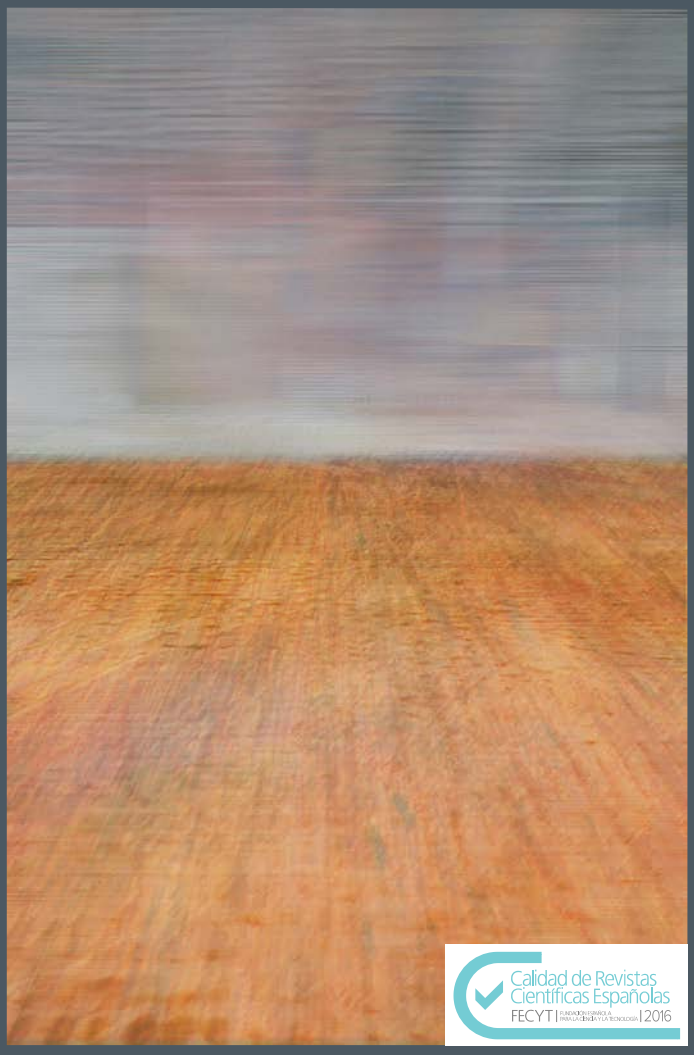
revista de EDUCACIÓN
Nº 377 JULIO-SEPTIEMBRE 2017

Incentivos salariales para el profesorado vinculados con el rendimiento de los estudiantes: análisis y propuestas, considerando el modelo catalán

Salary incentives for teachers linked to student outcomes: proposals based on an analysis of the Catalan model

Josep-Oriol Escardíbul Ferrá



GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

Calidad de Revistas Científicas Españolas
FECYT | 2016

Salary incentives for teachers linked to student outcomes: proposals based on an analysis of the Catalan model

Incentivos salariales para el profesorado vinculados con el rendimiento de los estudiantes: análisis y propuestas, considerando el modelo catalán

DOI: 10.4438/1988-592X-RE-2017-377-351

Josep-Oriol Escardíbul Ferrá
Universitat de Barcelona & IEB

Abstract

Salary incentives for non-university teaching staff, especially those linked to student outcomes, have been the subject of debate for some years. Their proponents highlight, above all, their ability to attract and retain talent in the education system, while their detractors claim that such incentives have no place in the profession and stress the methodological problems encountered in their application. The aim of this study is to analyse the suitability of introducing salary incentive systems for non-university teaching staff, focusing especially on bonus payments linked to student outcomes. To do so, first, different possible systems of salary incentives for teachers are described. Second, various evaluations of incentive systems employed both in the United States and elsewhere are reported. Third, the incentive system adopted in the Spanish autonomous community of Catalonia, formally regulated in 2014 and applied for the first time this school year, is critically evaluated. This autonomous community is a pioneer in applying this incentive system in Spain. Finally, an evaluation is undertaken of the suitability of operating a teacher assessment system and of basing it on student outcomes; and, taking these considerations into account, the Catalan program is analysed and a system of incentives is proposed for the whole of Spain.

Key words: international comparison, salary scale, assessment, salary incentives, pay per performance, student outcomes, Catalonia, Spain.

Received: 29-04-2016 Accepted: 03-03-2017

Resumen

Los incentivos salariales en el ámbito del profesorado no universitario, especialmente los vinculados al rendimiento de los estudiantes, son objeto de debate desde hace unos años. Principalmente, los defensores de su existencia destacan la capacidad de atraer y retener talento al sistema educativo. Los detractores no desean su aplicabilidad como incentivo entre el profesorado y enfatizan los problemas metodológicos existentes para ser aplicados. El objetivo del estudio que aquí se presenta es analizar la conveniencia de establecer sistemas de incentivos salariales al profesorado de enseñanzas no universitarias, con especial atención al denominado pago vinculado al rendimiento de los estudiantes. Por dicho motivo, en primer lugar se exponen las distintas posibilidades existentes a la hora de establecer incentivos salariales para el profesorado. En segundo lugar se muestran diversas evaluaciones realizadas de sistemas de incentivos existentes en varios países. Se consideran estudios tanto de Estados Unidos como las experiencias en otros países. En tercer lugar se valora el sistema de incentivos existente en Cataluña (regulado en 2014) y que se ha empezado a aplicar en el presente curso. Esta Comunidad Autónoma es pionera en aplicar este sistema de incentivos en España. Finalmente, se concluye con una valoración de la evidencia presentada que tiene diversos ámbitos: por un lado, la conveniencia de que exista un sistema de evaluación del profesorado y que dicha evaluación se base en el rendimiento de los estudiantes; por otro lado, y teniendo en cuenta las consideraciones anteriores, se analiza la experiencia catalana y se propone un sistema de incentivos para el conjunto del España.

Palabras clave: comparación internacional, escala salarial, evaluación, incentivos salariales, pago por rendimiento, rendimiento alumnado, Cataluña, España.

Introduction

Improving students' academic performance is high on the political agenda of most governments. To achieve this, empirical evidence suggests that it is better to invest in teacher quality than in quantity (Hanushek, 2011; author). Indeed, in the field of teacher policies, one measure that has recently received considerable attention has been the introduction of wage incentive programmes, known as 'merit pay' or 'pay per performance'. Such payments involve evaluating the performance of teachers and offering a monetary incentive to those that achieve certain levels. In these evaluations, student performance on external tests is

Received: 29-04-2016 Accepted: 03-03-2017

usually of high (or exclusive) importance, so that teacher assessments depend to a large extent, if not fully, on the performance of their students.

Around twenty developed countries offer wage incentives based on the outcomes of teacher assessments (Woessman, 2011), as do more than the half the states in the US (Berry and Eckert, 2012). However, not all attach the same weighting to student performance and, in fact, the characteristics of these programs are quite disparate. Following their lead, in 2014 the Catalan government incorporated performance-related pay in its legislation. Thus, to obtain the traditional *sexenio* (a wage supplement paid after each six years of service), teachers can ask to be assessed. One of the elements evaluated as part of this assessment is student performance.

This study analyses wage incentive systems for teachers that include some kind of bonus related to student performance. More specifically, it evaluates the system recently established in one of the autonomous communities of Spain, Catalonia, and assesses whether such incentives should be extended to the whole of the State. As there is not yet any empirical evidence about the impact of the Catalan pay-per-performance program, this study undertakes a critical evaluation of the model as it is being implemented.

The rest of the paper is structured as follows. The next section describes the various systems of teacher assessment that might be adopted and examines the opinions of their proponents and detractors. A review is then undertaken of studies that evaluate merit pay systems in operation in different countries. This is followed by an evaluation of the system introduced in Catalonia. The final section concludes by analysing the evidence presented and proposes a system of incentives for Spain.

Characteristics of teacher assessment systems based on performance

As noted, teacher performance assessment systems attach a significant weight to student performance. However, across the globe, teacher performance is assessed in many different ways. The different possibilities for designing an incentive system are outlined below:

- Performance index based on the inputs and outputs of the education production function, or a combination of the two:

Received: 29-04-2016 Accepted: 03-03-2017

  - In the case of inputs, this means including measures of the tasks performed by teachers: classroom management, development and production of class exercises, dealing with families, participation in school activities, etc. This index does not take student outcomes into account.
  - In the case of outputs, the results obtained by students on external tests or the gain recorded by the students on these tests are taken into consideration. Typically, the latter approach is adopted, applying one or other of two formats. First, a comparison is made between the students' gains on the external tests and their expected gains, where the latter are estimated by taking into account the students' previous test scores and their socioeconomic characteristics. Thus, the gain recorded by a student is compared with the expected gain given his or her circumstances and prior academic record. The difference between the two, i.e. the "value added", is then assigned to the teacher. Second, the improvement recorded by each student in their test results is compared to the gains of the rest of the students (or to those similar students), so that the gain is analysed in comparative terms. Each teacher is then assigned an average gain based on the relative improvement in the outcomes of each of their students compared to that of the other students.
- Individual or group incentives. In the first case, the bonus is awarded to each teacher according to the performance of his or her students. In the collective system, the incentive is awarded either to the teachers of a given school year, subject department or the whole school if their targets are achieved. The group then decides how the bonus should be distributed among the individual members.
- A one-off award (for a given year or period) or an incentive that is introduced as a permanent salary rise.
- A fixed, predetermined bonus (awarded when a given benchmark in the teacher assessment is reached) or a bonus that varies according to the results obtained (different bonuses linked to different benchmarks).
- Open (any teacher can win the bonus if they meet the fixed objectives) or closed (where only a few are eligible for the bonus – usually, taking the form of a competition or tournament).
- By external (conducted by an education authority) or internal (conducted by the school's management with possible participation of members of teaching staff) assessment, or a combination of the two.

Received: 29-04-2016 Accepted: 03-03-2017

Among the indicators used to assess teacher performance, the most controversial are those based fully or partially on students' outcomes, with strong arguments both in favour and against their use. These are set out below, although, briefly, the arguments in favour emphasise that a pay-per-performance system allows schools to attract and retain the best teachers as well as to increase students' results; in contrast, the arguments against stress the difficulty of linking student academic performance to teacher performance.

Arguments in favour of teacher assessment systems based on student performance:

- Such systems can lead to genuine improvements in student outcomes.
- Student performance is an objective indicator. It facilitates decision-making, because while school principals are able to detect quite readily the best and worst performing teachers, they have difficulty distinguishing between those in the middle.
- Teaching is one of the professions with the smallest range of salaries. Merit pay may represent the most effective form of wage increases.
- These systems can affect the selection and retention of teachers, attracting more competent teachers and ensuring the less competent take their talents elsewhere.
- They can improve the image of state schools by focusing greater attention on student outcomes.
- They are more cost-efficient than measures aimed at improving student-teacher ratios.
- Group incentives can encourage cooperation between teachers. Likewise, these systems promote a culture of continuous improvement in schools and encourage the participation of teachers in decision-making processes.
- The value-added method takes into account differences in student performance attributable to their socioeconomic situation, thus countering any criticism of the comparison of the performance of unlike students.
- Such systems may be more successful if introduced as a long-term plan and not as a pilot scheme or as a one-off award.

Arguments against:

Received: 29-04-2016 Accepted: 03-03-2017

- Students' academic results depend on several factors over which teachers have no control. Moreover, an (external) test only reflects one part of what has been learned.
- External test results are never available for all the students in each school year or in all subjects, thus making it impossible to apply objective methods.
- The assumption is made that the students' outcomes are the fruit of the work undertaken solely by that year's teacher, and that a teacher has no influence on the students' future learning.
- Student efforts can vary depending on the importance of a particular year in their overall school career, regardless of the work of their teachers.
- Student results on external tests may vary markedly from year to year and such variations cannot be attributed to the teachers (or even to student efforts).
- Teachers can adopt strategies to improve student outcomes but without necessarily improving their learning. These strategies can involve teaching students how to pass a particular test (teaching to the test) or cheating to improve results (revealing test questions, allowing copying, selecting the students that will sit the test, etc.).
- Individual performance payments may limit teacher cooperation, while group payments may lead to some teachers not committing themselves to the job at hand ('free-riders'), especially in schools with a large staff numbers.
- The costs of managing the incentive system may be high.
- Teachers may not be favourably disposed to a pay-per-performance system and union opposition can lead to potential conflict.

Review of studies evaluating performance payment systems

This section presents a review of studies that evaluate performance payment programs employing robust methodologies, that is, experiments or quasi-experiments (but above all the former) that allow causality to be inferred (for a detailed explanation of the analytical techniques used, see Schlotter et al., 2010). Results of the evaluations conducted in the United States are presented first, followed by those of programs employed in other countries. The review considers the analytical methodology applied in each case and Table 1 records their main characteristics.

Received: 29-04-2016 Accepted: 03-03-2017

TABLE I. Evaluations of pay-per-performance programs

Program	Incentive target	Simple index	Input/ Output	Comparative output assessment	Result
Experimental designs					
SPBP (New York)	School	No	15% input 85% output (60% value added)	Yes	Not significant
Chicago Heights	Individual and teacher group	Yes	Output (value added)	Yes	Only the fear of losing bonus was positive
Nashville (POINT)	Individual	Yes	Output (value added)	Yes	Not significant
Round Rock (Texas)	Teacher group	No	Output (value added)	No	Not significant
Kenya	School	Yes	Output (points and value added)	Yes	Doubtful positive outcome
India	Individual and school	Yes	Output (value added)	No	Positive. Individual improvement
Mexico	Individual and group	Yes	Output (value added)	No	Positive
Quasi-experimental designs					
North Carolina	School	Yes	Output (value added)	No	Positive
District of Columbia (IMPACT)	Individual	No	45/85% input 55/15% output (value added)	No	Positive
Austin (REACH)	Individual, group and school	No	Input and Output (value added)	No	Positive
TAP (Chicago)	Group	No	50% input 50% output (value added)	No	Not significant
Denver (ProComp)	Individual and group	Yes	Input and output (value added)	Yes	Positive
Little Rock (Arkansas)	Individual	Yes	Output (value added)	No	Positive
Israel (2002)	School	No	Output (test points and others)	No	Positive
Israel (2009)	Individual	Yes	Output (test points and others)	No	Positive
England	Individual	No	Input and output (value added)	No	Positive
México (Carrera Magisterial)	Individual	No	Input and output (test points)	No	Positive in part
Portugal	Individual	No	Input and output (test points)	No	Negative
Chile (SNED)	School	No	Input and output (points and value added: 65%)	Yes	Positive

Received: 29-04-2016 Accepted: 03-03-2017

The programs evaluated in the United States, while being quite distinct from one another, all link performance pay strongly, or even exclusively, to student outcomes (typically measured as gains on external standardized tests in maths and reading comprehension). The first three programs examined are experimental evaluations of specific programs that failed to find any positive results (see reviews of these studies in World Bank, 2015; Neal, 2011). First, the New York City Schoolwide Performance Bonus Program (SPBP), which provided incentives for underperforming schools in New York City, considered student scores on state tests (representing 25% of the total value of this indicator) and their progress in terms of outcomes (60% of the total). Assessments were comparative, partly with other schools throughout the city but, especially, with schools presenting similar characteristics (Fryer et al., 2012). Second, the Project on Incentives in Teaching (POINT), conducted in Nashville (Tennessee), in elementary and secondary schools, provided individual financial rewards for faculty conditional upon students presenting gains on state tests in maths higher than those obtained by other students with the same score in the previous school year. The incentive increased as the teacher's students rose into the top percentiles of the distribution of results. Third, the Round Rock District (Texas) introduced a pay-for-performance program that, unlike the other two, targeted high performing schools. In this case, the bonus was awarded according to the value added of a (multidisciplinary) team of teachers across several subjects (the group had to be placed in the top third to receive the reward). Thus, it was a group prize but individuals in the winning team might not receive it if their individual value-added score was low. Value added was calculated as the difference between the students' performance on external tests and their expected performance (obtained from an estimate that considered both the students' results in prior years and their personal and socioeconomic circumstances).

In contrast, the program implemented in Chicago Heights, a low-income locality in Illinois, was partially successful. This program created a performance payment system with four types of incentive for high school teachers: two individual and two group, conditional on the students obtaining good relative results (above the average results of the other schools) on external tests. The program compared the different behaviour of teachers towards the will to win a bonus and the fear of losing it. Thus, for each type of incentive (individual and group), one

Received: 29-04-2016 Accepted: 03-03-2017

system allowed teachers to earn up to $8,000, while the other paid half the bonus upfront, but this could be forfeited if the students failed to meet their objectives (or raised if the outcomes were positive). The results indicated that teacher performance only improved in the case of those in the system at risk of losing their incentive (moreover, greater effects were obtained when the incentive method was collective).

Overall, the results of the programs are quite disappointing. However, various elements might account for the poor outcomes. In the case of the New York and Texas programs, the complexity of the indicator (which made it difficult to know *a priori* the amount of effort required on the part of the teacher to receive the incentive) and the fact that a good many of the teachers were not in favour of a performance payment as an incentive may account for the negative results. Moreover, in the cases of New York and Nashville, all underperforming schools had an additional incentive to improve – namely, the threat of sanctions or closure they were under following the passing of the so-called No Child Left Behind Act of 2002 (which meant all teachers had an incentive to improve their results). Finally, the goals set in Nashville may have been too high to serve as a motivating factor (Dee and Wyckoff, 2013).

Among the quasi-experimental studies, we find analyses using techniques of regression discontinuity design, matching and differences-in-differences (Schlotter et al., 2010). Among the first of these, evaluations of two long-term programs stand out. First, the ABC School-wide Bonus Program, in North Carolina, awarded a bonus to teachers (in elementary and secondary schools) if the school's student body improved their maths and language scores by the expected amount, and doubled that bonus if they exceeded that score. The students' expected results were calculated in terms of a prediction that took into account the students' prior performance and their socioeconomic circumstances. The program evaluation recorded a positive effect, especially in the case of mathematics (Vigdor, 2008). Second, the IMPACT program (in the District of Columbia) also managed to improve the outcomes of students in public schools in what is a fairly conflictive district. This program placed teachers in five performance bands, with quite specific features: for example, high-performing teachers were able to earn a significant bonus, which could be consolidated in their salary if they achieved a certain benchmark during two consecutive years; low-performing teachers were dismissed, while those in the level immediately above them (classed as "minimally

Received: 29-04-2016 Accepted: 03-03-2017

effective") ran the risk of dismissal if they did not improve their results during the school year. The program evaluated the outcomes of students whose teachers were placed in the first and penultimate bands (the statistically most relevant points) and recorded significant gains in their performances in both cases. The incentive was measured in a multidimensional way, with the weight of the output being higher in those subjects for which student results on state tests were available. The incentive was calculated in a similar way to that employed in the previous program (Dee and Wyckoff, 2013).

Among the other quasi-experimental analyses, three studies seek to overcome selection bias by matching similar schools in terms of their observable characteristics. The REACH Program in the school district of Austin (Texas) included several incentive systems (individual and group, input- and output-based) for elementary and secondary schools. An evaluation highlighted the program's ability to increase the value added of student outcomes, measured by comparing the actual result with the expected result taking into account the students' socioeconomic characteristics. Gains were only recorded in the first year, although these were maintained into the second year (Balch and Springer, 2015). In contrast, an evaluation of the application of the Teaching Advancement Program (TAP) in Chicago's elementary schools showed no improvement in outcomes. In this case teachers were remunerated (collectively) using a method that gave equal weight to the results of the teachers' classroom observations and the gain in student scores on a state test. A quite distinct example is provided by Denver's ProComp program, initiated in 2006. This program did not offer incentives but rather to change the entire pay system, so that wage increases (not the base salary) became totally dependent on stimuli linked to a series of input and output indicators. In the case of outputs, the improvement in student results was taken into account. Thus, teachers received the incentive if at least 50% of their students were in the 55th percentile or higher of the gain in the results of the state's students, in the subjects of mathematics and language. The evaluation of this program revealed an improvement in the students' results taught by teachers adhering to the ProComp system compared to those who chose not to opt in (Goldhaber and Walch, 2012).

Finally, a differences-in-differences analysis was used to evaluate a performance-pay program implemented in Little Rock (Arkansas). In this study, Winters et al. (2008) report an improvement in the outcomes of

Received: 29-04-2016 Accepted: 03-03-2017

students at the participating elementary schools, compared to those of non-participants, especially in schools presenting the worst results before the program was introduced. Here, (individual) financial awards depended on student achievement on the state test in maths and reading comprehension. The gain recorded by each student over one school year was considered with the teachers' bonus increasing as these outcomes improved. The final bonus was based on the learning gain observed across their group of students.

A number of experimental and quasi-experimental evaluations have been undertaken outside the United States. Among the former, those conducted in the developing countries of Kenya, India and Mexico stand out. In the first case, Glewwe et al. (2010) analysed an incentive program implemented at the school level in 50 primary and secondary schools, where teachers obtaining the best results or the greatest gains in results were rewarded with a bonus. The study reports an improvement in student performance, although this gain only occurred in the subjects included under the incentive program and only while the program was in operation. As such, the results can be interpreted as indicating that the performance payment did not improve broad-based learning, but simply the teachers helped their students to prepare the evaluation tests. The second program was introduced in primary schools in a region of India. In this case, the incentive was given for improvements (greater than 5%) in student performance on maths and language tests. The results point to the success of the program in both the short and long term, with individual incentives having a greater effect than those at the school level (Muralidharan and Sundararaman, 2011). Finally, the Aligning Learning Incentives Program, introduced in 88 secondary schools in Mexico, offered three types of incentive: to students only (depending on their results in an end-of-year, curriculum-based maths exam); to teachers only (depending on the performance of their students); and to both students and teachers (in the case of the former depending on their results and the score of their classmates, and in the case of the latter depending on their students' results and on those of the other maths students in the school). The incentive depended on the students improving their level (with three levels being fixed), while teachers were penalized if their students fell to the lowest level. The results showed a positive effect of the program in the case of the first and, especially, the last treatment types, when the reward targeted both students and teachers and

Received: 29-04-2016 Accepted: 03-03-2017

depended on the results of all the school's students (Behrman et al., 2011).

Among the studies employing quasi-experimental methodologies, two conducted in Israel (Lavy, 2002, 2009) stand out for their methodological rigour. The first, undertaken in 2002, considers the effect of a program introduced in upper secondary schools with underperforming students. The reward in this instance was for the whole school and was received only by teachers in schools lying in the upper third of gains on various indicators linked to national university entrance exams. The program evaluation showed that the students at the participating schools improved their results and that their dropout rates fell. The second evaluation, reported in 2009, considers a similar program but on this occasion the incentives were given to teachers individually. The program was deemed positive, insofar as the probability of students sitting the national exam at the end of high school and their grades both increased. In both programs performance was measured by considering the difference between the result obtained and the expected result based on the characteristics of the school, in the first case, and on those of the students and their teachers, in the second.

Other, less methodologically sophisticated, studies have analysed programs implemented in England, Mexico and Portugal. While the programs are different, they all incorporate an individual merit pay system in which student progress (results or value added on external tests) is only one of the, often minor, elements included, together with indicators of inputs in the educational process (such as classroom teaching or professional development, usually linked to training). Likewise, these programs seek to establish permanent salary improvements (not just a one-off award). In the first two cases, positive effects of the pay-per-performance systems in terms of improved student outcomes are reported, although they were only quantitatively important in the case of England (Atkinson et al., 2009). In contrast, the Portuguese salary improvement program caused a significant decline in student achievement (Martins, 2009). Finally, Contreras and Rau (2012) evaluated an incentive program in Chile, introduced in 1997, and known as the *Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educacionales* (SNED). In this program, socially similar schools in each region competed with each other and the evaluation took into consideration a mean index score for each school that included educational measures of input and

Received: 29-04-2016 Accepted: 03-03-2017

output. The latter referred both to student scores on maths and language tests and to gains over time (equivalent to 65% of the total index value). The incentive was given to schools in the top 25% of this indicator value. The evaluation concluded that the program had significantly improved student outcomes.

Finally, we highlight the regression analysis reported by Woessmann (2011), with PISA data for 28 OECD countries, of which 13 had implemented some type of pay-per-performance program. Although the methodology prevents inferences of causality, the analysis is valuable because it looks at several countries at once. The results show a positive relationship between having a merit payment system (of any type) and student outcomes in mathematics and reading comprehension.

In short, the incentive systems described above are characterised by the following main elements. In the United States, programs employing collectively based incentives (that is, either group or school-wide) predominate over those offering individual incentives. Moreover, measures of student outcomes form a part of all the programs. This output indicator, however, is not usually comparative (outcomes being compared, for example, with those of other schools), rather there is a predominance of systems that analyse a student's performance relative to his or her potential. We also find as many simple as complex indexes. A comparison with the systems employed in other countries shows that output indicators are also always included and that they are not usually assessed on a comparative basis (as in the United States). Regarding other characteristic, namely incentive levels (individual or collective) and the nature of the index (complex or simple), we find an almost fifty-fifty divide.

The following conclusions can be drawn about the effect of pay-per-performance systems. The experimental evaluations undertaken in the United States highlight the virtual absence of any effects. In contrast, evidence from developing countries points to more positive outcomes. Likewise, the quasi-experimental evaluations undertaken both in the United States and elsewhere tend to show positive results. This suggests that performance pay can be successful in improving student outcomes but that it does not offer any *a priori* guarantees. As for the characteristics of the pay-per-performance systems that appear to be effective, the evidence presented allows us to identify a number of common elements. Obviously, causality cannot be inferred, but we can conclude that,

comparatively speaking, successful programs make more frequent use of simple indexes, applied totally or partially at the individual teacher level and employ non-comparative output indicators.

The Catalan program

Catalonia has recently adopted regulations whereby teachers can obtain bonus payments subject to a favourable assessment in a system based, in part, on their student outcomes. The assessment is voluntary – being open to *funcionarios docentes de carrera* (that is, qualified civil servants with a fixed teaching post), *docentes en prácticas* (teacher trainees), *interinos en activo* (working, fixed-term supply teachers), and teachers of Religious Education, which means the same bonuses can, in fact, be obtained by other means (see detailed explanation below). The introduction of the system has not been without its problems as it has had to overcome a general sense of mistrust among teachers and educational authorities alike due, in the main, to its introduction coinciding with the economic crisis and cutbacks throughout the sector.

The program was officially introduced by *ORDRE ENS*/330/6 November 2014, outlining regulations for promotion to a higher pay scale and the process by which a teacher can opt for a permanent pay rise. The system includes five ascending pay scales, with a gross monthly bonus of between 102.97 and 132.96 euros being paid for each, depending on the specific scale reached. A teacher can ask for a scale rise every six years of teaching experience, the payment being known as a *sexenio* (although, since 2012, this was extended to nine years for the first scale rise). Teachers have to accumulate a total of ten points to be awarded the scale rise. Each year of teaching is equivalent to one point, while to obtain the remaining four points (or one in the case of the first scale), teachers have to present certified evidence of their having undertaken various activities. The activities include having held positions or carried out duties of responsibility; having undertaken training courses; and having engaged in complementary activities in the school (note, the regulations assign a specific score to each of these). Similarly, teachers can earn points by asking to be assessed. The regulation provides for two types of assessment (which can be requested in alternate years): on the one hand, the role played by the teacher in improving the school's outcomes; and, on the other, the teacher's individual teaching work.

Received: 29-04-2016 Accepted: 03-03-2017

If opting for the first type of assessment, the procedure is as follows. First, the school's outcomes are evaluated. This includes various elements (performance on the external evaluations of students' basic competencies, the numbers of students completing each educational stage as well as those staying on into post-compulsory education). It is the responsibility of the Education Inspectorate to determine the extent to which the annual objectives set for each school have been achieved and to evaluate the improvement in its outcomes, based on an analysis of the indicators of educational performance. This assessment takes into account the self-evaluation completed by the school itself, in combination with the socio-economic context in which the school works. The Inspectorate ranks the schools according to four levels of achievement and outcomes. Second, any teacher working in a school awarded a level three or four on this assessment may request that their role in achieving this be evaluated. To do so, a commission (made up of the school inspector and members of the school staff) has to be set up. Third, this commission has the task of determining the degree of involvement of the school's teachers in achieving the objectives set for the school and in improving its outcomes (or maintaining its good results) and of awarding a score (zero, half a point or one point) to each teacher that asked to be assessed. This score is added to the total number of points that a teacher can certify in seeking entry to the new pay scale.

If opting for the second type –the voluntary assessment of the individual's teaching– the evaluation is carried out essentially by the Education Inspectorate. However, the school's management team also participates in the assessment by writing a report on the teacher's participation in the school's activities and about his or her involvement in the implementation of the school's educational project and general curriculum. To safeguard the teacher's rights and the transparency of the procedure, prior to the evaluation process, the criteria for each of the elements and descriptors on which the evaluation of the teaching performance are to be based are made public. They comprise the following:

- Planning of teaching activities (representing 10% of the overall evaluation).
- Implementation of teaching activities, including teaching and learning tasks (25%).

- Student evaluation, including the definition and application of the evaluation criteria adapted to the students, their academic performance in the areas of the syllabus or in the subjects taught by that teacher, as well as the adoption of measures that favour the improvement of student outcomes (25%).
- Classroom management, including being able to create a working environment that is conducive to coexistence and learning (20%).
- Participation in the annual implementation of the school's educational project (20%).

To complete the assessment, the inspector has access to all relevant information, including the materials used by the teacher in the implementation of his or her teaching. The inspector also interviews the teacher and school principal and, where deemed appropriate, any other persons responsible for the management and coordination of the school. In addition, the inspector has to carry out a direct classroom observation, that is, of any activity performed by the teacher in the school for and in the presence of students. The Inspectorate issues a report with an overall assessment that awards the teacher zero, half a point or one point.

Employing the same criteria as those adopted in Table 1, the Catalan program constitutes both an individual and group (school-wide) system, with a complex index (given that it is difficult for a teacher to know *a priori* their possibilities of being awarded the bonus payment), employing what is exclusively an output indicator (i.e. in the case of the school-wide system) and a combination of input and output indicators (although there is a predominance of inputs) in the case of the individual system. Likewise, a comparative assessment of outcomes in the school-wide system and a non-comparative assessment in the individual system are undertaken.

Conclusions

The empirical evidence as it stands does not allow any definitive conclusions to be drawn regarding the effectiveness of pay-per-performance programs, with some studies showing that merit pay systems fail to improve student outcomes and others reporting some positive outcomes. Likewise, no conclusions can be drawn as to which pay-per-performance mechanisms are best, although among the systems reported as being successful, we find

Received: 29-04-2016 Accepted: 03-03-2017

a predominance of simple indexes, incentives paid at the individual level and the inclusion (exclusively or otherwise) of output measures. Moreover, the research conducted to date does not always enable us to verify if the increase in outcomes are real improvements in learning and not just a reflection of teachers 'teaching to the test', or even the result of practices that see teachers *modifying* test outcomes so as to be able to obtain their incentive pay (Podgursky and Springer, 2007). Thus, the only thing we can affirm is that the success of a pay-per-performance program is not guaranteed. This does not mean that it cannot be achieved, but if a program is introduced it has to be very carefully monitored and the necessary changes made to ensure that it is effective.

This school year, in the autonomous community of Catalonia, teachers were given the possibility of asking to be assessed in order to be promoted to a higher pay scale. This assessment can be conducted by taking into account the role played by the teacher in the students' outcomes at the school level (a clear example of an output indicator) and/or by evaluating the teacher's individual performance (employing some output indicators, but dependent above all on input measures).

The Catalan incentive system would appear to have made the right decision in opting to include indicators related to teacher assessment, especially, if we bear in mind that 36% of Spanish teachers claim never to have been assessed, compared to a mean of just 9% in the OECD countries (see OECD, 2014). Likewise, it can also be considered a positive step to include both input and output indicators combined with individual and school-wide incentives – with the individual inputs a teacher can receive an incentive for having undertaken some specific action, while the collective output is linked to the overall outcomes of the school's students. Finally, the fact that the bonus is permanent, and not a one-off payment, serves to stimulate teacher involvement. As such, the Catalan system of incentives could be usefully extended across the whole of the Spanish state (with suitable adaptations to the specific characteristics of each of the autonomous communities).

However, such a system of incentives for the rest of Spain would need to take into account a series of additional elements. First, its application requires a maximum degree of transparency so as to avoid generating feelings of mistrust among teachers and education authorities. In this regard, the index needs to be as transparent as possible to ensure that teachers know which actions will be rewarded, and the work of the

education authority inspectors needs to be extremely clear to all schools. Second, the system would need to undergo a re-evaluation each year so that constant improvements might be introduced. In this regard, it is worth noting that a number of successful programs in other countries (including England) have made such modifications over time until an efficient model was achieved. Finally, teacher assessment has to become a mandatory (although not the sole) factor for climbing to a higher pay scale (*sexenio*). This means that this incentive should not be linked solely with being (a teacher) but also with doing (the job well).

Obviously, teacher motivation cannot be increased by means of wage incentives alone. Other complementary policies need to be introduced, including training for future teachers at university, the selection on merit of elementary and secondary teachers by the schools, non-monetary benefits and a greater social prestige for the teaching profession. But it is essential that teacher assessments be introduced as a common tool for the awarding of certain bonus payments. Such assessments can also be beneficial for students, not only in the short but also in the long term, by boosting their opportunities in the labour market (Lavy, 2015).

In short, the debate turns on the following questions: Should teachers be assessed? Should the results of this assessment generate benefits (including increased pay, among others) for the teacher's professional career? And should the assessment be based on pay-per-performance systems linked to student outcomes? My answer is yes to assessment (the first question) and yes to the fact that the assessment should have consequences (the second question). As for the third question, student outcomes should be taken into consideration in teacher assessments so as to make it quite clear that improving efficiency is a major objective of the system. However, student performance should only represent a small part of the mechanism, especially when first introducing the assessment system, so that teachers can recognize that the incentive depends to a large extent on what they do and because of the methodological difficulties associated with this mechanism.

Acknowledgments

The author gratefully acknowledges funding received from IVÀLUA (Catalan Institute for Public Policy Assessment - *Institut Català*

d'Avaluació de Polítiques Públiques) which has enabled this project to be carried out. He also wishes to thank analyst, Miquel Ángel Alegre, for his support.

References

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H. y Wilson, D. (2009). Evaluating the impact of performance-related pay for teachers in England. *Labour Economics*, 16, 251–261.

Balch, R. y Springer, M. G. (2015). Performance pay, test scores, and student learning objectives. *Economics of Education Review*, 44, 114-125.

Banco Mundial (2015). *Impact Evaluations in Education*. Washigton: The World Bank. Recuperado de http://datatopics.worldbank.org/EdstatsApps/Edu%20Evaluation/qEvaluation.aspx

Behrman, J. R., Parker, S. W., Todd, P. E. y Wolpin, K. I. (2011). Aligning learning incentives of students and teachers: Results from a social experiment in Mexican high schools. *Working Paper Penn Institute of Economic Research (PIER) n. 13-004*. Recuperado de http://papers.ssrn.com/sol3/JELJOUR_Results.cfm?form_name=journalBrowse&journal_id=265679&Network=no&SortOrder=ab_title&stype=asc

Berry, J. y Eckert (2012). Creating teacher incentives for school excellence and equity. *National Education Policy Center,* n. January, Recuperado de http://nepc.colorado.edu/publication/creating-teacher-incentives

Contreras, D. y Rau, T. (2012). Tournament incentives for teachers: Evidence from a scaled-up intervention in Chile. *Economic Development and Cultural Change*, 61(1), 219-246.

Dee, T. y Wyckoff, J. (2013). Incentives, selection, and teacher performance: Evidence from IMPACT. *NBER Working Paper 19529*. Recuperado de http://www.appam.org/assets/1/7/Incentives,_Selection,_And_Teacher_Performance_Evidence_From_Impact.pdf

Fryer, R. G. Jr., Levitt, S. D., List, L. y Sadoff, S. (2012). Enhancing the efficacy of teacher incentives through loss aversion: A field experiment. NBER Working Papers 18237. Recuperado de http://www.nber.org/papers/w18237

Glewwe, P., Nauman, I. y Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, 2(3), 205-27.

Goldhaber, D. y Walch, J. (2012). Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative. *Economics of Education Review*, 31(6), 1067–1083.

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

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

Lavy, V. (2009). Performance pay and teachers' effort, productivity and grading ethics. *American Economic Review*, 99(5), 1979-2011

Lavy, V. (2015). Teachers' pay for performance in the long-run: Effects on students' educational and labor market outcomes in adulthood. *NBER Working Papers 20983*. Recuperado de http://www.nber.org/papers/w20983.pdf

Martins, P. S. (2009). Individual teacher incentives, Student achievement and grade inflation. *IZA Discussion Paper Series No. 4051*. Recuperado de http://ftp.iza.org/dp4051.pdf

Muralidharan, K. y Sundararaman, V. (2011). Teaching performing pay: Experimental evidence from India. *Journal of Political Economy*, 119(1), 39-77.

Neal, D. (2011). "The design of performance pay in education". *NBER Working Papers 16710*, Recuperado de http://www.nber.org/papers/w16710.pdf

OECD (2014). *TALIS 2013 Results: An International Perspective on Teaching and Learning*. París: Organisation for Economic Co-operation and Development.

Podgursky, M. y Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, 26(4), 909-950.

Schlotter, M., Schwerdt, G. y Woessmann, L. (2010). Econometric methods for causal evaluation of education policies and practices: A non-technical guide. *IZA Discussion Paper 4275*. Recuperado de http://ftp.iza.org/dp4275.pdf

Received: 29-04-2016 Accepted: 03-03-2017

Vigdor, J. L. (2008). *Teacher salary bonuses in North Carolina*. Nashville, Tn: National Center on Performance Incentives. Vanderbilt University Peabody College.
https://my.vanderbilt.edu/performanceincentives/files/2012/10/200803_Vigdor_TeacherBonusesNC.pdf
Winters, M., Greene. J., Ritter, G. y Marsh, R. (2008). *The effect of performance-pay in Little Rock, Arkansas on Student achievement*. Nashville, Tn: National Center on Performance Incentives. Vanderbilt University Peabody College.
https://my.vanderbilt.edu/performanceincentives/files/2012/10/200802_WintersEtAl_PerfPayLittleRock.pdf
Woessmann, L. (2011). Cross-country evidence on teacher performance pay. *Economics of Education Review*, 30, 404-418.

Contact address: Universitat de Barcelona. Vicerrectorado de Economía. Edificio Histórico, 1ª planta. Gran Vía Corts Catalanes, 585. 08007. Barcelona. E-mail: oescardibul@ub.edu