14,6%	129,4
QUANTO	153	0,2%	118	16,3%	120,6
RAGGIUNTO	150	0,2%	138	19,0%	108,1
SIMILE	147	0,2%	122	16,8%	113,8
VARIABILITÀ	146	0,2%	92	12,7%	130,9
AD	144	0,2%	123	17,0%	110,9
RIFERIMENTO	144	0,2%	96	13,2%	126,4
SOPRA	137	0,2%	99	13,7%	118,5
DATO	136	0,2%	76	10,5%	133,2
T	136	0,2%	19	2,6%	215,1
DAGLI	135	0,2%	123	17,0%	104,0
REGIONALI	135	0,2%	85	11,7%	125,7
MEDIE	133	0,2%	85	11,7%	123,8
SOCIO	128	0,2%	113	15,6%	103,3
O	125	0,2%	101	13,9%	107,0
UNIFORMI	122	0,2%	118	16,3%	96,2
MENO	119	0,2%	113	15,6%	96,1
SUD	119	0,2%	70	9,7%	120,8
VARIANZA	117	0,2%	85	11,7%	108,9
APPRENDIMENTO	115	0,2%	94	13,0%	102
VALORI	115	0,2%	58	8,0%	126,1
NORD	113	0,2%	64	8,8%	119,1
CORSO	109	0,2%	104	14,3%	91,9
CULTURALE	108	0,2%	97	13,4%	94,3
RISULTANO	108	0,2%	95	13,1%	95,3
ECONOMICO	107	0,2%	98	13,5%	93,0
ED	107	0,2%	91	12,6%	96,4
AFFIDABILE	106	0,2%	106	14,6%	88,5
BACKGROUND	106	0,2%	88	12,1%	97,1
6	102	0,2%	63	8,7%	108,2
VARIE	102	0,2%	97	13,4%	89,1

LORO	100	0,1%	90	12,4%	90,6
MENTRE	100	0,1%	73	10,1%	99,7
QUELLA	100	0,1%	75	10,3%	98,5
7	99	0,1%	64	8,8%	104,4
HANNO	99	0,1%	78	10,8%	95,9
PRIMO	99	0,1%	86	11,9%	91,7
RIGUARDA	99	0,1%	80	11,0%	94,8
ANALISI	96	0,1%	77	10,6%	93,5
DALLA	96	0,1%	84	11,6%	89,9
TERZE	96	0,1%	78	10,8%	93
LICEO	94	0,1%	48	6,6%	110,8
ESCS	93	0,1%	70	9,7%	94,4
DALL	91	0,1%	78	10,8%	88,1
MOLTO	90	0,1%	75	10,3%	88,7
QUELLO	90	0,1%	67	9,2%	93,1
TUTTE	86	0,1%	71	9,8%	86,8
NELL	85	0,1%	65	9,0%	89
RAGGIUNTI	85	0,1%	74	10,2%	84,2
9	84	0,1%	46	6,3%	100,6
INTERNO	84	0,1%	80	11,0%	80,4
POSITIVI	84	0,1%	78	10,8%	81,3
PUNTI	84	0,1%	55	7,6%	94,1
HA	83	0,1%	71	9,8%	83,8
MA	82	0,1%	70	9,7%	83,2
INFERIORI	81	0,1%	67	9,2%	83,8
RISULTATO	79	0,1%	58	8,0%	86,7
DOCENTI	77	0,1%	64	8,8%	81,2
ANNO	75	0,1%	61	8,4%	80,6
RAGGIUNGE	75	0,1%	71	9,8%	75,7
ALCUNE	74	0,1%	68	9,4%	76,1
ANNI	73	0,1%	66	9,1%	76
CONFRONTO	73	0,1%	55	7,6%	81,8
2014	72	0,1%	55	7,6%	80,6
PERCENTUALI	72	0,1%	54	7,5%	81,2
8	70	0,1%	52	7,2%	80,1
BASSA	70	0,1%	53	7,3%	79,5
DUE	70	0,1%	59	8,1%	76,3
SU	70	0,1%	48	6,6%	82,5
PERMANENZA	69	0,1%	68	9,4%	70,9
TUTTI	68	0,1%	65	9,0%	71,2
COMPETENZE	67	0,1%	58	8,0%	73,5
ANDAMENTO	66	0,1%	63	8,7%	70

<i>CONTESTO</i>	65	0,1%	58	8,0%	71,3
<i>SIGNIFICATIVAMENTE</i>	65	0,1%	47	6,5%	77,2
<i>OTTENUTI</i>	63	0,1%	49	6,8%	73,7
<i>SEZIONI</i>	63	0,1%	51	7,0%	72,6
<i>COLLOCA</i>	62	0,1%	44	6,1%	75,4
<i>DENTRO</i>	62	0,1%	55	7,6%	69,4
<i>OVEST</i>	62	0,1%	33	4,6%	83,2
<i>S</i>	61	0,1%	33	4,6%	81,9
<i>D</i>	60	0,1%	43	5,9%	73,6
<i>REGRESSIONE</i>	60	0,1%	59	8,1%	65,4
<i>0</i>	59	0,1%	34	4,7%	78,4
<i>61</i>	59	0,1%	30	4,1%	81,6
<i>ASSICURARE</i>	59	0,1%	57	7,9%	65,2
<i>LO</i>	59	0,1%	46	6,3%	70,7
<i>LEGGERMENTE</i>	58	0,1%	45	6,2%	70
<i>RITENUTO</i>	58	0,1%	58	8,0%	63,6
<i>V</i>	58	0,1%	41	5,7%	72,4
<i>DISTRIBUZIONE</i>	57	0,08%	41	5,66%	71,1

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella. 3.9 - Parole a frequenza elevata presenti nei punti di debolezza dell'area Risultati nelle prove standardizzate nazionali

	FREQ.	% FREQ.	N. CASI	% CASI	TF • IDF
<i>DI</i>	2494	4,4%	628	86,6%	155,6
<i>E</i>	1492	2,7%	549	75,7%	180,2
<i>IN</i>	1474	2,6%	511	70,5%	223,9
<i>CLASSI</i>	1328	2,4%	522	72,0%	189,5
<i>LA</i>	1015	1,8%	469	64,7%	192,0
<i>LE</i>	995	1,8%	473	65,2%	184,5
<i>MATEMATICA</i>	741	1,3%	401	55,3%	190,6
<i>PER</i>	731	1,3%	345	47,6%	235,8
<i>RISULTATI</i>	706	1,3%	440	60,7%	153,1
<i>I</i>	678	1,2%	395	54,5%	178,8
<i>SI</i>	655	1,2%	371	51,2%	190,6
<i>TRA</i>	645	1,2%	394	54,3%	170,8
<i>DELLE</i>	634	1,1%	364	50,2%	189,7
<i>ITALIANO</i>	630	1,1%	360	49,7%	191,5
<i>È</i>	605	1,1%	324	44,7%	211,6
<i>DELLA</i>	604	1,1%	350	48,3%	191,0
<i>SCUOLA</i>	601	1,1%	345	47,6%	193,8
<i>A</i>	592	1,1%	337	46,5%	197,0
<i>CHE</i>	583	1,0%	332	45,8%	197,8
<i>PROVE</i>	556	1,0%	362	49,9%	167,7
<i>IL</i>	527	0,9%	303	41,8%	199,7
<i>NON</i>	509	0,9%	322	44,4%	179,4
<i>UNA</i>	507	0,9%	308	42,5%	188,5
<i>SONO</i>	497	0,9%	317	43,7%	178,6
<i>MEDIA</i>	496	0,9%	270	37,2%	212,8
<i>CON</i>	453	0,8%	278	38,3%	188,6
<i>DEL</i>	437	0,8%	247	34,1%	204,4
<i>NELLE</i>	423	0,8%	270	37,2%	181,5
<i>ALLA</i>	403	0,7%	253	34,9%	184,3
<i>UN</i>	395	0,7%	253	34,9%	180,6
<i>ALUNNI</i>	386	0,7%	256	35,3%	174,5
<i>LIVELLO</i>	379	0,7%	230	31,7%	189,0
<i>RISPETTO</i>	373	0,7%	243	33,5%	177,1
<i>DEI</i>	371	0,7%	252	34,8%	170,3
<i>AL</i>	367	0,7%	229	31,6%	183,7
<i>PRIMARIA</i>	353	0,6%	225	31,0%	179,4
<i>NAZIONALE</i>	335	0,6%	214	29,5%	177,5
<i>NELLA</i>	334	0,6%	214	29,5%	177,0
<i>PIÙ</i>	320	0,6%	198	27,3%	180,4

<i>GLI</i>	295	0,5%	220	30,3%	152,8
<i>DEGLI</i>	280	0,5%	194	26,8%	160,3
<i>L</i>	279	0,5%	198	27,3%	157,3
<i>SIA</i>	278	0,5%	177	24,4%	170,2
<i>LIVELLI</i>	259	0,5%	180	24,8%	156,7
<i>STUDENTI</i>	254	0,5%	169	23,3%	160,6
<i>NEL</i>	251	0,5%	188	25,9%	147,1
<i>ESITI</i>	250	0,5%	198	27,3%	140,9
<i>CLASSE</i>	243	0,4%	157	21,7%	161,5
<i>INVALSI</i>	234	0,4%	184	25,4%	139,4
<i>DELL</i>	230	0,4%	176	24,3%	141,4
<i>INFERIORI</i>	225	0,4%	175	24,1%	138,9
<i>ALLE</i>	212	0,4%	159	21,9%	139,7
<i>2</i>	211	0,4%	134	18,5%	154,7
<i>DATI</i>	210	0,4%	162	22,3%	136,7
<i>SECONDARIA</i>	207	0,4%	157	21,7%	137,5
<i>VARIABILITÀ</i>	194	0,4%	141	19,5%	138,0
<i>PROVA</i>	193	0,3%	120	16,6%	150,8
<i>DA</i>	192	0,3%	148	20,4%	132,5
<i>1</i>	190	0,3%	132	18,2%	140,6
<i>ALL</i>	187	0,3%	144	19,9%	131,3
<i>ISTITUTO</i>	184	0,3%	131	18,1%	136,7
<i>NEI</i>	179	0,3%	144	19,9%	125,7
<i>SECONDE</i>	176	0,3%	130	17,9%	131,4
<i>DISPARITÀ</i>	173	0,3%	149	20,6%	118,9
<i>ANCHE</i>	167	0,3%	134	18,5%	122,4
<i>NAZIONALI</i>	157	0,3%	129	17,8%	117,7
<i>INFERIORE</i>	149	0,3%	109	15,0%	122,6
<i>QUINTE</i>	147	0,3%	112	15,5%	119,2

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015

Tabella 3.10 - Parole a frequenza media presenti nei punti di debolezza dell'area Risultati nelle Prove standardizzate nazionali

	FREQ.	% FREQ.	N. CASI	% CASI	TF • IDF
5	144	0,26%	88	12,1%	131,9
VARIANZA	144	0,26%	108	14,9%	119,1
DATO	143	0,25%	96	13,2%	125,6
PERCENTUALE	142	0,25%	103	14,2%	120,3
SOTTO	142	0,25%	106	14,6%	118,6
STANDARDIZZATE	139	0,25%	125	17,2%	106,1
PUNTEGGIO	138	0,25%	93	12,8%	123,1
RISULTA	134	0,24%	101	13,9%	114,7
SCUOLE	127	0,23%	104	14,3%	107,1
AI	126	0,22%	102	14,1%	107,3
MOLTO	126	0,22%	95	13,1%	111,2
QUELLI	124	0,22%	103	14,2%	105,1
RIFERIMENTO	122	0,22%	85	11,7%	113,6
SUPERIORE	121	0,22%	91	12,6%	109,1
ALCUNE	120	0,21%	105	14,5%	100,7
AD	119	0,21%	101	13,9%	101,9
QUANTO	117	0,21%	100	13,8%	100,7
DUE	113	0,20%	81	11,2%	107,6
MEDIO	111	0,20%	80	11,0%	106,3
PUNTEGGI	111	0,20%	77	10,6%	108,1
REGIONALE	109	0,19%	85	11,7%	101,5
MEDIE	105	0,19%	68	9,4%	107,9
3	104	0,19%	73	10,1%	103,7
GRADO	103	0,18%	86	11,9%	95,4
NELL	103	0,18%	87	12,0%	94,8
INTERNO	102	0,18%	89	12,3%	92,9
DOTATI	101	0,18%	51	7,0%	116,4
UNIFORMI	99	0,18%	94	13,0%	87,8
4	98	0,17%	64	8,8%	103,3
ITALIA	98	0,17%	58	8,0%	107,5
SOCIO	98	0,17%	90	12,4%	88,8
PLESSI	96	0,17%	76	10,5%	94,0
CHEATING	93	0,17%	80	11,0%	89,0
MA	93	0,17%	84	11,6%	87,1
ALTA	91	0,16%	74	10,2%	90,2
APPRENDIMENTO	90	0,16%	74	10,2%	89,2
SOPRATTUTTO	90	0,16%	81	11,2%	85,7
DOCENTI	88	0,16%	76	10,5%	86,2
ESCS	88	0,16%	71	9,8%	88,8

HA	88	0,16%	72	9,9%	88,3
DALLA	86	0,15%	74	10,2%	85,2
VARIE	86	0,15%	78	10,8%	83,3
MENO	83	0,15%	74	10,2%	82,3
RISULTANO	83	0,15%	68	9,4%	85,3
SU	82	0,15%	55	7,6%	91,8
SIMILE	81	0,14%	74	10,2%	80,3
HANNO	79	0,14%	64	8,8%	83,3
O	78	0,14%	67	9,2%	80,7
ANALISI	76	0,14%	66	9,1%	79,1
CULTURALE	76	0,14%	73	10,1%	75,8
DALL	75	0,13%	68	9,4%	77,1
PARTE	75	0,13%	67	9,2%	77,6
QUESTO	74	0,13%	66	9,1%	77,0
MENTRE	73	0,13%	60	8,3%	79,0
ED	72	0,13%	64	8,8%	75,9
PUNTI	71	0,13%	52	7,2%	81,2
BACKGROUND	70	0,12%	59	8,1%	76,3
PLESSO	70	0,12%	45	6,2%	84,5
SOLO	70	0,12%	64	8,8%	73,8
CUI	69	0,12%	61	8,4%	74,2
ECONOMICO	69	0,12%	67	9,2%	71,4
FRA	68	0,12%	55	7,6%	76,2
II	68	0,12%	41	5,7%	84,8
LORO	68	0,12%	64	8,8%	71,7
PRESENZA	68	0,12%	63	8,7%	72,1
CORSO	67	0,12%	58	8,0%	73,5
LICEO	67	0,12%	43	5,9%	82,2
PRIMO	67	0,12%	59	8,1%	7,03
ANNO	64	0,11%	52	7,2%	73,2
RIGUARDA	64	0,11%	51	7,0%	73,8
SE	64	0,11%	58	8,0%	70,2
DIVERSI	63	0,11%	59	8,1%	68,6
NUMERO	63	0,11%	56	7,7%	70,1
QUELLA	63	0,11%	53	7,3%	71,6
BASSI	62	0,11%	54	7,5%	69,9
COME	61	0,11%	54	7,5%	68,8
CONTESTO	61	0,11%	55	7,6%	68,3
DIFFERENZA	61	0,11%	47	6,5%	72,5
MAGGIORE	61	0,11%	50	6,9%	70,8
SUD	61	0,11%	43	5,9%	74,8
DENTRO	60	0,11%	51	7,0%	69,2

ESSERE	60	0,11%	58	8,0%	65,8
TECNICO	60	0,11%	35	4,8%	79,0
DIFFICOLTÀ	59	0,11%	54	7,5%	66,5
LINEA	58	0,10%	46	6,3%	69,5
SEMPRE	58	0,10%	51	7,0%	66,9
SEZIONI	58	0,10%	49	6,8%	67,9
V	57	0,10%	35	4,8%	75,0
VALUTAZIONE	57	0,10%	50	6,9%	66,2
EVIDENZA	56	0,10%	49	6,8%	65,5
VALORI	56	0,10%	45	6,2%	67,6
ALCUNI	55	0,10%	50	6,9%	63,9
DAGLI	55	0,10%	49	6,8%	64,4
DIFFERENZE	55	0,10%	49	6,8%	64,4
TUTTE	55	0,10%	53	7,3%	62,5
BASSA	54	0,10%	45	6,2%	65,2
DAL	54	0,10%	46	6,3%	64,7
NEGATIVO	54	0,10%	42	5,8%	66,8
NORD	54	0,10%	37	5,1%	69,8
TERZE	54	0,10%	43	5,9%	66,3
ALTO	53	0,09%	43	5,9%	65,0
COMPETENZE	53	0,09%	47	6,5%	63,0
EMERGE	53	0,09%	47	6,5%	63,0
PARTICOLARE	53	0,09%	51	7,0%	61,1
REGIONALI	53	0,09%	48	6,6%	62,5
8	52	0,09%	30	4,1%	71,9
SUPERIORI	52	0,09%	41	5,7%	64,9
BASSO	51	0,09%	47	6,5%	60,6
6	50	0,09%	35	4,8%	65,8
9	50	0,09%	35	4,8%	65,8
CONFRONTO	50	0,09%	41	5,7%	62,4
DEBOLEZZA	50	0,09%	47	6,5%	59,4
RISULTATO	49	0,09%	41	5,7%	61,1
SCOLASTICO	49	0,09%	45	6,2%	59,1
ANNI	48	0,09%	45	6,2%	57,9
ELEVATA	48	0,09%	39	5,4%	60,9
RAGGIUNTI	48	0,09%	46	6,3%	57,5
DISTRIBUZIONE	47	0,08%	43	5,9%	57,7
INOLTRE	47	0,08%	45	6,2%	56,7
NEGLI	47	0,08%	45	6,2%	56,7
TUTTI	47	0,08%	44	6,1%	57,2

Fonte: elaborazione INVALSI, dati RAV a.s. 2014/2015