

ECONOMIC THEORY AND DIFFICULTIES WITH TEACHERS' PAY- FOR-PERFORMANCE SCHEMES

MARAYSA RIBEIRO ALEXANDRE

RICARDO SEQUEIRA PEDROSO DE LIMA

FÁBIO DOMINGUES WALTENBERG

TRANSLATED BY Marina Waltenberg Mauritz

ABSTRACT

Relating teacher payment to pupils' standardized test scores is spreading in Brazil. Such policies do not find their theoretical roots in the field of education science, but rather in the economic-management literature, especially in the so-called "principal-agent model". While they are regarded by some as a cornerstone for improving education quality, they are rejected by non-economists. The empirical evidence is ambiguous: both positive and non-positive effects have been documented. The contribution of this paper is to revisit the theoretical framework on which pay-for-performance schemes lay, aiming at testing the hypothesis that inconclusive effects could have been predicted by the economic literature itself. Complementarily, we investigate whether the theory sheds light on the reasons why such policies are strongly rejected in certain circles. We provide positive answers to both questions.

For their generous comments, critiques and suggestions, the authors are grateful and exempt from any responsibility for mistakes and omissions in this text: Ariana Martins de Britto, Érica Pereira Carneiro, Aloysio de Carvalho, Marco Tulio França, Danielle Carusi Machado, Jesus Alexei Obregón and the participants of seminars in *Centro de Estudos sobre Desigualdade e Desenvolvimento* [Center for Studies on Inequality and Development], at Universidade Federal Fluminense, Niterói, Brazil). For the financial support along 12 months, Ricardo Lima thanks the Pibic/CNPq/UFF Program. Maraysa Alexandre thanks Capes for the 24 months of masters' scholarship.

TEACHER ACCOUNTABILITY PROGRAM • TEACHER PAY •
PRINCIPAL-AGENT MODEL • ECONOMICS OF EDUCATION

IN THE 1990S, standardized learning assessment tests were created and spread throughout Brazil. Over the following decade, this process intensifies, gains popularity and leads to a virtually inevitable corollary: the introduction of policies linking teacher payment to pupil's performance in standardized tests. As reported in Andrade (2008), Ferraz (2009) and Bruns *et al.* (2011), following an international trend, a number of such policies are being implemented in many states and municipalities.

Such policies, here referred to as “accountability” or “incentive” policies, do not find theoretical root in the field of educational science. They are usually suggested by economists, administrators, academics or managers in departments of education, deliberately or inadvertently inspired by an economic-management literature called “information economics”, or “economics of contracts”, developed over the last 50 years.

More specifically, these policies are inspired by the “principal-agent model”, which studies situations in which an “agent” (e.g., a teacher) is hired by a “principal” (e.g., a commissioner of education) to perform a set of tasks (e.g., prepare for classes, motivate pupils, organize classroom time, etc.), aiming at producing a good or service of interest to the principal (e.g., that pupils learn). In general, the agent's effort in performing the set of tasks assigned to him cannot be perfectly observed by the principal, but some result of the process can (e.g.,

pupils' grades). Under these circumstances, it is suggested that teachers are paid according to such outcomes, assuming this would motivate them to act as desired by the principal (e.g., by increasing efforts). It is believed that, even if the principal does not know exactly how the agent should act in order to achieve the aspired results, under an incentives regime, the agent himself would search and find solutions to problems – whether pedagogical, disciplinary, administrative or of other nature – faced daily, leading pupils to learning more (TRANNOY, 1999).

On the one hand, these policies are usually not well received by education professionals, especially by teacher unions, as recently regarded in the teachers' strikes in Rio de Janeiro. By expressing their loathing of “meritocracy”, they were truly stating their disavowal of accountability policies. Resistance to this practice can also be seen in other countries, as reported, for instance, by Diane Ravitch (2010), researcher converted to criticizing and denouncing such policies in the USA, after years endorsing them.

On the other hand, such policies continue to spread in Brazil, under the administration of varied parties,¹ supported by prestigious economists, managers or other members of the national *intelligentsia*.² The same applies to other countries. In the United States of America, for instance, the Obama administration has sustained teacher accountability programs inherited from the Bush administration as key pieces of its education policy – and for doing so, received endorsement in editorials of important newspapers, even some with non-conservative bias, such as *The New York Times*.³

Ideally it would be possible to diminish the conflict between opinions *a priori* in favor or against accountability policies by means of the analysis of concrete assessments of results. For example, should an important part of empirical studies conclude that these policies are ineffective in improving pupils' learning, even their most ardent proponent could surrender to the evidence – and vice versa, in the opposite situation. However, results have been inconclusive, adding fuel to the fire: positive effects have been recorded (e.g., LAVY, 2002), as well as null or even surprisingly negative ones (e.g., FRYER, 2013). At times, even one single program presents positive results for one grade and null or negative results for another (e.g., OSHIRO; SCORZAFAVE, 2011; ALEXANDRE, 2013). In this context, both supporters and opponents tend to prize and emphasize results aligned with their own opinions, respectively.

Given this background, the main contribution of this article is to revisit the theoretical framework on which accountability policies lay, aiming at testing the hypothesis that inconclusive effects could have been predicted by the *economic literature itself*. In other words, that a careful analysis of the theoretical fundamentals of the principal-agent

1 Maybe Brazilian governing authorities only indulge their voters' will. A recent survey suggests that the majority of Brazilians look favorably upon teachers' pay schemes based on pupils' outcomes (DOLTON; MARCENARO-GUTIERREZ, 2013).

2 As illustrated by the letter of endorsement to the Municipal Department of Education of Rio de Janeiro, signed by various leaderships in the academic circle and the civil society in response to *inter alia* critics to the merit-based pay scheme introduced in the municipality. The motion, entitled “Claudia Costin and the education in the city of Rio de Janeiro” (“Claudia Costin e a educação na cidade do Rio de Janeiro”) can be read in Simon Schwartzman's blog, posted on November 26th, 2013: <<http://www.schwartzman.org.br/sitesimon/?p=4385&lang=pt-br>>.

3 For example, the editorials from February 4th, 2009, “A vital boost for Education”, and from March 18th, 2010, “Mr. Obama and No Child Left Behind”.

model, with all specificities of its application to the education field regarded, would make it possible – or rather would have made it possible – to foresee difficulties in the implementation of such programs. This analysis of the theoretical framework also has a complementary purpose: understanding whether the theory sheds light on the reasons behind many teachers', unions' and intellectuals' strong repulse of accountability programs. The hypothesis presented here is that this rejection is not owed to a merely dogmatic or ideological opposition (though these ingredients may be present), but also to reasons that reverberate economic principles and findings, even with the debate in the aforementioned circles clearly not being carried out and expressed in the economic jargon.

In order to illustrate the imbroglia in the evidence, it seems unnecessary to present a long survey of the literature. Instead, the next section brings a short synthesis of emblematic case studies of accountability programs that succeeded, failed or presented ambivalent results. It is followed by the truly theoretical investigation, reviewing assumptions and results of the basic principal-agent model, as well as implications to the analysis of teachers' pay systems. After that, some specificities of the application of the general model to the education scenario are focused. The last part is devoted to this article's conclusions, the main of which is the impossibility to refute either of the hypotheses presented.

DESIGN AND RESULTS OF EMBLEMATIC ACCOUNTABILITY PROGRAMS

There are so many teacher accountability programs, either in progress or discontinued, in Brazil and abroad, that thoroughly summarizing them would be a difficult task.⁴ Furthermore, such task would also be untimely, since this is not the goal of the present article, devoted to a mainly theoretical analysis. Therefore, the choice was to present in this section only four emblematic case studies of teacher accountability programs, both “successful” and “failed” ones, in order to establish the foundation for the sections that follow.

One of the most prominent and one of the first programs for which data was recorded was the successful accountability program carried out in Israel in the 1990s, reported in Lavy (2002). Out of a group of 62 schools non-randomly selected,⁵ it was decided that those schools with pupils among the top third performers in a multidimensional ranking⁶ would receive a bonus. The socio-economic background of the student body was controlled. The goal was to improve learning achievement and reduce dropout rate. Out of the total bonus, three fourths were destined to teachers' pay (collective incentive) and the rest

⁴ Detailed descriptions of designs and main results of these programs can be found in Alexandre (2013) and Lima (2012).

⁵ In order to be referred to the program, a school had to be the only one of its type in the region and go from 7th to 12th grade. From the 320 schools at the time, 170 met the requirements. Among these, 62 were chosen to the program (37 mixed Hebrew schools, 18 Hebrew schools and 7 Arab schools).

⁶ Average of credits per pupil, proportion of pupils to receive matriculation certificate and dropout rate.

should be used to improve faculty facilities. The bonuses per teacher ranged from 1 to 3 percent of the average teacher annual salary.

Making use of conventional econometric tools for the program evaluation, Lavy (2002) identifies positive and statistically significant effects for both assessed years in religious schools and only for the second year in other schools, except for the proportion of pupils to receive a matriculation certificate. Effects were more pronounced among lower performers. Such auspicious results would motivate a number of new experiments and assessments.

A recent successful example is found in a context very different from the one in Israel: it took place in the Indian state of Andhra Pradesh in the 2000s and was reported by Muralidharan and Sundararaman (2011). One of the reasons for the great appreciation for this article is the high scrutiny applied in each step during the implementation of the program, aiming at producing a perfect assessment from an econometric point of view. A sample of 500 schools was randomly selected. Bonus was to be linearly related to improvement on pupils' scores in multiple learning assessment tests administered on different dates (to minimize measurement errors) and under a sophisticated anti-fraud system. In order for a school to be granted the bonus, pupils' test scores should present a minimum improvement of 5%. The mean bonus was of approximately 3% of the average annual salary.

In both first and second years of the test relevant improvement has been recorded in the basic subjects (relevant to the bonus), as well as in other subjects. It was observed that teachers in the awarded schools started conducting extra classes, giving extra tests and assigning more homework. According to the authors, the possibility of pupils being taught to the tests has been verified and rejected.

Another renowned program was introduced in New York public schools in 2007 and 2008 and reported by Fryer (2013). Unlike the former two, this program has not presented successful results, despite its very careful implementation, as in the India case study. Following specific requirements for participation,⁷ 198 schools were selected for the program, which was carried out by the teachers' union and the New York City Department of Education – DOE. Each school had its own performance target, established by a formula that related, with different weights, criteria, such as: grades, grade variation (“added knowledge”), elementary graduation rates, attendance, among others. Should the target be fully achieved, the school would be entitled to a bonus (of up to 4% of the annual teacher's salary), which could be internally distributed at the school's own discretion, following certain rules, such as not to distribute rewards unevenly based on seniority. The program received around \$75 million and rewarded approximately 20,000 teachers.

7 In terms of average grades, percentage of foreign and special pupils and poverty rate, besides agreement of the majority of teachers to take part in the program.

Fryer (2013) found no evidence that teacher incentives would have improved pupils' achievements – on the contrary, he came across negative results, one of which statistically significant – nor substantially affected pupils' or teachers' behavior. Disconcerted, he enumerates reasons that could account for such unexpected results, such as: not large enough incentives; complexity of the incentive formula; teachers' lack of knowledge of how to improve pupil performance; incentives unevenly granted based on seniority, despite previous agreement and concealed behind different job titles; lack of effectiveness of group-based incentives.

Lastly, it is worth mentioning a national example of accountability program, carried out in the state of São Paulo, that has already undergone at least two assessments (OSHIRO; SCORZAFAVE, 2011; ALEXANDRE, 2013), having presented similar results: positive effects for the 5th grade and either null or negative effects for the 9th grade. The variable compensation system developed in São Paulo was linked to a school-specific target based on pupils' grades in Portuguese and Mathematics and the school's average pass rates.⁸ Granted annually, the bonus is proportional to the percentage of the target accomplished and may be as high as 20% of the annual salary. In 2009 total bonuses added up to R\$ 650 million (around US\$175 million), granted to approximately 210,000 employees of the educational system of the state of São Paulo. Even though the index takes into account only Mathematics and Portuguese, teachers of all subjects, as well as school principals and other employees are entitled to the bonus. There are restrictions to the payment of bonuses to teachers who are often absent.

Oshiro and Scorzafave (2011) compare the progress in pupils' grades in schools run by the state of São Paulo (affected by the program) and in other schools, that were not affected by it (São Paulo municipal schools, schools ran by other Brazilian states, etc.), always pairing up schools with similar characteristics. They come to the conclusion that the program has had positive (significant) impact over 5th graders' proficiency in both subjects, but negative impact (sometimes statistically significant) for 9th graders. In order to investigate whether these results were due to a too short period of time between the beginning of the program and the assessment, Alexandre (2013) repeated the analysis, taking 2011 as object, achieving, however, similar results: Positive effects for the 5th grade, negative ones for the 9th grade (all of them significant).

Based on the analysis of these paradigmatic case studies, one may come to the conclusion that, as announced in the introduction, evidence on the effectiveness of accountability programs is still far from being conclusive and more studies are required. Lastly, it is worth to emphasize that nothing is known about the actual long term motivation

8

Translator's Note: Test scores were collected through Sistema de Avaliação de Rendimento Escolar de São Paulo [System for Evaluating Academic Performance in the State of São Paulo] - Saresp -, which, together with pass rates, composes the Índice de Desenvolvimento da Educação de São Paulo [Index of Educational Development of the State of São Paulo] - Idesp.

of teachers or about the long term learning levels of pupils, since all available assessments are restricted to the short term.

REVISITING THE BASIC PRINCIPAL-AGENT MODEL

The principal-agent model is used to analyze conditions for the establishment and maintenance of contracts between economic agents in a situation of *asymmetric information*. The principal depends on tasks to be performed by an agent, who disposes of information on her own behavior, type or environment, which are not available to the principal nor to a third party (e.g., the courts). The first versions of the model were developed to describe labor relations in the private sector and between a single agent and a single principal. Variations were later proposed, for instance, adapting the model to the public sector, to the peculiarities of the education sphere, to situations involving multiple agents or principals, among many others.

The most acclaimed microeconomic theory handbook (MAS-COLELL; WHINSTON; GREEN, 1995) was used as source for the analysis of the basic model. However, the mathematical language, profusely used in the section devoted to the formal model in the aforementioned handbook (section 14.B), has been here relegated to footnotes, followed by explanations. Along the presentation of the basic model in this section of this article, emphasis is given to aspects relevant to better outlining and understanding labor relations in the education sphere.

For our purposes, a commissioner of education may be taken as “principal” and a “teacher” as agent. The principal’s challenge is to design a monetary compensation system that motivates the agent to behave accordingly (e.g., by preparing for classes, motivating pupils, organizing classroom time, etc.), so as to achieve the goals longed by the principal (e.g., that pupils learn). This result, longed by the principal and expressed by π , should be noticeable to both, as well as to third parties. On the other hand, the agent’s actions, expressed by e , are known by the agent but, according to the full version of the model, cannot be observed by others at a low cost. In fact, it is impossible for a commissioner of education to thoroughly monitor the work of each teacher under his administration, making this model at first sight appropriate to describe labor relations in the education branch.

The letter e was chosen to denote actions performed by the agent as an analogy to the idea of *effort*. Even though such literal association between e and *effort* was not necessary, nor was it necessary to limit e to a one-dimensional metric, it is a fact that these two simplifications are often made, to some extent to bypass mathematical difficulties in the resolution of the model. The imposition of such simplifications may also be due to the nature of the labor relations that originate the

9

This aspect of the problem can be formally expressed by: $F(\pi|e_a) \leq F(\pi|e_b), \forall \pi \in [\underline{\pi}, \bar{\pi}]$, or, in equivalent manner, by $\int \pi f(\pi|e_a) > \int \pi f(\pi|e_b), \forall \pi \in [\underline{\pi}, \bar{\pi}]$, where $F(\cdot)$ and $f(\cdot)$ represent the cumulative distribution function and the probability distribution function, respectively, and $\bar{\pi}$ and $\underline{\pi}$ represent the lowest and the highest limits of outcomes' distribution. This footnote, as well as the ones that follow are based on the presentation form of the principal-agent model, including its notation, as adopted by Mas-Colell, Whinston and Green (1995).

10

Mas-Colell, Whinston and Green (1995) assign the agent a simple Bernoulli utility function: $u(w, e) = v(w) - g(e)$, where $v(w)$ and $g(e)$, with $i = a, b$, respectively represent the utility of the wage w and the disutility of the effort e , with the following characteristics: $v'(w) > 0; v''(w) \leq 0; g(e_a) > g(e_b)$.

11

In the general model, the relevant outcome for the principal is profit, expressed monetarily. In the case of education, the relevant outcome is not monetary - it is pupils' test scores. Nothing would prevent the assumption that the principal is risk neutral regarding grades (not expressed monetarily), but not regarding profit (expressed monetarily). The hypothesis that the principal be neutral to the risk of test scores' oscillation would make sense, if such grades were one among many sources of utility to the principal, i.e. if they were one of the components of her "utility portfolio", whose risk would be minimized by means of diversification. Should this situation not be reasonable, i.e. should the principal's well-being strongly depend on pupils' scores in standardized tests, then the hypothesis of him being risk neutral would be implausible, making the mathematics of the model considerably more complicated.

theory – such as landlord-tenant, owner-manager, employer-employee and alike –, with the principal literally worried about the agent's level of dedication in performing his tasks, which does not lose on realism by being translated in a one-dimensional scale (e.g., number of worked hours adjusted by the intensity devoted to each hour). As shown ahead, the reduction of a teacher's work to a one-dimensional effort metric may not be reasonable, but let's overlook this for the moment.

The result, π , is presumably correlated to the actions of the agent, e , according to a probability distribution function, $f(\pi|e)$. It is worth emphasizing the importance of this assumption, which may be expressed as follows: it is assumed that pupils' performance is affected by actions taken by the teacher, but *not in a deterministic manner, but rather in a probabilistic one*. It is also assumed that a higher level of effort – say, e_a – results, on average, in higher levels of learning achievement, whereas a lower level of effort – say, e_b – leads, on average, to lower performance levels by pupils. Therefore, it is assumed that, for a higher level of dedication, a higher result is *expected*.⁹

Performing all the required tasks to ensure higher levels of pupils' learning achievement is very laborious to the teacher. Therefore, devoting higher levels of effort is considered costlier to the agent than devoting lower levels of effort. The potential conflict of interest between the two characters of the model is noticeable, once the principal is likely to desire high levels of effort by the agent (in order for pupils to get good grades) paying as little as possible (to minimize costs), whereas the agent would prefer to make little effort, but receive a good compensation. Non-economists may find these behavioral assumptions strange or inappropriate, but, right or wrong, the fact is that they are at the core of conventional economic models. The following section discusses to what extent they may or may not be adequate, in view of the singularities of the labor relations in the educational field.

The agent is also presumably "risk averse" – in economics, this means, among other things, that the uncertain and unpredicted fluctuation of a specific monetary gain, such as a monthly salary, causes discomfort to this individual.¹⁰ In order to simplify the analysis and because this is not its focus, the principal is assumed to be risk neutral, i.e. he is not negatively affected by uncertainty concerning the financial compensation to be earned (or, in the situation at hand, oscillation in pupils' grades).¹¹

The principal should ensure a minimum level of expected satisfaction to the agent in order to keep the labor contract reasonable for him, who would not be willing, for instance, to work for a too low a salary. Therefore, the principal's challenge is to establish a compensation formula that helps maximize what is in his best interest (pupils' grades)

and remains attractive to the agent – in economic jargon, a formula that considers the “participation constraints”.¹²

RESULTS WITH OBSERVABLE EFFORT

Momentarily leaving aside the difficulty to observe effort, the problem of optimization under restrictions described above could be solved, in order to achieve the ideal formula for the agent’s compensation.¹³ The result shows that the principal should offer the agent a fixed compensation. As usual in the hypothetico-deductive models applied in economic theory, the result logically follows from the set of assumptions chosen: assuming that the principal is risk neutral, he will provide the agent – risk averse – with full insurance, by not linking his compensation to the observed performance (potentially variable). Thus, for every level of desired effort, the principal would have to offer a fixed wage payment, w_e^* .¹⁴ Once known the relationship between effort and compensation and given that, for now, effort is considered observable, the principal would have to specify the optimal effort level of the agent to be demanded by contract, e^* , and subsequently establish the salary appropriate to such effort level.¹⁵

The compensation contract or formula derived from the solution of this problem simultaneously serves the interests of both principal – who ensures that the desired effort level will be achieved by a reasonable cost – and agent – who receives monetary compensation compatible with the effort put into the work and, an aspect of the highest importance, is exempt from the inconvenience caused by unexpected income variation.

Even if compensation contracts composed of a fixed and a variable part, linked to results – described later in this section – are common in other sectors, mainly in commercial activities, in the education sector, labor contracts with fixed compensation are still prevailing. There are two possible reasons for that. First, the potential contracting parts may not yet have realized the alleged advantages of variable compensation contracts. However, if that were the case, contracting parts who realized this would have a competitive advantage over others. The conventional economic theory rests upon the hypothesis of the rational choice; therefore, it would make no sense to assume that owners of private schools would not rationally choose the best contract. The enigma to be answered would then be why they would not have chosen it long ago. As for the public sector, in the absence of profit pressure (see the next section), it would be reasonable that pay-for-performance schemes did not gain popularity, as this would require pioneer initiatives of commissioners of education familiar with the model described here.

However, the prevalence of fixed compensation contracts in the education sector could also be due to an inadequacy of contracts

12

In order to define the optimal contract, the principal will have to solve the following problem of optimization under constraints:

$$\text{Max}_{e \in \{e_0, e_1\}, w(\pi)} \int (\pi - w(\pi)) f(\pi|e) d\pi$$

$$\text{s.t. } \int v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \bar{u}$$

where \bar{u} is the minimum level of satisfaction demanded, so that the agent accepts the contract.

13

The problem of outcome maximization could be simplified by means of a two-stage-problem: first, finding the salary that minimizes costs for every effort level, and second, establishing the optimal effort level. The optimal compensation should satisfy the first order condition provided by:

$$-f(\pi|e) + \gamma v'(w(\pi))$$

$f(\pi|e) = 0$, where γ represents the Lagrange multiplier. The result obtained is:

$$\gamma = \frac{1}{v'(w(\pi))}$$

14

The agent will receive a salary to meet the participation constraint “without leftovers”: $v(w_e^*) - g(e) = \bar{u}$, therefore $v(w_e^*) = g(e) + \bar{u}$. The wage should cover the inconvenience caused by the effort, $g(e)$, in addition to represent an option at least as good as those offered by the available alternatives, summarized by \bar{u} .

15

The optimal effort level is provided by: $\int (\pi f(\pi|e)) d\pi - v^{-1}(\bar{u} + g(e))$.

with a variable salary part, given the own nature of labor relations in this sector. For example, either because the association between result and effort is deterministic (rather than probabilistic), or because effort is observable at a low cost, constituting a situation where a fixed compensation contract would be optimal, according to the initial version of the model. In education, nevertheless, one must acknowledge that: (i) the association between π and e is not deterministic (as shown in the next section); and (ii) effort is not observable at a low cost, as previously argued, making it necessary to improve the model, taking this assumption into account.

In the framework of the theoretical model, even after this first partial presentation, it is already possible to start interpreting some of the reasons why many people despise accountability programs. This could be attributed, at least partly, to the inconvenience caused by the prospect of variable compensation, explained precisely by the agents' risk aversion. The level of risk aversion may vary from person to person and there is evidence, both from laboratorial experiments and from the analysis of household surveys (FALK; DOHMEN, 2008), of two associated stylized facts: (i) that women tend to be, on average, more risk averse than men; (ii) that teachers tend to be, on average, more risk averse than the average employee. The loathing of salary uncertainty would explain part of the resistance to accountability, especially in a predominantly female labor market, such as the one of educators (CHEVALIER; DOLTON, 2005).

RESULTS WITH UNOBSERVABLE EFFORT

Agent's and principal's goals will come into conflict when the latter wishes the former to put effort into her work, but the level of effort is not observable, or is only observable at a high cost. A mechanism that would possibly motivate an agent to putting the appropriate amount of effort into her work would be to tie her compensation to the observable result (pupils' grades), which, in turn, even if related to the effort, would still have a random component. In this case, incentives for a high level of effort, e_a , would inevitably bring about a given level of fluctuation to the agent's income.

At this point, aside from the participation constraint previously elucidated, a new need emerges: to motivate the agent to choose the desired level of effort, i.e. making her interested in putting this level of effort into her work without the need for constant monitoring by the principal. This new constraint, called "incentive compatibility constraint", aims at insuring that choosing the desired level of effort provides the agent with a higher level of net expected satisfaction than not choosing it.¹⁶ (Net satisfaction can be understood as the difference between the well-being provided by the compensation and

16

Formally, the new optimization problem becomes:

$$\begin{aligned} & \text{Min}_{w(\pi)} \int w(\pi) f(\pi|e) d\pi, \text{ subject to:} \\ & \text{i. } \int v(w(\pi)) f(\pi|e) d\pi - g(e) \geq \bar{u} \\ & \text{ii. } \int v(w(\pi)) f(\pi|e_a) d\pi - g(e_a) \geq \\ & \quad \int v(w(\pi)) f(\pi|e_p) d\pi - g(e_p) \end{aligned}$$

Where (i) is the participation constraint, and (ii) is the incentive compatibility constraint, in which a denotes the effort level one is willing to choose, and b denotes a level lower than that.

the inconvenience brought about by the effort). Should this constraint be met, the interests of principal and agent would then be “aligned”, to use the jargon of the economics of contracts.

The solution to this new optimization problem, now subject to two constraints, leads to a new optimal compensation scheme, $w(\pi)$, which, different from what happened in the case of observable effort, will depend on the *observed result*.¹⁷ If, under the previous conditions, a fixed wage was ideal, now the ideal compensation will require a more complicated formula. On a simplified version of the model, with only two possible effort levels,¹⁸ “high” (e_a) and “low” (e_b), the optimal compensation will depend, notably, on the ratio between probability distribution functions for each effort level, $\frac{f(\pi|e_b)}{f(\pi|e_a)}$, i.e. on relations between result and low effort (numerator) and between result and high effort (denominator).

Translating from Mathematics to English, the compensation’s subordination to this ratio means that, should either high or low effort levels by teachers lead to identical results, or even should a lower effort lead to a better result than a higher one, then the optimal compensation will again be the fixed one. However, in the third – and most interesting – possibility, with pupils’ expected grades higher for higher levels of teacher’s effort, *a performance based compensation would be valid* – typically comprised by a fixed and a variable shares.

It is important to stress a technical detail concerning this result: *the assumptions employed do not insure that the compensation should necessarily increase along with the observed grades, π* .¹⁹ This would only be true if, as pupils’ grades increased, the ratio between the likelihood that the agent had a low effort level and the likelihood that he had a high effort level reduced. This can be valid for certain grade ranges – e.g., comparing high with very high performance levels, or low with very low – but not for others, such as, say, intermediate levels. The implication is that it is possible that the optimal wage should be higher if more intermediate grade levels were analyzed, rather than high or low levels

A consequence of the technical aspect described above is that the optimal contract may not be a simple – e.g., linear and increasing – but rather a complex formula. The implementation of a complex system may however hinder the comprehension of the program’s rationale by the same agents whose behavior the program wishes to shape, making the incentive system unviable. On the search for reasons why certain programs may have failed, some authors – such as Fryer (2013), mentioned in the previous section – raise the hypothesis of lack of understanding of the formula by teachers. This leads to a dilemma between an easy and inappropriate formula, and a complex and unintelligible one.

17

The first order conditions will lead to: $-f(\pi|e_a) + \gamma v'(w(\pi)) f(\pi|e_a) + \mu [f(\pi|e_a) - f(\pi|e_b)] v'(w(\pi)) = 0$, where γ and μ represent the constraints’ Lagrange multipliers, leading to the following result:

$$\frac{1}{v'(w(\pi))} = \gamma + \mu \left[1 - \frac{f(\pi|e_b)}{f(\pi|e_a)} \right]$$

18

In general, infinite effort levels would be possible.

19

Formally, the first-order stochastic dominance assumption, $F(\pi|e_a) \leq F(\pi|e_b)$, does not imply that an increase in π produces an increase in the ratio $\frac{f(\pi|e_a)}{f(\pi|e_b)}$, as exemplified in

Figure 14.B.1

(MAS-COLELL; WHINSTON; GREEN, 1995, p. 486).

DISCUSSION

Nonetheless, even more important is the realization that the ideal compensation now depends on something difficult – if not impossible – to find out, namely the probability relations between different levels of pupils’ performance and diverse effort levels by teachers. Note that the problem here does not reside on the probabilistic nature of the relationship between π and e – as there are many reasons to believe it is indeed so –, but on the ignorance of the relevant probability distribution functions. In other words, the analysis of certain grades *generally does not accurately show the likelihood of the teacher’s effort being high, medium, low, etc.*

In practice, linear contracts are widespread – be it by chance or under the deliberate supposition that they constitute a good proxy of the optimal contract –, establishing a directly proportional relationship between pupils’ outcomes and teachers’ bonuses. Thus, notwithstanding its opacity, especially that of its result, the model described above may be of use to better found a frequent critic made to accountability programs, for instance, by Ravitch (2010), stating that the full responsibility for pupils’ performance cannot lie solely on the teacher.

It is indeed possible to envisage that a teacher does all in her power to improve the learning achievements of a group of pupils with certain characteristics – a teacher with a “high effort level”, to use the model’s jargon – without actually obtaining good results. Despite having acted as wished by the principal, this teacher will not receive her bonus, which will cause immediate frustration, as well as a decrease in future motivation. This may undermine the program’s own legitimacy. The opposite situation – low effort levels, good outcomes and bonus granted – would certainly not produce immediate frustration, but could equally diminish future motivation.

In order to carry on with the theoretical assessment, this new information requirement, resulting from the analysis of the models own result, should be disregarded, assuming that the principal may be able to estimate with reasonable accuracy the probability distribution functions mentioned above, or that the usual linear contract may represent a good proxy of the ideal contract. Even so, the principal would be faced with a dilemma: by offering compensation strictly linked to pupils’ performance – with low fixed salaries and a variable compensation highly sensitive to test scores – he would be providing powerful incentives, but causing great uneasiness, due to the high risk introduced to the compensation scheme. In doing the opposite, he would face the reverse problem (lower risk but feeble incentives). In wondering why certain accountability programs may have failed, some authors consider precisely the hypothesis of weak incentives, as happens when collective incentives are chosen, instead of individual ones (topic mentioned above and revisited in the next section).

Whatever the ideal compensation scheme, given the variability that is introduced and the assumption that the agent is risk averse, his expected salary should be fairly higher than the fixed salary he would receive for high levels of effort in a context of observable effort.²⁰ It is therefore noticed that taking the principal's difficulty in observing the agent's effort level into consideration leads to an increase in the expected implementation cost of a high effort level, e_a , compared to the former hypothetical situation, with perfect information.

A program carried out in the United Kingdom and evaluated by Atkinson *et al.* (2009) was relatively successful in the short term. At the first moment, the precaution of raising uniformly the salaries of virtually all teachers was taken, to later provide some of them with the possibility of receiving a bonus. Thus, everyone experienced a raise in the expected salary relative to the previous compensation, even those who ended up not taking part in the accountability program or those who took part in it, but were granted no bonus. Though no long term effect of this program is known, its positive short term effects may be due to this precaution, recommended by the theoretical results.

Even when faced with a raise of the expected salary, teachers' discomfort regarding accountability programs may persist due to a number of reasons, such as: (i) risk aversion levels may vary from teacher to teacher, so that the most risk averse would require a large raise of the expected salary to make up for the risk introduced with the new compensation system; (ii) a high *expected salary* does not eliminate salary fluctuations and therefore does not insure a constantly high *effective salary*, which may account for frustrations, should the bonus not be granted, with a possible impact over future motivation and effort level (leading to discouragement in extreme cases); (iii) because the relationship between e and π is probabilistic and, therefore, subject to "good luck" or "bad luck", and – even more importantly – for all intents and purposes this relationship is *unknown*, teachers will always have an intuitive perception that a bad result may occur, despite high effort levels, creating a sense of "injustice" in certain years, that might not be compensated by a sense of "good luck" in other years.²¹

SPECIFICITIES OF AGENCY RELATIONSHIPS IN THE PUBLIC SECTOR AND IN EDUCATION

Even though we've tried to understand the problematics concerning teachers' compensation in the light of the basic principal-agent model, developed for labor relationships in the private sector – where the result π is usually a sum of money (e.g., sales revenue) instead of pupils' learning achievements – we are aware that a great share of labor contracts in basic education takes place in the public sector, which has

20

Formally, given that, when the agent's effort is higher, his expected utility should be at least equal to the cost of this effort plus its reservation utility, leading to: $E[v(w(\pi))|e_a] = \bar{u} + g(e_a)$. Recalling that $v''(\cdot) < 0$, so, according to Jensen's inequality: $v(E[w(\pi)|e_a]) > \bar{u} + g(e_a)$. However, it is known that $v(w_{ea}^*) = \bar{u} + g(e_a)$, leading to the following conclusion: $E[w(\pi)|e_a] > w_{ea}^*$.

21

As documented in the behavioral economic literature, the decrease in well-being caused by the loss of a certain amount is usually higher than the increase of well-being caused by a gain of the same amount. See, for example, Kahneman (2012).

certain specificities. Moreover, even the work of teachers in the private sector differs from other occupations, for which incentive contracts were conceived in the first place.

Some studies, especially those of Dixit (2002) and Larré and Plassard (2008), point to characteristics of labor relationships in public organizations and in schools, respectively, with relevance to the theory of contracts. We have selected some of these characteristics that, in our understanding, stand out. They are presented in increasing order of importance and of practical consequences as follows:

- i. Absence of competition and profit.
- ii. Multiple agency relationships.
- iii. Likelihood that agents respond to nonmonetary motivation.
- iv. Multidimensional sets of tasks and goals, as well as limits to specialization.
- v. Education is teamwork.
- vi. Difficulties in assessing the “outcome” of the education process.

Each of these specificities is successively presented next, followed by a discussion of their relevance to the purposes of this article.

LACK OF COMPETITION AND OF CONCERN ABOUT PROFIT IN PUBLIC SCHOOLS

This article has previously referred to a difference between public and private schools. In the private sector, an environment with competition, there is always an external incentive, from which no company willing to survive in the market can escape, namely the search for profit. In the public sector, on the other hand, this search is not imperative, and neither is the impact of the competition. Suppose the potential profit of a private school is associated with its pupils' scores in standardized tests, which would take place by means of the attraction for a larger demand, willing to pay expensive tuition fees to schools well placed in rankings, such as *Exame Nacional do Ensino Médio* [High School National Exam] – Enem, for instance. This association seems to exist, given the publicity fuss made by well-placed schools on the occasion of the release of such rankings. Assuming also that, under certain conditions, accountability mechanisms may be able to promote enhancement in pupils' performance, it becomes likely that private schools show interest in implementing them.

In the public sector there is no pressure for profit – making pay for pupils' performance schemes less likely to spread naturally –, thus the common goal every stakeholder's effort should be directed to remains less evident. Establishing incentives based on pupils' scores in standardized tests would serve the purpose of converging all interests, which, at first, seems desirable – a conclusion used as argument by

supporters of teacher accountability systems. Nevertheless, matching up the interests of principal and agent may happen in a less congruous manner, which would add to explaining the inconclusive results of programs, as well as their rejection by teachers and unions. This lack of harmony is a consequence of other features of the public sector and of the education sector, discussed next.

MULTIPLE AGENCY RELATIONSHIPS

In the public sector as a whole, and in education in particular, there may be multiple “agency relationships” (as a synonym of relationships between principals and agents). As it is possible, as done so far, to take commissioner of education and teacher as principal and agent, respectively, one could also take the same commissioner of education as the agent of another principal, say a mayor (or a governor), who, in turn, would also serve as agent to other principals, such as his voters or groups with organized interests (unions, lobbies, etc.). Therefore, this would not be a one-to-one relationship as described in the basic model, but rather a network of relationships between agents and principals with different goals, possibly conflicting or categorically incompatible.

Due to goals extraneous to education, there may be vices in the very origin of the design of an incentives mechanism, should it try to answer, for instance, the demands of certain political parties, unions or lobbies, or even to fulfill campaign promises. Even a well-designed program at first may deteriorate. Should we assume a mayor’s or governor’s ultimate goal to be maximizing the own reelection chances (or those of a party colleague as successor), the actors involved in the web of agency relationships could distort incentive systems, for instance, by arbitrarily changing the rules in election years, in order to please teachers and unions, or by cutting back on their budget in favor of more popular measures among voters.

This relationship web would help explain why many programs, in Brazil as well as abroad, have been interrupted and why many others had their rules arbitrarily modified, as documented by Andrade (2008). In general, the existence of multiple agency relationships constitutes another factor that explains the inconclusive results of accountability programs – not every program would have been able to align the interests of all those involved – as well as the antagonistic opinions about them. By implication, even programs that succeeded in the short term may be subject to instability in the middle term.

AGENTS WITH NONMONETARY MOTIVATIONS

In the basic model, the agent’s well-being comes entirely from her monetary compensation (“extrinsic motivation”), effort is considered a burden and net satisfaction is the difference between the former benefit and the latter cost. Such assumptions are inadequate

whenever the agent takes pleasure in performing her tasks. Dixit (2002) states that the nonmonetary motivation (“intrinsic motivation”) could be rare among employees of the private sector – hence the behavioral assumption chosen for the basic model, designed for this sector –, but more frequent among those who choose careers in the public sector, possibly driven by some kind of idealism or vocation – for instance, among education or health professionals.

It makes no sense to idealize teachers, *a priori* picturing them as singular human beings, motivated solely by noble ideals and purposes, nor is it reasonable to treat them as a homogenous group of individuals, assigning all equal sources of motivation. However, one may not ignore a literature branch that suggests intrinsic motivation has in fact a higher importance for the average individual who chooses certain occupations. In the model developed by Heyes (2005), for example, raising nurses’ wages *reduces* the quality of the work done, because it would attract to the profession a higher proportion of individuals without the required vocation, essential for any good nurse.

Should the hypothesis that intrinsic motivation has great importance for an average teacher be valid, a cynical implication would be that the teacher’s expected salary may be relatively low and still meet the participation constraint, since the effort will not be regarded as arduous (AKERLOF; KRANTON, 2010). Another interpretation would sustain that, if nonmonetary incentives were strong, possibly even prevailing over monetary ones, the latter would not be that effective. Bénabou and Tirole (2003) even argue that the individual’s motivation may be reduced in the presence of an external form of motivation, such as a monetary one.

Belfield (2000) states that intrinsic motivation, or vocation, would be better shaped in education than in other sectors. Nonetheless, the author highlights that the evidence shows teachers do react to financial incentives, even if not always as wished by the designers of programs, i.e. they sometimes have “strategic” or “opportunistic” behavior, searching for loopholes or vices in programs (this topic will be revisited ahead).

The relative importance of intrinsic and extrinsic motivation for teachers demands further studies. For the purpose of this article, what matters is that the possible prevalence of intrinsic motivation brings about still another factor that may be able to explain the failure of some accountability programs, as well as the resistance of “agents” in the real world.

MULTIPLICITY OF TASKS AND GOALS AND LIMITS TO SPECIALIZATION

It is expedient to further analyze a previously mentioned aspect: contrary to assumptions of the basic model, teachers and school employees, be they public or private, do not perform one single task

having one single purpose at sight, so that considering effort and product as one-dimensional variables is unrealistic and inadequate.

It is reasonable to expect from a system that its pupils acquire knowledge, develop logical thinking and enhance their communication and expression skills. One may also wish schools prepare pupils for the labor market and for life, as well as help them to think critically about the society they live in, to develop emotionally, to cultivate notions of citizenship and responsibility, among many other goals.

In the presence of various tasks viewed as relevant, theory suggest equally compensating the effort put into each of these tasks. Assume there are only two goals – raising pupils’ average test scores and enhancing the self-esteem of pupils in emotional need – and that the teacher should put energy and time into different tasks in order to achieve them. So as to encourage teachers to care about both tasks it would be necessary to reward them for achieving both goals.

However, this suggestion of rewarding for multiple outcomes, aiming at stimulating the accomplishment of multiple tasks, comes across a very important obstacle: in education, and more specifically in a teacher’s work, some key outcomes are either intangible or unmeasurable (more to come on this subject). For instance, how to quantify at a low cost pupils’ level of self-esteem? That is why the proposal of compensating outcomes deriving from each and every task becomes unfeasible for all intents and purposes. In the will to implement accountability policies, important goals difficult to be measured may end up being left aside, in favor of more tangible ones, such as scores in standardized tests (BARR, 2012; BAKER *et al.*, 2010).

It just so happens, though, that different actors of the agency relationships web in the education sector may diverge as to the relative importance of each of the goals. What was understood as an accomplishment by some (e.g., an increase in pupils’ test scores) may be seen by others, in the best case scenario, as an incomplete achievement. Exclusively emphasizing one single goal that may not be considered the most important one for some constitutes another reason based on the economic literature for teachers’ rejection of accountability programs.

Before the inevitable multiplicity of tasks and goals, besides equal compensation for all tasks considered important – unfeasible, as shown above –, theory suggests yet another path: specialization, i.e. the definition of the tasks to be performed by each educator and of the results relating to each set of tasks. Thus, effort and responsibility would be, in a way, divided among educators, avoiding the “negligence” of any goal. A concrete example of specialization is the assignment of one teacher per subject. Another example is the assignment of a teacher to work as school principal or in any given managerial position. Finally, going back to the former scenario, with two goals – grades and pupil’s

emotional development –, theory would recommend putting a teacher in charge of the first goal and a psychologist, for instance, in charge of the second.

Highlights among the advantages of specialization are one's opportunity to become more efficient in one's position and – even more importantly when dealing with incentives – the possibility to separately compensate those responsible for each task. Educators specialized in tasks with more easily observable outcomes would receive an outcome-based compensation, whereas those specialized in tasks with less easily observable outcomes would receive fixed compensation.

Nonetheless, contrarily to many sectors, education presents many important boundaries to specialization. It is not always possible to unmistakably subdivide assignments. Going back to the example of the two goals, actions of a teacher “in charge of conveying knowledge” (a measurable goal, be it noisily, as discussed in the following section) will certainly have a potential impact over pupils' self-esteem (a goal which is more difficult to measure). Therefore, it would not be possible to mince responsibilities and outcomes, so as to have specialization solve the multiplicity problem outlined here. Such impossibility is related to the following specificity.

TEAMWORK

A pupil's grade in a subject's test may be directly related to knowledge acquired in other subjects. For instance, learning how to interpret texts in a language course in fourth grade helps understand mathematical problems and history or geography texts in the fourth grade, but also in the fifth and the following grades. If education is teamwork, how to grant each member of the team a compensation compatible to his relative contribution? How should incentives be designed, in order to suitably encourage teamwork?

Individualized compensation systems, matching a bonus to a single individual according to an observed outcome, supposedly deriving only from her own effort (e.g., physics teacher awarded for pupils' physics outcome at the end of a school year), tend to produce more powerful incentives than collective compensation systems, in which groups are rewarded for global outcomes. However, in a teamwork context, individualized systems may lead to a sense of injustice, due to an unsuitable perception of ownership regarding the outcome of someone else's work (e.g., a charismatic mathematics teacher of a certain grade may claim share of the bonus paid to a physics teacher who is often absent, or believe that it should rather be granted to the assiduous physics teacher of the previous grade). In extreme cases, individual incentives could discourage teachers to partner up with peers, or impair the school's work environment, damaging pupils.

The alternative of collective incentives, on its turn, leads to the traditional problem of provision of public goods, described in any microeconomics handbooks. In this scenario, everyone wishes the other would do their share – i.e. work hard, aiming at enhancing pupils' learning achievements – and that, as a consequence, the teaching staff could “produce” the public good at issue – i.e. good grades and the deriving collective bonus. The difficulty here is that, in the absence of a certain degree of mutual trust among teachers and due to the wide set of information difficulties already described in this article, those usually called “free-riders” may come up – those who do not work hard enough, not fulfilling their share in the agreement, but end up profiting from the public good obtained through the effort of others.

A possible solution would be to give one teacher the status of group supervisor and attach his compensation to outcomes obtained by the whole team – making him also subject to an incentives system. In the presence of a supervisor, each team member would then feel more under pressure to properly fulfill her duties, reducing the tendency toward idleness. For such cases, literature trusts the team size to be a key variable: the larger the group, the more difficult it is to control its members. Furthermore, it is worth recalling that the principal-agent model and the outcome-based compensation are resorted precisely because observing effort is too costly, making it reasonable to assume that, the larger the group, the higher the monitoring costs.

The difficulties that stem from teamwork, well mapped in the economic literature, add to the others previously discussed in this article as an additional contribution to explain the inconclusive results of compensation programs – which may employ either individual or collective contracts, with their respective advantages and disadvantages – and as a way to explain the inconvenience these programs produce in teachers and their unions.

DIFFICULTIES IN ASSESSING THE “OUTCOME” OF THE EDUCATION PROCESS

The basic model lays on the assumption that it is not possible to observe effort, e , but that some outcome, π , may be easily observed, such as the sales revenue of a specific salesman. However, it is not consensual which outcome is relevant in the education process. In other words, the evaluation of a teacher's performance is an important challenge in the model's transposition to education.

Larré and Plassard (2008) discuss details of different types of employee performance evaluation described in the economic literature – subjective *versus* objective, absolute *versus* relative, based on demand in quasi-market systems, etc. – transposing them to the context of education. For this article's purposes, it suffices to summarize the contrast between subjective and objective evaluations. Subjective

evaluations are performed by an immediate hierarchical superior. Their advantages include the provision of a broad evaluation of the teacher's effort, covering many aspects of the daily work (not limited to pupils' test scores) and the fact of being performed by someone who follows the teacher's activities from up close.

Nonetheless, the proximity between the parts and the subjectivity make room for the evaluator and the teacher under evaluation to collude, for inflation of the average "grade" attributed to teacher by evaluator, for compression of grade distribution, or, in another level, for too severe evaluation of subordinates not dear to superiors. Furthermore, none of these aspects can be objectively verified by a third party. Evaluators could report a great effort by a certain teacher for reasons unrelated to teaching effort; teachers aware of being under evaluation, by knowing the evaluator's personality, could relocate their effort from their end purpose to tasks aiming at pleasing or flattering their hierarchical superior, reducing the evaluation's quality and credibility.

Before such deficiencies of subjective evaluations, it would seem reasonable to resort to objective ones, generally based on test scores – precisely what was seen as the "outcome" while presenting the basic model. An important advantage of the objective evaluation is the usage of previously established criteria, which may be observed by agent, principal and third parties, fulfilling, at first sight, the requirements to take over the position of outcome, π , in the basic model.

On the other hand, the objective evaluation ignores dimensions of a teacher's work that are not related to conveying knowledge (as previously mentioned). Besides, even if conveying knowledge were the sole important aspect, questioning would come up concerning the quality of the information included in tests taken by pupils. It is very hard to determine which exact fraction of a pupil's grade is merit of a certain teacher's effort, given that, as attested in the literature on education production function, the grade is a reflex of a set of factors, current or past, school-related or not, deterministic or random (WALTENBERG, 2006). Therefore, a pupil's grade may depend on the quality of his nourishment, on the emotional and material support given by his parents, on the quality of teachers of other subjects and previous grades (as previously discussed), on classmates' motivation, on school facilities, etc.²²

Under these circumstances, it would be necessary to at least take as outcome a performance measure as "clean" as possible, i.e. obtained by means of the inclusion of many control variables, free from effects beyond teacher's reach, also taking into consideration some previous performance measure, given the cumulative nature of the learning process. This "clean" measure would be required in order to moderate mistakes and biases embedded in a gross measure of current

22

At this point, it is worth recalling a difficulty pointed out as of the presentation of the basic model. It was then said that a requirement of the model to determine the formula optimal compensation formula was that the probability distribution functions relating effort and outcome were known. For example, in the simplified version, with only possible effort levels, low (e_l) and high (e_h), one would have to know $f(\pi|e_l)$ and $f(\pi|e_h)$, but this information is not available. Thus, even if one decides to make use of an objective assessment of the teachers' effort level, its proper deduction based on scores in standardized tests will be unviable.

performance. Otherwise, for comparable effort levels, a teacher facing a difficult class, for example, or in a decayed school, would be less likely to make her pupils learn more. In fact, the true incentive provided in this case would be a side effect: school principals and teachers would avoid pupils, classes or schools considered more “difficult”. For that matter, Schookaert and Ooghe (2013), have demonstrated that *any accountability program* – be it weak (only through disclosure of test scores) or strong (including bonus or penalty for teachers and schools) – *has the side effect of producing incentives for the selection of pupils*.

Should the measurement of the outcome be too noisy, as seems to be the case for test scores, the theory of contracts suggests that the outcome be used in moderation in the compensation formula, or even ruled out, in extreme cases (PRENDERGAST, 1999). For a different reason, closer to the previous discussion on multiplicity of tasks and goals – that tests do not thoroughly assess the work done by teachers –, Baker *et al.* (2001) converge to the same recommendation: that standardized tests should be used only as part of a broader assessment process.

Lastly, Baker *et al.* (2010) point out that tests have statistical limitations, such as inaccuracies due to small samplings (e.g., small school or classes), yearly fluctuations in teaching staff or student body, as well as measurement errors. Menezes-Filho and Tavares (2011) investigate precisely the magnitude of measurement errors in the present accountability program in the State of São Paulo. They make use of two performance measures, the results of Saresp (that define who will be granted bonus and its value) and of Prova Brasil (Brazil’s national standardized test, which is irrelevant for the definition of São Paulo’s bonus) and come to the conclusion that not less than 35% of grade variance are due to sampling features and random factors, including measurement errors. By implication, they conclude that around 11% of the schools that have reached the goals would not have done it, had the measurement errors been left out, and around 11,5% of the schools that did not reach their goals would have done it and received the bonus. In short, even the seemingly straightforward task of measuring the outcome, π , is also full with difficulties.

CONCLUSIONS

This article has presented a scrutiny of the theoretical framework on which teacher accountability policies lay, as to test a main hypothesis – that the inconclusive effects of such policies would find an explanation in the economic literature itself – as well as a complementary hypothesis, that economic theory also sheds light on the reasons for the considerable rejection of such programs in certain circles. The

observations presented along the article lead to the conclusion that there are not enough elements to reject either of the hypotheses.

As for the first hypothesis, thorough analyses of the basic principal-agent model and of the adaptations required to transpose it to the public sector and to education show that it was to be expected that some programs, as carefully designed and conducted as recommended by the economic theory, were to be successful – at least according to the definition of success based on the conventional parameters, namely increase of average test scores in the short term. Yet there are so many requirements for them to work and such high chances that some flaw interposes or a serious side effect arises, that the failure of some programs does not constitute a surprise, be it according to the conventional measurement, focused on the average score, or to more comprehensive evaluation criteria.

As for the second hypothesis, many of the teachers', unionists and academics' critiques may have ideological roots – perhaps even corporatist, as some imply –, but the truth is that some of them relate almost literally to difficulties already well mapped in the economic literature – which is precisely the primary theoretical source on which accountability programs lay. Thus, these critiques should not be summarily disdained by economists and managers in the field of education.

What are the implications of this report and of its conclusions? Fierce defenders of accountability programs will infer it is possible to design successful programs, taking into consideration all the warnings listed here, as well as the examples of successful programs. The most strident opponents, in turn, will find in this article a good foundation or reinforcement for their critics.

We are not in a position to state that either one or the other view is correct. But we must admit that, based on what was presented here, we are very reticent to the idea that accountability programs are a key mechanism to improve the quality of our educational system. It seems to us that, at best, an accountability program, even when extremely well-conceived and implemented, will only be able to inadequately or roughly reward teachers' effort. Despite that, nothing prevents pupils' average test score to increase as a consequence of the implementation of a program – so that one of the goals is reached. Nonetheless, at the expense of side effects still not fully known, of the spread of various forms of resentment and of resistance that could undermine the reform's legitimacy in the long run.

As economists, we are very in favor of the idea of comparing costs and benefits in the decision making. For us, it does not seem to be evident, at the current stage of accumulated knowledge on the subject,

that the potential benefits of accountability programs are sufficiently higher than the costs involved.

REFERENCES

- AKERLOF, G. A.; KRANTON, R. E. *Identity economics: how our identities shape our work, wages, and well-being*. New Jersey: Princeton University Press, 2010.
- ALEXANDRE, M. R. *Programas de responsabilização de professores: quais são seus reais efeitos?* Dissertação (Mestrado em Economia) – Universidade Federal Fluminense – UFF, Niterói, 2013.
- ANDRADE, E. “School accountability” no Brasil: experiências e dificuldades. *Revista de Economia Política*, São Paulo, v. 28, n. 3, jul./set. 2008.
- ATKINSON, A. et al. Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, n. 16, p. 251-261, 2009.
- BAKER, E. L. et al. Problems with the use of student test scores to evaluate teachers. *EPI Briefing Paper*, n. 278, 2010.
- BARR, N. *Economics of the welfare state*. 5. ed. Oxford: Oxford University Press, 2012.
- BELFIELD, C. R. *Economic principles for education: theory and evidence*. Northampton, MA Edward Elgar, 2000.
- BENABOU, R.; TIROLE, J. Intrinsic and extrinsic motivation. *Review of Economic Studies*, v. 70, p. 489-520, 2003.
- BRUNS, B.; FILMER, D.; PATRINOS, H. A. *Making schools work: new evidence on accountability reforms*. Washington, DC: World Bank, 2011.
- CHEVALIER, A.; DOLTON, P. The labour market for teachers. In: MACHIN, S.; VIGNOLES, A. (Ed.). *What's the good of education? The economics of education in the United Kingdom*. Princeton: Princeton University Press, 2005.
- DIXIT, A. Incentives and organizations in the public sector: an interpretative review. *The Journal of Human Resources*, v. 37, n. 04, p. 696-727, 2002.
- DOLTON, P.; MARCENARO-GUTIERREZ, O. *Global teacher status index*. Varkey-Gems Foundation, out. 2013.
- FALK, A.; DOHMEN, T. You get what you pay for: incentives and selection in the education system. In: ECONOMIC INCENTIVES: DO THEY WORK IN EDUCATION? INSIGHTS AND FINDINGS FROM BEHAVIOURAL RESEARCH. 16-17 may 2008. Munique, Alemanha: CESifo/PEPG, 2008.
- FERRAZ, C. Sistemas educacionais baseados em desempenho, metas de qualidade e a remuneração de professores: os casos de Pernambuco e São Paulo. In: VELOSO, F. et al. (Org.). *Educação básica no Brasil: construindo o país do futuro*. Rio de Janeiro: Campus/Elsevier, 2009.
- FRYER, G. R. Teacher incentives and student achievement: evidence from New York city public schools. *Journal of Labor Economics*, v. 31, n. 2, p. 373-427, 2013.
- HEYES, A. The economics of vocation or why is a badly paid nurse a good nurse? *Journal of Health Economics*, v. 24, n. 3, p. 561-569, 2005.
- KAHNEMAN, D. *Rápido e devagar*. Duas formas de pensar. Rio de Janeiro: Objetiva, 2012.
- LARRÉ, F.; PLASSARD, J. M. Quelle place pour les incitations dans la gestion du personnel enseignant? *Recherches Économiques de Louvain: Louvain Economic Review*, v. 74, n. 3, 2008.
- LAVY, V. Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy*, v. 110, n. 6, 2002.
- LIMA, R. *Programas de responsabilização de professores: análise crítica dos fundamentos teórico-conceituais e das evidências empíricas*. (Trabalho de conclusão de curso) – Faculdade de Economia, Universidade Federal Fluminense – UFF, Niterói, 2013.

MAS-COLELL, A.; WHINSTON, M.; GREEN, J. *Microeconomic theory*. Oxford: Oxford University Press, 1995. cap. 13, 14 e 23.

MENEZES-FILHO, N.; TAVARES, P. A. Noise in education pay for performance programs: evidence using independent measures of school outcomes. In: ENCONTRO BRASILEIRO DE ECONOMETRIA, 33., 2011, Foz do Iguaçu. *Anais...* Foz do Iguaçu: SBE, 2011.

MURALIDHARAN, K.; SUNDARARAMAN, V. Teacher performance pay: experimental evidence from India. *Journal of Political Economy*, v. 119, n. 1, p. 39-77, 2011.

OSHIRO, C. H.; SCORZAFAVE, L. G. Efeito do pagamento de bônus aos professores sobre a proficiência escolar no Estado de São Paulo. In: ENCONTRO NACIONAL DE ECONOMIA, 39., Foz do Iguaçu, Brazil, 2011.

PRENDERGAST, C. The provision of incentives in firms. *Journal of Economic Literature*, v. 37, n. 1, p. 7-63, mar. 1999.

RAVITCH, D. *Death and life of the great American school system: how testing and choice are undermining education*. New York: Basic Books, 2010.

SCHOKKAERT, E.; OOGHE, E. School accountability: can we reward schools and avoid pupil selection? Bonn: IZA, may 2013. (Discussion Paper Series, n. 7420)

TRANNOY, A. L'égalisation de savoirs de base: l'éclairage des théories économiques de la responsabilité et des contrats. In: MEURET, D. (Éd.). *La justice du système éducatif*. Bruxelles: De Boeck Université, 1999. (Coleção Pédagogie en développement)

WALTENBERG, F. Teorias econômicas de oferta de educação: evolução histórica, estado atual e perspectivas. *Educação e Pesquisa*, São Paulo, v. 32, n. 1, p. 117-136, jan./abr. 2006.

MARAYSA RIBEIRO ALEXANDRE

Executive Analyst of the Rio de Janeiro State Department of Education;
M. A. in Economics from Universidade Federal Fluminense – UFF
maraysaribeiro@gmail.com

RICARDO SEQUEIRA PEDROSO DE LIMA

M. A. in Economics student at the Stockholm University (Sweden);
B. A. in Economics from Universidade Federal Fluminense – UFF
rise0313@student.su.se

FÁBIO DOMINGUES WALTENBERG

Assistant Professor at the Department of Economics and at the PhD
program in Economics; researcher at the Center for Studies on Inequality
and Development – CEDE – from Universidade Federal Fluminense – UFF
waltenberg@economia.uff.br