



Casi concreti di analisi dei dati restituiti nelle scuole

Pietro Bonfanti

30.10.2015

Diesse Lombardia

Come sono costruiti i quesiti delle prove Invalsi?

- Le prove Invalsi sono costruite allo scopo di rendere i dati forniti direttamente confrontabili e comparabili. Prove *valide, attendibili e oggettive*.
Tempo richiesto per la costruzione: 15-18 mesi.
- I quesiti delle prove Invalsi si caratterizzano per la **chiusura degli stimoli e delle risposte**.
 - **Obiettivo:** ridurre l'ambiguità interpretativa, facilitare il lavoro di correzione, ottenere un lavoro di correzione univoco.
 - Anche nelle domande a risposta aperta, ci sono domande aperte a risposta univoca (*cloze*)
 - Le domande aperte a risposta articolata vengono corredate da precise indicazioni per la correzione.
- Il processo di costruzione delle domande richiede particolare attenzione se si vuole ottenere una prova che abbia una robustezza dal punto di vista psicometrico.

Come sono costruiti i quesiti delle prove Invalsi?

- Le **domande a risposta chiusa** rappresentano la tipologia di domande più utilizzata nella costruzione di prove standardizzate perché:
 - Le modalità di correzione soddisfano il criterio di riproducibilità
 - Riducono il problema delle omissioni e gli studenti le percepiscono come più agevoli
 - Ogni domanda sottoposta ad analisi statistica fornisce una serie di dati che consentono di capire più facilmente il perché degli errori
 - Consentono di valutare anche processi cognitivi complessi

Come sono costruiti i quesiti delle prove Invalsi?

- Il ruolo dei **distrattori**:
 - Vengono introdotti in modo tale che la risposta fornita dallo studente rappresenti il risultato di un articolato processo di discriminazione. Bisogna evitare che lo studente arrivi alla soluzione corretta per approssimazioni successive; al contrario, un quesito ben formulato dovrebbe **far attivare allo studente un procedimento logico** per giungere alla risposta corretta.
- Per questo:
 - non troppo vicini alla risposta corretta
 - abbastanza attrattivi e plausibili (ad es. evitare che possano essere esclusi anche senza leggere il testo)
 - non costruiti per trarre in inganno il rispondente

Come sono costruiti i quesiti delle prove Invalsi?

- Lo studio delle proprietà psicometriche dei test è una fase fondamentale, non solo durante il *pre-testing*, ma anche *ex post*, e cioè quando la prova è già stata somministrata agli studenti.
 - Verifica empirica dell'adeguatezza dello strumento rispetto alle finalità per il quale è stato concepito
- La valutazione dell'adeguatezza delle prove Invalsi passa attraverso due domande: **cosa** vogliamo misurare e **come** vogliamo farlo, e cioè attraverso la valutazione della *validità* (il grado con cui uno strumento misura quello che ritiene di misurare) e dell'*attendibilità* (la precisione con cui lo misura).

Come sono costruiti i quesiti delle prove Invalsi?

- Indici utilizzati, derivati dalla Teoria Classica dei Test:
 - **Indice di difficoltà degli item** (percentuale di risposte corrette)
 - Nel processo di selezione delle domande, vengono generalmente incluse nella prova solo quelle domande con una percentuale di risposte corrette che oscilla tra 10% e 90%.
 - Suggerisce anche una prima ipotesi di posizionamento di ciascun item all'interno del fascicolo.
 - **Indice di discriminatività** (misura la capacità di ciascun item di distinguere studenti con livelli diversi di abilità)
 - Quanto più una domanda è discriminativa, tanto più essa è in grado di misurare la variazione di probabilità di fornire la risposta corretta anche per piccole variazioni di abilità del rispondente
 - Viene utilizzato *l'indice di correlazione punto-biseriale*, definito come la correlazione tra i punteggi ottenuti dai rispondenti su un item e il punteggio totale su tutti gli item
 - **Alpha di Cronbach** (valuta la coerenza interna degli item che compongono una prova)

Come sono costruiti i quesiti delle prove Invalsi?

- La Teoria Classica dei Test presenta dei limiti, innanzitutto l'impossibilità di tenere separate le caratteristiche dei soggetti (in termini di abilità) da quelle degli item (in termini di difficoltà).
- Vengono pertanto utilizzati anche strumenti tipici dell'*Item Response Theory*:
 - **Modello di Rasch**: permette di stimare l'abilità dei soggetti indipendentemente dalla difficoltà degli item, e viceversa
 - 1) Maggiore è l'abilità, maggiore è la probabilità di rispondere correttamente agli item contenuti nella prova
 - 2) Qualsiasi individuo dovrebbe superare più facilmente un item semplice che uno difficile
 - Permette di verificare anche che le domande si distribuiscano lungo l'intera scala di difficoltà

Come sono costruiti i quesiti delle prove Invalsi?

- L'utilizzo di questi indici (in particolare dopo il *pre-test*) permette di scartare alcuni quesiti, o di modificare le opzioni di risposta di altri quesiti che vengono comunque proposti
- Esempio: quesito della prova di Matematica 2013 della classe I Secondaria di I Grado:

D4 Il nonno Andrea e sua nipote Marta stanno camminando vicini lungo un sentiero. Ogni due passi che fa il nonno Marta ne deve fare 3 per restare affiancata al nonno. Dopo che il nonno Andrea ha fatto 40 passi quanti passi ha fatto Marta?

- A. 40
- B. 60
- C. 80
- D. 120

Come sono costruiti i quesiti delle prove Invalsi?

- Analisi IRT delle risposte al quesito D4 (*pre-test*)

item:43 (M6RF036)

Cases for this item 851 Discrimination 0.26
Item Threshold(s): 0.52 Weighted MNSQ 1.04
Item Delta(s): 0.52

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	30	3.53	-0.16	-4.71(.000)	-0.49	0.36
2	1.00	324	38.07	0.26	7.84(.000)	0.20	0.69
3	0.00	85	9.99	-0.12	-3.55(.000)	-0.15	0.46
4	0.00	328	38.54	-0.03	-0.97(.334)	-0.02	0.58
7	0.00	1	0.12	0.01	0.18(.856)	-0.11	0.00
9	0.00	83	9.75	-0.15	-4.43(.000)	-0.24	0.63

Come sono costruiti i quesiti delle prove Invalsi?

- Analisi IRT delle risposte al quesito D4 (*pre-test*)

Numero di alunni a cui è stato somministrato

Buona discriminazione

item:43 (M6RF036)

Cases for this item 851 Discrimination 0.26
 Item Threshold(s): 0.52 Weighted MNSQ 1.04
 Item Delta(s): 0.52

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	30	3.53	-0.16	-4.71(.000)	-0.49	0.36
2	1.00	334	38.07	0.26	7.84(.000)	0.20	0.69
3	2.00	300	9.99	-0.12	-3.55(.000)	-0.15	0.46
4	0.00	328	38.54	-0.03	-0.97(.334)	-0.02	0.58
7			0.12	0.01	0.18(.856)	-0.11	0.00
9			9.75	-0.15	-4.43(.000)	-0.24	0.63

Possibili risposte

Risposte non valide

Mancate risposte

Come sono costruiti i quesiti delle prove Invalsi?

- Analisi IRT delle risposte al quesito D4 (*pre-test*)

Correlazione *punto-biseriale* item-test (correlazione tra la probabilità di scegliere una data opzione e l'abilità complessiva del rispondente)

Deve essere positiva per l'opzione corretta e negativa per le opzioni errate

item:43 (M6RF036)

Cases for this item: 51
 Item Threshold: 0.52
 Item Delta(s): 0.52
 Frequency absolute: 51
 Frequency relative percentage: 0.52
 Point Biserial Correlation: 0.26
 MNSQ: 1.04

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	30	3.53	-0.16	-4.71(.000)	-0.49	0.36
2	1.00	324	38.07	0.26	7.84(.000)	0.20	0.69
3	0.00	85	9.99	-0.12	-3.55(.000)	-0.15	0.46
4	0.00	328	38.54	-0.03	-0.97(.334)	-0.02	0.58
7	0.00	1	0.12	0.01	0.18(.856)	-0.11	0.00
9	0.00	83	9.75	-0.15	-4.43(.000)	-0.24	0.63

Come sono costruiti i quesiti delle prove Invalsi?

- Analisi IRT delle risposte al quesito D4 (*pre-test*)

Livello medio di prestazione dei rispondenti nella *scala di Rasch*

item:43 (M6RF036)

Cases for this item 8

Item Threshold(s): 0

Item Delta(s): 0.52

Significatività delle correlazioni punto-biseriali
 n 0.26
 SQ 1.04

Variabilità del dato della colonna precedente

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	30	3.53	-0.16	-4.71(.000)	-0.49	0.36
2	1.00	324	38.07	0.26	7.84(.000)	0.20	0.69
3	0.00	85	9.99	-0.12	-3.55(.000)	-0.15	0.46
4	0.00	328	38.54	-0.03	-0.97(.334)	-0.02	0.58
7	0.00	1	0.12	0.01	0.18(.856)	-0.11	0.00
9	0.00	83	9.75	-0.15	-4.43(.000)	-0.24	0.63

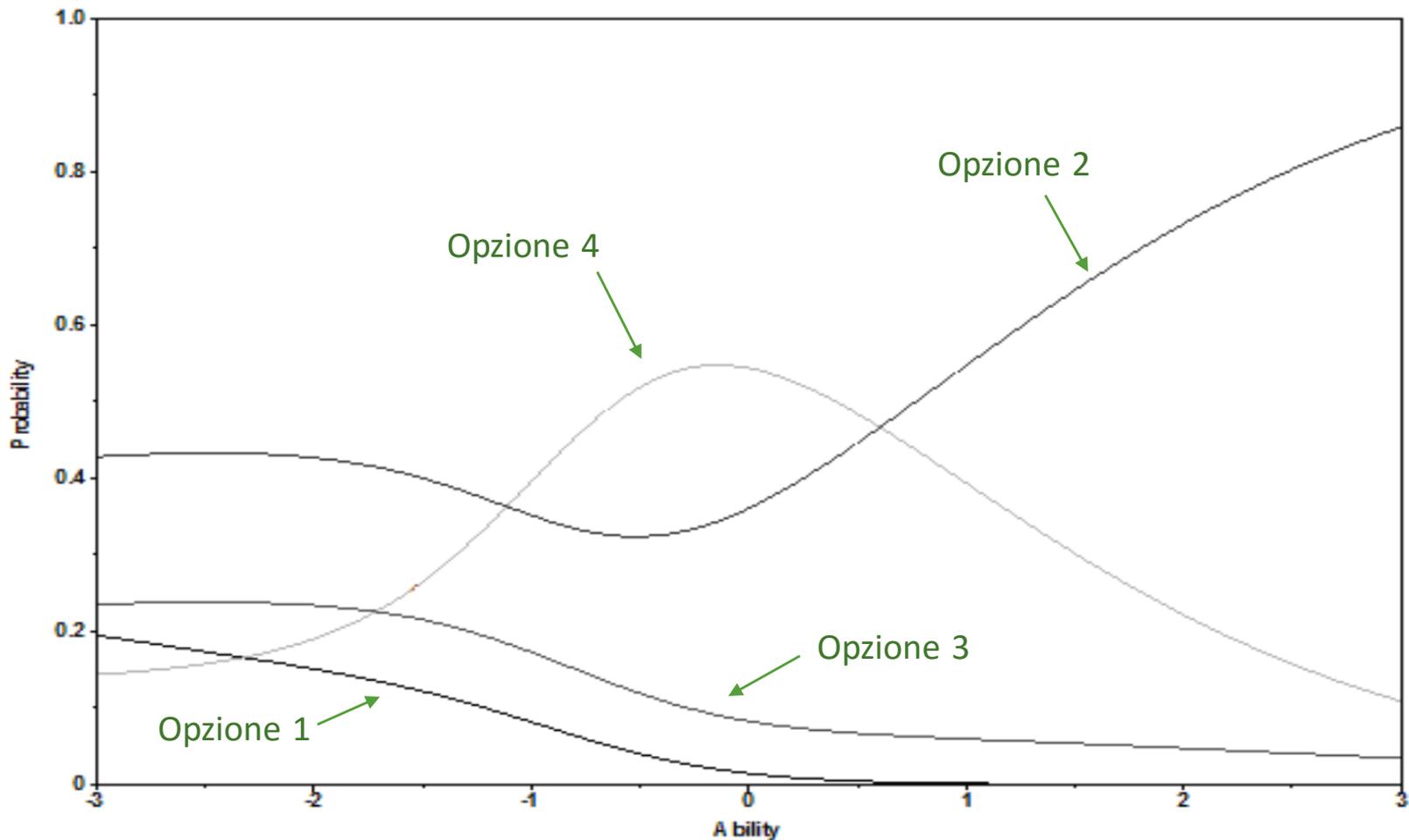
Come sono costruiti i quesiti delle prove Invalsi?

- Il problema di questo quesito è il quarto distrattore:
 - coefficiente Pt-Bis debolmente negativo, troppo vicino allo 0 e non significativo statisticamente ($t=.335$, il massimo accettabile è pari a 0,05)
 - Nel gruppo di alunni che hanno scelto questa opzione di risposta (il 38,5%) ci sono studenti con un livello di abilità discreto: il livello di abilità medio dei rispondenti al quarto distrattore è -0,02, molto vicino all'abilità media nella scala di Rasch (0)

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	30	3.53	-0.16	-4.71(.000)	-0.49	0.36
2	1.00	324	38.07	0.26	7.84(.000)	0.20	0.69
3	0.00	85	9.99	-0.12	-3.55(.000)	-0.15	0.46
4	0.00	328	38.54	-0.03	-0.97(.334)	-0.02	0.58

Come sono costruiti i quesiti delle prove Invalsi?

- Curva caratteristica del quesito D4 nel *pre-test*:



Come sono costruiti i quesiti delle prove Invalsi?

- Nella versione definitiva è stata cambiata l'opzione 4 e in parte la domanda:

D4. Marta e il nonno camminano insieme lungo un sentiero. Ogni 2 passi fatti dal nonno, Marta ne fa 3 per restargli al fianco. Quando il nonno ha fatto 40 passi, quanti passi ha fatto Marta?



- A. 80
- B. 60
- C. 40
- D. 20

Come sono costruiti i quesiti delle prove Invalsi?

- Analisi IRT delle risposte al quesito D4 (definitivo):

Item 8

item:8 (D4)

Cases for this item 27416 Discrimination 0.33

Item Threshold(s): -0.98 Weighted MNSQ 1.03

Item Delta(s): -0.98

Label	Score	Count	% of tot	Pt Bis	t (p)	PV1Avg:1	PV1 SD:1
1	0.00	5069	18.49	-0.16	-27.12 (.000)	-0.26	0.72
2	1.00	19189	69.99	0.33	57.81 (.000)	0.17	0.83
3	0.00	1194	4.36	-0.16	-26.62 (.000)	-0.60	0.74
4	0.00	1433	5.23	-0.20	-34.12 (.000)	-0.69	0.71
7	0.00	23	0.08	-0.02	-2.93 (.003)	-0.44	0.59
9	0.00	508	1.85	-0.08	-12.93 (.000)	-0.48	0.85

Quali dati restituisce l'Invalsi?

- L'Invalsi restituisce alle scuole tanti dati, alcuni in tabelle, altri nei grafici. Per certi versi, i dati in nostro possesso sono sovrabbondanti, sta a noi scegliere.
- Per la stesura del RAV le indicazioni fanno riferimento principalmente a tre dati:
 - **Punteggi generali** (confronto con dato nazionale, regionale e di area; confronto con scuole con contesto socio-economico-culturale simile -> indice ESCS)
 - La **varianza** tra le classi (e la varianza interna alla classe)
 - La percentuale di studenti che si colloca nei **livelli bassi** all'interno della scala dei risultati

N.B. Parliamo sempre di confronto, non di dati assoluti!

- Questi dati sono importanti, ma insufficienti a dare una valutazione. Cominciamo però a studiare questi.

I punteggi generali

Vengono forniti:

- i punteggi assoluti al netto del ***cheating***
- il confronto con i dati nazionali, regionali e di area (freccie, significativamente sup, inf...)
- il confronto con l'**indice ESCS**
- la percentuale di risposte corrette (al lordo del *cheating*)
- il cheating

Questo per ogni classe e per la scuola nel suo complesso.

I punteggi generali

Cheating:

- è un fenomeno rilevato attraverso un controllo di tipo statistico sui dati e si riferisce a quei comportamenti *impropri* tenuti nel corso della somministrazione: risposte copiate da altri studenti, libri o fonti, o suggerite più o meno esplicitamente dai docenti.
- L'effetto del *cheating* è misurato mediante un indicatore percentuale che esprime quale parte del punteggio osservato è mediamente da attribuire alle predette anomalie.

È necessario conteggiare il *cheating* per poter avere una garanzia dell'attendibilità dei risultati.

Noi sappiamo se gli studenti hanno effettivamente copiato, o se hanno risposto molto bene in autonomia

I punteggi generali

Indice ESCS:

- La differenza in tabella è calcolata rispetto al risultato medio delle 200 classi/scuole con *background* socio-economico-culturale più simile a quello della classe/scuola considerata.
- Si considerano principalmente il titolo di studio dei genitori degli studenti, la loro condizione occupazionale e la disponibilità di risorse economiche. Viene utilizzato il Questionario Studente per reperire tali dati.
- Per le classi di II Primaria e III Secondaria di I Grado tale indice non è fornito.
- I livelli di *background* sono definiti rispetto alla distribuzione nazionale dell'indicatore ESCS:
 - Primo quartile (fino al 25%): livello basso
 - Secondo quartile (dal 25% al 50%): livello medio-basso
 - Terzo quartile (dal 50% al 75%): livello medio-alto
 - Quarto quartile (dal 75% al 100%): livello alto

I punteggi generali

Confronto con dato nazionale, regionale e di area:

- Su internet abbiamo **frecche verso l'alto o verso il basso**; sul file excel le frecce vengono sostituite dalla dicitura *significativamente superiore e significativamente inferiore*. Indicano una differenza rispettivamente positiva e negativa statisticamente significativa, ossia con una probabilità superiore al 95% di verificarsi anche nella popolazione e non solo nel campione.
- Le **frecche orizzontali** (nel file excel la dicitura *non significativamente differente*) indicano, invece, una differenza positiva o negativa statisticamente non significativa.

Vale la pena calcolare anche la differenza numerica tra i punteggi!

N.B. Non viene restituito il punteggio relativo alle classi con più del 50% di assenti, o il punteggio di scuole con dati validi in meno del 50% delle classi.

La varianza

Il dato sulla varianza tra i punteggi nella scuola si trova nei grafici.

- **Varianza TRA/TOT:** indica la variabilità dei risultati tra una classe e l'altra (prendendo in esame il punteggio o la differenza con l'indice ESCS); vengono confrontate la varianza della scuola e quella nazionale.
- **Varianza DENTRO/TOT:** indica la variabilità dei risultati all'interno delle singole classi; vengono confrontate la varianza della scuola e quella nazionale.

La varianza

Ai fini del RAV, ma anche di una valutazione da parte del DS, è opportuno tenere presente che:

- la **varianza TRA/TOT** dovrebbe essere il più possibile **bassa**, indicando così che **tra una classe e l'altra c'è omogeneità**
 - Buoni criteri di formazioni delle classi; nella scuola non sono presenti sezioni *più forti* o *meno forti*, o almeno le differenze sono contenute
 - Lavoro comune del collegio docenti
- La **varianza DENTRO/TOT** dovrebbe essere il più possibile **alta**, indicando così una **eterogeneità interna alla classe**: la scuola favorisce il processo di personalizzazione insegnando a studenti nella stessa classe distribuiti in livelli di apprendimento anche molto diversi tra loro

I livelli

Gli studenti vengono suddivisi in 5 livelli, in base al loro punteggio confrontato con il punteggio medio nazionale:

- **Livello 1:** punteggio minore o uguale al 75% della media nazionale
- **Livello 2:** punteggio compreso tra il 75% e il 95% della media nazionale
- **Livello 3:** punteggio compreso tra il 95% e il 110% della media nazionale
- **Livello 4:** punteggio compreso tra il 110% e il 125% della media nazionale
- **Livello 5:** punteggio maggiore del 125% della media nazionale

Altri dati?

Come già detto, sicuramente questi dati non bastano. L'Invalsi fornisce altri strumenti e altri dati. A noi la scelta di cosa studiare.

Alcuni documenti utili per proseguire:

- Il *Rapporto Risultati*
- La *Sintesi del Rapporto Risultati*
- *I primi risultati delle Prove Invalsi 2015 in dieci punti*
- Il *Rapporto Tecnico*
- I *Quadri di Riferimento* delle prove Invalsi
- **I testi delle prove**
- Le *guide alla lettura* delle diverse prove

I primi cinque documenti sono reperibili cercando da Google (consigliato) o sul sito www.invalsi.it (sconsigliato); i testi e le guide alla lettura sul sito www.engheben.it

L'andamento

- Da quest'anno l'Invalsi restituisce una tabella che registra l'andamento della scuola negli ultimi tre anni in quella materia. Questo è un dato utile, soprattutto se negli anni precedenti non è stato fatto un lavoro serio di lettura dei dati Invalsi.
- Più utile ancora (per chi può) è studiare l'andamento di quel gruppo di classi durante il percorso scolastico.
 - Il dato della V Primaria può essere confrontato con il dato della II Primaria di tre anni prima.
 - Questo dato non viene fornito dall'Invalsi, dobbiamo reperirlo noi.
 - Fino all'anno 2014-2015 è possibile anche il confronto della III Secondaria di I Grado con la I Secondaria di I Grado del 2012-2013; il prossimo anno non sarà più possibile perché dall'anno 2013-2014 non è più stata effettuata la rilevazione nelle classi prime.

Parti della prova, ambiti, dimensioni

È opportuno studiare le tabelle che riportano i risultati della scuola suddivisi nelle varie **parti della prova** per Italiano (testo narrativo, testo espositivo, ecc; l'elenco varia a seconda del grado di scuola) e in **ambiti** e **dimensioni** per la prova di **Matematica** (ambiti: numeri, spazio e figure, ecc; dimensioni: conoscere, risolvere problemi, argomentare, ecc; l'elenco varia a seconda del grado di scuola).

Questo lavoro è interessante per tutti i docenti, non solo quelli di area: se una classe è debole nella comprensione di un testo, tutti i docenti sono interessati, non solo quello di italiano; allo stesso modo, se c'è difficoltà nell'argomentare, non è coinvolto solo il docente di matematica.

Dati suddivisi per genere, regolarità e cittadinanza

Si può decidere, per effettuare uno studio più approfondito, di visionare i risultati suddivisi per genere in alcuni **grafici**.

Allo stesso modo, scuole con un elevato numero di studenti ripetenti (o di anticipi) possono ritenere utile visionare i **grafici** in cui i risultati vengono suddivisi in studenti regolari, anticipatari e posticipatari; oppure studiare **tabelle** con i risultati esclusivamente degli studenti regolari. Questo vale anche per le tabelle relative alle parti della prova, agli ambiti e alle dimensioni.

Dati suddivisi per genere, regolarità e cittadinanza

Una scuola con un elevato numero di studenti stranieri vorrà studiare i **grafici** in cui gli studenti vengono suddivisi in:

- **Nativi** (o italiani): alunni nati in Italia e con almeno un genitore nato in Italia.
- **Stranieri di I generazione**: alunni nati all'estero e nessuno dei due genitori nato in Italia.
- **Stranieri di II generazione**: alunni nati in Italia da genitori stranieri.

Oppure vorrà prendere visione delle **tabelle** in cui vengono riportati i **risultati dei soli nativi**.

I microdati

Da quest'anno, l'Invalsi fornisce un file in cui sono contenute le risposte dei singoli studenti.

Le tavole sono di difficile lettura, ma possono dare spunti interessanti:

- È possibile vedere, magari nell'item che abbiamo deciso di studiare, chi ha effettivamente risposto giusto e chi sbagliato; chi non ha risposto...
- È possibile vedere (più arduo!) se i punteggi molto bassi sono di persone che andavano male, o di persone che non si sono impegnate...

Il dettaglio degli item

Lo studio più difficile e lungo, ma anche quello più interessante per il singolo docente, è quello del **punteggio** (percentuale di risposte corrette) **della classe** suddiviso nei **singoli item**.

L'Invalsi fornisce a questo scopo, per ogni classe:

- Una **tabella** con, per ogni item, le **percentuali** di risposta corretta/errata/mancante, o le percentuali di risposta per ogni opzione di risposta (la risposta corretta è quella con la casella colorata)
- Un **grafico** in cui in ascissaci sono gli item, e in ordinata la **differenza** tra la percentuale di risposte corrette in quell'item nella classe e la percentuale di risposte corrette in quell'item in Italia. Quindi un punto alto indica un'eccellenza, un punto sotto il livello 0 indica un item in cui la classe ha ottenuto risultati inferiori alla media nazionale

Il dettaglio degli item

Nel **grafico** è possibile cercare gli item su cui lavorare, fare delle riflessioni... perché la classe ha risposto mediamente peggio dei coetanei.

Nella **tabella** è possibile ricercare in quali item l'attenzione degli studenti si è focalizzata su un distrattore. È utile in questa analisi farsi aiutare dalla *Guida alla lettura* della prova in questione.

N.B. Percentuali di risposte basse possono non allarmare: non è vero che «*mal comune, mezzo gaudio*», ma dobbiamo anche verificare se quel particolare item aveva un elevato indice di difficoltà. In questo senso, per evitare lunghe ricerche, il grafico costituisce già di per sé un buon indicatore di questo fatto, poiché non analizza la percentuale di risposte ma la differenza con il dato nazionale.

L'uso del *Rapporto Risultati*

Per le Scuole Secondarie di II Grado, lo strumento della tabella in cui sono riportate le percentuali di risposta nei singoli item ha un'ulteriore utilità.

Nel *Rapporto Risultati* è presente infatti l'elenco delle percentuali delle risposte per item, non soltanto in generale in Italia, ma anche suddiviso per Licei, Tecnici e Professionali.

Allo stesso modo, nel *Rapporto Risultati* è presente anche il confronto dei punteggi generali suddiviso per area, per regioni e per tipologia di Scuola Secondaria di II Grado.

Questo permette di confrontare i risultati della propria scuola con le altre scuole della stessa tipologia (ma non di indirizzo: non è possibile attualmente separare i risultati in matematica dei licei scientifici da quelli dei licei delle scienze umane, per esempio).