

Institution: Queen Mary University of London		
Unit of Assessment: 16 Economics and Econometrics		
Title of case study: School Performance, Score Inflation and Education Policy in Italy		
Period when the underpinning research was undertaken: 2011-2017		
Details of staff conducting the underpinning research from the submitting unit:		
Name(s): Erich Battistin	Role(s) (e.g. job title): Professor of Economics	Period(s) employed: Sept 2013–Feb 2018
Period when the impact occurred: 2016–2020		
Is this case study continued from a case study submitted in 2014? N		
1. Summary of the impact		
<p>Research by Professor Battistin has substantially influenced the evaluation procedures adopted by INVALSI, the Italian Government agency in charge of assessing educational attainment in primary and secondary schools. The research fundamentally shaped policy-makers' understanding of the flaws in the existing assessment which had produced score inflation in the early years of the tests. Specifically, the research highlighted inherent weaknesses in grading accountability and protocols that created incentives for teachers to inflate scores to economize on effort. The research provided the basis for INVALSI to reform monitoring systems to tackle improper practices in the administration and grading of national exams. These reforms have eliminated score inflation, increased confidence in INVALSI's evaluation of student performance, and provided fairer assessments of learning achievement nationwide.</p>		
2. Underpinning research		
<p>Standardized testing has become the accepted protocol to measure the quality of teaching and learning for schools and students alike. Globally, performance-based accountability policies are used to allocate public funding, design pay incentives for teachers, and target underperforming areas of a country's education system. Test-based accountability may create incentives for schools, teachers, and students to manipulate results, as evidenced by cases across Europe, the UK and North America.</p> <p>Italian schools have long used internal exams to track students' transition from elementary-to-middle, and middle-to-high school, but standardized testing is a more recent development. National, standardised tests from the government agency INVALSI became mandatory for all primary and secondary schools in 2009, with the purpose of informing and guiding national education policy.</p> <p>Comparative studies of student achievement using international standardized tests from the <i>Trends in International Mathematics and Science Study</i> (TIMSS) and the <i>Progress in International Reading Literacy Study</i> (PIRLS) show that Italian primary schools perform at a lower level than many other European counterparts, with students in the country's Southern regions well behind those in the North. However, INVALSI's testing suggested an opposite regional pattern, with students in the South outperforming their Northern peers. INVALSI data used for this case study also show that lower per capita income in the South is associated with higher scores, and that public spending is inversely related to achievement. Taken at face value, this evidence undermines the need for proactive EU involvement to support the modernization of education in Southern Italy, as defined in the Italian National Operative Programme scheme.</p> <p>In research started in 2013 in collaboration with Angrist (MIT, USA), De Nadai (UNSW, Australia) and Vuri (University of Rome, Italy) and published in 2017 ([3.1] and [3.2]), Battistin has shown that score inflation was more frequent in smaller classes relative to larger ones. This explained significantly higher score inflation in the INVALSI tests in the South compared to the North, due to a higher prevalence of small classes. Battistin's contribution stems from his expertise in the economics of education and econometrics. His role in the research was to gather the administrative records and design an empirical strategy for complex policy evaluation, and to coordinate and</p>		

manage the relationship with public institutions including access to data, agreement of protocols, timeline and collaborations. He also led the dialogue on the implications of the research findings for the design of future INVALSI assessments.

Battistin's research [3.1] demonstrated the relevance of social constraints imposed by peer monitoring in the administration of INVALSI tests. Specifically, local teachers are in charge of proctoring and grading of tests. Until 2013 (the period covered by the research), teachers were required to hand copy students' original answers onto machine-readable score sheets, a burdensome clerical task that had to be completed within a few days of testing, typically after school hours and without overtime compensation. As score sheets for larger classes were transcribed by a team rather than an individual, score manipulation to economize on grading effort was more prevalent in small classes because of lower peer monitoring. While score inflation was already a cause for concern for INVALSI, the underlying causes were unknown. Battistin's research showed how the extra uncompensated work required for grading exams, combined with limited monitoring, led to upwardly-biased grading. Specifically, evidence shows that teachers in small classes often forwent individual score transcription and pasted entire sheets of official answers onto the machine-readable score report forms, regardless of students' original answers. Indeed, improved academic achievement in smaller classes found in INVALSI data was due to a higher incidence of score misreporting in those classes. Due to shortcomings in INVALSI's grading protocols, score misreporting was the result of under-provision of grading effort, rather than cheating and score embellishment motivated by motives of self-promotion. The research [3.2] also shows that adjusting for score manipulation explains the apparent discrepancy between conclusions based on data from INVALSI and international surveys (TIMMS and PIRLS).

3. References to the research

[3.1] Battistin, E., De Nadai, M., & Vuri, D. (2017). Counting rotten apples: Student achievement and score manipulation in Italian elementary Schools. *Journal of Econometrics*, 200(2), 344-362. doi.org/10.1016/j.jeconom.2017.06.015

[3.2] Angrist, J. D., Battistin, E., & Vuri, D. (2017). In a small moment: Class size and moral hazard in the Italian Mezzogiorno. *American Economic Journal: Applied Economics*, 9(4), 216-49. doi.org/10.1257/app.20160267

4. Details of the impact

This research on school testing regimes in Italy has resulted in three primary impacts:

Informing INVALSI's grading practices to combat score inflation

About 500,000 primary school students participate in the national test every year. It is estimated that, prior to the changes discussed below, between 15% and 20% of the exam results in some Southern regions were inflated. Battistin's work was used by INVALSI as evidence in their internal presentations to set out the support for revising monitoring programs and tackle improper practices in the administration and grading of national exams [5.1, paragraph 3]. As a direct response to Battistin's findings, INVALSI moved to a system that limits the scope for score manipulation during the transcription of students' responses. Specifically, as of November 2017, INVALSI introduced a formal obligation to disclose results to several bodies within and outside the school.

Moreover, direct electronic transmission to INVALSI of students' original responses has replaced the practice of manual transcription by teachers onto machine-readable score sheets. This change improved the transparency of the grading process, as well as the agency's ability to monitor the quality of data transmitted. Finally, multiple versions of the same exam with questions presented in random sequence are now administered during testing [5.1, paragraph 3]. As a result of these changes, exams in recent years have seen a much lower incidence of score manipulation (see Section 4.3 below).

Enhancing policymakers' understanding of educational testing in Italy

Battistin's analysis of the INVALSI data indicated at an early stage that there was evidence of score manipulation in educational testing in Italy. Battistin communicated the preliminary findings of his research to policymakers at INVALSI and solicited their feedback on the implications for the design of standardised testing. Preliminary results were discussed with the INVALSI President, who authorised Battistin to link INVALSI data to administrative records on schools and students from the Ministry of Education to carry out this research (the matching is mentioned in [3.2]).

In September 2016, results from Battistin's work were presented at a public event organised by INVALSI, aimed at school managers, teachers and policy makers, to promote test-based accountability and explore the factors shaping education inequalities across Italian regions ([3.1], paragraph 2). Finally, in October 2016, Battistin expounded his research and the role of INVALSI testing to a broader non-technical audience at the Senate of the Italian Republic, where he was invited to deliver a training course on the design and evaluation of education interventions in Italy. About 40 officers of the Senate and representatives of Italian Regions and Provinces enrolled on this course. Battistin presented findings from his research, shaping government departments' understanding of monitoring and evaluation protocols for the school system and the associated design of public funding ([5.1], paragraph 2).

Informing the national discourse on the discrepancy between international assessments and INVALSI assessments

The latest waves of INVALSI assessment data provide a picture of regional differences in attainment in line with international surveys (as conjectured by [3.1]). While INVALSI's monitoring systems are set up as low-stakes assessments, whereby school funding is not directly tied to achievement, they remain the main source for reporting on educational attainment at the national and sub-national levels and provide grounds to motivate the need for EU schooling investments in underperforming areas. Battistin's work has identified the causes of the discrepancy between international assessments and INVALSI assessments. Further, the changes introduced by INVALSI as a result of this research helped to rebuild the reliability of INVALSI's assessment [5.1, paragraph 3].

5. Sources to corroborate the impact

[5.1] [Testimonial] President of INVALSI.