

Performance Decline and the Stake of Tests

Discussion

Flavia Coda Moscarola – Planning, Research and Evaluation Department
Compagnia di San Paolo and CeRP-Collegio Carlo Alberto

Summary

- The paper analyzes the effect of question position on the performance of students in a test
 - Focusing on INVALSI Maths tests (wave 2015) for grades 6(5?) and 8
 - Exploiting the fact that the allocation of booklets is randomized
 - Highlighting the difference:
 - btw grade 6 (low-stake) and grade 8 (high-stake) tests
 - by gender
 - (by age/experience)
- They find that
 - Students' performance declines with the increase in the position of questions, but the decline is lower
 - in grade 8 (high-stake) test than in grade 6 (low-stake) test → due to increased effort and commitment
 - for girls than for boys in grade 8 (high stake) test → due to the supposed higher effort and commitment of girls w.r.t. boys

Comments

- The paper is a spin-off of the seminal paper by Brunello et al. (2021).
- The research question is highly relevant. Investigating (*ceteris paribus*) :
 - the effect of questions position on the students' performance in high-stake vs low-stake tests
 - whether there is evidence of a differential effect by gender or by age/experience

Contribute to many strands of literature (e.g: on the determinants of student's performances in tests; on the gender gaps; on the way to estimate the impact of different programs using effective tests) and potentially provide interesting suggestions for new program interventions/new approaches to education and testing.

- In this preliminary version of the work, the effect of questions position on students' performance in tests (detected in Brunello et al., 2021) is confirmed also for grade 8 students. However, the differential effect observed in high vs low-stake tests still mixes up, in my opinion, a few different effects: the "effort and commitment" as suggested by the authors, but also maybe: a cohort effect, a test/type effect and a maturity/age effect.

Comments

The cohort effect: Actually, the authors compare the test for Grade 6 (5?) (low-stake) with the one for Grade 8 (high stakes) of the same INVALSI wave (the 2015 one). Students do indeed pertain to two different cohorts. Could be possible to run an analysis using (more recent and) longitudinal data (at least on a sample of students) and control in this way for cohort effect?

(Related to this: in the slide with “the summary of results”, it is not clear to me what “we look at the same population” means).

The type of test effect: It seems to me that Grade 6 and Grade 8 tests do actually differ (but I don't know about the year 2015) because one is Paper&Pencil and the second is Computer based. How can this affect results? In Paper&Pencil questionnaires, in principle, students can answer to questions in a different order w.r.t. the one proposed. Can the effect of the Position be affected by that? In Brunello et al. (2021) this hypothesis is tested. What about Computer assisted tests? Is it still possible to answer in a different order?

Organizzazione delle Prove nazionali INVALSI			
Grado scolastico	Classe	Tipo di prova	Discipline esaminate
2	II Scuola Primaria	PPT - Prova cartacea	Italiano Matematica
5	V Scuola Primaria	PPT - Prova cartacea	Italiano Matematica Inglese
8	III Scuola Sec. I Grado	CBT - Prova al computer	Italiano Matematica Inglese
10	II Scuola Sec. II Grado	CBT - Prova al computer	Italiano Matematica
13	Ultimo anno Scuola Sec. II Grado	CBT - Prova al computer	Italiano Matematica Inglese

Comments

- A maturity/age effect: it seems to me a good idea to exploit the information related to the months of birth to disentangle maturity/age effect from the effort/commitment effect, but I was wondering whether it is not preferable and possible to use it as an IV, instead of as a simple interaction.
(by the way, how do you exactly interpret the estimates you get for “position*months” in the two equations?)
- In the paper by Brunello et al. (2021) the list of personal characteristics that can be exploited to control for the influences of environmental factors (trained to test, class dimension, etc...) and soft-skills is longer than the one presented in this (preliminary) analysis. Is it possible to add (those additional) controls to the analysis?
 - Is information related to geographic area (or, even better, the city and the quartier in which the school is) available?
 - Is information related to school overall performance/reputation available?
- Perhaps, it can be interesting to analyse the effect determined by the “language” and the “gender-specific conceptual frames” used in the booklets
 - Boggio et al. (2020) investigated the effect of non gender-neutral conceptual frames in pupils’ performance in a test/game built up to elicit patience levels.