

LE PAROLE DEL RAV

La metavalutazione del RAV:
obiettivi e metodologia





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

Il capitolo è stato redatto da Isabella Quadrelli

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.

1.1. Metavalutazione: cos'è e a che cosa serve

La metavalutazione è stata definita da Scriven come “la valutazione della valutazione”. Come riportato da Stufflebeam, Scriven introdusse questo termine nel 1969 applicandolo alla valutazione di un protocollo di *assesment* di programmi in ambito educativo (Stufflebeam [1974]2011). Da allora diversi studiosi si sono posti il problema di utilizzare modelli e criteri per effettuare valutazioni di secondo livello, ovvero di esprimere giudizi su programmi che si prefiggono di valutare specifici oggetti o politiche e su processi e strumenti di valutazione¹.

Diverse sono le accezioni e le definizioni di metavalutazione.

Stufflebeam ne ha proposto una definizione operativa: «the process of delineating, obtaining, and applying descriptive information and judgmental information—about the utility, feasibility, propriety, and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness, and social responsibility—to guide the evaluation and/or report its strengths and weaknesses» (Stufflebeam 2001: 185). Secondo questa definizione, la metavalutazione fornisce indicazioni sul merito e il valore della valutazione. La metavalutazione può svolgere una funzione formativa – quando i risultati vengono utilizzati per migliorare programmi o strumenti valutativi, o sommativa – quando si devono fare considerazioni sui costi e i vantaggi degli

interventi di valutazione. Essa può realizzarsi in itinere, per guidare e orientare il programma di valutazione oppure ex post facto.

Il Joint Committee on Standards for Educational Evaluation considera la metavalutazione come uno standard di riferimento per testare l'accuratezza del proprio operato da parte dei valutatori stessi. Essa pertanto si configura come un'attività interna al programma di valutazione, uno standard di qualità che dovrebbe essere applicato dagli stessi valutatori.

Bustelo distingue la metavalutazione, che ha come obiettivo la valutazione del processo di valutazione, da altri strumenti, come la meta analisi e l'*evaluation synthesis* che sono interessati ai risultati della valutazione. La metavalutazione si interessa quindi del modo in cui è stato condotto il processo di valutazione. Può interessarsi anche dei risultati raggiunti ma focalizzandosi sulla qualità degli stessi (le evidenze raccolte sono sufficienti? Le interpretazioni, conclusioni e raccomandazioni sono state effettuate a partire da un processo logico, tenendo in considerazione le evidenze disponibili?). Secondo l'autrice, la metavalutazione si riferisce a una sorta di controllo di qualità del processo di valutazione e alla comprensione della funzione della valutazione nel ciclo delle politiche (Bustelo 2002). La metavalutazione può essere effettuata sia da soggetti interni, ovvero dai valutatori impegnati a realizzare il programma di valutazione, sia da soggetti esterni. Essa inoltre può avere come oggetto

¹ Si pensi alla pubblicazione da parte di diversi enti e associazioni deputati alla valutazione di criteri per valutare la qualità degli strumenti di valutazione,

l'utilizzo di forme di auditing e di disegni sperimentali o quasi sperimentali per valutare modelli valutativi alternativi.

il disegno dello studio valutativo, il processo di valutazione e i risultati della valutazione (Bustelo 2002).

I criteri che dovrebbero orientare una buona metavalutazione, secondo Stufflebeam, sono, in parte, quelli che riguardano la valutazione di ogni progetto di ricerca: la validità interna ed esterna e la possibilità di generalizzare i risultati. A ciò si aggiungono criteri specifici relativi alle finalità della valutazione: i risultati sono utili per il pubblico al quale sono destinati? I benefici dei risultati della valutazione sono maggiori dei costi necessari per ottenerli? (Stufflebeam [1974]2011).

Bustelo individua criteri specifici in relazione alle due funzioni della metavalutazione: controllare la qualità del processo di valutazione e comprendere la funzione della valutazione. Rispetto alla prima funzione i criteri riguardano l'adeguatezza della metodologia utilizzata e la correttezza e qualità dei risultati della valutazione. Per quanto riguarda la comprensione della funzione della valutazione occorre considerare il grado di integrazione della valutazione nelle istituzioni e nelle organizzazioni coinvolte nel programma e l'utilità della valutazione per il miglioramento, l'accountability e la comprensione delle politiche pubbliche (Bustelo 2002).

1.2. Obiettivi della metavalutazione del RAV

La metavalutazione del RAV si colloca nell'ambito del progetto PON Valu.E (Valutazione/Autovalutazione Esperta), realizzato grazie ai fondi PON 2014-2021, articolato in azioni di studio, ricerca e servizio, con la finalità complessiva di promuovere e sostenere la qualità del Sistema Nazionale di Valutazione. In particolare, la metavalutazione rientra tra i programmi di ricerca dell'Azione 1

“Valutare la valutazione delle scuole” e si prefigge l'obiettivo specifico di approfondire la conoscenza dei processi di autovalutazione e l'uso degli strumenti di valutazione da parte delle scuole.

In questo contesto la metavalutazione si prefigge di valutare la qualità del processo di autovalutazione realizzato dalle scuole. Nell'ambito di tale finalità rientra anche lo studio dell'efficacia degli strumenti realizzati dall'INVALSI per supportare le scuole nella conduzione del percorso di autovalutazione.

Il focus dello studio metavalutativo è principalmente formativo. L'obiettivo è infatti di comprendere se gli strumenti hanno funzionato, se hanno orientato le scuole nella realizzazione di un percorso di autovalutazione *data-driven* e se sono stati in grado di stimolare un processo riflessivo, nell'ottica di un miglioramento continuo degli strumenti e delle misure di sostegno rivolte alle scuole.

La prospettiva che abbiamo utilizzato è quella della valutazione realistica (Pawson & Tilley 1997) che verifica non solo se un programma o un intervento abbia funzionato ma anche in che misura, per quali soggetti e in quali circostanze. Secondo la valutazione realistica un programma funziona (ha risultati positivi) solo se è in grado di introdurre le idee e le opportunità appropriate ('meccanismi') per determinati gruppi nei contesti sociali e culturali precisi (contesti). In luogo di una causalità di tipo lineare e sequenziale, la valutazione realistica considera i meccanismi generativi che in un determinato contesto portano a un cambiamento, in linea con gli obiettivi del programma di intervento.

Un meccanismo permette di spiegare quale elemento o aspetto del programma lo rende efficace poiché è in grado di tenere conto della natura situata del programma stesso, dei processi micro e delle condizioni

macrosociali, che costituiscono la realtà nella quale esso opera, e di dimostrare come i risultati del programma stesso derivino dalle scelte degli stakeholders (reasoning) e dalla loro capacità (risorse) di mettere tali idee in azione.

Le scuole, pertanto, sono considerate come soggetti dotati di agency, che interpretano le richieste e gli stimoli provenienti dall'esterno (dai policy maker), a partire dai significati condivisi localmente e dagli scopi che si prefiggono, che possono anche differire da quelli previsti dalla valutazione. Il programma oggetto di valutazione, nel nostro caso, è l'autovalutazione d'istituto come definita dal D.p.R. 80/2013 che ha istituito il Sistema nazionale di valutazione in materia di istruzione e formazione. Il Decreto ha attribuito una grande rilevanza all'autovalutazione facendola diventare il punto di partenza dell'intero percorso di valutazione definito dall'art. 6. Con l'autovalutazione le scuole individuano le priorità e gli obiettivi di miglioramento; le azioni realizzate a seguito dell'autovalutazione rappresentano l'oggetto della rendicontazione sociale delle scuole. La valutazione esterna si concentra sugli obiettivi e le priorità individuate dalle scuole per confermarle oppure reconsiderarle.

L'INVALSI ha realizzato alcuni strumenti di supporto all'autovalutazione. In linea con gli scopi del D.p.R. 80/2013 e dei successivi documenti di orientamento prodotti dal MIUR² essi si pongono l'obiettivo di realizzare un'autovalutazione *data-driven*, ovvero una valutazione che parta dall'analisi di evidenze osservabili, e di favorire la riflessione interna alla scuola per l'individuazione degli obiettivi di miglioramento a partire da un'analisi critica dei punti di forza e di debolezza. Tali

strumenti comprendono: un Quadro di riferimento articolato in macroaree: contesto, esiti e processi; un sistema di indicatori riferiti a tutte le dimensioni del quadro di riferimento e un format strutturato (RAV) che dovrebbe rappresentare una guida per le scuole per: riflettere sui dati, esprimere un giudizio sul livello di qualità raggiunto in ciascuna area di esito e di processo e, infine, individuare le priorità e gli obiettivi di miglioramento.

Il modello di autovalutazione proposto dall'INVALSI si configura come *multiple measure* (Ehren & Swanborn 2012) poiché considera diversi tipi di indicatori, sia di esito, sia di processo. L'articolazione in esiti e processi del quadro di riferimento elaborato dall'INVALSI è stata utilizzata per evitare forme di riduzionismo che potrebbero far coincidere la qualità delle scuole unicamente con i risultati ottenuti dagli studenti; sono infatti anche considerati i processi didattici, quelli gestionali e organizzativi.

Inoltre, si tratta di un modello di autovalutazione semistrutturato in quanto prevede sia strumenti maggiormente strutturati (un set di indicatori predefinito, le rubriche di valutazione, il format per l'individuazione delle priorità e degli obiettivi di processo) sia spazi per l'espressione delle scuole. Le scuole, infatti, possono inserire ulteriori dati e indicatori, hanno la possibilità di descrivere in spazi aperti i propri punti di forza e di debolezza e di giustificare la scelta del livello di qualità in ciascuna area. Questo tipo di impostazione ha l'obiettivo sia di strutturare un processo riflessivo a partire da evidenze osservabili, sia di promuovere la partecipazione delle scuole attraverso l'inserimento di informazioni ritenute

² In particolare, la direttiva n. 11 del 18 settembre 2015 e le linee guida Orientamenti per l'elaborazione del Rapporto di Autovalutazione, prot. 1738/15.

rilevanti e la narrazione di alcuni aspetti relativi al proprio contesto e alle azioni realizzate.

Il risultato atteso dal programma è la produzione di un rapporto di autovalutazione che illustri il percorso di riflessione effettuato dalla scuola, nel quale sono indicati, a partire dai punti di debolezza individuati con l'autovalutazione, i traguardi di sviluppo a medio termine e gli obiettivi operativi a breve termine che la scuola intende realizzare per migliorare la propria situazione.

I meccanismi ipotizzati per raggiungere questo obiettivo riguardano gli aspetti legati all'implementazione del percorso autovalutativo, e più in generale all'organizzazione scolastica, gli elementi di tipo culturale e attitudinale, nonché le caratteristiche dello strumento di autovalutazione. Questi fattori sono stati individuati come rilevanti dalla letteratura che si è occupata di analizzare la qualità della valutazione e l'utilizzo di strumenti di autovalutazione da parte delle scuole. Ad esempio, nello studio di Schildkamp e Vissher (2009), i fattori legati ai processi di implementazione e organizzazione delle scuole rivestono un ruolo rilevante nell'influenzare l'utilizzo degli strumenti di autovalutazione nel corso del tempo. Cooksey e Mark (2012) ritengono che la disponibilità di risorse e supporto per la valutazione siano fattori che influenzano la qualità della valutazione. Anche la disponibilità di competenze nell'ambito della valutazione rappresenta un fattore associato alla qualità dei risultati realizzati (Cooksey & Mark 2012). Ad esempio, Schildkamp e Vissher (2009) ritengono che l'utilizzo dello strumento di autovalutazione ZEBO³ in Germania sia stato influenzato da aspetti di tipo culturale, ovvero dal

significato attribuito alla valutazione da parte del personale scolastico e dalla disponibilità dello stesso all'innovazione e al cambiamento. Nello stesso studio è emerso che la percezione rispetto alla capacità dello strumento di intercettare aspetti ritenuti fondamentali per il funzionamento scolastico è associata a un utilizzo continuativo dell'autovalutazione nel corso del tempo.

Nel nostro studio abbiamo ipotizzato che gli strumenti di autovalutazione proposti rappresentino una risorsa per il supporto del percorso di autovalutazione. Essi infatti mettono a disposizione delle scuole un quadro di riferimento che individua le variabili e i criteri per definire la qualità del servizio scolastico, dati e benchmark di riferimento per valutare la posizione della singola scuola e un percorso semistrutturato per orientare la riflessione della scuola nelle diverse aree di interesse. L'utilizzo ottimale degli strumenti presuppone la presenza di ulteriori meccanismi, tra i quali:

- l'implementazione di un percorso di autovalutazione partecipato che coinvolga esponenti chiave delle diverse componenti scolastiche (docenti, personale ATA, genitori, studenti, eventuali stakeholder esterni);
- la disponibilità di risorse da dedicare alla valutazione (tempo, organizzazione funzionale del lavoro dei docenti, possibilità di convogliare risorse economiche interne per supportare la valutazione, disponibilità di supporto esterno...);
- la disponibilità a livello di scuola di competenze nell'ambito della

³ L'acronimo fa riferimento all'autovalutazione nella scuola primaria

- valutazione e dell'analisi e interpretazione dei dati;
- la presenza di un atteggiamento positivo verso la valutazione e l'innovazione.

1.3 Il disegno della ricerca

Con questo studio si è inteso valutare la qualità del percorso di autovalutazione partendo dai “prodotti” dell'autovalutazione stessa, ovvero dai rapporti realizzati dalle scuole (RAV). L'analisi si concentrerà quindi sui risultati dell'autovalutazione (output) con il fine di valutare la loro qualità e indirettamente l'appropriatezza e coerenza del percorso di analisi e interpretazione realizzato dalle scuole.

La ricerca ha riguardato i RAV prodotti nell'anno scolastico 2014-2015 e pubblicati nel portale del MIUR *Scuola in chiaro*. L'analisi è stata condotta su un campione casuale, stratificato per macroarea geografica, composto da 725 istituzioni scolastiche di primo e secondo ciclo. Il metodo di elaborazione dei dati utilizzato è l'analisi testuale computer assistita dei campi aperti del RAV.

Le domande di ricerca che hanno orientato la metavalutazione riguardano tre aspetti principali, ossia:

- 1) l'efficacia degli strumenti. Poiché l'obiettivo dell'insieme degli strumenti prodotti dall'INVALSI era quello di supportare le scuole nella realizzazione di un percorso di autovalutazione *data-driven*, riflessivo e coerente, sono state considerate la validità interna e l'affidabilità dei risultati dell'autovalutazione. Pertanto, lo

scopo dell'analisi è stato quello di verificare quanto l'autovalutazione realizzata dalle scuole sia risultata coerente con gli standard previsti dalla comunità di esperti. Le domande alle quali abbiamo cercato di dare risposta sono: le scuole hanno utilizzato i dati per individuare i propri punti di forza e di debolezza? Il percorso autovalutativo risulta coerente?

Studi precedenti condotti sui rapporti di autovalutazione realizzati dalle scuole partecipanti al progetto Vales⁴ hanno evidenziato che poco più di un terzo delle scuole aveva fatto riferimento ai dati, utilizzandoli in una prospettiva comparativa, per individuare i propri punti di forza e di debolezza, mentre la maggior parte aveva motivato il giudizio espresso senza fare riferimento a precise evidenze empiriche (Muzzioli et al. 2016);

- 2) il grado di riflessività e articolazione dell'analisi. L'attivazione di un processo riflessivo a livello di scuola è un presupposto per la realizzazione di azioni di miglioramento a partire dai risultati della valutazione. Il modello di autovalutazione realizzato dall'INVALSI ha la finalità di stimolare la riflessione attraverso domande guida e dati di riferimento per confrontare la situazione della scuola; esso intende anche promuovere un orientamento attivo delle scuole mediante la narrazione di aspetti ritenuti significativi e l'inserimento di informazioni aggiuntive considerate rilevanti.

⁴ Per approfondimenti sul progetto Vales (Valutazione e Sviluppo della Scuola): <https://www.invalsi.it/invalsi/ri/vales/>

Con l'ausilio dell'analisi testuale si cercherà di valutare il livello di approfondimento dell'analisi ovvero la capacità di leggere i dati in relazione ai benchmark di riferimento, alle caratteristiche del contesto e di inserirli all'interno di un quadro che contempla la possibilità del cambiamento e del miglioramento. Inoltre, si cercherà di individuare dai testi la presenza di temi, problemi e aspetti proposti dalle scuole, non previsti dal quadro di riferimento e dagli indicatori definiti centralmente;

- 3) l'impatto degli strumenti, intesi come testi che devono essere a loro volta decodificati e interpretati, sul processo di autovalutazione.

Un precedente studio esplorativo condotto sull'utilizzo del RAV da parte delle scuole ha evidenziato che le rubriche di valutazione svolgono una funzione di strutturazione e di "frame" sulle considerazioni proposte dalle scuole per motivare il punteggio autoattribuito nella scala di valutazione. I descrittori delle rubriche di valutazione svolgono infatti un ruolo importante nel definire i temi che vengono ripresi dalle scuole nelle motivazioni dei giudizi e nell'influenzare il lessico utilizzato nei campi aperti del RAV (INVALSI 2016). Sono necessarie ulteriori analisi per comprendere l'impatto di questi strumenti; infatti, se da un lato la riproposizione da parte delle scuole dei temi proposti dalle rubriche rappresenta un indicatore di validità del processo di autovalutazione (abbiamo indicazioni che le scuole hanno riflettuto proprio sui temi ritenuti rilevanti dal quadro di riferimento), dall'altro la scarsa articolazione e

contestualizzazione dell'analisi possono essere un indicatore di "ritualismo" (le scuole possono aver utilizzato un approccio maggiormente burocratico all'autovalutazione riducendo al minimo il processo di riflessione sui dati).

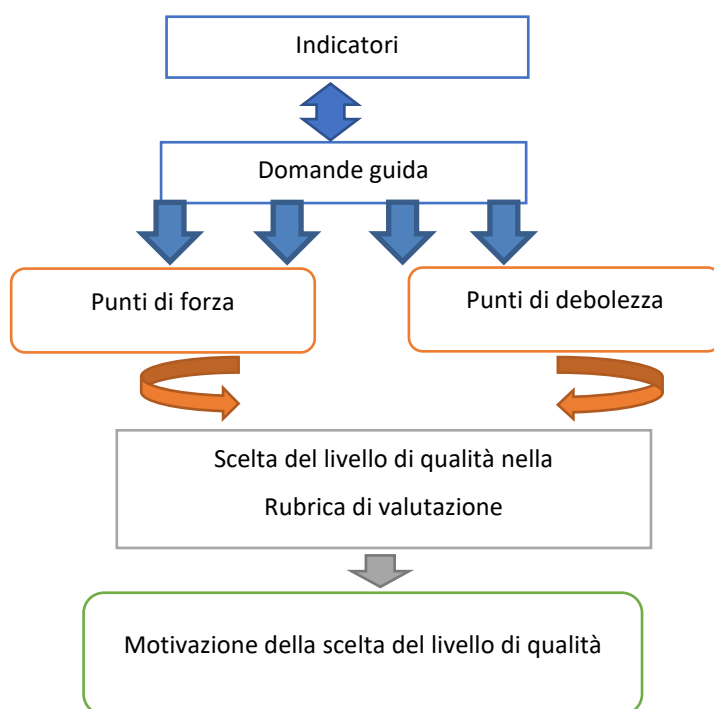
La metavalutazione proposta prende in considerazione i principali campi aperti del RAV e in particolare:

- i campi dedicati alla descrizione dei punti di forza e di debolezza;
- i campi dedicati alla motivazione dei giudizi attribuiti nelle rubriche di valutazione;
- il campo aperto della sezione 5 del RAV dedicato alla motivazione della scelta delle priorità.

I testi prodotti dalle scuole e inseriti in questi campi permettono di comprendere se e come le scuole abbiano utilizzato i dati (campi punti di forza e di debolezza ed campo motivazione del giudizio); il ruolo svolto dagli strumenti (che ruolo hanno svolto le domande guida? e i descrittori delle rubriche?); infine permettono di cogliere la coerenza del percorso di autovalutazione orientato al miglioramento (campo aperto dedicato alla motivazione della scelta delle priorità) e di individuare il contributo specifico portato dalla scuola.

Il modello di autovalutazione prevede, infatti, che le scuole descrivano nei campi aperti "Punti di forza" e "Punti di debolezza" gli aspetti positivi e le criticità relativi ad ogni area e subarea del Quadro di riferimento, partendo dall'analisi degli indicatori forniti dall'INVALSI, di altre evidenze empiriche individuate dalle scuole e dalle domande guida proposte. Dalla considerazione dei punti di forza e di debolezza dovrebbe scaturire la scelta del livello di qualità nella rubrica di valutazione (scala da 1 – molto critica a 7 – eccellente) e la motivazione di tale scelta (Fig. 11).

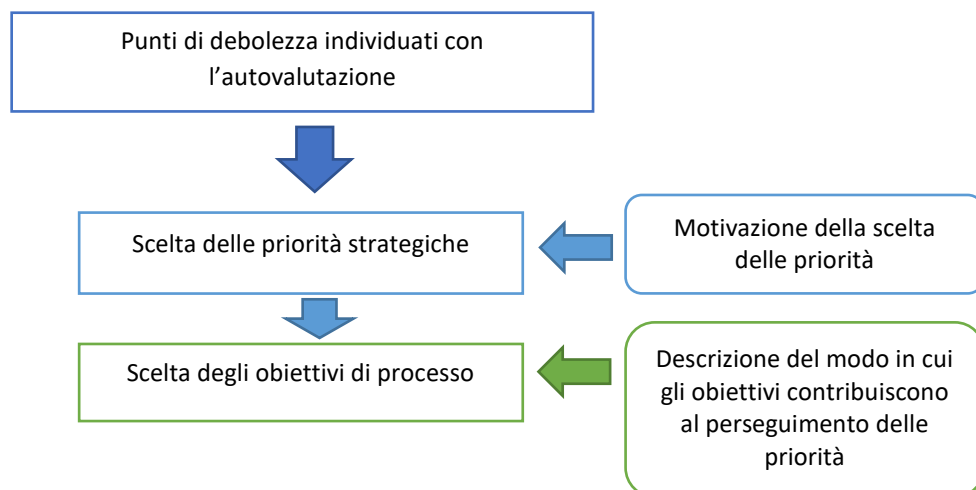
Fig. 1.1 La struttura del RAV: dai dati alla motivazione del giudizio



I risultati dell'autovalutazione dovrebbero rappresentare il punto di partenza per l'individuazione delle priorità strategiche e

degli obiettivi operativi che contribuiscono alla realizzazione delle priorità stesse (Fig. 1.2).

Fig. 1.2 La struttura del RAV. Dall'autovalutazione alla scelta degli obiettivi e delle priorità



Data la vastità del corpus del testo, l'analisi è stata condotta su alcune aree del Quadro

di riferimento, come evidenziate in Tab. 1.1, a partire da alcuni criteri teorici.

Tab. 1.1 Aree del Quadro di riferimento selezionate per l'analisi testuale

Esiti	Risultati scolastici
Processi – Pratiche educative e didattiche	Risultati nelle prove standardizzate nazionali Curricolo, progettazione, valutazione Ambiente di apprendimento
Processi – Pratiche gestionali e organizzative	Orientamento strategico e organizzazione della scuola Sviluppo e valorizzazione delle risorse umane

In primo luogo, abbiamo ritenuto necessario considerare aree di risultato e aree di processo e, tra queste, sia quelle relative alle pratiche didattiche ed educative, sia quelle che riguardano le pratiche gestionali ed organizzative. Ciò permette di offrire una visione dei principali ambiti semantici rilevanti per la valutazione delle scuole, inseriti nel Quadro di riferimento. Alcuni studi precedenti, inoltre, hanno mostrato che le scuole utilizzano con maggiore frequenza dati e indicatori relativi ai risultati scolastici rispetto ai dati riguardanti i processi (Muzzioli et al. 2016); includere nell'analisi le aree di processo è, quindi, necessario per rilevare come le scuole si dispongono verso l'utilizzo dei dati di processo. L'area Risultati nelle prove standardizzate nazionali è stata considerata nello studio in quanto rappresenta per le scuole un ambito particolarmente rilevante dal punto di vista simbolico – visto il significato associato alle prove stesse e alla percezione di “oggettività” dei dati sui risultati degli studenti. D'altro canto, per il lavoro di formazione svolto da tempo sull'utilizzo dei dati delle prove INVALSI, in quest'area le scuole dovrebbero possedere maggiori competenze di analisi e interpretazione dei dati. Altre aree, come Curricolo, progettazione e valutazione e Sviluppo e valorizzazione delle risorse umane sono importanti perché toccano aspetti centrali della vita scolastica e del

lavoro degli insegnanti e perché interessate da recenti cambiamenti normativi.

1.4 L'analisi testuale: metodologia e fasi dell'analisi

In questo paragrafo sono presentate le finalità e le caratteristiche dell'analisi testuale quale metodologia che si basa sull'apporto di molteplici discipline e che integra diverse tecniche e strumenti al fine di realizzare un'analisi automatica o semiautomatica dei testi. E' inoltre descritta l'applicazione dell'analisi testuale al progetto di metavalutazione del RAV, illustrando le caratteristiche del corpus testuale e le diverse fasi dell'analisi realizzata con il software Wordstat di QDAMiner.

1.4.1 L'analisi testuale

L'analisi testuale si riferisce al complesso dei metodi e delle tecniche utilizzati per effettuare un'analisi automatica o semiautomatica di un insieme di testi, con lo scopo di individuare informazioni rilevanti. Essa si sviluppa grazie al contributo di diverse discipline tra cui la linguistica, l'analisi del discorso, la statistica, l'informatica e la metodologia della ricerca socio-economica.

Nei suoi utilizzi più recenti l'analisi testuale si fonda su un approccio formale, ovvero sull'analisi delle forme grafiche contenute

nel testo. Questo approccio, sviluppato a partire dagli studi di Lebart e colleghi alla fine degli anni '80 (Lebart, Salem & Berry 1998), è indipendente dalla lingua utilizzata poiché privilegia le forme grafiche e i loro rapporti – considerati su un piano statistico - per arrivare ai significati (Alivernini 2008). Nelle versioni più recenti i software disponibili per l'analisi testuale permettono non solo di quantificare le occorrenze delle parole ma di considerare il contesto di utilizzo delle parole stesse, con evidenti vantaggi sulla possibilità di estrarre i significati presenti in un corpus testuale (Wiedemann 2013). Il significato di una parola è dato dalle parole che la circondano o dalla selezione di quelle che possono sostituirla nella frase; «il senso sottostante un testo/discorso, di cui s'intende dare una rappresentazione con metodi statistici, è costituito dal sistema dei significati che 'si tiene' (come una sorta di ecosistema) sulla base dell'insieme delle co-occorrenze nell'intero corpus di dati testuali» (Bolasco 2003:2)

L'analisi testuale, basandosi sull'analisi automatica dei testi, può essere effettuata su insiemi di testi di grandi dimensioni. Essa infatti non presuppone la lettura dei materiali da analizzare ma si basa sull'applicazione di modelli (metriche) per l'individuazione di informazioni rilevanti. Secondo Bolasco, una metrica è una «rappresentazione, ora di tipo lessicale, e quindi paradigmatica del linguaggio utilizzato, ora di tipo testuale e quindi sintagmatica del discorso, inteso come

'senso' d'insieme, presente nel corpus» (Bolasco 2013:42).

L'analisi testuale si basa sull'utilizzo di molteplici tecniche di tipo statistico, di *information retrieval*⁵ e di *information extraction*⁶. L'utilizzo di queste tecniche, insieme alle risorse messe a disposizione dall'informatica, permettono di applicare criteri di ricerca e di osservazione omogenei su tutto il corpus testuale e rendono possibile quantificare la diffusione di alcuni elementi e concetti.

L'obiettivo dell'analisi testuale è, in ultima istanza, ridurre la complessità del corpus testuale identificando le parole chiave rilevanti che consentono di ricostruire il significato sottostante. In questo processo vengono utilizzati criteri quantitativi per supportare le scelte qualitative, come la selezione delle unità lessicali, l'attribuzione di pesi alle parole dei testi, l'individuazione di relazioni tra le forme grafiche, l'estrazione di temi e concetti.

L'analisi testuale è quindi anche il terreno di incontro e di integrazione tra approcci quantitativi e qualitativi (Dalla Ratta Rinaldi 2007; Wiedemann 2013). Nelle riflessioni più recenti di carattere metodologico, è evidenziato come l'analisi automatica e semiautomatica dei testi, anche arricchita dai contributi più recenti provenienti dal *text mining*, offra occasioni concrete di integrazione – e non solo di giustapposizione – tra le due famiglie di approcci. La natura esplorativa dell'analisi testuale frequentemente pone il ricercatore di fronte a scelte che hanno un carattere qualitativo, come nell'individuazione di

⁵ L'information retrieval (IR) è l'insieme delle tecniche utilizzate per gestire la rappresentazione, la memorizzazione, l'organizzazione e l'accesso ad oggetti contenenti informazioni quali documenti, pagine web, cataloghi online e oggetti multimediali. Generalmente le informazioni sono ricercate mediante queries, parole o stringhe di parole chiave che rappresentano l'informazione richiesta. Con l'IR si individua un insieme

di oggetti (documenti) rilevanti nei quali è contenuta l'informazione richiesta.

⁶ L'information extraction (IE) è l'insieme delle tecniche che permettono di individuare informazioni utili da un insieme di documenti. L'obiettivo dell'IE è l'estrazione di informazioni utili per descrivere un documento.

pattern per i quali non esistono standard di riferimento (Yu, Jannasch-Pennell & DiGangi, 2011); oppure nella scelta delle singole tecniche da applicare per fornire una rappresentazione adeguata del corpus testuale. L'utilizzo di una specifica tecnica dipende dagli obiettivi dell'analisi e dalle scelte che vengono compiute dal ricercatore nel corso di un processo iterativo, di natura tipicamente qualitativa, che tiene insieme aspetti teorici, caratteristiche del corpus testuale e scelte metodologiche. Allo stesso tempo, strumenti come il *text mining* richiedono al ricercatore un'apertura mentale nella ricerca di significati e categorie emergenti dai dati molto simile all'attitudine richiesta dall'applicazione dei principi della *grounded theory* (Ibid.). Più in generale, queste tecniche permettono di soddisfare due richieste fortemente sentite dai ricercatori sia di orientamento qualitativo sia quantitativo, ossia l'esigenza di quantificare e quella di comprendere il "come" dei fenomeni (Silverman 2005). Le tecniche di statistica testuale e i software che ne supportano l'analisi possono fornire un contributo rilevante se permettono di migliorare la qualità di quanto è possibile inferire dai risultati (Aprea e Sappa 2014).

1.4.2 Caratteristiche del corpus testuale

Nel presente lavoro, l'analisi testuale è stata realizzata con il software WordStat. WordStat è un modulo per l'analisi testuale progettato per analizzare materiali testuali di vario tipo: risposte a domande aperte, interviste, articoli, titoli e altri materiali testuali non strutturati. Può essere utilizzato per la categorizzazione automatica dei testi attraverso l'applicazione di vocabolari o per applicare diversi metodi di text mining. WordStat permette di sviluppare e validare vocabolari semantici o tematici; se utilizzato come

supporto alla codifica manuale consente di applicare in maniera sistematica le regole di codifica. Infine, permette di evidenziare le differenze nell'utilizzo delle parole tra sottogruppi di individui o casi.

Nell'analisi automatica dei testi, il corpus è costituito dalla raccolta di materiale testuale oggetto di studio. Ogni testo è ulteriormente scomponibile al suo interno in unità più piccole; si possono immaginare ulteriori suddivisioni dettate dagli obiettivi delle analisi, sino ad arrivare alle più piccole unità del testo rappresentate dalle parole.

Poiché ai fini dell'analisi automatica i testi vanno acquisiti all'interno del programma, nell'identificazione delle partizioni interne al corpus risultano importanti i criteri che il ricercatore decide di utilizzare per analizzare il materiale. «Normalmente», afferma Tuzzi (2003, 29), «quando è stata decisa una ripartizione in testi, si presuppone che esistano anche (uno o più) criteri di raggruppamento in grado di classificarli, ossia si suppone che esista una variabile esterna al corpus che assegna ogni testo a un gruppo o a un individuo».

Nel nostro caso, la definizione dei criteri per l'acquisizione nel programma dei testi delle scuole non ha generato particolari problematiche poiché il corpus è stato organizzato tenendo conto della struttura e della divisione in aree derivante dal format del rapporto di autovalutazione (RAV).

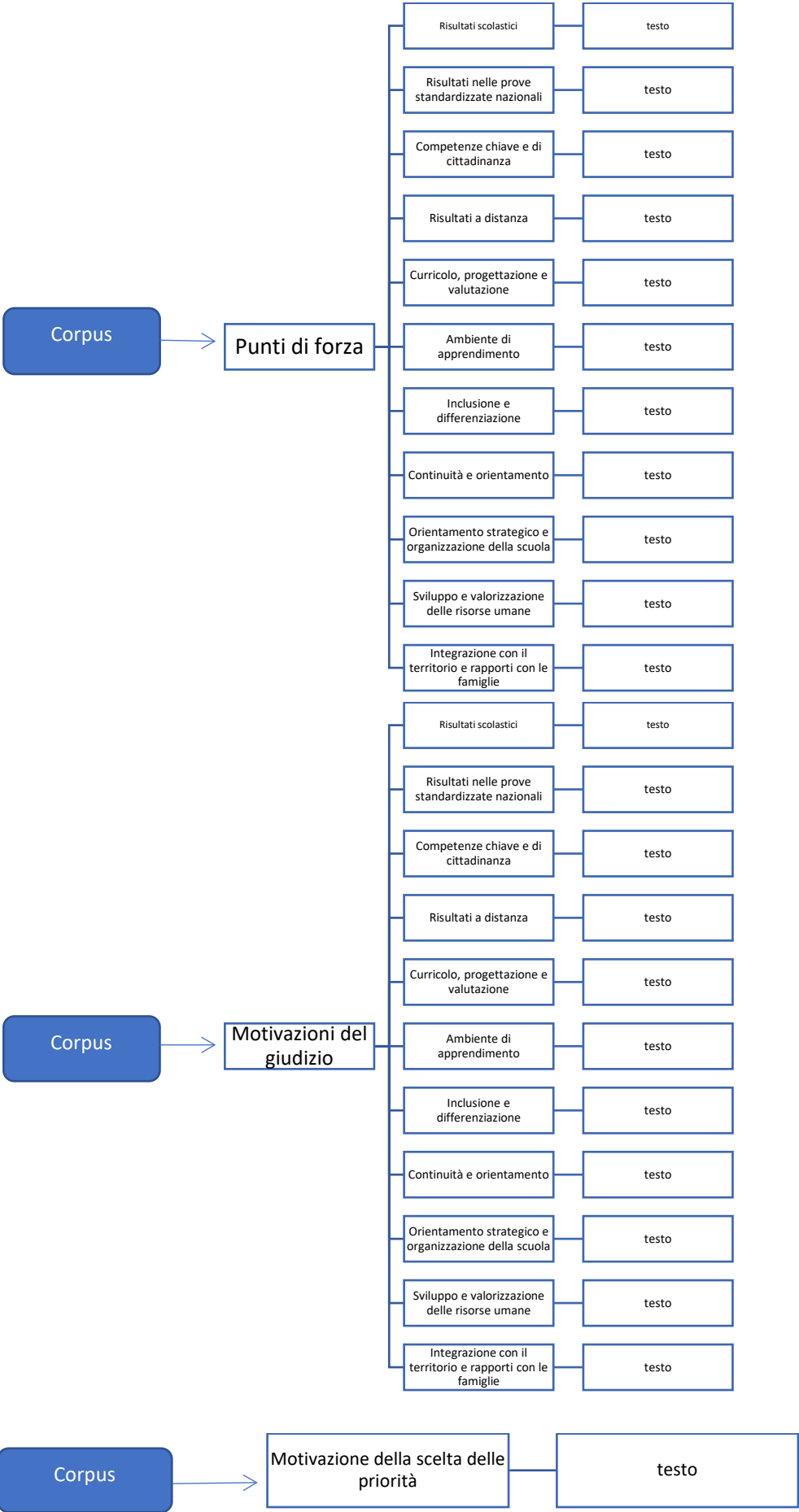
Se partiamo dalla definizione di corpus che evidenzia che esso è rappresentato da «una collezione di unità di contesto, o frammenti, che si ritengono fra loro coerenti e pertinenti per essere studiate sotto un qualche punto di vista o proprietà» (Bolasco 2013: 46), possiamo considerare come corpus dell'indagine l'insieme dei campi aperti del RAV selezionati per l'analisi testuale (ovvero i campi aperti denominati: punti di forza; punti di debolezza; motivazioni del giudizio; motivazione della

scelta delle priorità sulla base dei risultati dell'autovalutazione).

Il corpus è frazionato al suo interno in parti (o testi) a partire da alcune variabili di categorizzazione che riprendono la struttura del RAV. La prima variabile riprende la struttura in sezioni del RAV e ripartisce il corpus in: "Punti di forza", "Punti di debolezza", "Motivazioni del

giudizio" e "Motivazione della scelta delle priorità". La seconda variabile è denominata "Aree del Rav" e riprende la suddivisione in aree di esito e aree di processo del Quadro di riferimento del Sistema nazionale di valutazione. La maggior parte delle analisi è effettuata su parti del corpus selezionate a partire dalla combinazione di queste due variabili.

Fig. 1.3 Partizioni del corpus testuale del RAV



Alcune partizioni, come quelle relative ai punti di forza e di debolezza, possono essere suddivise in sottopartizioni per riprodurre la struttura del RAV che in alcune aree di processo prevede una suddivisione interna in subaree, ove questo sia ritenuto opportuno per i fini delle analisi. Ogni parte è suddivisa al suo interno in frammenti. I frammenti possono essere "interi testi, documenti, loro sezioni o semplici frasi" (Bolasco 2013: 46). Definiamo "frammento" il singolo testo o documento analizzato (es. la motivazione dell'area Risultati scolastici prodotta dalla scuola y). Ogni frammento può essere scomposto in "sezioni" (es. le subaree previste nel RAV che vengono riprese nei campi aperti Punti di forza e Punti di debolezza).

Come affermato da Bolasco, quando si considerano le varie possibilità di analisi del materiale testuale, risulta prezioso il concetto della collezione che è possibile "sfogliare" da diversi punti di vista, anche in relazione alle variabili esterne associate ai testi. «In tal modo, il corpus è assimilabile a un database 'sfogliabile' in differenti modi, a seconda degli obiettivi, in funzione di queste informazioni strutturate» (Bolasco 2013: 47).

Le suddivisioni interne ai frammenti, come i paragrafi e le frasi, sono utili per effettuare alcuni tipi di analisi, come la ricerca delle co-occorrenze o la clusterizzazione, per le quali è possibile prendere quali unità di contesto le frasi, i paragrafi, o l'intero frammento.

Il corpus complessivo relativo al RAV è composto da 3.280.267 token, ovvero occorrenze delle forme grafiche, e da 28.552 types, cioè forme grafiche diverse. Esso può essere definito un corpus di dimensioni molto grandi poiché supera le 100.000 occorrenze (Bolasco 2013).

Le analisi sono condotte su singole partizioni del corpus che hanno le seguenti dimensioni:

- Punti di forza: 1.618.190 token e 21.306 types
- Punti di debolezza: 801.097 token e 16.598 types
- Motivazioni giudizio: 716.023 token e 12.748 types
- Motivazioni priorità: 71.646 token e 5.467 types

Tutte le partizioni e sottopartizioni del corpus presentano dimensioni adeguate per l'analisi di tipo statistico.

1.4.3 Fasi dell'analisi e ipotesi

Nel presente lavoro, la scelta delle tecniche di analisi testuale è stata effettuata tenendo in considerazione gli obiettivi della ricerca. Nel corso dell'analisi, attraverso il confronto continuo tra le domande di ricerca, le caratteristiche e le potenzialità offerte dagli strumenti disponibili in WordStat sono state individuate le tecniche più adeguate e le fasi di analisi necessarie per perseguire gli obiettivi prefissati.

L'approccio utilizzato è quello dei mixed methods basato sull'integrazione dell'approccio lessicometrico e di quello ermeneutico in fasi alternate (Niglas 2000). L'approccio lessicometrico si basa sul trattamento automatico e quantitativo dei testi ed è orientato all'analisi della distribuzione delle parole nel corpus testuale.

L'approccio ermeneutico, tradizionalmente utilizzato nell'analisi qualitativa, si basa sulla lettura dei testi ed è orientato all'individuazione delle relazioni tra le categorie.

Nel presente lavoro, l'integrazione dei due paradigmi permette di sfruttare le potenzialità derivanti dalla quantificazione (es. conoscere la frequenza e co-occorrenza di specifiche parole) per supportare decisioni di tipo qualitativo (costruire le

categorie, assegnare le parole alle categorie). Nella fase finale, l'analisi di tipo lessicometrico è stata utilizzata per evidenziare differenze nei profili lessicali dei testi prodotti da diversi gruppi di scuole. Nella prima fase si è proceduto quindi all'analisi lessicometrica. I testi sono stati sottoposti a una normalizzazione "leggera" (Bolasco 2013) che ha riguardato sostanzialmente l'individuazione e correzione degli errori ortografici e l'eliminazione di possibili fonti di sdoppiamento dei dati testuali. Non si è proceduto alla lemmatizzazione poiché, ai fini dell'analisi, non è stato ritenuto opportuno modificare i testi originari prodotti dalle scuole. Questa scelta è motivata dagli obiettivi dell'indagine e, precisamente, dall'esigenza di confrontare il corpus con risorse testuali esterne rappresentate dagli stimoli presenti negli strumenti di autovalutazione utilizzati dalle scuole. Si tratta dei testi delle domande guida presenti in ogni area e subarea del RAV e dei testi dei descrittori delle rubriche di valutazione di ciascuna area. Tra le domande di ricerca vi è infatti l'analisi dell'impatto di tali risorse sull'utilizzo degli strumenti da parte delle scuole che può essere osservato attraverso la presenza di eventuali "contaminazioni" tra il linguaggio degli strumenti e quello dei RAV (cfr. sotto). Le unità di analisi a questo livello sono rappresentate dalle singole parole e dalle *multiwords*⁷. Sono state effettuate, dapprima, annotazioni di tipo grammaticale, individuando all'interno del corpus le forme grafiche riconducibili a diverse categorie grammaticali: sostantivi, aggettivi e forme verbali. In questa fase sono state utilizzate tecniche quantitative: l'analisi delle forme grafiche a

partire dalla loro frequenza, l'analisi delle parole rilevanti e, ove opportuno, l'analisi di quelle specifiche.

L'analisi delle frequenze è stata effettuata considerando specifiche fasce di frequenza (alta, media e bassa) in considerazione della specificità della tipologia di parole che ricorrono con frequenze diverse. Infatti, nella fascia alta di frequenza (solitamente le prime 30-50 parole) è ragionevole aspettarsi la presenza di parole-funzione, ovvero di preposizioni, articoli e congiunzioni, che hanno un significato puramente grammaticale, e di alcune, poche, parole tema. Queste ultime sono parole molto frequenti, che solitamente riprendono i temi e gli argomenti principali, ma spesso poco utili per ricostruire in maniera più analitica il contenuto dei testi. Esse sono spesso parole ambigue «poiché il loro utilizzo così intenso nasconde molti usi e quindi spesso varie funzioni e accezioni» (Bolasco, 2013: 129). Nella fascia bassa si trovano gli hapax, ovvero le parole con occorrenza pari a 1; si tratta di parole che contribuiscono alla ricchezza lessicale di un corpus ma che, a causa della loro scarsa diffusione, sono poco significative ai fini della costruzione di metriche con le quali interrogare il corpus. Nella fascia media si riscontrano le parole di maggiore interesse per l'individuazione di parole chiave a partire dalle quali è possibile ricercare nei testi, attraverso la costruzione di categorie e di modelli, specifiche informazioni.

L'analisi delle parole rilevanti del vocabolario è stata effettuata a partire dall'indice tf-idf. Si tratta di un indice elaborato nell'ambito dell'*information retrieval*, molto utilizzato nei motori di ricerca per ordinare, secondo criteri di rilevanza, i risultati di una ricerca sul web.

⁷ Si tratta di segmenti ripetuti piuttosto frequenti che compongono espressioni di uso comune come ad es. Prove Invalsi; Certificazione delle competenze, ecc.

Nell'analisi testuale viene utilizzato come criterio interno al corpus a partire dal quale definire la rilevanza di parole o *multiwords*. Il principio su cui si basa il funzionamento dell'indice tf-idf è di pesare la frequenza di una parola con il grado di diffusione nei frammenti del corpus. Pertanto, si attribuisce rilevanza alla frequenza ma anche al fatto che le parole siano concentrate in pochi frammenti piuttosto che in molti. In questo senso, i termini con un valore elevato dell'indice tf-idf sono rilevanti o caratterizzanti perché permettono di discriminare i documenti del corpus, ovvero di caratterizzarli per la presenza di specifiche parole rispetto a tutti gli altri. In questo contributo, individuare le parole e i segmenti rilevanti o caratterizzanti è fondamentale per comprendere, ad esempio, se i rapporti di autovalutazione prodotti dalle scuole sono coerenti, ovvero se i temi trattati nei campi aperti delle diverse aree rispecchiano il contenuto delle aree stesse. Inoltre, l'analisi della rilevanza ci ha permesso di individuare alcune parole caratterizzanti utili per comprendere la struttura del corpus e delle sue partizioni e per orientare le analisi sui testi.

Il criterio della specificità è utilizzato per individuare il sovra o sottoutilizzo di alcune parole nelle singole partizioni rispetto al vocabolario dell'intero corpus. A differenza dell'analisi della rilevanza, l'analisi della specificità si avvale del confronto con un criterio esterno all'insieme dei testi oggetto di analisi. Sono così individuate delle forme specifiche di alcune parti di un corpus, ovvero di alcuni frammenti omogenei rispetto alle modalità di una variabile di classificazione. Le parole specifiche sono quelle forme grafiche che risultano sovra o

sotto-utilizzate sulla base di un test statistico rispetto a un modello di riferimento (Bolasco 2013). Per essere significativo, dal punto di vista statistico, il confronto deve considerare partizioni o corpora di dimensioni sufficientemente ampie⁸. Nel nostro caso, le singole sottopartizioni del corpus (es. le motivazioni dell'area Risultati scolastici) hanno dimensioni più che sufficienti poiché tutte superano le 50.000 occorrenze.

Infine, partendo dalle finalità dell'indagine, tra le quali rientra anche lo studio dell'impatto del lessico prodotto dagli strumenti di valutazione sul lessico utilizzato dalle scuole nel redigere i campi aperti del RAV, abbiamo proceduto con la ricerca delle parole "originali" presenti nei testi che compongono il RAV. Come abbiamo anticipato, precedenti studi di carattere esplorativo condotti sul RAV (INVALSI 2016; Quadrelli & Perazzolo 2018) hanno evidenziato l'utilizzo frequente nelle motivazioni di parole e segmenti ripetuti presenti negli stimoli proposti dagli strumenti di autovalutazione. La maggior parte delle scuole utilizza il lessico veicolato attraverso gli strumenti; gli stimoli testuali contenuti negli strumenti rappresentano una sorta di repertorio dal quale attingere parole ed espressioni sulla valutazione. In taluni casi, la riproposizione di parole ed espressioni utilizzate negli strumenti assume maggiormente il carattere di un approccio "ritualistico" all'autovalutazione. La ricerca di parole originali è utile per comprendere il grado di riflessività nel processo di autovalutazione. Infatti, tali parole possono veicolare concetti, temi e prospettive interpretative autonomamente elaborati dalle scuole oppure possono essere indicatori di una

⁸ Per l'analisi testuale Bolasco considera corpora di piccole dimensioni quelli costituiti da almeno 15.000 occorrenze (Bolasco 1999).

maggiore articolazione dell'analisi dei dati da parte delle istituzioni scolastiche. L'ipotesi di partenza è che tanto più è elevata la quota di parole originali, maggiore è il grado di elaborazione, articolazione e approfondimento dei testi prodotti, ovvero maggiore è la probabilità che le scuole abbiano rielaborato i temi proposti, prospettato temi innovativi o contestualizzato l'analisi dei dati. Per testare tale ipotesi, sono stati dapprima costruiti dei vocabolari costituiti da tutte le parole presenti nei descrittori e nelle domande guida, uno per ciascuna area del quadro di riferimento; successivamente le parole utilizzate dalle scuole nei campi aperti delle singole aree sono state messe a confronto con quelle dei vocabolari elaborati.

Per quanto riguarda l'analisi a livello testuale, l'obiettivo è stato l'individuazione dei nuclei tematici principali e la loro distribuzione nel corpus. A tale scopo è stata utilizzata una tecnica di riduzione della dimensionalità del materiale testuale, ovvero la procedura *Topic extraction* di WordStat. Questa tecnica, basata sull'analisi delle componenti principali, permette di individuare nuclei semantici e/o tematici attraverso l'estrazione di componenti cui sono associate alcune parole chiave. La specificità di questa tecnica è che la stessa parola può essere inclusa in più di una componente, rendendo così possibile l'individuazione del carattere polisemico delle parole. La *Topic extraction* può essere applicata a diversi livelli del testo: la frase, il paragrafo e l'intero documento. Nel nostro caso, l'unità di analisi utilizzata è stata la frase in considerazione delle caratteristiche specifiche dei testi analizzati. I testi del RAV, infatti, sono tendenzialmente brevi e contengono riferimenti a molteplici tematiche. L'analisi a livello della singola frase permette di avere informazioni sulla struttura "fine" del testo, poiché vengono

evidenziati i diversi nuclei semantici contenuti nella medesima proposizione. A partire dai risultati della *Topic extraction*, si è proceduto alla costruzione dei vocabolari tematici. Sono state costruite categorie che includono parole e segmenti ripetuti considerati come indicatori parziali dei principali temi contenuti nei testi. Il vocabolario tematico è stato costruito partendo sia dai nuclei semantici individuati con la *Topic extraction*, sia dalla considerazione degli obiettivi specifici della ricerca. Per esempio, come precedentemente illustrato, uno degli obiettivi della metavalutazione riguarda la valutazione della coerenza del processo di autovalutazione: per raggiungere questo obiettivo nell'analisi dei campi aperti "Motivazioni del giudizio assegnato", nell'analisi a livello testuale si è prestata attenzione alla ricerca della coerenza tra i termini utilizzati dalle scuole per qualificare il proprio operato e il punteggio autoattribuito nella scala di valutazione. In particolare, sono state costruite categorie che includono i termini a valenza positiva e negativa, ovvero parole, per lo più aggettivi, utilizzati per qualificare in senso positivo o negativo i risultati o i processi analizzati.

La costruzione dei vocabolari è stata effettuata a partire da una prospettiva ermeneutica ed è stata supportata dall'analisi delle concordanze (keyword in context e keyword retrieval) per rilevare i contesti di utilizzo delle parole.

La costruzione del vocabolario tematico ha la finalità di ricercare la frequenza e diffusione di specifici temi nel corpus testuale e di rilevare la struttura complessiva dei documenti prodotti dalle scuole. Pertanto, oltre all'analisi della frequenza dei termini contenuti nelle singole categorie e della diffusione delle categorie nei documenti, è stata effettuata l'analisi delle co-occorrenze a livello di documento. L'analisi delle co-occorrenze

permette di individuare i rapporti tra le categorie tematiche utilizzando una varietà di tecniche (cluster analysis, link analysis, proximity plot).

Nella terza fase, le tecniche di tipo quantitativo sono state utilizzate al fine di verificare eventuali differenze nel profilo lessicale di diversi gruppi di scuole (scuole del primo e del secondo ciclo) o tra i testi riconducibili alle diverse macroaree e aree del Quadro di riferimento e alle diverse sezioni del RAV (es. tra le partizioni Punti di forza e Punti di debolezza o tra le Motivazioni delle diverse aree).

Inoltre, sono state considerate le differenze tra le scuole autocollocate in livelli diversi della scala di valutazione⁹ per indagare la coerenza tra i testi delle Motivazioni del giudizio, nelle quali le scuole dovevano giustificare la scelta del livello di qualità, e il punteggio autoattribuito. L'ipotesi di partenza è che le scuole che si sono collocate nei livelli più elevati della scala abbiano utilizzato con maggiore frequenza termini a valenza positiva rispetto alle scuole collocate nei livelli più bassi e viceversa.

In questa fase sono state utilizzate le tavole di contingenza e l'analisi delle corrispondenze lessicali.

Riferimenti bibliografici

Alivernini, F. (2008) 'La statistica testuale'. In F. Lucidi, F. Alivernini, A. Pedon, Metodologia della ricerca qualitativa. Bologna: Il Mulino.

Aprèa, C. and Sappa, V. (2014) 'Utilizzo dei metodi misti per lo studio delle concezioni individuali in ambito educativo. Esempi di applicazione in un approccio fenomenografico'. *Formation et pratiques d'enseignement en questions*, 17, pp. 169-191.

Bolasco, S. (2003) 'L'analisi statistica dei testi: intrecci problematici e prospettive'. Giornata di studio su Applicazioni di analisi testuale, Roma 16

dicembre 2003. <http://www.taltac.it/file/Bolasco-AST%202004%20introduz%20vers%20abstract%20esteso.pdf>

Bolasco, S. (2013) L'analisi automatica dei testi. Fare ricerca con il text mining. Roma: Carocci

Bustelo, M. (2002) 'Meta-evaluation as a tool for the improvement and development of the evaluation function in public administrations'. In Meta-evaluation as a tool for the improvement and development of the evaluation function in public administrations, The 2002 EES Conference: three movements in contemporary evaluation learning, theory and evidence: University Complutense de Madrid.

Cooksy, I. J. and Mark, M. M. (2012) 'Influences on Evaluation Quality'. *American Journal of Evaluation*, 33, pp. 79-87.

Della Ratta Rinaldi, F. (2007) 'L'analisi multidimensionale dei testi'. In L. Cannavò, L. Frudà, Ricerca sociale. Dall'analisi esplorativa al data mining. Roma: Carocci

Ehren, M. C.M. and Swanborn, M. S.L. (2012) 'Strategic data use of schools in accountability systems'. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 23(2), pp. 257-280.

INVALSI (2016) Le Rubriche del RAV Prime analisi, validità e affidabilità, uso da parte delle scuole delle Rubriche del Rapporto di Autovalutazione. https://www.invalsi.it/snv/docs/121216/Rapporto_rubriche.pdf

Lebart, L., Salem, A., and Berry, L. (1998) Exploring textual data. Dordrecht-Boston-London: Kluwer Academic Publishers.

Muzzioli, P., Perazzolo, M., Poliandri, D., and Quadrelli, I. (2016) 'La qualità del percorso di autovalutazione'. *Scuola Democratica*, special issue, 2-3, pp. 421-438.

Niglas, K. (2000) 'Combining quantitative and qualitative approaches'. European Conference on Educational Research, Edinburgh, 20-23 September 2000. <http://www.leeds.ac.uk/educol/documents/00001544.htm>

⁹ A tale scopo sono state prodotte delle tavole di contingenza utilizzando le categorie del vocabolario tematico e i termini in esse contenute per evidenziare eventuali differenze statisticamente significative ($p <$

0,05) nell'utilizzo di termini da parte delle scuole collocate in diversi livelli della scala. Successivamente, è stata effettuata l'analisi delle corrispondenze lessicali.

Pawson, R. and Tilley, N. (1997) Realistic evaluation. Sage, Thousand Oaks.

Quadrelli I. and Perazzolo M. (2018) 'Analisi testuale delle motivazioni della rubrica Risultati nelle prove standardizzate nazionali'. Working paper INVALSI, n. 34

https://www.invalsi.it/download2/wp/WP34_Quadrelli_Perazzolo.pdf

Schildkamp, K. and Visscher, A. (2009) 'Factors Influencing the Utilisation of A School Self-evaluation Instrument'. Studies in Educational Evaluation, 35(4), pp. 150-159.

Silverman, D., (2005) 'Doing qualitative research: a practical handbook'. London: Sage.

Stufflebeam, D. L. ([1974]2011) 'Meta-evaluation'. Journal of MultiDisciplinary Evaluation, 7:15, pp. 99-158.

Wiedemann, G. (2013) 'Opening up to big data: computer-assisted analysis of textual data in social sciences'. Historical Social Research, 38(4), pp. 332-358. <https://doi.org/10.12759/hsr.38.2013.4.7332-358>