

IL CONFRONTO TRA LE COORTI 2019 E 2021: LE BASI PSICOMETRICHE

Nota Metodologica

Marta Desimoni, PhD
INVALSI

Le rilevazioni degli
apprendimenti
A.S. 2020-21

Il confronto tra le coorti 2019 e 2021: le base psicometriche

Uno degli obiettivi più importanti nelle indagini su larga scala in campo educativo è il delineare l'andamento degli apprendimenti nel tempo. Tale obiettivo è particolarmente rilevante negli ultimi anni scolastici, dato il possibile impatto delle misure di contrasto della diffusione della pandemia globale dovuta al COVID-19 sul mondo della scuola. Il presente documento si pone l'obiettivo di illustrare la procedura metodologico-psicometrica utilizzata da INVALSI al fine di comparare gli esiti delle rilevazioni nazionali degli apprendimenti degli allievi della scuola primaria in due anni scolastici (a.s.): l'a.s. 2018-19 e l'a.s. 2020-21.

Nel disegno della rilevazione INVALSI per la scuola primaria, ogni anno sono somministrate agli allievi di seconda primaria prove standardizzate “carta e matita” di Italiano¹ e Matematica. In quinta primaria, dall'a.s. 2017-18 le prove standardizzate “carta e matita” sono tre, Italiano, Matematica e Inglese, per un totale di quattro scale (Italiano, Matematica, Inglese-Ascolto e Inglese-Lettura). Date le caratteristiche della rilevazione, censuaria e con rilascio contestuale delle prove, il disegno della rilevazione INVALSI per la scuola primaria prevede la somministrazione ogni anno di prove diverse, al fine di evitare eventuali distorsioni legate alla sovraesposizione degli item che le compongono. Tale disegno consente di ottenere ogni anno un quadro del sistema scolastico. Non consente, tuttavia, di comparare direttamente i risultati conseguiti per ogni grado di scolarità (per es. la quinta primaria) da coorti diverse di allievi, dunque il passaggio da una valutazione di tipo sezionale dei risultati ad una valutazione diacronica o di *trend*. Prove diverse costruite per misurare una stessa caratteristica non osservabile (variabile latente), seppure pensate per avere proprietà simili, possono infatti differire per caratteristiche psicometriche, tra cui la difficoltà. Se due soggetti rispondono a due diverse prove, ottenendo due punteggi diversi (per esempio, diverse percentuali di risposte corrette), in assenza di altre informazioni è impossibile sapere se la differenza nel punteggio ottenuto sia attribuibile a differenze nella caratteristica rilevata tra i due rispondenti o semplicemente al fatto che uno dei due test fosse più facile dell'altro. Il confronto tra le stime dell'abilità di soggetti differenti che rispondono a test differenti, seppure pensati come misure di uno stesso costrutto, non può essere effettuato neanche nell'ambito dei modelli di risposta all'item, nel caso in cui non sia stabilita una scala di misura comune (von Davier, Carstensen, von Davier, 2006).

¹ nelle sole classi campione è inoltre somministrata una prova di lettura

Data la rilevanza della comparazione degli esiti di apprendimento tra le coorti di allievi frequentanti la scuola primaria nell' a.s. 2018/19 e nell' a.s. 2020/21, l'INVALSI ha implementato un disegno di ancoraggio e delle procedure di *linking* finalizzati a rendere possibile il confronto tra la stima dell'abilità degli allievi delle due coorti. Il disegno di ancoraggio utilizzato ha previsto l'estrazione, dal campione delle rilevazioni INVALSI 2021, di un "sottocampione di ancoraggio", ossia un ampio gruppo di allievi cui sono state somministrate sia le prove INVALSI 2021 sia una prova INVALSI 2019 (Matematica, Italiano o Inglese). Lo schema del disegno di ancoraggio è rappresentato nella figura 1. Per maggiori informazioni sui possibili disegni di ancoraggio, si rimanda a Wright e Stone (1979), von Davier, Carstensen e von Davier (2006), Kolen e Brennan (2014).

Figura 1. Schema del disegno di ancoraggio per le rilevazioni INVALSI 2019 e 2021.

		Prova INVALSI 2019	Prova INVALSI 2021
Campione INVALSI 2019	Allievi delle classi campione		
Campione INVALSI 2021	sottocampione di ancoraggio		
	allievi delle classi campione (a esclusione del campione di ancoraggio)		

Nel sottocampione di ancoraggio, le prove delle rilevazioni INVALSI 2021 sono state somministrate nelle stesse date previste per tutti gli allievi dello stesso grado di scolarità e secondo la procedura prevista da INVALSI per la somministrazione nelle classi campione. La somministrazione della Prova INVALSI 2019 nel sottocampione di ancoraggio è avvenuta circa 15 giorni prima dell'inizio del calendario delle somministrazioni INVALSI 2021 per la primaria, dunque nello stesso periodo dell'anno scolastico della rilevazione nazionale, sempre alla presenza di un osservatore esterno. La scelta di somministrare le prove in due occasioni è stata effettuata al fine di evitare di affaticare gli allievi proponendo loro due prove nello stesso giorno.

La richiesta di adesione al progetto di ancoraggio da parte delle scuole è stata preceduta da incontri (a distanza) organizzati da INVALSI al fine di illustrare gli obiettivi del progetto e la procedura utilizzata. Per ogni classe partecipante al progetto di ancoraggio, è stata registrata

l'informazione rispetto all'esposizione pregressa degli allievi alla prova INVALSI 2019 (per esempio, per un uso della prova come strumento per la didattica); in base a quanto dichiarato dal personale docente si è proceduto con l'eventuale esclusione della classe dal campione di ancoraggio. Inoltre, le prove INVALSI 2019 sono state rese non disponibili sul sito INVALSI a partire dall'autunno 2020 fino al termine del progetto. Sono ovviamente rimaste a disposizione di coloro che fossero interessati le prove degli altri anni scolastici. La numerosità dei sottocampioni di ancoraggio e la relativa distribuzione per area geografica sono riportate nella tabella 1.

Tabella 1. Numerosità dei sottocampioni di ancoraggio per grado e scala

Grado scolastico	Scala	Nord ovest	Nord est	Centro	Sud	Sud e Isole	Totale
V primaria	Inglese-ascolto	383	454	295	337	323	1792
	Inglese-lettura	375	420	283	306	305	1689
	Italiano	385	585	345	360	267	1942
	Matematica	380	558	366	376	260	1940
II primaria	Italiano	398	515	374	347	235	1869
	Matematica	383	525	396	341	230	1875

Gli strumenti coinvolti nel progetto sono le prove INVALSI delle rilevazioni 2019 e 2021. Seppure diverse per item e stimoli che le compongono, tali prove hanno in comune alcune importanti caratteristiche. In *primis*, esse sono pensate per rilevare gli stessi costrutti teorici. Ossia si ipotizza che per ciascuna scala (Italiano, Matematica, Inglese-lettura, Inglese-Ascolto) una sola variabile latente influenzi le risposte degli allievi agli item (variabili osservate) entro ciascuna coorte e tra le coorti. I costrutti oggetto di indagine nelle prove INVALSI di Italiano e Matematica della scuola primaria sono definiti dai relativi Quadri di Riferimento (INVALSI, 2018a, 2018b), uno per ambito disciplinare, scritti in coerenza con le Indicazioni nazionali per il curriculum della scuola dell'infanzia e del primo ciclo di istruzione (D.M. n. 254 del 16/11/2012). Le prove INVALSI di Inglese, anch'esse costruite tenendo conto delle Indicazioni nazionali per il curriculum della scuola dell'infanzia e del primo ciclo di istruzione (D.M. n. 254 del 16/11/2012), hanno come principale riferimento il Quadro Comune di Riferimento per le lingue (QCER) del Consiglio d'Europa (Council of Europe, 2001). Per ogni grado e disciplina oggetto di indagine, inoltre, le prove proposte alle diverse coorti scolastiche hanno una struttura simile e sono coerenti nel tempo i criteri alla base della selezione degli stimoli e i formati delle domande, come definito nelle linee generali dai relativi QdR (INVALSI, 2018a; INVALSI 2018b) o descritto da materiali illustrativi prodotti da INVALSI (per

es. le Guide alla lettura della Prova di Inglese, INVALSI 2019; 2021). Un ulteriore elemento a garanzia della consistenza delle prove INVALSI nel tempo è il modello psicometrico di riferimento, ossia il modello di Rasch per item dicotomici (1960; 1980), ampiamente utilizzato nella ricerca educativa e considerato un *gold standard* per la costruzione di strumenti di rilevazione in ambito educativo.

Il processo di *linking* degli esiti delle rilevazioni delle due coorti è stato realizzato nella cornice psicometrica del modello di Rasch (1960; 1980) con il programma Acer ConQuest (versione professional 5.11.4, Adams, Wu, Cloney & Wilson, 2020). Il modello di Rasch (1960; 1980) descrive in termini probabilistici l'esito dell'interazione tra un rispondente e un item dicotomico sulla base dell'abilità del soggetto e della difficoltà dell'item. Se un test è costruito coerentemente al modello di Rasch (1960; 1980) e alle sue assunzioni (unidimensionalità, monotonicità e indipendenza locale), allora gli "oggetti della misurazione", item e soggetti, saranno "misurati" su una scala lineare a intervalli equivalenti, con un'unità di misura comune: il logit (Brodgen, 1977). Coerentemente alle proprietà delle scale a intervalli equivalenti, le differenze tra le posizioni dei soggetti sul continuum rappresentante l'abilità latente, così come le differenze tra le posizioni degli item, avranno un significato invariante nei livelli di abilità considerati; inoltre, sulla base del principio dell'oggettività specifica, sarà possibile confrontare gli "oggetti" di misurazione indipendentemente dalle condizioni specifiche di osservazione. Questo vuol dire che i confronti tra item saranno indipendenti dai soggetti ai quali gli item sono stati somministrati in fase di stima dei parametri (calibrazione), e che i confronti tra i soggetti saranno indipendenti dagli item somministrati ai soggetti per misurare la proprietà in esame. Il confronto tra le misure ottenute da due (o più) forme del test, tuttavia, richiede che esse siano espresse su una scala di misura comune.

Preliminarmente al processo di *linking* vero e proprio è stato testato empiricamente, per ciascun grado scolastico e disciplina:

- l'adattamento al modello di Rasch (1960; 1980) di ciascuna prova per ogni occasione di somministrazione (per es. attraverso gli indici di adattamento al modello *Weighted Mean Square*, *Weighted-MNSQ*, o *infit*, e *Unweighted Mean Square*, *Unweighted-MNSQ*, o *outfit*, Wright e Masters, 1982);
- per la sola prova della rilevazione INVALSI 2019, l'eventuale presenza di item che presentassero un funzionamento differenziale tra le coorti, con un *bias* noto in letteratura come *Item Parameter Drift* (Goldstein, 1983; O'Neill, Peabody, Tan & Du, 2013);

- che gli item della prova INVALSI 2019 e della prova INVALSI 2021 potessero essere considerati indicatori di uno stesso costrutto teorico, con analisi condotte sulla base dei dati raccolti nel sottocampione di ancoraggio, in linea con la letteratura scientifica sul *linking* di strumenti di rilevazione diversi tramite soggetti comuni (Bond & Fox, 2007).

Come si può osservare dalla figura 1, nella matrice dati della rilevazione 2021 sono dunque presenti allievi che hanno risposto sia agli item della prova 2019 sia agli item della prova 2021 (il sottocampione di ancoraggio). Attraverso la procedura nota in letteratura come “*Fixed parameter calibration method*” (FPC, per es. Kim, 2006; König, Khorramde, Yamamoto & Frey, 2021), nel processo di calibrazione condotto sulla matrice dati del campione INVALSI 2021, i parametri di difficoltà degli item della prova INVALSI 2019 sono stati fissati al valore stimato sul campione INVALSI 2019 (con l’esclusione degli item presentanti un funzionamento differenziale sostanziale tra le coorti, O’Neill *et al.*, 2013). Dunque i dati raccolti nel 2021 non contribuiscono alla stima dei parametri della prova INVALSI 2019, considerati come noti, ma alla calibrazione dei parametri della prova INVALSI 2021 e alla stima dell’abilità degli allievi della rilevazione INVALSI 2021. A esito del processo di *linking*, è stato possibile ottenere la stima della posizione relativa degli item non ancorati (i parametri di difficoltà degli item della prova INVALSI 2021) e di ciascun rispondente (la stima dell’abilità) su una scala rappresentante il *continuum* della variabile latente oggetto di indagine. Tale scala, identificata tramite l’ancoraggio dei parametri della prova INVALSI 2019, è espressa sulla stessa metrica della rilevazione INVALSI 2019, rendendo comparabili i risultati delle due rilevazioni. Tale procedura è stata confrontata con altre procedure di *linking*, per esempio l’ancoraggio psicometrico dei parametri dei soggetti comuni, verificando che esse producano risultati consistenti. Le differenze nella difficoltà relativa degli item della prova 2019 tra la somministrazione del 2019 e del 2021 sono state utilizzate per calcolare l’*errore di link* (Monseur & Berenitz, 2007).

RIFERIMENTI BIBLIOGRAFICI

Adams, R.J, Wu, M.L, Cloney, D., Wilson, M.R. (2020). *ACER ConQuest: Generalised Item Response Modelling Software* [Computer software]. Version 5. Camberwell, Victoria: Australian Council for Educational Research.

Bond, T. & Fox, C.M. (2007). *Applying the Rasch Model. Fundamental measurement in the human sciences*. (Second Edition). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Brogden, J. E. (1977). *The Rasch model, the law of comparative judgment and additive conjoint measurement*. *Psychometrika*, 42(4), 631-634.

Goldstein, H. (1983). *Measuring changes in educational attainment over time: problems and possibilities*. *Journal of Educational Measurement*, 20 (4), pp. 369-377.

INVALSI (2018a). Quadro di Riferimento delle prove INVALSI di Italiano. Disponibile da: https://invalsi-areaprove.cineca.it/docs/file/QdR_ITALIANO.pdf

INVALSI (2018b). Quadro di Riferimento delle prove INVALSI di Matematica. Disponibile da: https://invalsi-areaprove.cineca.it/docs/file/QdR_MATEMATICA.pdf

INVALSI (2019) Servizio Nazionale di Valutazione a. s. 2018/19. Guida alla lettura Prova di Inglese – Fascicolo 1 Classe Quinta – Scuola primaria. Disponibile da: https://invalsi-areaprove.cineca.it/docs/2019/Guida_lettura_G05_ENG_2019.pdf

INVALSI (2021) Servizio Nazionale di Valutazione a. s. 2020/21. Guida alla lettura Prova di Inglese – Fascicolo 1 Classe Quinta – Scuola primaria. Disponibile da: https://invalsi-areaprove.cineca.it/docs/2021/Guida_alla_lettura_ENG_G05_MS2021.pdf

Kim, S. (2006). *A Comparative Study of IRT Fixed Parameter Calibration Methods*. *Journal of Educational Measurement*, (43), 4, pp. 355-381.

Kolen, M.J. & Brennan, R.L. (2014). *Test Equating, Scaling and Linking. Methods and Practices* (Third Edition), New York, Ny: Springer.

König, C., Khorramde, L., Yamamoto, K., Frey, A. (2021). *The benefits of Fixed Item Parameter Calibration for Parameter Accuracy in Small Sample Situations in Large-scale Assessments*. *Educational Measurement: Issues and Practice*, 40, 1, pp. 17-27.

O'Neill, T., Peabody, M., Tan, R.B.J, Du, Y. (2013). *How much item drift is too much?* *Rasch Measurement Transactions*, 27, 3, pp. 1423-1424.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1980) *Probabilistic models for some intelligence and attainment tests*. (Expanded edition). Chicago: The University of Chicago Press.

von Davier, A.A., Carstensen, C. H., von Davier, M. (2006) *Linking Competencies in Educational Settings and Measuring Growth*. Princeton, NJ: ETS.

Wright, B. D. & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Wright, B.D. & Stone M.H. (1979). *Best Test Design. Rasch Measurement*. Chicago, Illinois: MESA PRESS.