

GLI APPROFONDIMENTI DEL SERVIZIO STATISTICO

La newsletter ufficiale del Servizio Statistico INVALSI



Gentilissimo/a,

come la precedente (dicembre 2024), anche questa newsletter de “*Gli approfondimenti del Servizio Statistico*” è dedicata all’esplorazione dei contenuti e delle potenzialità informative degli *Open Data* che l’INVALSI mette a disposizione di qualsiasi utente.

I dati in formato aperto che saranno presi in considerazione sono contenuti in un file denominato “Punteggi e percentuale di studenti nei livelli di competenza per ripartizioni territoriali e caratteristiche di contesto”, che può essere scaricato dalla sezione *open data* del portale del Servizio Statistico INVALSI.

Assieme ai dati sono forniti due ulteriori file che forniscono, rispettivamente, l’elenco (tracciato) delle variabili disponibili ed una serie di note esplicative.[1]

Il file che contiene i dati ([Punteggi e percentuale di studenti nei livelli di competenza per ripartizioni territoriali e caratteristiche di contesto – INVALSI Servizio Statistico](#)) è in formato CSV e può essere aperto con *Excel* o con un qualsiasi foglio di calcolo *Open Source*, ad esempio *OpenCalc*.

Un “mega” file

I dati del file “Punteggi e percentuale di studenti...” consentono di svolgere analisi molto dettagliate sotto il profilo territoriale, dei percorsi scolastici, delle caratteristiche soggettive degli studenti (cittadinanza, condizione economica familiare). Più precisamente, il file contiene l’informazione aggregata di maggior dettaglio mai rilasciata dall’INVALSI, naturalmente rispettando i limiti stabiliti dalla normativa sulla tutela della *privacy*.

Il risvolto della medaglia è la creazione e la conseguente diffusione di un file molto grande, di circa 75 Mb. Per dare un’idea più tangibile, le informazioni di cui stiamo parlando sono contenute in una tabella di 20 colonne – le colonne corrispondono alle variabili – e di ben 575.425 righe!

Il numero esorbitante di righe dipende dal fatto che ciascuna di esse contiene dati ottenuti dall’incrocio di più dimensioni. Per quanto attiene al piano territoriale, il livello più dettagliato che viene reso disponibile è quello del comune presso cui ha sede la scuola. Moltiplicando i comuni per i gradi scolastici delle prove INVALSI (II, V, VIII, X, XIII), per le caratteristiche di genere, cittadinanza e condizione socio-economica si ottiene un ordine di grandezza di alcune centinaia di migliaia di combinazioni. A queste vanno aggiunte poi le migliaia di righe che derivano dall’aggregazione dei dati ad un livello gerarchicamente superiore. Al di sopra dei comuni ci sono le province, le regioni e le ripartizioni; licei, istituti tecnici ed istituti professionali formano l’ordine delle scuole secondarie di secondo grado, e via di seguito. Per ciascun sotto-insieme definito da tali combinazioni di caratteri il file mette a disposizione dati sintetici sui risultati INVALSI, quali la media e la variabilità dei punteggi o la percentuale di studenti che si collocano nei vari livelli di competenze. C’è proprio (di) tutto!

[1] Si tratta dei metadati che consentono all’utente l’utilizzo e la piena fruizione dei contenuti informativi dei dati veri e propri.

Il dettaglio regionale rappresenta un utile punto di riferimento. Tuttavia, molte importanti informazioni che hanno ad oggetto la programmazione e la gestione delle scuole sono prese ad un livello territoriale ancor più disaggregato, provinciale o addirittura comunale.

Attingendo agli *Open Data* INVALSI, chi è interessato ad approfondire i risvolti territoriali delle questioni che attengono al mondo della scuola può trovare molti elementi utili.

Per farlo è sufficiente infatti recarsi sulla pagina web del Servizio Statistica INVALSI all'URL: <https://serviziostatistico.invalsi.it/catalogo-dati/>

Non solo senso comune

Il motivo per il quale si è deciso di assemblare tutto in un unico file è innanzitutto di convenienza: può essere utile e comodo sapere che le informazioni di cui possiamo aver bisogno risiedono in uno stesso spazio. Basta avere la pazienza di cercare attraverso i filtri e le funzioni di ricerca dei fogli di calcolo che rendono l'operazione estremamente semplice e veloce. In fondo, come recita un famoso adagio, nel più ci sta il meno.

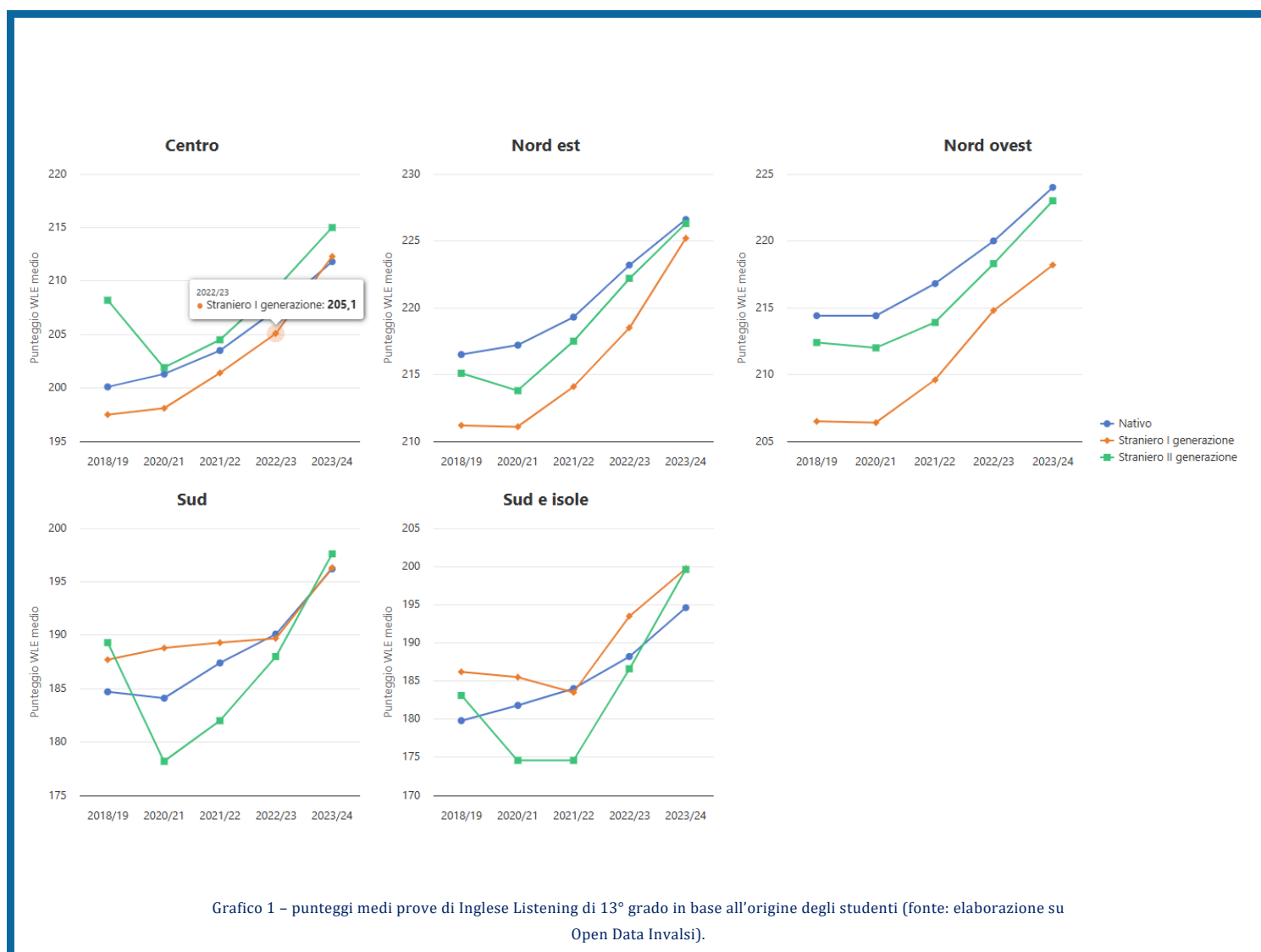
Ma c'è una seconda motivazione per fornire i dati in forma compatta. Suddividere fin dall'inizio le informazioni in base ad un criterio – ad esempio rendendo disponibile un file distinto per tutti i comuni, uno per tutte le province ecc. – equivale a suggerire un ordine all'analisi. È come se chi fornisce le informazioni conducesse l'utente a confrontare soltanto un comune con altri comuni, una regione con altre regioni e via di seguito. Ciò ovviamente ha una sua coerenza ma “stona” con la logica degli *Open Data*. Questi, infatti, vanno pensati come una sorta di materia prima un po' grezza, su cui ciascuno è autorizzato a condurre le proprie specifiche ricerche, piuttosto che come un manufatto già predefinito. Del resto, sul sito dell'INVALSI e in molti altri luoghi si trovano articoli, saggi, trattati che analizzano i dati delle prove fornendo molteplici direzioni interpretative. Per contro, gli *Open Data* sono un invito a sperimentare, uscendo anche un po' fuori dagli schemi consolidati.

Qualche spunto

Forniamo di seguito qualche esempio di esplorazione creativa dei dati contenuti nel file “Punteggi e percentuali di studenti...”. Il filo rosso che lega questi esempi, il tema a cui vogliamo accostarci da differenti punti di vista, è il confronto fra risultati degli studenti italiani e degli studenti stranieri di prima e di seconda generazione.

Nel primo grafico sono messi a confronto per le cinque ripartizioni territoriali considerate nel rapporto INVALSI (Nord Ovest, Nord Est, Centro, Sud, Sud e Isole[2]) gli andamenti medi nelle prove di Inglese Listening dell'ultimo anno delle scuole superiori.

Questo genere di confronto piuttosto tradizionale mette subito in evidenza non soltanto l'eterogeneità nazionale ma anche l'estrema varietà geografica dei risultati ottenuti dagli studenti.



[2] La ripartizione Nord Ovest comprende Valle d'Aosta, Piemonte, Liguria e Lombardia; la ripartizione Nord Est comprende Veneto, Trentino Alto-Adige, Friuli Venezia-Giulia ed Emilia Romagna; la ripartizione Centro comprende Toscana, Marche, Umbria e Lazio; la ripartizione Sud comprende Abruzzo, Molise, Campania e Puglia; la ripartizione Sud e Isole comprende Basilicata, Calabria, Sicilia e Sardegna.

Come sottolineato negli ultimi rapporti INVALSI, l'andamento delle competenze in Inglese ha risentito assai meno dell'impatto del covid di quanto non sia avvenuto per Italiano e Matematica. Il *trend* verso l'innalzamento dei risultati in Inglese sembra pertanto essere un tratto strutturale nel nostro Paese. Tuttavia, balza immediatamente agli occhi come nel Sud e nelle Isole i risultati degli stranieri di seconda generazione abbiano fatto registrare un calo vistoso durante il periodo della didattica a distanza – circostanza che, per contro, non si verifica nel Nord del Paese. Ciò solleva domande in ordine al grado di integrazione scolastica degli studenti stranieri nelle diverse aree. Ma i risultati potrebbero dipendere anche da altri fattori, ad esempio da differenze sostanziali in ordine alle caratteristiche ed all'origine nazionale degli stranieri che risiedono in aree diverse del Paese.

L'impatto dell'eterogeneità delle provenienze o ascendenze nazionali degli studenti stranieri può essere colto in modo assai più diretto nel grafico 2 in cui sono messi a confronto gli andamenti dei risultati medi nelle prove di Italiano dell'ultimo anno della scuola secondaria di primo grado per gli studenti della provincia di Prato e per gli studenti della Toscana.

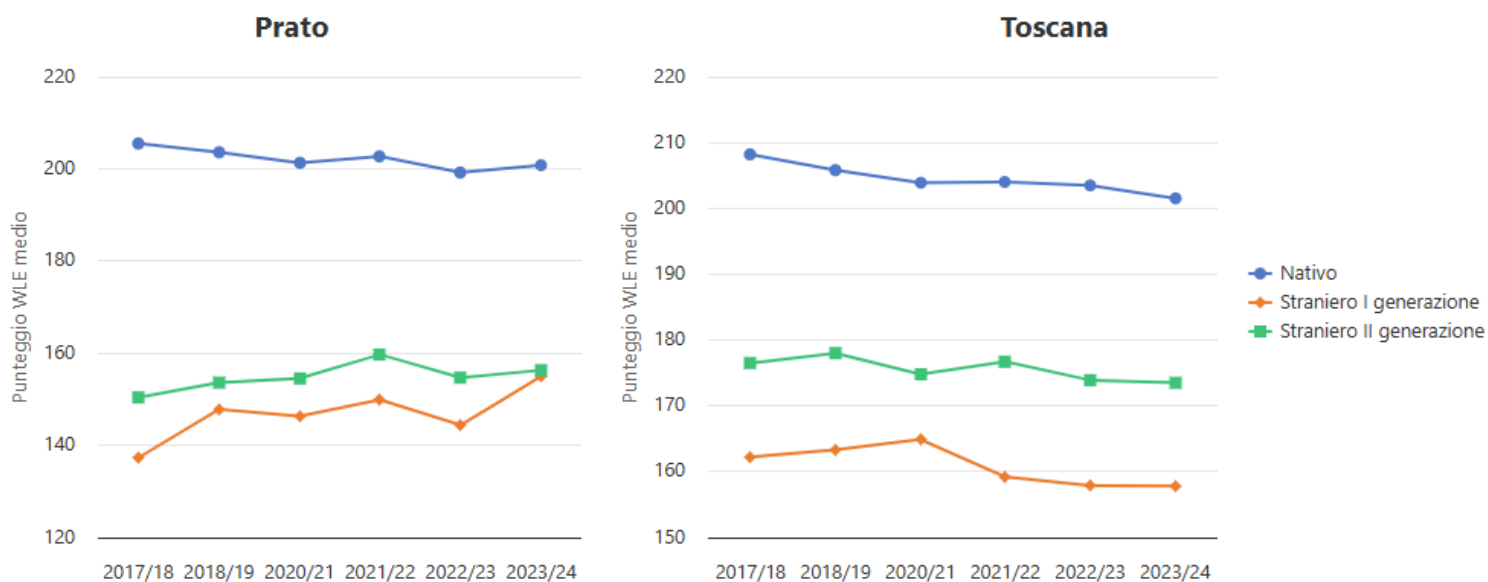


Grafico 2 – punteggi medi prove di italiano di 8° grado in base all'origine degli studenti, confronto fra comune di Prato e Toscana (fonte: elaborazione su Open Data Invalsi).

Il risultato nettamente peggiore degli studenti stranieri della provincia di Prato dipende in larga misura dal fatto che è nettamente prevalente la componente asiatica (tabella 1).

Per contro, nella Toscana presa nel suo complesso vi è una maggior incidenza di giovani rumeni ed albanesi – che incontrano ostacoli minori nell'assimilazione della lingua italiana rispetto alla nazionalità cinese più numerosa nel panorama demografico di Prato.

Prato			Toscana		
Nazionalità	n.	%	Nazionalità	n.	%
Cina	36551	63,6	Romania	71285	16,8
Albania	4583	8,0	Cina	68890	16,2
Romania	3482	6,1	Albania	56036	13,2
Pakistan	2925	5,1	Marocco	28299	6,7
Marocco	1672	2,9	Senegal	13689	3,2

Tabella 1 – Principali gruppi nazionali nella popolazione straniera residente, confronto fra provincia di Prato e Toscana
(fonte: elaborazione su Open Data ISTAT).

I dati circa la composizione nazionale della popolazione straniera che vive nei vari territori (regioni, province, comuni) possono essere consultati e scaricati in formato aperto dal portale demografico dell'ISTAT (<https://demo.istat.it/>). Questo è un ulteriore esempio tangibile del valore sistemico degli Open Data: ciò che viene messo a disposizione da un Ente può essere collegato ai dati rilasciati da altri soggetti, a formare un vero e proprio arcipelago da cui è possibile trarre le informazioni più utili ai fini dei vari tipi di analisi e ricerche che si vogliono effettuare.

Di seguito il link per esplorare [il catalogo degli Open Data](#)

e il link [Newsletter - Servizio Statistico INVALSI](#) dove è possibile recuperare i precedenti numeri della newsletter.