



What Works in Education?

Using Evidence to Improve Education

1

July 2015

What works in education?: The question that has to be asked

Miquel Àngel Alegre Canosa

Is the introduction of pay-for-performance salary incentives for teachers linked to students' academic performance advisable?

J. Oriol Escardíbul

“For too long, education has been based on inertia and tradition, and changes in educational intuitions or beliefs were unfounded. The “what works” movement enters into the world of education with a clear objective: to promote evidence-based educational policies and practices. Ivàlua and the Jaume Bofill Foundation join forces to promote the movement in this country.



What Works in Education?

Using Evidence to Improve Education

What works in education? The question that has to be asked



Miquel Àngel Alegre Canosa

Analyst at the Catalan Institute of Public Policy Evaluation (Ivàlua). Expert in the evaluation of education policies.

In most societies, the educational agenda is often driven by political ideology, conventional wisdom, folklore and wishful thinking as it strives to meet the needs and interests of the economy, business, employers, law and order, civil society, parental choice, and at least rhetorically, the children, young people and adults who make up the learning community [...]. Much of this impetus represents the triumph of hope over reason, sentiment over demonstrated effectiveness, intuition over evidence.

Davies, "What is Evidence-Based Education?" June, 1999 [1].

These are good times for evidence-based reform in education. Due in particular to Investing in Innovation (i3) and the Institute of Education Sciences (IES), the number of proven programs in all subjects and grade levels is increasing. [...] [This] creates new opportunities for policy and practice in education. Already, School Improvement Grants (SIG) are newly introducing an option for schools to choose proven, comprehensive reform models. Other areas of policy may also soon begin to encourage or incentivize use of programs with strong evidence.

Slavin, "Are Proven Educational Innovations Ready for Prime Time?",
In October 2014 [2].

In the first quotation, Philip Davies refers to the overall situation of education in developed countries (including the US); in the second fragment, Robert E. Slavin shines the spotlight on the United States. Fifteen years separate one remark from the other.

And yet we dare to say that the current situation in education policy here in Catalonia is better described by Davies than by Slavin. Indeed, in Catalonia, as well as in Spain and many other countries around us, educational policy decisions concerning the launch, maintenance, reform, or abolition of policy, are rarely based on solid empirical evidence regarding their effectiveness. In fact, it is quite common that such decisions are not based on any kind of sound empirical evidence at all.

Is it constructive to reduce the years of comprehensiveness of the educational system? Is it a good idea to expand the range of public education from 0 to 3 years? Is it effective to allow families to choose freely the school they want for their children? Is it negative to extend the maximum ratio of students per class? It is good to increase the number of hours students spend in school each day? Does the continuous school day work in secondary education? Does it work in primary education? Does operating a means-tested scholarship program for continuous post-compulsory education work? And, one based on merits? Do economic incentives for teachers work? What about school rankings? And, all together, why do they work? Or, for whom do they work?

Many of these questions have been discussed and debated in a wide range of forums. At the same time, they all tie in with aspects faced by education policymakers in recent years. However, we cannot say that the focus of public debate (more or less specialized) or the approach of the reforms in question have taken sufficient account of the accumulated empirical evidence regarding the effectiveness of the different options.

A similar situation occurs when what is at stake are programs, educational activities or practices of a more “micro” nature (teaching/pedagogical decisions, courses, organization of schools, etc.). Does using ICT in teaching work? Does grouping by levels work? And, what about flexible grouping? Cooperative learning? Project work? Intensive one-to-one tutoring? Is homework bad? Repeating a year? Are summer programmes beneficial? And the family-school programs? Here too, we must ask ourselves which groups and objectives to use when evaluating these initiatives?

In Catalonia, as well as in Spain and many other countries around us, educational policy decisions concerning the implementation, maintenance, reform, or abolition of policy, are rarely based on solid empirical evidence regarding their effectiveness.



Once again, the debates and practical decisions formulated around these issues have rarely been supported by any serious collection or analysis of evidence, exercises capable of providing information concerning the impact of these and other initiatives.

By this we do not mean to say that there is a lack of available knowledge in other countries and in different educational contexts about the effectiveness demonstrated by education policies of a more “structural” nature, or more “micro” type initiatives. On the contrary, very often the source of the problem is to be found elsewhere, and is usually associated with other kinds of difficulty: how do we identify a more robust insight into a particular subject? Where can we find the time to address and review it? How do we translate this knowledge and apply the benefits thereof in our immediate context? And, above all, how can it be suitably applied on the playing field, (that of program and policy development) where other factors such as political interest or commitments, stakeholders, institutional routine, certain cultural values, media pressure, etc., tend to play an important role?

Ivàlua and the Jaume Bofill Foundation wish to contribute their grain of sand in facing up to these challenges. Therefore, our proposed design for the publication of *What Works in Education?* is to produce a body of work capable of furnishing the debate and educational practice with scientific evidence and to do so in such a way that it has the potential to combine precision, dissemination and connection with current education challenges.



In the following publication, Ivàlua and the Jaume Bofill Foundation wish to contribute their grain of sand in facing up to these challenges. Therefore, our proposed design for the publication of ***What Works in Education?*** is to produce a body of work capable of furnishing the debate and educational practice with scientific evidence and to do so in a such a way that it has the potential to combine precision, dissemination and connection with the education challenges currently being faced in Catalonia.

Some context: “what works” in education

What works in education? This is the underlying question guiding this publication, and which, with each new issue in the series, we will answer, placing the focus on areas of particular relevance for education intervention. The publication therefore, serves as a link to another movement, generically referred to as the “what works” movement, whose reason for being is based on promoting evidence-based policy making.

It is beyond the scope of this introduction to provide an in-depth analysis of the origin, meaning and development of the entire “what works” movement, a movement that began within the health sector at the beginning of the 1970s [3]. As regards the field of education, it might be argued that the “what works” movement is the latest development in “school effectiveness research” (SER). This line of research began in the early eighties in reaction to literature that, basing its arguments on the conclusions presented in the well-known Coleman report (1966) [4] placed the socio-economic and ethnic composition of schools as the leading factor capable of conditioning students’ academic performance, far removed from the impact capacity of resources available to schools and how these resources are allocated. SER studies, on the other hand, seek to demonstrate that schools, how they work and are organized, can help student performance gains, including reducing academic inequality between one centre and another [5].

The “what works” approach can be interpreted as being derived from this school of research, with which it shares the objective to analyse the effectiveness of education intervention and the ultimate goal of providing empirical evidence for the decision-making process. Yet both perspectives differ from one another with respect to one central element: the type of evidence they tend to produce, or aspire to produce. SER studies are based on observational data and the use of multivariate statistical techniques. They provide evidence of the degree of association between independent variables (interventions or specific schools traits) and dependent variables (i.e. students or schools’ academic performance). [6] The “what works” approach however, aims to demonstrate causal relationships between intervention and the proposed outcomes. Moreover, it is difficult to infer causality without engaging in experimental assessment (randomized assignment between treatment group and control group) and, under certain circumstances, quasi-experimental studies¹.

It must be said that, traditionally, the field of education policy has been quite unresponsive to the developments and claims put forward by the “what works” perspective. Recently, especially in the English-speaking world, this has crystallized into what some have called a “quiet revolution” in the relationship between education and evidence. [7] The references shown in [Table 1](#) testify to some of these advances.

The “what works” movement’s reason for being is based on promoting evidence-based policy making.



We would like to highlight two especially paradigmatic initiatives, both examples of institutionalization of the “what works” perspective in the field of education policy. The first example can be found in the [What Works Clearinghouse](#) (WWC), an initiative promoted in the US by the federal government’s Institute of Education Sciences. The initiative was set up in 2002, in association with the No Child Left Behind program. This program emphasised the need for publicly funded education activities to incorporate, while at the same time generate, evidence of effectiveness. In this way, the WWC review existing assessments of a comprehensive set of education interventions, assess the methodological quality of these studies applying extremely stringent standards², and on this basis, establish the degree of effectiveness of the programs in question. All this information is accessible to policy makers, researchers and members of the educational community to promote evidence-based decisions and practices.

A second example can be found in the UK, with the establishment of the [Education Endowment Foundation](#) (EEF) in 2011. This foundation, recognized by the British government as the “what works” centre specializing in the field of education³,

- 1 We cannot overlook the fact that this new focus has been largely triggered by the entry of financial discipline (in particular, econometrics) into the analysis and evaluation of the effectiveness of educational policies areas until recently mainly dominated by science education.
- 2 Only studies of effectiveness based on experimental designs have the option of receiving the highest rating of WWC (“WWC Meets the Standards Without Reservations”).
- 3 The Education Endowment Foundation as part of the so called What Works Network, the UK Government’s Cabinet Office initiative designed to promote the use of evidence based decision making within about public services. In addition to the EEF, the ‘what works’ network is made up of five centres specialized respectively in, health and social protection, aging, local development, child care and crime.

focuses its mission on producing evidence regarding what works (and does not) in the fight against social inequalities in learning. One of the principal tools developed by the EEF is the so-called Teaching and Learning Toolkit (for primary and secondary education) and the Early Years Toolkit (for pre-school education). These platforms summarize and qualify current information regarding the effectiveness of different educational courses of action, drawing on evidence provided by studies and systematic reviews of particular relevance. This enables the EEF, as well as the WWC, to function as an interface or mediation between the academic world and the world of political decision-making, allowing the former to adapt its specific tempo and language to the need for immediacy and clarity of the latter.

In Catalonia, the field of educational policy is still a long way from what these developments represent and, as has already been mentioned, we believe it is worth trying to reverse this situation.

Institutions such as the What Works Clearinghouse (USA) and the Education Endowment Foundation (UK) work as an interface or mediation between the academic world and the world of politics and educational practice.



Table 1.
Principal repositories and periodicals having an impact on education

“What works” in education repositories	Institution	“What works in education” periodicals
What Works Clearinghouse http://ies.ed.gov/ncee/wwc/	Institute of Education Sciences, U.S. Department of Education, Estats Units	Practice Guides / Intervention Reports / Single Study Reviews / Quick Reviews http://ies.ed.gov/ncee/wwc/publications_reviews.aspx
Teaching and Learning Toolkit http://educationendowmentfoundation.org.uk/toolkit/	Education Endowment Foundation, Regne Unit	(Material available online on the website of the institution: range of different formats)
Best Evidence Encyclopedia (US) http://www.bestevidence.org/	Center for Data-Driven Reform in Education, Johns Hopkins University, Estats Units	Better: Evidence-based Education Magazine http://www.betterevidence.org/ Best Evidence in Brief http://education.jhu.edu/research/crre/newsletter.html
Best Evidence Encyclopedia (UK) http://www.bestevidence.org.uk/	Institute for Effective Education, University of York, Regne Unit	
Evidence 4 Impact, http://www.evidence4impact.org.uk/index.php		
Impact Evaluations in Education http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E	The World Bank	(Material available online on the institution website: range of different formats)
Social Programs That Work (topics: Early Childhood, Education K-12, Postsecondary Education) http://evidencebasedprograms.org/	Coalition for Evidence-Based Policy	(Material available online on the institution website: range of different formats)

Source: Author's own elaboration

What we collect: solid evidence on impact

It is clear that the question “what works?” gives rise to concerns about the effectiveness of policies and their impact on the reality they are designed to improve. We wonder, therefore, what exactly we know about the ability of educational policies and programs to make significant impacts on key aspects of the reality of education today. And what do we mean by impact? Well, exactly that: the changes to the situation which have been subject to intervention and are exclusively attributable to the intervention in question.

We are aware that identifying the impact of a program (educational or otherwise) is not as easy task. On the one hand, the aspect affected by the proposed change (for example, the school dropout rate) is subject to influence from many factors (political, social, labour, economic...), which exist concurrently with the programs. So, what happens to participants after program implementation is not necessarily an effect of the intervention itself.

On the other hand, an evaluation of the impact cannot also be based on the simple and direct comparison of the outcomes from participants in the program with those from non-participants. There is a strong possibility that participants and non-participants (be they schools, teachers or students) already possessed specific differential traits which might influence outcome in the area of interest. These differences are the result of what is known as “selection bias” in access to programs: either participants are actively chosen on the basis of certain objective requirements and/or biased assessment, or it is the participants who select themselves for the program (these might be the most highly motivated subjects or the most academically prepared). Therefore, a direct comparison between the status of participants and non-participants often results in a biased evaluation of the impact of the intervention.

So the in response to the question, “what works?”, there is no straightforward answer: only those based on solid methodology, i.e., those which establish a solid causal relationship between policy and changes in the issues addressed, thereby overcoming the two obstacles mentioned above -the issue of contemporary factors and selection bias. This premise restricts the focus to the evidence of impact obtained through experimental or quasi-experimental methods.

The question “what works?” gives rise to concerns about the effectiveness of policies and their impact on the reality they are designed to improve and focuses on the evidence of impact obtained through experimental or quasi-experimental methods.



In short, the goal of these methods is to identify a control group (non-participants in the program or non-beneficiaries of the education intervention) as similar as possible to the group of participants. Ideally, participants and non-participants should present identical sets of the characteristics which might be related to the probability of program success. This would allow us to state that any differences we may observe, once the intervention has been completed between one group and the other, represents its impact.

The most effective strategy is to randomize (a draw or raffle) to decide who, from the entire list of eligible candidates (schools, students, teachers...), would receive the program and who should be assigned to the control group. This procedure is the cornerstone of experimental design in assessment, a design which some authors have described as the “gold standard” when it comes to providing evidence of what works. [8] When applying this design is not a viable option, for one reason or another, it is then that specific assessment designs come into play.

We cannot dwell here on the explanation of the characteristics and applications of experimental and quasi-experimental methods in assessing the impact of education policies. This explanation is addressed in-depth in the guide, “How to assess the impact of education policies” (from the Ivàlua practical guidelines for evaluating public policy collection). [9]

Protocol for collecting and analysing evidence

Either way, the function of the article included in this publication is not that of directly assessing the program impact. Rather, its purpose is to collect and analyse what the data from other studies using rigorous methods have to say about their ability to make an impact. There are several methods which can be used when the objective is to produce a balance sheet of accumulated knowledge associated with a particular matter or issue. In addition, we have to bear in mind that the level of thoroughness and systematization of the aforementioned knowledge can be quite disparate.

Table 2 summarizes the main characteristics of the recognized methods used in collecting and reviewing evidence: from methods which utilise a minimum of systematisation and protocol (ad hoc review of the literature) to the most comprehensive methods (multiple systematic review).

The purpose of the articles in this publication is to gather and review what other studies using rigorous methods, have to say about the impact of education policies and practices. The “review of reviews” procedure is given priority as a strategy for gathering and reviewing evidence.



Table 2.
Standard protocol in evidence gathering and review

Review procedure	Nature	Definition	Limitations	Duration
Ad hoc literature review	Not systematic	Review of the most relevant/ illustrative studies Does not explain bibliography research process	High risk of selection bias: available studies or those with positive outcome Suitable in the context of limited resources	1 week to 2 months
Quick scoping review	Not systematic	Maps out the available literature on a subject	Responds to very descriptive questions Research from a very limited selection of databases Searches using few key words	1 week to 2 months
Review of reviews	Systematic	Prepares a synthesis of reviews conducted –rather than primary materials	Can be used only when there is sufficient accumulated and reviewed evidence Quality of reviews may be disparate	2 to 4 months
Rapid evidence assessment	Systematic	Restricted research into the effectiveness of a policy Has a search protocol Can be traced and replicated	Necessary to define a very specific question regarding effectiveness Timeframe limited and recent	2 to 6 months
Full systematic review	Systematic	Extensive review of literature on the effectiveness of a policy Uses systematic protocol Searches academic databases and gray literature	Document gathering process may be quite traditional Requires some time May require additional human resources -internal or external	8 to 12 months
Multi-arm systematic review	Systematic	Extensive review of the literature to address different issues regarding policy effectiveness Ibid, “full systematic review”	Increases limitations of the “full systematic review”	More than 12 months

Source: Adapted from <http://www.civilservice.gov.uk/networks/gsr/resources-and-guidance/rapid-evidence-assessment/what-is>

In general, the articles included in the *What Works in Education?* series give priority to the “review of reviews” process as a strategy for gathering and reviewing evidence. This undertaking, justified by adherence to standards of accuracy, coverage and agility to which we aspire, do not exclude the possibility of including references to primary studies or assessments which may be of particular relevance (thematic focus, methodology used, proximity to Catalan geographical or political context, etc.). Resorting to the reviewed primary studies will be especially necessary in the case of contributions addressing areas of intervention for which there is not sufficient hard evidence as yet accumulated and reviewed.

The outcomes we focus on

As could not be otherwise, concern about what works has been directed towards multiple fields and outcomes, (see [Table 3](#)): cognitive outcomes (performance competence), and non-cognitive (social, emotional and meta-regulative skills) education achievement (qualification attained) educational transitions (e.g. pathways beyond compulsory education), employment outcomes (insertion and working conditions) or even in the field of health (physical and mental), and safety (criminal practices). In other words, the evidence of effectiveness of one or another program may be reviewed according to multivariable interests.

Table 3
Outcomes usual impact assessment of educational policies

Scope	Outcomes	Assessment tools
Cognitive	Performance in key academic skills (language, mathematics, science, etc.)	School assessments (quantitative and qualitative) Standardized testing (internal or external) Ad hoc skills testing (program-specific)
Non-cognitive	Social and emotional development Self-confidence and personal autonomy School attitudes, resources and expectations	School assessment (quantitative and qualitative) Standardized questionnaires (approved metric scales) Ad hoc questionnaires (program-specific)
Achievements	Level of education attained Graduation/repeating at key stages	Administrative records (education) Longitudinal panel data
Transitions	Itineraries adhered to during stages of education	Administrative records (education /work) Longitudinal panel data
Labour	Job placement (speed of access to employment) Working conditions (stability and wages) Job suitability (in accordance with qualification and type of training)	Administrative records (work) Longitudinal panel data Ad hoc questionnaires (program-specific)
Health and safety	State of physical and mental health Relationship to criminal or risk practices	Administrative records (social and health) Longitudinal panel data Ad hoc questionnaires (program-specific)

Source: Author's own elaboration

In fact, the different evidence repositories shown in [Table 1](#) generate alternative interpretation:

1. *From outcome to program.*

Effective and non-effective programs are identified according to the selected outcome. This system is currently that used by the US Education Sciences Institute's What Works Clearinghouse.

2. *From program type to outcome of interest.*

Once a set of areas and intervention types have been identified, their effectiveness is synthesized based on a series of outcomes. For example, Education Endowment Foundation (UK) toolkits reviewed the impact capacity of different lines of action, in pre-school as well as in primary and secondary schools education on the performance of students who were academically and socially challenged.

The approach we have chosen in this publication shares the blueprint and focus of the EEF. Thus, additional articles in the series will be grounded in a thematic starting point, based on the type of intervention, and will then question their ability to effect a positive impact on the students' educational advancement and opportunities, particularly the most vulnerable. Obviously, the scope and specific characteristics of each area of intervention being dealt with will determine the final profile of this focus; that is, whether concern is focused on one or another type of vulnerability, one learning stage or another, or one school context or another.

The articles in the series will question the ability of interventions to impact positively on progress and educational opportunities for students, especially the most vulnerable.



We begin by reviewing teacher-incentive programs

The definition of the salary conditions and financial incentives for teachers is central to teacher policy, along with other areas such as health care, selection processes and access to teaching positions, working conditions and professional careers, initial and ongoing training, teacher assessment tools, support for professionalisation, and centre leadership, etc. In Catalonia, many of these areas have been subject to interventions and reforms in recent years.

Regarding economic incentives, in recent years the government has introduced reforms in the definition of teaching merits, which to a certain degree represent a tentative shift towards the introducing some class of performance-based payment rationale. This is, for example, the case of the latest measures introduced in connection with the granting of the promotion (and the associated salary supplements) ([Agreement GOV/29/2012](#)); in particular, the consideration of a teacher's implication in increasing student performance gains for the centre as a basic criteria meriting professional advancement.

In any event, we are still far from the level of development some western countries have achieved in terms of financial tools and incentives. One such paradigmatic case is

that of the United States, not only due to the huge diversity of teacher-incentive schemes in operation, but also the amount of empirical evidence that has been generated (often experimental evidence) regarding the effectiveness of these schemes. We will soon have the opportunity to test this out.

Here in Catalonia, the debate around the virtues and evils of financial incentive systems for teachers has been markedly ideological, heavily weighed down by resistance and preconceived ideas, a debate which has given little space for generating and benefitting from the wealth of scientific knowledge available on this issue.

We are aware that we are dealing with a very sensitive topic. And we share many of misgivings which are more than likely to raise their heads with the approach of certain proposed financial incentives: undermining intrinsic motivation of teachers (teaching vocation), introduction of competitive activity to what are in essence collaborative environments, encouraging teaching-to-test practices, etc. However, we believe it is worthwhile to discover whether different incentive schemes have worked in other contexts or not, and if so, to what extent. In other words, whether or not they have been able to affect any significant impact on students' academic outcomes. We believe that it is a worthwhile step to learn from the accumulated knowledge, so that if we ever do have the opportunity to expand or curtail progress towards one financial incentive formula or another, we may do so considering, among other things, what we know for sure about their effectiveness (or ineffectiveness).

So how effective are the systems of financial incentives for teachers? Do they work or not? Moreover, what pathway or expectation of effectiveness might they adopt in a social and educational context such as that in Catalonia? These are the questions posed by Oriol Escardíbul in the following article, and to which he responds by avoiding preconceptions and prejudices and relying mainly on what the empirical evidence tells us.

We want to know if different incentive schemes have worked in other contexts or not, and to what extent they have.



References

- [1] Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47(2), 108–121. http://www.jstor.org/stable/3122195?seq=1#page_scan_tab_contents
- [2] Slavin, R. E. (2014). Are Proven Educational Innovations Ready for Prime Time? [The Huffington Post]. http://www.huffingtonpost.com/robert-e-slavin/are-proven-educational-in_b_6074382.html
- [3] Cochrane, A. L. (1972). *Effectiveness and Efficiency: Random Reflections on Health Services*. Nuffield Provincial Hospitals Trust. http://www.nuffieldtrust.org.uk/sites/files/nuffield/publication/Effectiveness_and_Efficiency.pdf
- [4] Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of Educational Opportunity*. Washington, D.C.: US Congressional Printing Office. <http://files.eric.ed.gov/fulltext/ED012275.pdf>
- [5] Teddlie, C., & Reynolds, D. (2000). *The International Handbook of School Effectiveness Research*. Londres: Falmer Press.
- [6] Luyten, H., Visscher, A., & Witziers, B. (2005). School Effectiveness Research: From a review of the criticism to recommendations for further development. *School Effectiveness and School Improvement*, 16(3), 249-279. <http://www.tandfonline.com/doi/abs/10.1080/09243450500114884?journalCode=nse20>
- [7] Collins, K. (2013). "Show me the evidence" - *A quiet revolution - the activity and impact of the EEF*. Institute of Education, Londres.
- [8] Duflo, E., Glennerster, R., & Kremer, M. (2007). Using randomization in development economics research: A toolkit. A.T.P. Schultz i J.A. Strauss (Eds.) *Handbook of Development Economics – Vol. 4* (pp. 3895-3952). Amsterdam: North-Holland. <http://economics.mit.edu/files/806>
- [9] Alegre, M.A. (2015). *How to assess the impact of educational policies*. Ivàlua collection of practical guides on evaluation of public policies, no. 11. Barcelona: Catalan Institute of Public Policy Evaluation. <http://www.ivalua.cat/generic/static.aspx?ID=669&>

Is the introduction of pay-for-performance salary incentives for teachers linked to students' academic performance advisable?



J. Oriol Escardíbul

Professor of Applied Economics at the University of Barcelona and researcher at the Institute of Economics of Barcelona (IEB) and the Group of Interdisciplinary Educational Policy (GIPE).

Motivation

Increasing the academic performance of students is a feature on government agendas far and wide. Empirical evidence suggests that it is better to invest in the quality of teachers rather than the quantity [1] [2]. In this sense, one of the areas which has been the focus of attention in recent times, particularly in the United States, has been the establishment of salary incentive programs, called merit pay or pay-for-performance (PFP). PFP consists in assessing the performance of teachers and offering a financial incentive for reaching certain standards. In these evaluations an increasingly greater weight is given to the analysis of teacher performance linked to student output in external testing given that the objective is to reward initiatives that lead to gains in student achievement.

Approximately twenty developed countries as well as more than half of US states operate incentive pay based on teacher assessment, although not all assign the same level of weighting to student performance. Catalonia has recently passed legislation to introduce PFP, which includes partial consideration of student performance as an assessment factor.

The number of studies is insufficient and heavily focused on the United States but conclude that, although PFP systems can improve student performance, success is not guaranteed.

Questions influencing the review

This review evaluates existing pay-for-performance incentive programs. These can be linked to inputs or outputs of the education production function. Among the former taken into consideration are the tasks performed by teachers as part of their teaching role, among others. The latter includes the assessment of teachers' contribution on the basis of student performance, usually measured in terms of performance gains or gains in external testing. In this analysis we have mainly considered indicators based on outputs, given that these have recently been the focus of education authorities and are more controversial. The review sets out to address four basic questions: a) What impact has PFP methodology on student outcomes? b) What is its impact on the assignment of teachers to schools? c) How do teachers rate the use of these methods? d) What applicability might an incentive system have in Catalonia?

Review of the evidence

Purpose of the review

The object of this review is to ascertain whether the introduction of a system of financial incentives for teachers based on the evaluation of their performance, improves students' academic outcome. A sophisticated meta-analysis system was not developed given that the studies reviewed include very varied performance pay systems (PFP), as well as different variables related to the definition of the gains. Moreover, the number of studies is insufficient. However, the review did succeed in formulating certain conclusions. Specific program assessments as well as reviews of the programs themselves were analysed.¹ Before evaluating the assessments, the varying PFP formats, which are very diverse, are explained:

The object of this review is to ascertain whether the introduction of a system of financial incentives for teachers based on the evaluation of their performance, improves student performance.



- Performance index based on inputs, outputs, or a mixture of both.
- In the case of outputs, one can consider the students' performance or the performance gains. In the latter case, two types of methods are usually adopted. The former compares the performance of students in external tests with the anticipated performance, which is estimated taking into account the outcomes of previous

¹ Several institutions review single PFP studies, but almost no meta-analysis, but rather assessments of a methodological standard of the research examined. This document prefers referencing the online repository where these institutions deposit the assessments mentioned above but have reviewed also the primary studies (originals). The following repositories include PFP documentation:

- What Works Clearinghouse (WWC), the Institute of Education Sciences, the Department of Education of the Government of the United States. <http://ies.ed.gov/ncee/wwc/>
- Teaching and Learning Toolkit, the Education Endowment Foundation in the UK. <https://educationendowmentfoundation.org.uk/toolkit/>
- Impact Evaluations in Education, World Bank. <http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E>

tests and the students' socioeconomic situation. In this way, gains made by a student are compared with the expected performance in accordance with their circumstances and prior academic history; the difference or "value-added" is assigned to the teacher. The latter takes into account the gains in outcomes for each student compared to other students (or only the more similar students), so that gains are analysed in comparative terms. An average gain derived from the relative gains in outcomes for each of their students is assigned to each teacher.

- Individual, (incentive awarded to each teacher based on the performance of their students), or group. The second option bestows an award on a group of teachers in a course, subject or to the entire education centre if they reach the objectives. The group (or school) may then decide how to distribute the award among the teachers.
- Once-off award or one which becomes a permanent salary increase.
- Predetermined amount or a quantity which varies depending on the outcomes obtained.
- Open (any teacher can gain the award if they achieve the objectives) or closed (like a competition or tournament) where only a few can win.
- By external assessment (performed by an education authority), internal (developed by school management with possible participation of teachers) or mixed.

Firstly, a group of studies is reviewed evaluating PFP programs using optimal analysis methodology (i.e., allowing for causal inference). These are based on experimental designs where random selection takes place among the group in which the assessed policy has been implemented and the control group (or comparison, similar to the previous example but where policy is not implemented). Secondly, a series of studies using quasi-experimental methods is evaluated, where, as in the previous case, there is a treatment group and a control group but not random selection, so as to eliminate selection bias using certain econometric techniques [3]. In each group, firstly the outcomes of assessments for the United States are presented and then those obtained for other countries. (the [table 1](#) below contains the main features of each program reviewed)

Table 1.
Assessments for pay-for-performance programs (PFP) (N = 19)

Program	Incentive level	Simple index	Input/output	Comparative output assessment	Amount (\$)		Results	Standard deviation
					Min.	Max.		
Experimental Designs (N = 7)								
New York (SPBP)	School	No	15% input 85% output (60% value-added)	Yes	1.400	3.600	Not significant	
Chicago Heights	Individual and teacher group	Yes	Output (value-added)	Yes	0	8.000	Only fear of loss is positive	from 0.2 to 0.4
Nashville (POINT)	Individual	Yes	Output (value-added)	Yes	5.000	15.000	Not significant	
Texas (Round Rock)	Teacher group	No	Output (value-added)	No	5.400	5.900	Not significant	
Kenya	School	Yes	Output (points and value-added)	Yes	26	51	Dubious positive	0.13 to 0.22
India	Individual and school	Yes	Output (value-added)	No	2,25	450	Positive. Individual gain	0.17 to 0.27 0.35 to 0.52
Mexico	Individual and group	Yes	Output (value-added)	No	0	2.000	Positive	0.2 to 0.6
Experimental design (N = 12)								
North Carolina	School	Yes	Output (value-added)	No	750	1.500	Positive	0.09 to 0.13
District of Columbia (IMPACT)	Individual	No	Input 45% / 85% Output 55%/15% (value-added)	No	5.000	25.000	Positive	0.24 to 0, 27
Austin (REACH)	Individual, grupo y escuela	No	Input and output (value-added)	No	1.000	3.000	Positive	0.13 to 0.18
Chicago (TAP)	Group	No	50% input 50% output (value-added)	No	1.100	6.400	Not significant	
Denver (ProComp)	Individual and group	Yes	Inputs and output (value-added)	Yes	400	2.500	Positive	0.09 to 0.18
Arkansas, Little Rock	Individual	Yes	Output (value-added)	No	350	7.600	Positive	0.15 to 0.22
Israel (2002)	School	No	Output (test points and other)	No	1.000	2.500	Positive	0.02 to 0.04
Israel (2009)	Individual	Yes	Output (value- added and other)	No	1.750	15.000	Positive	0.04 to 0.09
England	Individual	No	Input and output (value-added)	No	2.000	2.000	Positive	
Mexico (Carrera M)	Individual	No	Input and output (test points)	No	No partially positive	0.03 to 0.15
Portugal	Individual	No	Input and output (test points)	No	Negative	-0.04 to -0.54
Chile (SNED)	School	No	Input and output (points and value-added 65%)	Yes	370	439	Positive	0.14 to 0.25

The impact of economic incentives on students' performance

In the first group, the programs evaluated in the United States are quite different, even though they all associate PFP, either significantly or exclusively, with student outcomes (usually as gain on external standardized testing in math and reading comprehension). The three programs presented below all showed a lack of positive performance. First, the New York City's Schoolwide Performance Bonus Program (SPBP) contains an incentive for underperforming schools in New York City which evaluates student performance in state exams (25% of the total value of the indicator) and progress in outcomes (60%). Assessments are of a comparative nature, in part with schools city-wide but especially with similar centres.

Secondly, the Project on Incentives in Teaching (POINT), in Nashville (Tennessee) for primary and secondary school courses established an individual award for teachers conditional on their students' gain in state exams of higher level mathematics greater than those achieved by other students with similar scores the previous year. The reward depends on the teacher being in the 80%, 90% and 95% percentile of outcome distribution. Thirdly, the District of Round Rock (Texas) established a PFP program, which unlike the previous ones, is aimed at schools with good outcomes.

In this case, the award is given based on the value-added of a team (multidisciplinary) of teachers in different subjects (the group should be placed in the top third to be rewarded). As such, this is a group prize, but a teacher on a winning team might not receive the award if they have a low individual outcome. The value-added is calculated as the difference between the performance of students in the group's external testing as well as expected performance (obtained from an estimate which takes into account the performance in previous years as well as personal and socioeconomic conditions).

By contrast, a program implemented in Chicago Heights, a low-income town in Illinois which operates a PFP system with four classes of incentive for teachers (secondary level) has been a partial success: two individual and two group incentives conditional on students performing well (above average outcomes compared with other schools) in external tests. The program sets out to compare the different behaviour of teachers working to win a reward with those who fear losing one. Thus, for each type of incentive (individual and group), one system rewards participants with up to \$8,000, while the other delivers half of the reward in advance, which may be lost if students fail to reach objectives (or be increased in the event of a positive outcome) - is what is known as a "fear of loss" scheme. The results indicate that only student performance in the "fear of loss" scheme improves and moreover, the effect is greater when the gain method was group oriented [4] [5].

Box 1.

The importance of incentives (I): fear of loss versus desire to win.

During the 2010-2011 academic year, Roland G. Fryer analysed a program of financial incentives for teachers linked to student performance in Chicago Heights, a low-income school district in the state of Illinois. These awards (individual and group) were conditional on student achievement gains (above average) on standardized tests. The Research involved 150 teachers in different schools, involving four treatment groups and one control. Two groups (one with an individual reward system and one group) earned rewards. Two different groups (individual or group rewards) receive advance payment of half the maximum amount (\$4,000) but they can lose this advance (in whole or in part) if their students do not reach the goals set (and in the same way can increase the reward to a maximum of \$8,000). The study then compares the behaviour of the desire to win and the fear of loss.

To compare the performance, a group of ten students was put together with similar outcomes before intervention, where one is a student of the teacher evaluated and the rest are from other schools. The position of each student within their group is calculated, on the basis of annual score variation on state tests. The teacher value-added is the average outcome obtained from the position occupied by each of the students within their groups. In the team awards, outcome depends on the performance of students from the teachers of the same grade and subject in each centre as a whole.

Fryer concludes that only groups subject to the fear of loss, and especially in group award schemes, show student performance gains.

More information:

Roland G. Fryer, Jr., Steven D. Levitt, John List, and Sally Sadoff (2012). "Enhancing the efficacy of teacher incentives through loss aversion: A field experiment." NBER Working Paper 18237. Cambridge, MA: National Bureau of Economic Research

All in all, the programs evaluated show quite disappointing outcomes. It should be mentioned however, that certain elements may explain the poor performance. In New York as well as in Texas, both the complexity of the indicator (which makes it difficult to know in advance the effort that teachers are required make in order to receive the incentive) and the fact that most of the teachers were not supporters of PFP as a means of payment. In New York and Nashville, all underperforming schools have an incentive to improve as there is a risk of receiving fines and closure after the adoption of the No Child Left Behind Act of 2002 (as such, all teachers had an incentive to improve outcomes.) Finally, the Nashville goals were perhaps too high to serve as a motivating factor [6].

All in all, the programs evaluated show quite disappointing outcomes. It should be mentioned however, that certain elements may explain the poor performance.



Outside the U.S. experimental evaluations in three developing countries stand out: Kenya, India and Mexico. In the first case, an incentive program at school level, applied to 50 primary and secondary schools, where a specific number of schools which obtain the best performance, or the most significant gains in outcomes, show student performance gains. However, this gain occurs

only in subjects in the incentive program and for the duration of the program; therefore, the performance can be interpreted as showing that PFP does not improve learning, but simply that teachers helped students prepare for the tests. The second experience was carried out in primary schools in a region of India. In this case, the incentive is for a performance gain of over 5% in language and math tests. The outcomes show the success of the program in the short and long-term, and that individual incentives have a greater effect than those at schoolwide level [5]. Lastly, the Alignment Incentive Program, in 88 senior high schools in Mexico, offering three types of incentives: one exclusively for students (depending on outcomes from a maths test at the end of the school year); one exclusively for teachers (depending on the performance of their students), and the third for students and teachers (in the first case depending on their outcomes and their classmates scores and, in the second case conditional on the outcome in mathematics of their students as well as all the other students in the centre). The award is conditional on students gain in level (three are established) and there is a penalty if they fall to the lowest. The outcomes show a positive effect of the program in the first case and especially the last, when the stimulus is aimed at students and teachers, and depends on group-wide outcomes of students at the school. [7]



Outside the U.S. experimental evaluations in three developing countries stand out: Kenya, India and Mexico. In the first case, an incentive program at school level, applied to 50 primary and secondary schools, where a specific number of schools who obtain the best performance or the most significant gains in student performance, show student performance gains.

A second group of studies constitute quasi-experimental evaluation methods. Methodologically rigorous research is collected but only admits causal inference if we accept the assumptions of the procedure. As in the case of experimental studies, experiences from the United States are shown first, followed by data for other countries.

Two long-term programs were evaluated using regression discontinuity design. On the one hand, the ABC School-wide Bonus Program, North Carolina, which awards teachers (primary and secondary) when the entire student body obtain expected level performance gains in mathematics and language, and twice this amount if this is exceeded. The expected outcome is calculated based on a prediction that takes into account students previous performance as well as their socioeconomic status. The evaluation shows the program's positive effect (especially in mathematics) [8]. Moreover, the IMPACT program (District of Columbia) also improves the outcome of public school students in a quite troubled district. This program establishes five levels of teacher efficiency with certain peculiarities: teachers in the higher level obtain a significant award, which can become a permanent payroll feature if gains are obtained for two consecutive years; teachers on the lower level are let go and those in the section immediately above the lowest can be let go if they fail to show performance gains within

a year. The program evaluates students of teachers in the top and penultimate category (most important) and shows how performance gains have increased in both cases. Suffice to say that the incentive is measured in a multidimensional way, where output in subjects where state performance data is available, carries greater weight and is calculated in a similar way to the previous program [9].

Box 2.

The importance of the incentive (II): rewards and punishments

Thomas Dee and James Wyckoff analyse the IMPACT program, implemented in the District of Columbia, an area with poor academic performance. The program was set up in 2009 and the authors evaluated performance after three years of implementation.

IMPACT establishes five categories of teacher; “ineffective”, “minimally effective”, “developing”, “effective” and “highly effective.” teachers in the two highest levels can be awarded a rise in salary. If a teacher spends two consecutive years in the highest category, the pay rise becomes a permanent salary increase (which varies depending on the job category and characteristics of the centre but can mean an increase in salary of up to 30%). Teachers in the lower category are dismissed while those in the penultimate group may be let go if they fail to produce gains within a year. The award is calculated from the score given to teachers through a multidimensional index made up of four components. The first is based on observation of a teacher in the classroom (carried out by school management as well as an independent expert). The second depends on the student performance on state tests. Since only the data for some subjects and courses is available, the results for the rest are generated on the basis of an assessment of the degree of compliance (as measured by the school principal) with learning objectives established between teachers and principals. The third element depends on the principal’s assessment of the teacher’s participation in school activities. The final factor is based on the school-wide performance gains on state tests. Finally, a bad review regarding professionalism (measured in terms of compliance with the rules of the centre) can result in points being deducted from the index. For teachers with students who are assessed using external tests, the first component accounts for 35%, the second, 50%, the third, 10% and the fourth, 5%. For other teachers, the criteria are, respectively, 75%, 10%, 10% and 5%. The assessment program contemplates two areas around two relevant points: students in the “minimally effective” teacher’s group and those in the “very effective” group. The outcomes show that in both cases there are significant gains in student outcomes. In addition, many teachers in the “minimally effective” category abandon the teaching profession.

More information:

Dee, T., and Wyckoff, J. (2013). “Incentives, selection, and teacher performance: Evidence from IMPACT.” NBER Working Paper 19529. Cambridge, MA: National Bureau of Economic Research.

Three studies set out to avoid selection bias by using matching techniques between similar centres (in observable characteristics). The REACH program, in the Austin school district has several incentive schemes (individual and group input and output) for primary and secondary schools. One assessment shows an increase in teacher value-added, measured by comparing the outcomes obtained by the students with respect to possible outcomes taking into account their socioeconomic characteristics. The award is only given the first year, when the highest level obtained in the first year remains present in the second [10]. However, assessment of the Teaching Advancement Program (TAP) implementation in primary schools in Chicago shows no performance gains. In this case, teachers are remunerated (collectively) on the basis of parity between the observation of teachers' performance in the classroom and outcome gains on state tests (real outcomes are also compared with expected outcomes) [5]. Launched in 2006, the Procomp Denver program shows a different experience. The program does not offer incentives, instead it provides a change in the system of pay, so that salary increases (not the base salary) rely entirely on stimuli associated with input and output indicators. In the case of the latter, a factor for achieving an increase in salary is linked to student performance gains when at least 50% of students are in the 55% student performance gains percentile or higher in state tests in mathematics and language. Program assessment shows gains in student outcomes associated with teachers in the Procomp method by comparison with those who chose not to adopt this system (joining the system was voluntary for teachers who joined the workforce prior to 2006) [11].

Finally, using a differences-in-differences approach, which analyses a PFP program in Little Rock (Arkansas), shows the improved performance of students in participating primary schools (compared to non-participants), especially those with the worst outcomes before joining the program. In this case, incentives (individual) depend on student performance on state tests in math and reading; gains of each student in a class are taken into account, in such a way that the reward increases in keeping with greater achievement gains and teachers receive incentives based on the gain observed in their group of students [12].

Outside the United States two studies implemented in Israel stand out for the meticulous methodology. The first study, carried out in 2002, examined the effect of a program implemented in secondary schools for students whose performance was below standard. The award was given to the entire school but only for teachers in the upper third of gains as shown by several indicators linked to nationwide university entrance exams. Analysis of the program reveals that students from participating schools show performance gains and a reduction in dropout rate. The second assessment, implemented in 2009, examines a similar program to the one mentioned above but where the incentives are for individual teachers. Program assessment is positive in terms of the likelihood of students taking national secondary level exams at the end of school and the performance score obtained. It must be pointed out that in both cases performance is measured taking into account the difference between the expected outcomes and those obtained depending on characteristics of the school in the first case, and of the teachers' students in the second example [5].

Studies implemented in England, Mexico, Portugal and Chile utilize a less accurate methodology. The programs vary, but all incorporate an individual PFP system in which student gains (value-added outcomes on external tests) are just one of the elements, often carrying less weight, along with education process input indicators (such as teaching activity or professional development normally associated with training). These studies also seek to establish permanent salary increases (not merely a single initiative-based award). The first two cases confirm the positive effects of the PFP system on improving student outcomes, although these are quantitatively important only in England. By contrast, the salary increase program implemented in Portugal shows negative student outcomes [5] [13] [14]. Finally, a study in Chile is worth mentioning, in which the National Assessment of Educational Establishments (SNED, as per the Spanish acronym) incentive program launched in 1997, where schools with similar social categories from each region compete. The evaluation analyses average scores for each school on an index which includes education input and output measures; the latter considers both students outcomes in math and language as well as their gains over time (equivalent to 65% of the total value of the index). Incentives are awarded to schools in the top 25% of the value indicator. The program produced significant student performance gains [15].



In conclusion, experimental evaluations from the United States show ineffective practice of PFP programs. By contrast, the evidence from developing countries presents more positive results. Also, quasi-experimental evaluations, both in the US and in other countries, tend to demonstrate positive outcomes.

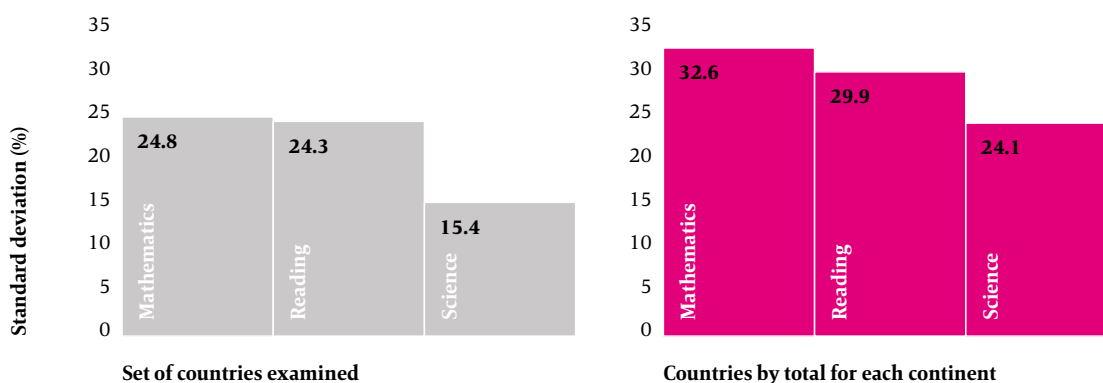
Finally, the review includes a regression analysis using PISA data from 28 OECD countries, of which 13 implement some type of PFP system. Even though the methodology prevents causal inference analysis, the data is valuable given that it examines several countries together. The outcomes show a positive correlation between those implementing PFP systems and outcomes in mathematics and reading comprehension: outcomes increase by around 25% standard deviation of mathematics and reading comprehension scores (15% on science tests). The effects are greater when we perform a continental fixed comparison only between countries on the same continent² (see Graph 1) [16].

In conclusion, experimental evaluations from the United States show ineffective practice of PFP programs. By contrast, the evidence from developing countries presents more positive results. Also, quasi-experimental evaluations, both in the US and in other countries, tend to demonstrate positive outcomes. Therefore, PFP can be successful in improving student outcomes, but is not guaranteed a priori.

² This 25% effect is relatively significant: a country in an intermediate gain position improves 10 places in the PISA classification; therefore, by example, the effect represents 72% of the impact of repeating a primary school year. The incidence is greater than that detected in the majority of evaluations carried out in the United States, when outcomes are positive and similar to the median effect in the rest of the countries (see Table 1).

Graph 1. Pay-for-performance and performance gains in PISA tests

Performance gains among students aged 15 in 2003 PISA test in countries implementing pay-for-performance systems



Source: Ivàlua, based on Ludger Woessmann (2011). "Cross-country evidence on teacher performance pay". *Economics of Education Review*, vol. 30, p. 404-418.

Using PFP to assign teachers to schools

The allocation of higher quality teachers (defined in terms of qualifications, knowledge and experience) to underperforming schools (often with a high percentage of students from low-income families and ethnic minorities) is an issue which concerns policymakers in many countries, given the uneven distribution of teachers to schools [17] [18]. The evidence shows that increased remuneration is no guarantee that schools with the greatest needs will attract (or retain) the best teachers, given that this is not a priority issue for many teachers in terms of mobility and permanency in centres. Teachers claim to be more concerned about working conditions, existing resources, teaching workload, school leadership quality and place of residence [19]. This debate has attracted input concerning an analysis of PFP systems as to whether they attract the most efficient teachers (defined in this case as teachers of students with higher gains in external tests) to the centres where they are most needed. There is practically no evidence in this regard, although a recent randomized experiment in the United States shows outcomes which are only partially positive and with modest effect [20]. Therefore, it appears that monetary incentives (PFP or others) are not very successful in effecting a more equitable distribution of the best teachers between centres.

Teachers' attitude towards PFP systems

The United States has produced studies into the degree of acceptance of PFP systems among teachers and society as a whole. The results show that they are not very popular among the former group, but clearly favoured by the latter. Among teachers, the greatest opposition is caused by methods based on the calculation of the teacher value-added by student performance outcomes on standardized tests [21]. However, the assessment is also influenced by the structure of PFR systems: when teachers feel that assessment is fair and realistic objectives have been set, acceptance

is greater. In any case, some teachers (and unions) reject the very philosophy of PFR, which they do not consider appropriate within the education environment. Nevertheless, surveys in the US show that the percentage of teachers in favour of PFR has grown over time (currently around 40%), receiving greater acceptance in areas where they already exist, as well as between teachers in the lower salary bracket or those belonging to ethnic minorities [22].

Limitations

The literature available is not sufficient to reach definitive conclusions regarding the effectiveness of pay-for-performance programs. Further studies are required (especially randomized experiments showing causality), which subsequently allow for meta-analysis. In addition, research should be able to ascertain whether gains are real performance gains in learning outcomes and not just because the teacher favours teaching to the test, or using practices which alter test results (filtering students who take the test, permitting copying, etc.) in order to be awarded incentives [23]. It is also necessary to include the long-term impacts, both in relation to the continuity of students in the educational system and its consequences on the labour market [24] [25]. Finally, to draw conclusions based on a more similar environment to that in Catalonia, it would be fitting to analyse experiences from more countries, especially those in Europe, where it is virtually non-existent.

Summary

The empirical evidence presented does not allow us to reach conclusive results regarding the efficacy of pay-for-performance programmes (PFP). Recent studies based on randomized experiments indicate that PFP generally does not improve student outcomes in the United States, but does so in developing countries. Nevertheless, assessments using other methodologies show mostly positive results. Regarding the characteristics of PFP, it is impossible to conclude which mechanism is best, both in relation to the level at which incentives are offered, (individual teachers, groups of teachers and the whole school), the award amount, or whether it only includes output indicators. A PFP system with a simple indicator not using comparative effectiveness outcomes (taking into account other students) appears to be more effective, even though further studies are needed to draw definitive conclusions.

The abovementioned elements, therefore, only allow us to state that the success of a PFP program is not guaranteed. This does not mean that success is unobtainable however, but that if implemented, extremely careful monitoring of the program is required in order to gradually implement the necessary changes in order to ensure its effectiveness.

Table 2.
Pay-for-performance systems (PFP) based on student performance

In favour	Against
Real gains can be generated in student outcomes.	Students' academic performance also depends on factors beyond teachers. In addition, a test only reflects part of learning.
It is an objective indicator. It helps in decision making, as principals can clearly detect teachers at the extremes (best and worst) but it is difficult to distinguish between those at intermediate levels.	No results of external tests are available for all students in all courses or subjects to apply objective methods. It is difficult to encourage teachers of subjects which are not assessed externally.
Teaching is one of the professions with lower wage dispersion. PFP can mean pay increases for the more effective.	It is assumed that the results of the students are attained solely by the teacher of the course, and that a teacher doesn't affect the future learning of students.
It can affect the selection and retention of teachers, attracting and retaining the more effective teachers letting go of less competent teachers.	Students can work in different ways depending on the importance of a particular course in their academic progress.
You can improve the image of public schools by giving more importance to students' performance and results.	Teachers can create strategies to improve student outcomes without improving learning.
They are more cost-efficient than measures linked to the system of providing more teachers per student.	The results of students in external tests may have no significant annual variations attributable to teachers.
Group incentives can encourage collaboration among teachers. You can create a culture of continuous improvement and teacher participation in decision making.	Individual PFP may reduce cooperation among teachers. The groups can create problems of non-implication (free-rider) by some teachers, especially in schools with a significant number on the staff.
The value-added method examines the difference in student performance derived from their socioeconomic situation.	The cost of administering the incentive system can be high.
It might be more successful if designed as a long-term plan instead of a pilot program.	Possible lack of a predisposition to PFP among teaching staff and teachers' unions giving rise to conflicts.

Implications for practice

The review of pay-for-performance (PFP) studies shows that such incentives can improve student performance, but are not a guarantee thereof. Therefore, if policy-makers wish to implement a PFP program, assessment should be incorporated from the very outset in order to introduce ongoing improvement to the design and implementation of policies to achieve the desired results.

Furthermore, the studies reviewed do not allow us to conclude categorically which features are required to facilitate the successful implementation of a PFP system, although certain elements do seem to favour achieving objectives of the goals of improving student performance and can facilitate its implementation.

Taking these precautions into account, we believe that the eventual entrenchment of a PFP system in Catalonia should consider the following questions:

1. Firstly, the performance indicator should be relatively simple, so that teachers can evaluate the effort required to gain the reward.

- 2 Secondly, one should consider whether it is convenient to incorporate input and output elements. With the first type of indicator, teachers can verify that certain items assessed are wholly dependent on them. With the second, the incentives are linked to the results of the students (this could carry greater weight increasingly over time).
- 3 In third place, we should reflect on the weighting of collective assessment elements (e.g. centre-fixed) to avoid hindering the necessary collaboration between teachers in the learning process, and what could be assigned as individual items.
- 4 Fourthly, we should consider whether awards are set at a certain level (whoever reaches this level receives the incentive) or by comparing results (giving the incentive to teachers or schools with the best results).
- 5 Fifth, we should assess the degree of stability assigned to the system (whether approached as a pilot or long-term scheme) and the funding incentive thereof. Ideally, the decision taken on this last point (incentive amount) should be based on a cost-effective analysis of the measure, in other words, taking into account the expected impact and gains and which are achievable by other teacher policies at a similar or lower cost.
- 6 Finally, policymakers should seek the involvement of teachers and their representatives, providing information and training on PFP systems to the former and inviting the latter to participate in system design.

It is fair to say that recent legislation passed in Catalonia regarding stages of education promotion (ORDER ENS/330/2014, November 6) establishes that, as of the 2015-2016 academic year, teachers can achieve salary increases depending on their involvement in improving the results of the centre, as well as through individual assessment of the teacher. In the first case, an indicator is established based on outputs, while in the second, inputs predominate. Therefore, it opens the door to the introduction of a PFP system which also contains most of the elements mentioned above. Meanwhile, under the current system, elements associated with assessment are essential requirements to achieve bonuses and the system is applicable only in part to personnel: state qualified career teaching staff, state qualified interns and currently working interns. It is significant that the new legislation has already incorporated a PFP system.

Clearly, it is not only salary incentives that increase motivation of teachers. Improvements are required in the areas of training, recruitment, remuneration and the social prestige of the teaching profession. We also insist that the evidence regarding their effectiveness is mixed. Therefore, any initiative along the lines of progressing towards a PFP system in Catalonia should be closely monitored, assessing design, implementation and impact, retaining potentially effective components and discarding those which are proven to be fruitless.

Bibliography

- [1] Hanushek, E. A. (2011). "The economic value of higher teacher quality". *Economics of Education Review*, vol. 30, p. 466-479.
[http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202011%20EER%2030\(3\).pdf](http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202011%20EER%2030(3).pdf)
- [2] Calero, J., Escardíbul, J.O. (2014). *Recursos escolares y resultados de la educación*. Madrid: Fundación Europea Sociedad y Educación.
<http://issuu.com/efse/docs/re-calero-escardibul?e=3027563/9412255>
- [3] Blasco, J. Casado, D. (2009). *Avaluació d'impacte*. Ivàlua Collection of practical guides for public policy evaluation, n. 5. Barcelona: Catalan Institute of Public Policy Evaluation.
http://www.ivalua.cat/documents/1/01_03_2010_11_33_12_Guia5_Impacte_Setembre2009_revfebr2010_massavermella.pdf
- [4] Institute of Education Sciences. "What Works Clearinghouse" (WWC).
<http://ies.ed.gov/ncee/wwc/topic.aspx?sid=17>
- [5] The World Bank (Banc Mundial). *Impact Evaluations in Education*.
<http://datatopics.worldbank.org/EdStatsApps/Edu%20Evaluation/evaluationHome.aspx?sD=E>
- [6] Neal, D. (2011). "The design of performance pay in education". *NBER Working Papers 16710*, Cambridge, MA: National Bureau of Economic Research. <http://www.nber.org/papers/w16710.pdf>
- [7] Behrman, J. R., Parker, S. W., Todd, P. E., 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. Philadelphia: University of Pennsylvania.
http://economics.wustl.edu/files/economics/imce/todd_paper.pdf
- [8] 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
- [9] Dee, T., Wyckoff, J. (2013). "Incentives, selection, and teacher performance: Evidence from IMPACT". *NBER Working Paper 19529*. Cambridge, MA: National Bureau of Economic Research.
http://curry.virginia.edu/uploads/resourceLibrary/16_De-IMPACT.pdf
- [10] Balch, R., Springer, M. G. (2015). "Performance pay, test scores, and student learning objectives". *Economics of Education Review*, vol. 44, p. 114-125.
<https://my.vanderbilt.edu/matthewspringer/files/2015/02/Balch-and-Springer-2015.pdf>
- [11] Goldhaber, D., Walch, J. (2012). "Strategic pay reform: A student outcomes-based evaluation of Denver's ProComp teacher pay initiative". *Economics of Education Review*, vol. 31(6), p. 1067-1083.
[http://www.cedr.us/papers/working/WP%202011-3%20Procomp%20Strategic%20Compensation%20\(9-28\).pdf](http://www.cedr.us/papers/working/WP%202011-3%20Procomp%20Strategic%20Compensation%20(9-28).pdf)
- [12] Winters, M., Greene, J., Ritter, G., 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
- [13] Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., Wilson, D. (2009). "Evaluating the impact of performance-related pay for teachers in England". *Labour Economics*, vol. 16, p. 251-261.
<http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/wp113.pdf>
- [14] Martins, P. S. (2009). "Individual teacher incentives, Student achievement and grade inflation". *IZA Discussion Paper Series No. 4051*. Bonn: The Institute for the Study of Labor (IZA).
<http://ftp.iza.org/dp4051.pdf>
- [15] Contreras, D., Rau, T. (2012). "Tournament incentives for teachers: Evidence from a scaled-up intervention in Chile". *Economic Development and Cultural Change*, vol. 61(1), p. 219-246.
http://eco3147.facea.puc.cl/economia_puc/images/stories/Instituto/Profesores/tournament.pdf
- [16] Woessmann, L. (2011). "Cross-country evidence on teacher performance pay". *Economics of Education Review*, vol. 30, p. 404-418. <http://repec.iza.org/dp5101.pdf>
- [17] Akiba, M., LeTendre, G. K., Scribner, J. P. (2007). "Teacher quality, opportunity gap, and national achievement in 46 countries". *Educational Researcher*, vol. 36(7), p. 369-387.
http://www.researchgate.net/profile/Gerald_Letendre/publication/43901719_Teacher_Quality_Opportunity_Gap_and_National_Achievement_in_46_Countries/links/02e7e52fce8c6aadd7000000.pdf
- [18] Berry, B., Eckert, J. (2012). *Creating teacher incentives for school excellence and equity*. Boulder, CO: National Education Policy Center. <http://files.eric.ed.gov/fulltext/ED528716.pdf>

- [19] Murnane, R. J., Steele, J. L. (2007). "What is the problem? The challenge of providing effective teachers for all children." *The Future of Children*, vol. 17(1), p. 15-43.
http://www.princeton.edu/futureofchildren/publications/docs/17_01_02.pdf
- [20] IES (2014). "Do disadvantaged students get less effective teaching? Key findings from recent Institute of Education Sciences studies". *NCEE Evaluation Brief*, vol. January. Institute of Education Sciences. National Center for Education Evaluation and Regional Assistance.
<http://ies.ed.gov/ncee/pubs/20144010/pdf/20144010.pdf>
- [21] Glazerman, S., A. Protik, B. Teh, J. Bruch, Max, J. (2013). *Transfer incentives for high-performing teachers: Final results from a multisite experiment (NCEE 2014-4003)*. Washington, DC: National Center for Education Evaluation and Regional Assistance. <http://ies.ed.gov/ncee/pubs/20144003/pdf/20144003.pdf>
- [22] Leigh, A. (2013). "The economics and politics of teacher merit pay". *CESifo Economic Studies*, vol. 59 (1), p. 1-33. <http://andrewleigh.org/pdf/PerformancePay.pdf>
- [23] Podgursky, M., Springer, M. G. (2007). "Teacher performance pay: A review". *Journal of Policy Analysis and Management*, vol. 26(4), p. 909-950.
<http://faculty.missouri.edu/~podgurskym/wp-content/uploads/2008/rep10.pdf>
- [24] Chetty, R., Friedman, J. N., Rockoff, J. E. (2014). "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood". *American Economic Review*, vol. 104(9), p. 2633-79.
http://obs.rc.fas.harvard.edu/chetty/value_added.pdf
- [25] 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, Cambridge, MA: National Bureau of Economic Research. <http://www2.warwick.ac.uk/fac/soc/economics/staff/vlavy/w20983.pdf>

Further reading

- Alger, V. E. (2014). *Teacher incentive pay that works: A global survey of programs that improve student achievement*. Vancouver, Canada: The Fraser Institute-Barbara Mitchell Centre for Improvement in Education.
<https://www.fraserinstitute.org/uploadedFiles/fraser-ca/Content/research-news/research/publications/teacher-incentive-pay-that-works.pdf>
- Lavy, V. (2007). "Using performance-based pay to improve the quality of teachers". *The Future of Children*, vol. 17(1), p. 87-109. http://www.princeton.edu/futureofchildren/publications/docs/17_01_05.pdf
- OECD (2009). *Evaluating and rewarding the quality of teachers: international practices*. París: OECD Publishing.
http://www.oxydiane.net/IMG/pdf_8709031E.pdf

First edition: July, 2015
© Fundació Jaume Bofill, Ivàlua, 2015
fbfill@fbfill.cat, info@ivalua.cat
<http://www.ivalua.cat>
<http://www.fbfill.cat>

Authors: Miquel Àngel Alegre, J. Oriol Escardíbul
Translator: Diarmuid Rory Magourty
Edition: Fundació Jaume Bofill
Design and layout: Enric Jardí
ISBN: 978-84-944534-2-7
LD: B 26124-2015

This publication is licensed under a Creative Commons **Attribution-Noncommercial-No Derivatives (by-nc-nd)**.
Reproduction, distribution and public communication of the work are allowed providing you give appropriate credit to the authors. You may not use the material for commercial purposes. If you remix, transform, or build upon the material, you may not distribute the modified material.



In collaboration with:



Obra Social "la Caixa"