The implementation of merit pay for schoolteachers in Italian schools, 2015-2016: a change of attitudes?

Abstract

In 2015, a new performance-related pay scheme was introduced for schoolteachers in Italy, as part of education accountability policies aimed at improving their performance. From that year, all Italian state school principals were offered the opportunity to distribute wage bonuses to deserving teachers. During the first year of implementation, 82% of the schools appointed ad hoc internal committees and distribute the premia, on average to one teacher out of three. A large majority of schools implemented the process of internal distribution of funds, choosing their own criteria for defining what merit was and how much it should be prized. Results suggest that merit pay was mostly uncorrelated to students' achievements, though more than one round of application would be required for more rigorous tests. However, the article argues that the 2015 reform aligned the Italian system of evaluation and assessment with other international accountability policies.

Keywords

Schoolteachers; education; performance-related pay; merit pay; Italian schools; implementation study

1. Introduction

The Italian government introduced in 2015 (law 107/2015 better known as "*la buona scuola*") a new component of merit-based salary (on average approximately less than 1% of the salary). For the initial year the funds allocated to merit based incentives to teachers corresponded to an average of 250 euro per teacher (a nationwide total of 200 million Euros). Locally formed evaluation committees had total freedom to select teachers and individual premia, though they were required to keep minutes of the adopted criteria (which are currently under study by Indire, an agency of the Ministry of Education) in order to identify best practices. The parliamentary debate signalled potential opposition from the left wing as well as from teachers unions. While unions were clearly aiming to an (almost) egalitarian distribution of the wage increase (and also collected signatures to call for a referendum, without being able to reach the validity threshold of half million), the center-left Renzi government was informally supporting a more skewed distribution, by introducing among the procedural indicators for the evaluation of school principals the degree of uniformity according to which these resources are to be distributed.

A previous experience of teachers evaluation was conducted under the center-right government in 2012 (Trelle Fondazione per le Scuole, 2013). A so-called reputational method of evaluation was applied to a self-selected sample of 33 schools: teachers elected a committee made of two colleagues, who joined the school principal in selecting the best 30% of teachers in each school. A parallel experiment was conducted during in the same period, where teachers were assessed as a group according to the value-added achieved in student competences (Fondazione Agnelli, 2015). The two experiments took rather different perspectives, since the former was a one-shot initiative, based on self-promotion of candidates (teachers self-proposed their candidature by compiling a self-assessment template, but adhesion rate was limited, reaching only 903 out of 2809 teachers

working in the school sample). The latter took three years to develop (in order to gauge the progress of students over a significant time interval) and the premium was a collective one (namely best schools were awarded a fixed number of premia – 20 premia corresponding to 65000 euro each) which were then distributed among teachers within the school according to local preferences (no information is available regarding the distribution criteria).

In the present paper, we conduct for the first time an empirical investigation of the implementation of the massive experiment conducted in 2015 in all state schools in Italy comprising preprimary, primary and secondary levels. Generally speaking, the "Buona Scuola Reform" is under researched and few academic studies have discussed it in-depth. More specifically, the implementation stage of the 2015 reform policy cycle has not yet been investigated systematically. The purpose of this paper is to offer fresh and original insights from the analysis of government data provided by the Ministry of Education and other government agencies. By doing so, we aim to contribute to the academic debates on the difficult implementation of education accountability policies and controversies associated with the introduction of performance-related pay schemes for teachers in public schools in Italy.

Performance-related pay (PRP)[1] was introduced as part of a broader education reform package aimed at improving the quality and performance of teachers and overall accountability of the Italian education system. Teacher's quality matters and it is one of the most important input factors (Braga et al. 2019). Linking rewards to individual performance, instead of a fixed salary, was an innovative aspect of the Renzi Government "modernizing" reform agenda. As Hood and Margetts argue (2010), performance-related pay is at the core of Bentham's eighteenth century

[1] By Performance-related pay (PRP), we refer in this paper to government programmes aimed at incentivise teachers to improve their performance. The rationale for these programmes is that incentive or merit pay may motivate teachers in exerting more effort in their activities, thus improving the quality of teachers input (Duflo et al, 2012).

principle of rational public management and a central tenet of the "Scientific management" movement developed by F. Winslow Taylor. It was introduced in 2014 in England and Wales and has been studied since then by public administration scholars with an interest in education (Marsden, 2015). Performance-related pay plans normally replace the administrative pay system based on grades, with one still based on grades but also on some forms of discretionary bonuses. They have been a global phenomenon and many countries around the world have adopted performance-related schemes for civil servants and schoolteachers. It has been put forward as a strategy to improve the school internal improvement and accountability in 22 out of 28 countries surveyed by the OECD (OECD, 2013).

This paper is a study in the design and implementation of a performance-related bonus for Italian teachers, as adopted in 2015. Linking reward to (past or future) performance of individual school teachers is notoriously difficult to design, and let alone to implement in an effective way. As noted in a recent systematic review of teacher merit pay (Ritter et al, 2019), the majority of programs, at least in the United States, are short-term experiments that have issues of long-term sustainability. They suggest that reforms are viewed as temporary by school staff and teachers and the scope for institutional learning is very low. In addition, these reforms are notoriously difficult to implement effectively due to the lack of evidence that proves any positive effects on motivation (Marsden, 2010). The expected improvement in teachers' motivation has proved elusive, possibly due to existing intrinsic motivation among them. However, the difficult implementation of pay for performance should not alienate scholars from studying this post-decisional process and mechanisms that lead to positive and negative effects. The next section of the paper will review the major theoretical arguments for and against school teachers performance-related pay, that will then be used as assessment criteria for this case study.

2. The Shift to Performance Accountability in Education

Educational accountability reforms, coupled with poor performance of students on international rankings, have attracted interest in the design and implementation of performance-related pay policies as instruments to raise standards and the quality of education system (Hanushek, 2011), the rational being that when teachers are rewarded for good performance, this will consequently improve the quality of teaching. In the United States, the 1983 report "A Nation at Risk" triggered the emergence of the educational accountability reform agenda, based on the introduction of strong assessment measures of teachers' performance and financial incentives to the best performing ones. The report by the National Commission on Excellence in Education warned that if the United States were to compete in a global economy, investment on skills formation was needed. Thus the emphasis of this movement was on academic achievement (Graham, 2005; Jennings, 2012). According to Hamilton and Koretz, accountability is "*a set of policies and procedures that provide rewards and/ or sanctions as a consequence of scores on large-scale achievement tests*" (Hamilton & Koretz, 2002, p. 3).

One of the most prominent policy changes in education accountability has been the issue of raising 'standards' and 'performance' (OECD, 2013). Beginning in the 1980s, especially in the US and UK, the underpinning assumption about the quality of an educational system came to be viewed as its capacity to improve students' achievements. The quality of education came to drive reform expectations and policy programs. Primary and secondary schools in Europe adopted new instruments to improve their accountability mechanisms, through performance indicators, management by objectives, business plans, quality and teachers assessment, and evaluation (author, 2009).

Accountability for performance is a new approach that emphasizes the use of performance mechanisms intended to improve the quality of public services (Hanushek and Raymond, 2001). Performance-related pay schemes are illustrations of this new approach to administrative reforms. Accountability for one's action came to be defined as demonstrating one's performance (Mulgan, 2000). This new approach implies the need to produce new data and information regarding teacher's performance and evaluation (Behn, 2001). It also triggers the adoption of new organizational values.

Performance accountability finds its intellectual and historical tradition in the good estate management. As Day and Klein (1987) discuss it, one of the key historical developments in the system of financial accountability in England was the creation of the statutory Commission for Examining the Public Accounts in 1780, which integrated the notion of balancing the entry with the management of resources, namely economy with efficiency (Day and Klein, 1987: 14). In its literal sense of book keeping, accountability is an Anglo-Norman practice. It can be traced to the Norman conquest of England after 1066, when William I required all the property holders in his realm to render a count of what they possessed. As Day and Klein suggest: "*'managerial accountability' is about making those with delegated authority answerable for carrying out agreed tasks according to agreed criteria of performance*" (1987: 27). According to this conceptualization, managerial accountability is foremost a technical process by which governments ensure that fiscal regularities, efficiency and value for money have been achieved. Accountable management means holding individuals responsible for performance measured on the basis of objective and agreed criteria. Individual managers and their units are answerable for their performance to the actors who have delegated authority to them, such as politicians and citizens. Managerial models of administrative reforms require those with delegated authority, such as heads of hospitals or schools, to be answerable for producing outputs and meeting targets. The values embodied in

managerial accountability are cost effectiveness, efficiency and managerial autonomy (Sinclair, 1995: 222).

Managerial accountability has different dimensions, discussed by Day and Klein (1987). Mainly they refer to financial accountability, as public managers have become responsible for devolved budgets to schools and hospitals. The delivery of public social services in recent decades has been marked by greater devolution of budgets to lower level units within the public sector (Glennerster, 2002). For instance, some schools in England were allowed to 'opt out' of local education authority's control and became more autonomous. Financial accountability consists of spending the allocated money according to appropriate rules within agreed legislative framework. Another important component of managerial accountability is efficiency accountability, which refers to the process of generating value for money. Input and output oriented managerial types of accountability depend ultimately on measuring performance and assessing organizational efficiency. In the case of most social services, outputs are not easily measurable nor are production processes clearly understood.

Insert Table 1. here

Professional accountability, based on professional expertise and specific skills, does not fit the notion of performance accountability. It is a horizontal type of mechanism whereby professionals are answerable to peers rather than principals (according to principal-agent theory). Teachers define their own pedagogic activities with a large degree of 'discretion' from local or central bureaucracies. Being professionally accountable for one's work means to represent the interests and values of particular occupational groups, like teachers, rather than the public interest (Sinclair, 1995). Therefore, performance accountability and professional accountability rest on different mechanisms.

The web of accountability is more extended than a simple hierarchical model, and more so in the case of the delivery of public services (author, 2009). School teachers are located at the centre of a very complex network of account holders, to the extent that performance accountability is multidimensional (Romzek, 1996; author, 2012a). Teachers need to be responsive to peers, to parents, to head teachers, to local and central bureaucracies, and to the society at large. It is an ever-expanding terrain of social relationships. Adding one layer of accountability (performance) over the others (professional) increases the complexity, but whether it makes accountability more effective remains to be tested. More research on the evaluative frameworks of accountabilities is needed. It may be that in fact each line of accountability is weakened in a multi-layered system, whereby teachers and their associations can resourcefully play one accountability holder against the other.

3. Open controversies in the literature

Performance-related pay schemes have been discussed widely in the academic literature, not only by economists of education, but also public policy and political scientists. The general consensus presumes that enhanced accountability in education puts pressures on the teaching professions and school organizations that will eventually promote appropriate behaviors on the part of teachers and improve the overall quality of education, and possibly for this reason it has been ironically defined the "promise of performance" (Dubnick, 2005). However, there is no clear-cut evidence of a positive impact of performance-related pay. Table 2 presents a summary of theoretical arguments in favor and against the use of performance-related pay for school teachers. In the sequel we will shortly review the controversial arguments underneath these propositions.

Insert Table 2 here

Pro-reform Claims:

Performnce-related pay for school teachers has been put forward as a strategy to serve both the quality of teaching (Murphy, 2011; Braga et al., 2019; Atkinson et al, 2004) and the internal management of the school as organization (Marsden, 2015; author, 2012). Financial rewards linked to performance are viewed as appropriate instruments to change behavior and encourage teachers to work better and improve their productivity (Lavy, 2007). Their impact may be long-lasting on students' returns in the labour market, as Lavy shows in a study of Israel policies (Lavy, 2015). Quality of teaching matters as it is positively correlated to students' achievements. Therefore, incentives that change behavior of teachers have a positive impact on the overall quality of education (Lazear, 2003). As far as the internal management of schools is concerned, Marsden (2015) argues that pay for performance is effective in the alignment of individual and organizational goals. He presents robust evidence from the analysis of the British case showing how pay for performance brings positive effects when it interacts with other processes, such as organizational values and organizational goal setting. Performance appraisal is useful as a goal-setting instrument. Performance based management is also more likely to improve accountability for public services, as it generates data and information publicly available to users and decision makers (author, 2012). Evidence to the contrary (that input-based pay systems work more effectively) is challenged by the study of Hanushek and Rivkin (2004). They provide empirical evidence that students' achievements do not improve in the case of Knowledge and Skills pay plans.

Anti-reform claims:

One of the most vociferous arguments against the introduction of pay for performance schemes is that teachers don't support them and that trade unions oppose them. The opposition of teachers'

unions to instruments intended to measure performance has been extensively documented and researched (Ballou, 2001). Merit pay reforms are expected to fail due to strong trade union's opposition. Moreover, in 1986, Murnane and Coehen published one of the fiercest criticisms against linking pay of schoolteachers to performance. Their argument is based on the "nature of teaching" which is different from the private sector. They point to the fact that these plans are introduced as experiments and they never spread widely because they are not suited for the collegial and teamwork characteristic of the teaching profession. Pay for performance encourages isolation and diminishes teachers' morale (Dolton et al, 2003). Richardson (1999) likewise criticizes the performance-related pay scheme because they underpin the cooperation, which is typical of the teaching profession. The output of teachers is not marketed, and thus business-like methods of performance pay are not suitable in the public education sector.

A second line of argument emphasizes the perverse effects of performance accountability, such as teaching to the test and cheating (Koretz, 2002). A third line of criticism concerns the effects of performance pay on motivation. A growing body of empirical evidence shows that pay levels in the teaching professional are not a source of motivation (Vaarlem, 1992). Marsden corroborates these findings by showing that the motivational positive consequences of pay for performance have proved elusive. Results from a survey conducted in England and Wales by Marsden (2015) show that pay for performance, introduced in 2014, was viewed negatively by teachers and that motivational effects were not positive as anticipated by the reform. This is also confirmed by the study by Ritter et al. (2019). Having said that, we note that the arguments put forward by Murname and Cohen in 1986 did not take into account the technological developments obtained in the last twenty years, and the progress governments have made in collecting performance measurement data and monitoring evaluation. It is also worthwhile considering that context matters when assessing the implementation of pay for performance schemes.

4. **The Italian experiment in merit-pay for teachers (2015)**

In 2015, a new political consensus triggered the center-left Renzi Government in Italy to re-introduce in education the mechanisms of performance-based accountability, drawing from international experiences and policy learning from other countries. According to the proponents, a more accountable teaching profession would perform better as it responds to external pressures.

The Renzi government could benefit from a few historical antecedents in Italy. The link between financial rewards and merit-based evaluation was not new in 2015, but had been discussed and introduced for the first time in 1958 by the then-Minister of Education Aldo Moro. He introduced the instrument of career and salary promotion for schoolteachers based on their performance to be assessed through a national public examination (so-called "*concorso per merito distinto*" – see Previtali 2018). This was cancelled in 1974 because it was held not egalitarian. It resurfaced again in 1999, when Luigi Berlinguer was Minister of Education (1996-2000). Art. 29 of the 1999 national contract of schoolteachers introduced merit-pay based for 20% of all schoolteachers (approximately 150.000). It established that 1 out of 5 teachers could be eligible for a monthly pay rise equivalent to 353 Euro (the total annual payment would have been 4.236 euro, gross of taxes) subject to the results obtained at a national public examination (so-called *"concorsone").* Only teachers with 10 years of seniority could participate to the public examination. This system was intended to reward best teachers on the basis of their teaching practices. The reward was significant and based on the need to promote a more meritocratic system.[2] In the summer of 1999, the national contract of teachers spelled out the instruments to be used in the "concorsone" (25% of the mark was linked to the Curriculum Vitae, 25% to the national exam's results, and 50% to the teaching observation and/or mock lesson in front of the examining committee, indicated as

[2] See a recent interview with Luigi Berlinguer, in Corriere della Sera, on July 12th, 2019.

"*verifica in situazione*"). The public examination was called for in December 1999. In February 2000, it was suspended and never implemented thereafter. Minister Berlinguer was forced to resign in April 2000, after fierce opposition by all trade unions of schoolteachers, and direct demonstrations that took the streets of Rome and other Italian cities on February 17th, 2000. Initially, the opposition came from the most radical trade unions, such as Gilda and Cobas, but soon after all major national trade unions opposed vehemently the adoption of merit pay.

The 2015 "Buona Scuola" reform marked a significant shift towards performance accountability, namely a shift from input to output based performance management. "Input" criteria in schools refer to teaching practices, knowledge and skills of teachers, teachers' training and qualifications; "output" mainly concerns students' learning outcomes and student achievements. Pay for performance schemes can be divided in two categories, though there exists a wide variety of designs and forms. On one hand, there are Merit-based Pay schemes that reward teachers in relation to improved output-factors, such as students' achievements and learning outcomes; on the other, there are Knowledge and Skills based Pay plans that emphasize the knowledge and training acquired by teachers, the use of innovative pedagogic methods, but have no relevance for students' outcomes (Podgursky and Springer, 2007; Hanushek and Rivkin, 2004). In the empirical analysis of the 2015 reform, we have classified schools along the input-output based criteria, in order to trace the extent of the shift from input to output accountability mechanisms.

In order to study the Italian implementation of a new merit pay scheme (which involves less than 1% of an average yearly wage) we combined three different datasets. The first one contains information collected in 2014-15 and 2015-16 and made available to schools to endorse self-assessment reports (called *RAV-Rapporto di autovalutazione*, available online for each school). This database contains information on teachers and school principals (numerouness, type of

contract, seniority and level of qualification) as well as on internal management procedures. The second database consists of student achievements according to *prove SNV-Invalsi*, a sort of PIRLS/PISA test which is taken at grades 2, 5, 8 and 10; finally the third database has been directly created by the schools reporting their accomplishment of the new pay scheme (called *Monitoraggio della valorizzazione professionale*[3] - *Merit* for short). As often the case, when working with datasets derived from administrative sources, the main problem is defining the boundaries of the relevant population.[4]

With an overall population of 13040 schools recorded in the school year 2015-16, a fraction (1814) consists of schools where the tests were administered but self-assessment reports are not available.[5] This includes autonomous regions (Valle d'Aosta – 33 schools; Trentino Alto Adige – 92 schools), having local students sitting the national tests but not teachers and school principals compiling the self-assessment reports or distributing merit premia. Similarly, 2137 private schools are excluded, since they were not required to introduce merit premia for teachers, who are paid according to school budgets. Thus our working sample consists of 8997 schools distributed in 18 regions. For 92.7% of them we do have all information about the merit distribution, while for the remaining 653 schools we presume they were unable to appoint a committee leading to the distribution of the merit premia. One should also note that there are few cases (33 schools) where premia have been distributed without information available in the self-assessment report.

Net of this small group, the number of teachers involved in the merit assessment is reported in Table 3: over a total of 668.979 tenured teachers, only a small fraction (corresponding to 2.4% of

[3] The forms were available to schools in the website.
[4] In Table A.1. in the Appendix we report the number of schools that are available in different datasets.
[5] This should not come as a surprise, since merger of schools are rather frequent, especially in regions experiencing declining fertility rates.

the total) are working in schools that were (self) excluded from the merit assessment and/or premia distribution. Temporary teachers were excluded from the premia distribution by law.

Insert Table 3 here

Within the school population under analysis we observe three behaviours, which are worth interpreting: schools that have been unable to start the process, because there is no evidence of appointment of the assessment committee (corresponding to 10% of the schools – see Table A.2. in the Appendix – but only to 5.5% of teachers); schools that have appointed the committee, but for various reasons were unable to conclude the assessment, and therefore to reach approved minutes for bonus distribution (corresponding to 7.6% of schools and 8.1% of teachers); finally, schools that distributed the bonus, though with different outcomes (82.4% of schools and 86.4% of teachers). This is a first indication of an open support to the initiative: the limited number of schools unable to reach an agreement and/or the limited number of teachers excluded from the potential distribution stays in sharp contrast to what has happened two decades earlier, when a Minister of Education was forced to resign on a similar proposal (though supported by a larger amount of bonuses to be distributed).

5. Patterns of distribution of resources

The patterns of distribution are nevertheless rather different among schools. By taking the ratio between teachers obtaining a bonus and total number of tenured teachers in the school, the mean is 41% over 7281 schools where we have information about successful distribution, while the median is 36% (meaning that in half of the schools one out of three teachers obtained a premium). As a consequence, the distribution is rightward skewed, as noticeable from graph 1 where a normal distribution is overimposed. Very restrictive distributions (one could be tempted

to name them “meritocratic” if we could know the criteria adopted in the distribution) are present in a minority of schools (at 5% of the distribution the fraction of awarded teachers is 18%); one fourth of schools have distributed to one teacher out of four (28% corresponds to the boundary of the first quartile – see also table A.3. At the other tail of the distribution, one fourth of schools have awarded the premium to half of the teachers (or more), and there is also a minority of schools (less than 3%) having awarded the premia to nine teacher out of ten.

This can be taken as an indicator of success of this innovation in performance-related pay for teachers: the vast majority of the schools endogenously adopted a policy that allocated a bonus to only one third of them. Though the list of the beneficiaries was generally not made public, it seems impressive that teachers did not reject the implicit message of the policy that “one teacher out of three provides better quality teaching than the rest”. Looking backward to previous attempts to promote merit assessment, this can be counted as the first experience of policy induced wage differentiation in the Italian educational system. One may object that the lack of opposition derives from the small size of the bonus (if the per-teacher allocation was 200 euro per year and only one third of them received the bonus, the average bonus amounts to nearly 600 euro per year. In addition this measure was perceived as temporary, and the anticipation was confirmed the following year (2016-17) when the unions obtained the transfer of the funds earmarked for merit evaluation to the general fund for wage bargaining within the schools.

Insert Graph 1 here

We investigated the existence of potential correlation between observable features at school level and fraction of teachers receiving the bonus. Apart from school size (the larger is the school, the smaller is the fraction of awarded teachers), we were unable to detect any statistically significant

correlation with gender composition, teachers' average age or level of qualification. In order to test the potential existence of union opposition, we asked for the membership rate at school level, which however was refused by the Ministry for privacy protection; as an alternative, we proxied this effect with the school turnout rate for election of union representatives at school level: the correlation is positive but statistically insignificant.

Some evidence of the effect of the school climate can be inferred by the correlation between attitude towards student testing and generosity of the bonus allocation. The Italian testing agency (Invalsi) estimate for each class taking a test either in literacy or numeracy a probability of cheating, based on covariances between students' answers to different items (Bertoni et al, 2013). When a school displays a high level of cheating (the horizontal axis in figure 2, it may be indicative of a relaxed attitude towards outcomes. And if results are irrelevant (or even opposed as criterion for accountability), then is does not come as a surprise if a positive correlation emerges between a generous (i.e. not selective) allocation of bonuses and cheating to the test.

Insert Graph 2 here

It is not therefore surprising that on average we do not find evidence of correlation between school outcomes (as measured by student test achievements) and premia distribution.[6] In Table 4 below, one can detect a slightly more restrictive attitude in best performing schools (and vice versa, more generous distribution in less performing schools). But correlations are weak from a statistical point of view, and they even disappear once we replace student achievements with

[6] Since test are taken at different grades and are not strictly comparable, we have standardised test score at grade level, and then averaged them across grades, in order to get a single score at school level.

parental background measures (typically the ESCS index measuring economic, social and cultural status of the family of origin – not reported in the table).

Insert Table 4 here

6. The process of bonus distribution

From the third dataset we get some information on the process of merit assessment within participating schools. When we tabulate the composition of the assessment committee by different stages attained in the premia distribution, we notice that the presence of external members (despite one member being required by law to be external to the school) make the achievement of an agreed distribution more difficult and/or more restrictive (see table 5). Overall in such cases we are dealing with a small number of schools, such that this does not alter the general process of distribution.

When considering the procedures followed by the committees in order to achieve a distribution (see Table 5) one can observe that explicit framing of the criteria seems to help achieving the final outcome: whenever the committee has made explicit reference to the three criteria suggested by the law (paragraph 129 of the law 107/2015 known as “Buona scuola”) the probability of achieving an effective distribution raises by 6 percentage points, though not necessarily becoming more generous in terms of fraction of teachers awarded the premia. The qualitative analysis of the criteria designed by individual schools show that the committees followed closely the central government’s blueprint.

Insert Table 5 here

The framing of the procedure was guided by the law, but local committees at the school level were free to adopt additional criteria and/or assign different weights to the proposed criteria. There was ample flexibility in the implementation of criteria of evaluation by the school committee. As

expected, the vast majority of the school committees followed the law requirements: 7345 schools out of 7388 (99%) made explicit reference to the three criteria, though half of them (57%) used different weights for different items.

In order to understand how the national framework setting out the three performance criteria was implemented by schools' committees, we conducted a qualitative text analysis of the minutes in a sample of 70 schools provided by the Italian Ministry of Education. The Ministry collected data from self-evaluation of schools as part of a monitoring exercise of the implementation of the 2015 reform. We analyzed the committee formal minutes, which detailed extensively in the vast majority of cases the merit pay criteria applied to the school. It is worth noting that schools' participation to the monitoring exercise was voluntary. Although the committees had to follow the national legal framework, the process of elaboration of merit pay criteria was very resourceful and creative. In the vast majority of cases (97.7%), the schools used all the three criteria provided by the law. In the vast majority of the cases, schools embarked upon a process of designing performance pay criteria for their school teachers. In two third of the cases, the school committee assigned different weights to different criteria. In order to test the significance of the input-output types of accountabilities presented earlier in the theoretical framework of the paper, we created a coding frame to analyze all the committees' minutes. We focused exclusively on teaching activities and not organizational and management activities (such as mentoring a junior colleague, or attendance to collegial meetings). We developed the coding frame deductively from the existing literature on merit pay, and we used Podgursky's (2007) classification. The results of the qualitative analysis we conducted led to a classification of schools in the sample along three categories:

a) predominantly input-based criteria for bonus distribution;

b) predominantly output-based criteria; and

c) balanced mixing of input and output based criteria.

Output-based criteria refer mainly to the use of standardized tests administered by INVALSI on students' achievements as measures of teachers' performance. Input-based criteria refer to the acquisition of teachers' new skills in ICT, teachers' training and time allocated to acquiring new qualifications. Input also referred to developing new pedagogic instruments, or new pedagogic methods, which were not explicitly linked to students' achievements and results. Among the 70 schools we have analysed, more than half (37 schools corresponding to 52.8% of the total) distributed the bonus predominantly on input criteria and one fourth (17 schools corresponding to 24.3%) adopted a mixed approach. Only a small fraction followed an output approach (10 schools equivalent to 14.3%).[7] The limited role assigned to output measures in this new pay policy introduced in Italy is indicative, at best, of the predominant orientation towards the traditional content of professional accountability: the best teachers are identified as those who improve their qualification, knowledge, teaching methods, independently of their impact and level of effectiveness in teaching, as measured by students achievements. The multiple types of accountabilities discussed earlier in the paper are institutionally entrenched in complex institutional networks that make the implementation of reforms and radical change very difficult to achieve (Hill and Hupe, 2014).

We then analyse the correlation between the input-output classification which we created and the distribution of the premium using two measures, the share of teachers receiving the bonus and

[7] There are 6 schools that we were unable to classify according to our scheme because the lacked details in the process of bonus attribution.

the Gini index in the bonus distribution within the schools.[8] As we can see from Table 6, the share of teachers obtaining the bonus is consistently lower for schools that adopted the output-based categories; symmetrically the Gini index is constantly higher in the same schools. This shows that the distribution of the premium was more "meritocratic" and less egalitarian for schools that adopted an output-based accountability model whereby performance was assessed on the basis of students' achievements. This is also consistent with the aggregate evidence of positive correlation between generosity of the distribution and probability of cheating in test taking.

Insert Table 6 here

7. Conclusions

In this paper, we have examined in depth the implementation of an innovative teachers' incentive programme in all Italian public schools in the school year 2015-2016. This reform is path breaking in two fundamental ways: first, it departs from the historical path dependency of failed reforms concerning teachers' performance; second, it aligns the Italian educational accountability system with international education reforms. From 1958, when Aldo Moro introduced the first attempts to assess teachers, to date, education experiments intended to evaluate teachers' performance have never been fully implemented in Italy. In some cases, ministers had to withdraw their government plans and suffered a very high political cost for attempting to introduce performance pay schemes for school teachers. In 2000, the Minister of Education, Berlinguer resigned from office in the midst of strikes, public demonstrations, and teachers' fierce opposition to his plans. Our expectation was that the 2015 Buona Scuola reform would have had the same unfortunate destiny.

[8] For all the schools that have filled the monitoring forms we know the fraction of teachers receiving the bonus, but we do not have information about the internal distribution. For a subset of schools (those who were quick enough to conclude the procedure and pay the bonus by August 2016) we also have information about the internal distribution, that we describe using a known inequality index (the Gini concentration index).

Against the backdrop of repeated aborted attempts to introduce merit pay for teachers in Italy, the data presented in this paper shows that in 2015-2016, the *Buona Scuola* reform, introduced by the Renzi government in Italy, has been implemented by the vast majority of Italian schools, allocating a (rather low level of) salary bonus to one third of teachers per school. This suggests that the professional and organizational values of Italian teachers and schools have changed, to the point that it becomes openly acceptable that one teacher out of three is better qualified than other two colleagues.

However, the Italian reform remains an unfinished business and a rather timid one, in so far as monetary incentives are less focused on the improvement of students' achievements (outputs). Our qualitative analysis of the minutes of a sample of less than a hundred schools selection committees suggests that input-based principles were prevalent in the process of defining the performance criteria, pushing for a more generous distribution of the bonus. Input-based measures are traditionally linked to professional accountability that favours peer reviews, qualifications, teaching practices, and collegiality. As Lavy showed in his study of teachers' performance incentives in Israel (2002), teachers' incentives programmes focused on students' achievements are more cost effective than resource-based programmes (such as additional teaching time and teacher's training).

This paper does not suggest that the 2015 experiment under study is the optimal incentive structure in schools or the only performance system to improve students' achievements. On the contrary, our analysis of the reform implementation suggests that individual teachers' efforts are difficult to measure, and individual performance is technically problematic to assess with a high level of accuracy. Generally, monetary incentives to teachers as a function of the improvement of students' achievements are rare and remain at the level of local experiments that are not suitable for scaling up. For this reason, we believe that group incentives of schoolteachers in Italy may

represent in the future a better promising avenue to explore, while the new Italian government discusses programme design and how best to correlate teachers' performance with students' achievements. There are two clear advantages of group interventions: first, competences formation is a cumulative effect and, second, group evaluation is more in line with organizational values. We need further research on group performance interventions for teachers in Italy, though there is evidence of their success in other countries (Lavy, 2002).

Despite the future political uncertainty, the importance of our study here lies in the fact that the power of teachers' performance incentives is evident in Italian schools, even in the case of a low performance bonus. Teachers are not only motivated by financial rewards, but also by public recognition of their efforts. For instance, the symbolic power of publishing the results of high achieving groups of teachers on the school public board, or in the local press, should not be estimated by policymakers. The reputational impact of publicly rewarding good teachers has not been tested in the Italian context yet, but it may be considered in the future process of designing and ameliorating the existing programmes. Moreover, high-achievers would be encouraged to take on management roles, such as school principals.

References

Atkinson, A., S. Burgess, B.Croxon, P. Gregg, C. Propper, M. Slater and D. Wilson (2004). *Evaluating the impact of performance-related pay for teachers in England*. University of Bristol: Centre for Market and Public Organization.

Ballou, D. (2001). Pay for performance in public and private schools. *Economics of Education Review*, 20(1): 51-61.

Behn, R.D. (2001). *Rethinking democratic accountability*. Washington, DC: Brookings University Press.

Bertoni, Marco, Brunello, Giorgio and Rocco, Lorenzo (2013). When the cat is near, the mice won't play: The effect of external examiners in Italian schools, *Journal of Public Economics*, 104(C), 65-77.

Braga. M., D. Checchi, C. Garrouste, and F. Scervini (2019). Selectivity or Rewarding Teachers? International evidence from primary schools, Bonn: IZA DP 12020.

Day, P. and R.Klein (1987). *Accountabilities*. London: Tavistock Publications

Dolton, P., S. McIntosh, and A. Chevalier (2003). *Teacher pay and performance*. Bedford Way Papers. London: Institute of Education.

Dubnick, M (2005). Accountability and the promise of performance: in search of the mechanisms. *Public performance and management review*, 28(3): 376-417.

Duflo, E., R. Hanna and S. Ryan (2012). Incentives work: getting teachers to come to schools. *American Economic Review*, 1241-78.

Glennerster, H. (2002). United Kingdom Education 1997-2001. *Oxford Review of Economic Policy*, 18 (2), 120-136.

Graham, P. (2005) Classroom-based assessment: changing knowledge and practice though preserving teachers education, *Teaching and Teacher Education*, 21(6), 607-621

Hamilton, L. and D. Koretz (2002). Tests and their use in test-based accountability systems, in L. Hamilton, B. Stecher and S. Klein (eds.) *Making Sense of Test-based Accountability in Education*, Santa Monica, California: RAND.

Hanushek, E. and M. Raymond (2001), The confusing world of education accountability, *National Tax Journal*, 54(2): 365-384.

Hanushek, E. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30 (4), 466-479.

Hanushek, E. and S.Rivkin. (2014). Teacher quality. in Eric A. Hanushek and Finis Welch (eds) *Handbook of the Economics of Education*, Volume 2, Amsterdam Elsevier: 1051-1078

Hill, H. and P. Hupe (2014). *Implementing Public Policy*, Sage, 3rd edition.

Hood, C. Perry, and H. Margetts (2010). *Paradoxes of Modernization: Unintended consequences of public policy reform*. OUP: Oxford.

Jennings, J. (2012). The effect of accountability design on teachers' use of test score data, *Teachers College Record*, vol. 114.

Koretz, D. (2002). Limitations in the use of achievement tests as measures of educators' productivity. *The Journal of Human Resources*, 37(4), 752-777.

Lazear, E. (2000), Performance Pay and Productivity, *The American Economic Review* 90(5): 1346-1361

Lazear, E. (2003). Teacher incentives. *Swedish Economic Policy Review*, 10: 197-213.

Lavy, V. (2002). Evaluating the effects of teachers' group performance incentives on students achievements. *Journal of Political Economy*, 10(6), Dec: 1286-1318.

Lavy, V. (2007). Using performance-based pay to improve the quality of teachers. *Future of Children*, 17(1): 87-109

Lavy, V. (2015).Teachers' Pay for Performance in the Long-Run: Effects on Students' Educational and Labor Market Outcomes in Adulthood. NBER Working Paper no. 20983. Cambridge, MA.

Marsden, D. (2010), The paradox of performance-related pay systems (chapter 10), in C. Hood, Perry 6, and H. Margetts, *Paradoxes of Modernization: Unintended consequences of public policy reform*. Oxford: Oxford University Press.

Marsden, D. (2015). Teachers and performance pay in 2014: first results of a survey. CEP Discussion Paper No. 1332.

Mulgan, R. (2000). Accountability. An-ever expanding concept?, *Public Administration*, 78: 555-73

Murphy, R. (2011). Improving the impact of teachers on pupil achievement in the UK - interim findings. Sutton Trust, London.

Murname, R.J. and D.K. Cohen. (1986) Merit pay and the evaluation program: why most merit pay plans fail and a few survive. *Harvard Education Review*, 56(1): 1-18

OECD (2005). *Performance related pay policies for government employees*. Paris: OECD

OECD (2013). *Synergies for Better Learning. An International Perspective on Evaluation and Assessment*. Paris: OECD

Podgursky, M and M. Springer (2007). Teacher performance pay: a Review, *Journal of Policy Analysis and Management*, 26(4), 909-949

Previtali, Damiano (2018). *Il sistema nazionale di valutazione. Una rilettura*. Utet Università.

Richardson, R. (1999). Performance related pay in schools: an assessment of the green papers. National Union of Teachers, London.

Ritter, G, J. Trivitt, L. Foreman, and C. DeAngelis (2019). Making Sense of the Existing Evidence: A Meta-Analysis of Teacher Merit Pay, University of Arkansas

Romzek, B.S. (2000). Dynamics of Public Sector Accountability in an Era of Reform, *International Review of Administrative Sciences,* 66: 21-44.

Sinclair, A. (1995). The chameleon of accountability. *Accounting, Organisations and Society*, 20, 219-37.

Trellle Fondazione per la Scuola. (2013). *Valorizza - A Holistic and Contextual Approach for Teacher Appraisal.* http://www.treellle.org/files/lll/Valorizza.pdf

Varlaam, A., D. Nuttall, and A. Walker (1992). *What makes teachers tick? A study of teacher morale and motivation.* Clare Market Papers, No. 4. London: Centre of Education Research (London School of Economics)

Tables in Manuscript

Table 1. Input and Output-Based Types of Accountabilities in Schools

	Performance (Output) Accountability	**Political Accountability (Input)**	**Professional Accountability**
Accountability holders			
External	Ministry of Education and/or local government	Voters	None
Internal	Governing bodies and head teachers	None	Peers (teachers)
Accountable actors	School principals and teachers	Elected representatives	Teachers
Content	Performance and Output	Representation (Input)	Teaching practices, qualifications, knowledge
Mechanism	Output Measurement	Election	Collegiality

Source: own elaboration

Table 2. Theoretical Controversies related to Performance-related pay for schoolteachers

Arguments For Performance-related pay	**Arguments Against** Performance-related pay
• Improves the quality of teaching (Murphy, 2011; OECD, 2005; Atkinson et al, 2004)) • Improves the recruitment of high-productivity employees (Lazear, 2000) • Changes teachers' behavior with positive effects (Lavy, 2002; Atkinson et al, 2004; Podgursky, 2007) • Improves goal setting within schools (Marsden, 2015) • Improves quality of education system (Braga et al, 2019) • Enhance public accountability (author, 2012b) • Knowledge and skills based reward plans are ineffective (Hanushek and Rivkin, 2004) • Gains in labour market by students (Lavy, 2015)	• Not suited to the "nature of teaching" (Murname and Cohen, 1986) • Pay levels are not a source of motivation (Varlaam et al, 1992) • Increases cheating, and teaching to the test (Dolton et al, 2003) • Teachers don't support them (Murname and Cohen, 1986) • Evidence that compensation motivation mechanisms don't work (Marsden, 2010) • Triggers trade unions' opposition (Ballou, 2001)

Source: own elaboration

Table 3 – Number of tenured teachers (*insegnanti di ruolo*) in state schools by involvement or not in bonus distribution – school year 2015-16

	in schools not involved in the bonus distribution	in schools involved in the bonus distribution	number of schools involved in bonus distribution	tenured teachers per school involved
Piedmont	1585	44348	566	78.35
Liguria	377	14834	192	77.26
Lombardy	2182	93431	1131	82.61
Veneto	1024	48418	597	81.10
Friuli V.G.	590	12349	168	73.51
Emilia Romagna	559	43347	540	80.27
Tuscany	812	39450	483	81.68
Marche	231	18415	242	76.10
Umbria	395	10191	139	73.32
Lazio	2399	60031	719	83.49
Abruzzo	598	15805	202	78.24
Molise	425	3761	55	68.38
Campania	1967	79793	1000	79.79
Basilicata	58	8226	123	66.88
Apulia	235	49656	675	73.56
Calabria	801	26595	383	69.44
Sicily	1501	64648	854	75.70
Sardinia	598	19344	275	70.34
Total	**16337**	**652642**	**8344**	**78.22**

Table 4 – Schools by student achievement and fraction of teachers obtaining the premium

	0-28% restrictive	28-36% rather restrictive	36-50% rather generous	50-100% generous	Total	number of schools
test score (Invalsi)	literacy					
low	24.23	22.63	25.83	27.30	100	1564
middle-low	25.67	22.85	26.43	25.05	100	1593
middle-high	24.25	23.51	26.24	26.00	100	1608
high	26.99	23.08	25.22	24.72	100	1586
missing	*22.26*	*20.43*	*27.20*	*30.11*	*100*	*930*
Total	24.90	22.69	26.10	26.32	100	7281
	numeracy					
low	25.39	22.34	25.13	27.14	100	1540
middle-low	23.39	23.83	27.52	25.27	100	1599
middle-high	25.11	24.12	24.31	26.46	100	1625
high	27.28	21.74	26.78	24.20	100	1587
missing	*22.26*	*20.43*	*27.20*	*30.11*	*100*	*930*
Total	**24.90**	**22.69**	**26.10**	**26.32**	**100**	**7281**

Table 5 – Bonus distribution and criteria adoption

	premia have been distributed ?			% teachers obtaining premium
Has the committee made explicit the criteria for premia distribution ?	% no	% yes	number of schools	
no	10.37	89.63	548	42.34
yes	8.44	91.56	7417	41.64
Total	8.57	91.43	7965	41.69
Has the committee followed the indications of the law 107/2015 criteria (paragraph 129 point 3 letters a, b, c) ?	% no	% yes	number of schools	
no	14.39	85.61	140	43.56
yes	8.32	91.68	7277	41.61
Total	8.44	91.56	7417	41.64

Table 6 – Bonus distribution by Input-Output classification and salary inequality

(first figure: share of teachers receiving a bonus (58 schools)

– second figure: Gini index in salary distribution (27 schools))

	input	mixed	output	Total
generous	0.69	0.67	0.54	0.67
	0.53	*0.67*	*0.77*	*0.61*
rather generous	0.43	0.45	0.40	0.43
	0.70	*0.69*	*0.69*	*0.70*
rather restrictive	0.37	0.30	0.32	0.35
	0.78			*0.78*
restrictive	0.24	0.36	0.21	0.27
	0.81	*0.80*	*0.87*	*0.82*
Total	0.49	0.54	0.37	0.48
	0.67	*0.69*	*0.76*	*0.69*

Graphs in Manuscript

Graph 1. Distribution of the bonus selectivity index among schools awarding the bonus

(see attached file)

Graph 2. Bonus selectivity and cheating

(see attached file)