



# When the cat is near, the mice won't play: The effect of external examiners in Italian schools <sup>☆</sup>



Marco Bertoni <sup>a</sup>, Giorgio Brunello <sup>b,\*</sup>, Lorenzo Rocco <sup>c</sup>

<sup>a</sup> University of Padova and Centre for Economic Performance, LSE, Italy

<sup>b</sup> University of Padova, IZA and CESifo, Italy

<sup>c</sup> University of Padova, Italy

## ARTICLE INFO

### Article history:

Received 5 June 2012

Received in revised form 20 January 2013

Accepted 25 April 2013

Available online 7 June 2013

### JEL classification:

C31

H52

I2

### Keywords:

Education

Testing

External monitoring

Indirect treatment effects

## ABSTRACT

We use a natural experiment to show that the presence of an external examiner has both a direct and an indirect negative effect on the performance of monitored classes in standardised educational tests. The direct effect is the difference in the test performance between classes of the same school with and without external examiners. The indirect effect is the difference in performance between un-monitored classes in schools with an external examiner and un-monitored classes in schools without external monitoring. We find that the overall effect of having an external examiner in the class is to reduce the proportion of correct answers by 5.5 to 8.5% – depending on the grade and the test – with respect to classes in schools with no external monitor. The direct and indirect effects range between 4.3 and 6.6% and between 1.2 and 1.9% respectively. Using additional supporting evidence, we argue that the negative impact of the presence of an external examiner on measured test scores is due to reduced cheating (by students and/or teachers) rather than to the negative effects of anxiety or distraction from having a stranger in the class.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

A problem with test-based accountability systems in education is that they generate incentives for teachers, students and school administrators to “game” the system in order to obtain better results. The manipulation of test outcomes generates efficiency losses both when these outcomes are used to allocate resources to schools and teachers and when – more modestly – they provide valuable benchmarking information which can affect the choices of schools and their stakeholders.

One mechanism for inflating test scores is outright cheating. Empirical analysis of cheating behaviour is scarce.<sup>1</sup> In their influential study, Jacob and Levitt (2003) develop an algorithm for detecting

teachers' cheating that combines information on unexpected test score fluctuations and suspicious patterns of answers for students in a class. They find that a small fraction of Chicago teachers responded to accountability pressures by completing student examinations in an attempt to improve outcomes.

A possible deterrent of forms of cheating that may occur during the test – e.g. students copying from one another or teachers communicating the correct answers – or during the scoring – e.g. teachers changing students' answers or filling in missing answers – is monitoring by external examiners. External monitoring has costs and benefits. Costs increase with the desired level of coverage. Benefits depend both on the efficiency gain associated to a reduction in cheating and on how effective monitoring is in influencing test scores and reducing cheating.

In this paper, we estimate the impact of external monitoring on test scores, using a rather unique natural experiment designed by the Italian central test administrator (INVALSI), which assigned external examiners to randomly selected classes and schools with the task of monitoring students taking the test and reporting results.<sup>2</sup> We compare test outcomes in the classes with an external examiner

<sup>☆</sup> The authors are grateful to Erich Battistin, Thomas Breda, Daniele Checchi, David Figlio, Ifty Hussain, Edwin Leuven, Marco Manacorda, Guy Michaels, Hessel Oosterbeek, Steve Pischke, Olmo Silva and to the audiences at the 2012 LSE-CEP Annual Conference, the 2012 HECER Economics of Education Summer Meeting in Helsinki, the APPAM-INVALSI conference “Improving Education through Accountability and Evaluation” in Rome, and at seminars in Padova and CIDE (Bertinoro) for comments and suggestions. We also thank Patrizia Falzetti, Roberto Ricci and Paolo Sestito (INVALSI) for helping us with data collection and for explaining several technical features of the administration of the SNV tests. Financial support by the Ministry of Italian Universities (PRIN contract n. 2009MAATFS\_002) is gratefully acknowledged. All errors are our own.

\* Corresponding author at: Via del Santo 33, 35100 Padova, Italy.

E-mail address: [giorgio.brunello@unipd.it](mailto:giorgio.brunello@unipd.it) (G. Brunello).

<sup>1</sup> See Figlio and Loeb (2011), for a review of the recent literature.

<sup>2</sup> These tests are taken by the universe of primary second and fifth grade students. INVALSI sampled a number of classes and schools for external monitoring to obtain reliable data, speed up data collection and verification and prepare an annual report on the state of primary education in Italy.

with the outcomes in other classes, where the test was administered by a local teacher, and find that the rate of correct answers is lower in the former than in the latter. Using additional supporting evidence, we argue that the negative impact of the presence of an external examiner on measured test scores is due to reduced cheating (by students and/or teachers) rather than to the negative effects of anxiety or distraction from having a stranger in the class.

Our study contributes to the literature on school accountability in two main directions. First, we show that the introduction of external examiners has a significant effect on measured test scores in an environment where there are incentives to manipulate results. Second, we document that the monitoring effects of having an external examiner spill over to un-monitored classes of the same school. We decompose the overall effect of external monitoring – which we measure as the difference in the average rate of correct answers in monitored classes and in classes of un-monitored schools – into a direct and an indirect effect. The direct effect is the difference in the test performance between classes with and without external examiners belonging to schools selected for external monitoring. The indirect effect is instead the difference in performance between un-monitored classes in a school with an external examiner and un-monitored classes in schools without external examiners.

We estimate that having an external examiner reduces the percentage of correct answers by 3.6 to 5.4 percentage points – depending on the grade and the test – which corresponds to 5.5 to 8.5% of the average score in classes belonging to schools with no external examiner. The estimated direct effect ranges from 2.8 to 4.2 percentage points (4.3 to 6.6%), and the residual indirect effect from 0.8 to 1.2 percentage points (1.2 to 1.9%). We discuss two alternative reasons why the effects of monitoring spread from the monitored class to the other classes in the same school. The first is that the presence of an external examiner in the school acts as a disciplinary device also on students and teachers in other classes of the same school because of the fear that the examiner may roam about. The second is that teachers dislike excessive dispersion in average class scores within the same school, because of the conflicts it could generate.

We find that the estimated overall effect of external supervision is significantly higher in the schools located in Southern Italy than in Northern schools and in schools where class size is smaller and the proportion of tenured teachers is higher. We show that territorial differences are associated to differences in social capital, even after controlling for territorial differences in GDP per capita and unemployment rates.

Studying the Italian experience with external monitoring has both advantages and disadvantages. The key advantage is that the random allocation of examiners to schools and classes allows us to bypass the selection problems that typically plague the evaluation of monitoring effects. A potential disadvantage is that in the Italian context there is limited accountability of schools and teachers. In this environment, the incentives to cheat may be weaker than in high-stakes contexts. In this case, our estimates can be interpreted as lower bounds of the effect of external monitoring in contexts where the incentives to manipulate results are stronger.

The paper is organized as follows: [Section 2](#) reviews the relevant literature and [Section 3](#) describes the design of the INVALSI test and the dataset. The empirical strategy is presented in [Section 4](#). The main empirical results, a few robustness checks and extensions are reported in [Sections 5, 6 and 7](#), respectively. Conclusions follow.

## 2. Review of the literature

Aside from outright cheating studied by [Jacob and Levitt \(2003\)](#), the literature has identified several indirect ways that teachers and school administrators can use to manipulate student results. On the one hand, [Jacob \(2005\)](#), [Figlio \(2006\)](#), [Figlio and Getzler \(2006\)](#), [Cullen and Reback \(2006\)](#) and [Hussain \(2012\)](#) investigate whether

schools engage in strategic manipulation of the composition of the pool of tested students by excluding low ability students, either by reclassifying them as disabled or by strategically using grade retention and disciplinary suspensions. On the other hand, [Figlio and Winicki \(2005\)](#) show that during testing periods some schools increase the caloric intake provided by school cafeterias so as to boost students' performance. Attempts to increase test scores by taking psycho-stimulant drugs are documented for the US by [Bokhari and Schneider \(2011\)](#), who show that the diagnosis of "attention deficit/hyperactivity disorder" is more frequent in states where there are stronger accountability laws.

To our knowledge, we are the first in this literature to investigate both the direct and the indirect effects of external examiners as deterrents of cheating in standardised tests. That indirect treatment effects can occur has been already pointed out by a broader literature. [Heckman et al. \(1999\)](#), for instance, discuss how policy effects may spread to those not directly participating in the programme mainly because of general equilibrium or spill-over effects. [Miguel and Kremer \(2004\)](#) evaluate both direct and external effects of a Kenyan programme aimed at treating intestinal worm infection among primary school kids. In a similar fashion, [Angelucci and De Giorgi \(2009\)](#) evaluate the effects of *Progesa*, a Mexican aid programme based on cash transfers, and stress the importance of estimating indirect treatment effects on the ineligible when there are social interactions between eligible and ineligible individuals.

## 3. The design of INVALSI Servizio Nazionale di Valutazione (SNV) tests and the data

INVALSI<sup>3</sup> standardised tests in Italian and Math were introduced in Italian primary schools in 2008<sup>4</sup> to evaluate school productivity. The purposes of the evaluation<sup>5</sup> are to inform the central government about the general performance of the school system, and to offer schools a standardised reference to self-assess their strengths and weaknesses, using a value added approach. These tests are not formally high-stakes, because the allocation of resources to schools, the salary of teachers and the school career of students do not explicitly depend on test outcomes. Even so, pressure to perform well in the tests has been high because of the widespread expectations that they might be used at some point to evaluate teachers and schools. These expectations were fostered by the Ministry of Education, who in an intervention at the Lower House of the Italian Parliament (June 10th 2008) when the tests were introduced, made explicit reference to the need to establish within a few years a system of evaluation and incentives for teachers and schools based on student performance in the tests. Schools have an incentive to perform well also because results affect their reputation. Although the outcomes of the tests are not made public by INVALSI, schools have access to the results of their own students and can disclose them to parents and other stakeholders, in an effort to build their reputation and attract good students.<sup>6</sup>

Since 2008 the tests have been administered every year. In this paper, we focus on the 2010 wave because of its peculiar design

<sup>3</sup> INVALSI is the National Institute for the Evaluation of the Education System, in charge of the design and administration of standardised education tests in Italy.

<sup>4</sup> See Law n.147–2007, and the Ministry of Education and Research Decree n.74 and 76–2009.

<sup>5</sup> See article 2 of the INVALSI statute (Ministry of Education and Research Decree n. 11–2011) and the Ministry of Education and Research Directive n. 88–2011.

<sup>6</sup> "INVALSI does not provide public rankings of schools based on the outcomes of the test. The main purpose of the tests is to provide each single school and its stakeholders with valuable information that can help them to benchmark and improve their performance. Each school is free to advertise its own results, using the tools provided by the Ministry of Education..." (free translation by the authors of [Ricci and Sestito, 2012](#)).

Download English Version:

<https://daneshyari.com/en/article/969832>

Download Persian Version:

<https://daneshyari.com/article/969832>

[Daneshyari.com](https://daneshyari.com)