



# Teacher Development

An international journal of teachers' professional development

ISSN: 1366-4530 (Print) 1747-5120 (Online) Journal homepage: <http://www.tandfonline.com/loi/rtde20>

## Within the teacher evaluation policies black box: two case studies

Denise Vaillant & Gabriela Gonzalez-Vaillant


To cite this article: Denise Vaillant & Gabriela Gonzalez-Vaillant (2017) Within the teacher evaluation policies black box: two case studies, *Teacher Development*, 21:3, 404-421, DOI: [10.1080/13664530.2016.1259649](https://doi.org/10.1080/13664530.2016.1259649)

To link to this article: <http://dx.doi.org/10.1080/13664530.2016.1259649>



Published online: 04 Dec 2016.




[Submit your article to this journal](#) 



Article views: 39



[View related articles](#) 



[View Crossmark data](#) 

## Within the teacher evaluation policies black box: two case studies

Denise Vaillant<sup>a</sup>  and Gabriela Gonzalez-Vaillant<sup>b</sup> 

<sup>a</sup>Department of Education, Universidad ORT-Uruguay, Montevideo, Uruguay; <sup>b</sup>Department of Sociology, Stony Brook University, New York, NY, USA

### ABSTRACT

This paper analyzes the relationship between teacher evaluation policies in Chile and Uruguay, teacher professional development, and teacher performance effectiveness. The analysis tries to understand the extent to which the existing processes for teacher evaluation stimulate an improvement in the quality of teaching. The comparison of two national cases with similar educational results but radically different educational policies allows the authors to take a close look within the 'black box' of educational policy-making. The paper is based on in-depth analysis of the PISA survey to heads of schools and in-depth interviews with teachers, heads, and key informants. The paper finds that policies aiming to strengthen teacher development and teacher performance through teacher evaluation cannot be understood in dichotomist ways; lack of attention to teachers' perspectives and unintended effects of policy implementation can lead to marginal effects. Uruguay, where teacher evaluation policies have weak practical effects, has managed to obtain similar educational results to Chile. However, the experience of Chile shows that teacher evaluation can have an effect on teacher professional development and, in the long run, this may have an effect on the quality of teaching and student results.

### ARTICLE HISTORY

Received 11 March 2014  
Accepted 29 January 2016

### KEYWORDS

Teacher evaluation; teacher performance; professional development; Latin America; educational results

This paper aims to understand the extent to which teacher evaluation may be an effective tool to support teacher development and improve performance effectiveness. To analyze the effect of teacher evaluation on performance, the paper proposes the in-depth analysis of two national cases: Chile and Uruguay. These countries provide rich comparisons due to the differences in the organization of their respective educational systems. While in Chile teacher evaluation is part of the educational debate and it is included in the agenda, in Uruguay there is little debate on the topic and educational policies have somewhat ignored teacher evaluation. Despite their prominent differences regarding teacher evaluation policies, both countries reach similar educational results in the student evaluation conducted by the Programme for International Student Assessment (PISA). The analysis proposed will allow us to enter the 'black box' of the implementation of educational policies regarding teacher evaluation. Instead of looking at cases that have radically different outcomes and different educational systems, we propose to unpack the factors that can explain similar

results in the context of radically different policy scenarios. In so doing, this study seeks to improve the current understanding of the effects of evaluation systems on teacher development and therefore on the students' learning. By exploring the question of what is the effect of teacher evaluation on educational results, the study provides empirical insight into the intricate mechanisms that mediate political policies seeking to improve teacher development and educational results.

The national context is relevant, but there are also regional trends that should be taken into consideration. This justifies a double perspective throughout this article; while acknowledging the need for contextualization for each individual country to understand the particularities that are underscoring results, it also presents a comparative analysis regarding teacher evaluation, and their relationship with teacher performance effectiveness to extract some more general conclusions. This topic cannot be analyzed in isolation, it should be considered within the broader network of factors that defines teacher professional development in Latin America. Despite regional heterogeneity and this study's specific focus on Chile and Uruguay, it is possible to identify several issues and dilemmas to be solved in the upcoming years and some perspectives that might enlighten the debate about teacher development and improve performance effectiveness across the board.

The paper is organized in five main sections. The first section begins by placing the problem in a regional perspective. It then presents some important international debates around the issues of teacher evaluation and teacher professional development. The second section succinctly presents the two national cases under study: Chile and Uruguay. The third section presents the methodology employed by the research, along with the nature of the data analyzed in the results section. The findings are organized into two main sections that analyze the quantitative and the qualitative findings, respectively. The discussion section concludes by presenting some policy implications of the findings and by offering some possible lines for future research.

## 1. Teacher evaluation: a regional perspective

When discussing teacher evaluation in Chile and Uruguay the topic needs to be analyzed within the broader context of the teaching career in Latin America, taking into account the important variations among countries in the region. It is important to consider the uneven situation of teachers, and the differences in the political agendas of the different Latin American countries. Despite these differences, there are trends and challenges which are shared by the nations in the continent. This is clearly shown in a study by UNESCO/OREALC (2012) indicating that the current contexts of Latin American countries, even considering diversity of regional situations, are defined by the difficulty of keeping good teachers within the educational environment. There is little incentive for the teaching profession to be a career of choice. There are also issues regarding inappropriate working conditions, and serious problems regarding remuneration and incentives.

From a wider perspective, the teaching profession should be framed within a specific environment and the legal regulations of, among other aspects, the selection process, practice, mobility, development, promotion, and retirement of people practicing the profession (Terigi 2009). The teaching profession in most Latin American countries is regulated by the Teachers Statute that legitimates the teaching activity defining rights and obligations (Vaillant and Rossel 2006). In the majority of Latin American countries, teachers work in state

primary schools, they are public employees, and, as a consequence, enjoy stability in their positions. Some of the most relevant characteristics of the teaching profession in Latin America are related to its structuring in levels, importance of seniority, and the tendency to leave the classroom as a main way of promotion (OECD 2012; UNESCO/OREALC 2012). In general, there is one modality of vertical promotion and another one of horizontal promotion in the teaching profession. The first one is related to the possibility of leaving aside the teaching role to take up other responsibilities, and the second one refers to professional development options without giving up teaching.

According to Morduchowicz (2009), many countries in Latin America seem to give seniority a special place as the main component for the teacher to move forward in a professional career that finishes, at the highest position, outside the classroom with administrative and management responsibilities.

Many countries in Latin America have recently made progress in policies regarding the transformation of the teaching profession and, in some cases, they have generated important social mobilizations (Vaillant 2012). These new models of the teaching career are strongly influenced by the introduction of the so-called horizontal promotion in opposition to the traditional vertical promotion that historically ruled in the region (Morduchowicz 2009). The horizontal promotion refers to the possibility of managing positions by having access to other working positions in the school without it being necessary to give up teaching. The experiences in Chile, Mexico, Colombia, Peru, and the State of Sao Paulo in Brazil, among others, are examples of these new models of the teaching career (UNESCO/OREALC 2012).

All of the available evidence suggests that the quality of teachers in Latin America is the decisive constraint on the region's progress toward a higher performance of the educational systems: according to Burns and Luque (2014), low standards for entry into teaching; low-quality candidates; salaries, promotions, and job tenure delinked from performance; and weak school leadership are the main causes underlying low professionalism in the classroom and poor education results. The teaching career is mediated by a wide range of factors, among which the evaluation systems are highlighted. These will be analyzed in depth in the following sections.

## 2. Teacher evaluation process: an international debate

Comparative studies show that top education systems invest heavily in the evaluation of teacher performance. Burns and Luque (2014) suggest that teacher evaluation plays two critical roles: improving teachers' quality and holding them accountable. In this regard, Singapore, Japan, Korea, and China's Shanghai have all developed effective systems for assessing their teachers' performance and progress.

The OECD's *Building a High-Quality Teaching Profession – Lessons from around the World* report (2011) emphasized the importance of effective teacher evaluation and continuous professional development in sustaining teacher quality. The report made the following observations on teacher evaluation: (a) credentials and years of experience are weak indicators of teacher effectiveness and performance-based rewards should go beyond these; (b) effective teaching is something that can be evaluated in a compensation scheme using teacher and student performance as a basis; (c) measures of teacher performance need to be valid, reliable, and also agreed upon by teachers themselves; (d) teachers appreciate appraisal and feedback even though it usually does not lead to any reward.

The experience in Latin America is considerably less extensive to date. While Mexico, in the 1990s, and Colombia in 2002 introduced the region's first teacher evaluation systems, implementation issues have undermined their impact and Mexico's system is currently being redesigned. According to Burns and Luque (2014), Chile's 2003 system remains the region's best practice example to date. Ecuador began implementing teacher performance evaluations in 2007, and Peru is now introducing a comprehensive system similar to Chile's. Elsewhere in the region, less comprehensive and systematic experiments are currently underway.

Teacher evaluation has not been a priority in many Latin American countries, which does not mean that there is no practice or regulations regarding this aspect. Schools, supervisors, and heads, as well as students and their families, assess the behavior of teachers (Murillo and Román 2010; Vaillant 2012). However, the criteria and evaluation perspectives vary widely among them.

Teacher evaluation is a subject that produces discussion among educational authorities and teacher unions; its implementation is mediated by negotiations that do not always take into account good performance technical criteria. The initiatives have been fostered by the authorities and are not necessarily easily accepted by teachers, there is a general sense of dissatisfaction regarding the capacity the educational systems have had to effectively use the evaluations' results in guiding policies and national reforms and especially in reaching the school and classroom in a constructive and practical way (Navarro 2003; Burns and Luque 2014).

A study on performance evaluation and teaching career carried out in 50 countries of America and Europe (Murillo 2006) states that one of the principal concerns of educational systems in America and Europe is to develop systems of teaching career and teacher performance evaluation that contribute to the professional development of teachers and therefore to the improvement in teaching quality. However, it is recognized that the evaluation of teacher performance is a highly conflictive topic due to the opposing interests and opinions of politicians and administrators, teachers, unions, and scholars.

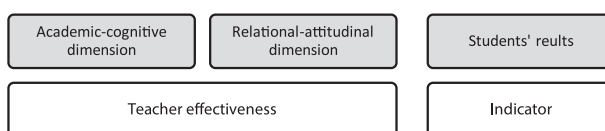
### 3. Teacher performance effectiveness

How to define the 'effective teacher' is a question that refers to the ability of getting good results but also to the ability of achieving more or better goals with limited resources. Thus, teacher effectiveness seems to be directly linked to the teaching activity but also to the situation of teachers, factors affecting their profession, and the results achieved (Darling-Hammond et al. 2005; Gordon, Kane, and Staiger 2006). The notion of teacher effectiveness seems to be associated with different issues such as a deep knowledge of the subject taught, the capacity to structure and explain new knowledge, the ability to show its sense and value, the capacity to organize a situation or teaching and/or learning process, the capacity to establish a minimum rapport with students, previous training, and students' grades.

Literature about the topic (Danielson 2006; Goldhaber 2006; Hunt 2009) reflects that teacher skills are a key element in students' learning effectiveness but it is not always easy to isolate them for description and analysis. Likewise, elaborating a list of basic capacities for good teaching is not an easy job since the teacher works in a context with a certain number of special features. Darling-Hammond and Sykes (2003) summarize the features of an effective teacher: verbal skills and knowledge of the subject taught, academic skills,

**Table 1.** Dimensions and components of teacher effectiveness.

Dimension	Components
Academic-cognitive dimension	<ul style="list-style-type: none"> <li>• Training</li> <li>• Continuous education and professional development</li> <li>• Teaching strategies</li> <li>• Ability to work with colleagues</li> <li>• Knowledge of students</li> <li>• Skills to teach different contexts</li> </ul>
Relational-attitudinal dimension	<ul style="list-style-type: none"> <li>• Moral and ethical values</li> <li>• Motivation</li> <li>• Leadership</li> <li>• Empathy</li> <li>• Entrepreneurial spirit</li> <li>• Readiness and positive attitude</li> <li>• Intellectual stimulation</li> <li>• Comprehension</li> </ul>

**Figure 1.** Teacher performance effectiveness indicator.

professional knowledge, and experience. Other characteristics which are important for good practice are also recognized, such as enthusiasm, flexibility, perseverance, and concern for students. Other authors state that an effective teacher is one who achieves the learning objectives set for their students (Rivkin, Hanushek, and Kain 2005; Darling-Hammond 2006; Day et al. 2007).

These capacities have to do with the cognitive-academic dimension (Vaillant 2010) and with the relational-attitudinal dimension, which refer to a series of factors exemplified in the following Table 1.

All in all, the evidence suggests that dimensions and components that make up a good guide to teaching capacities should work as training tools and should take into account the knowledge and attitudes that in the case of teaching are a key element for students' learning (see Figure 1).

It is especially important to highlight the great influence the analysis of students' results has had as a key indicator to measure teacher effectiveness. In fact, educational researchers from several countries have analyzed school results identifying factors that explain learning achievements, and teacher effectiveness has been found to be a key element in the classroom (Wenglinsky 2002). The effectiveness of teacher performance is still the object of much debate (Goldhaber 2006), since there are few clues on the specific parameters could be used to apply this notion to real contexts and for the development of public policies.

#### 4. The case of Uruguay and Chile

Teacher evaluation systems in Chile and Uruguay are related to the current modality of educational governance in each of these countries. In the first case, we have one of the most

decentralized systems among Latin American countries. The transformation took place during the Eighties under the military regime and implied the transfer of responsibility for schools to municipal districts (Gershberg and Meade 2006). The Ministry of Education (MINEDUC) is responsible for the service funding, definition, passing, and supervision of educational policies, texts distribution, and assessment of educational achievement. This model has provided great importance to the Assessment of Educational Quality (SIMCE), managed by the MINEDUC, a system that tries to guarantee the right of parents/consumers to freely choose among educational units competing in the market with minimum intervention by the central government. In Chile, the Good Teaching Framework (MBE, created in 2003), which works as a standard for teacher performance, is an example of the use of guiding criteria for teacher evaluation and continuing education. In September 2009, the new General Law of Education was passed in Chile, which established a new institutional framework for education. Additionally, two new regulatory bodies were created in 2011 to oversee the quality of education and act as a watchdog for the correct use of public funds and deepen the mechanisms of school accountability (see Figure 2).

Uruguay has taken a different way when compared to Chile. Since democracy was restored in 1985, the educational system has been working in an extremely centralized and hierarchical way. All decisions, from administrative matters to curricular frameworks, are taken in the capital city, Montevideo, and are homogeneously applied across the country. In this centralized model, all the administrative roles, management, and administrative processes take place within the national government jurisdiction. There is systematic evaluation of students in public and private educational centers but the analysis of socio-economic factors has been privileged over variables such as school management or internal efficiency of the educational system (Vaillant 2012).

Ravela (2002) compares national systems of evaluation with 'strong' consequences that imply sanctions, such as the Chilean case, with national systems of evaluation with 'weak' consequences, such as the case of Uruguay, where results have 'just an informative and formative role.' The results are handled in a general way or aggregated and therefore they affect the definition of educational policies at a higher or intermediate level.

Uruguay makes a low-risk analysis of evaluations results keeping the school's results confidential while Chile makes a high-risk analysis, designing a ranking according to an index that includes educational results of the school. Such index is disseminated via the written press and the Internet for the families to know the educational system reality. As well, policies of economic incentives aimed at the teacher take into account the results their students get in the tests (Katzman, Aristimuño, and Monteiro 2003, 21).

Chile	Uruguay
<ul style="list-style-type: none"> <li>• decentralized</li> <li>• privileges internal efficiency</li> <li>• subsidies to demand</li> <li>• teacher evaluation= incentives</li> <li>• teacher and center accountability</li> <li>• national evaluation with "strong" consequences</li> </ul>	<ul style="list-style-type: none"> <li>• centralized and hierarchichal</li> <li>• privileges external efficiency</li> <li>• subsidies to supply</li> <li>• teacher evaluation≠incentive</li> <li>• social accountability</li> <li>• national evaluation with "weak" consequences</li> </ul>

**Figure 2.** Comparison of educational systems in Chile and Uruguay.

Even though there are clear differences between Chile and Uruguay regarding policies of teacher evaluation, however, both countries obtain relatively similar results in the PISA triennial international survey in 2006, 2009, and 2012. It is therefore worth considering why such radically different systems in terms of teachers' evaluations get similar results in international evaluations. How do systems of evaluation affect teacher professional development and the effectiveness of teacher performance?

## 5. Methodology

The article is based on a mixed approach of quantitative and qualitative techniques and it was carried out in the context of a broader investigation funded by the Global Development Network.<sup>1</sup> Using multiple approaches helps capitalize the strengths of the different research strategies employed and transcend their limitations (Tashakkori and Teddlie 2003, 11). We use the different methods sequentially; while the quantitative methods provided us with statistically significant general findings and gave us breath, the qualitative part of the research provided us the necessary depth and allowed us to account for the complexities and nuances of each case. Rocco et al. remind us that 'many researchers find that to conduct this level of research involves mixing methods and perhaps also mixing paradigms' (2003, 19).

An analysis of PISA to understand teacher evaluation practices and their relation with educational results proves to be a privileged way of analyzing relationships across our two cases in statistically significant ways. PISA is an evaluation program of students at age 15, using a representative sample of the population attending secondary education, carried out in OECD countries (currently 30) and also other countries of the rest of the world. Both Chile and Uruguay participated in the 2009 round. PISA provides information on student achievement at age 15, student characteristics, family background, and school and institutional characteristics. For the purpose of this paper, we focus on the questionnaire to the heads of schools as it has interesting data on teacher evaluation. According to PISA 2009 samples, there were 200 questionnaires from Chilean head teachers and 232 from Uruguayan head teachers.

Table 2 summarizes the dimensions that we focused on for analyzing the data.

The quantitative analysis was carried out in two sequential stages. First we undertook a descriptive analysis to characterize each type of school in Chile and Uruguay and identified sources of variation to be exploited. Second, using the PISA 2009 pooled database of Chile and Uruguay, we estimated separate regressions for each set of variables, using the full sample. We also separated regressions for each country, with necessary controls. The original research project that focused on educational governance included several other variable

**Table 2.** Summary of variables.

Variable	Dimensions studied	Source
Teachers' evaluation	<ul style="list-style-type: none"> <li>• Object of evaluation</li> <li>• Evaluating actors</li> <li>• Instruments and procedures</li> </ul>	PISA – Questionnaire to center's head (2009) Interviews Literature review
Teachers' performance	<ul style="list-style-type: none"> <li>• Students' educational results</li> <li>• Involvement with the center</li> <li>• Satisfaction and care for their job</li> </ul>	PISA – Students' test (2009) PISA – Questionnaire to center's head (2009)



sets that are not reported here for the sake of clarity. Controls in the regression estimates included gender, the PISA index of socioeconomic status, grades achieved by the students; school controls include the PISA index of quality of school resources, school size, a set of dummies describing school admission policies, school socioeconomic status, a set of dummies indicating private/public provision, and private/public financing.

As rich as the PISA database is, it does not allow by itself the apprehension of such complex phenomena as the ones the research has set out to explore. Quantitative explanation does not necessarily account for causality, nor does it illuminate the mechanisms under which relationships take place. For this reason, 60 semi-structured interviews were completed in Chile and Uruguay. Two types of interviews were carried out: on the one hand, we interviewed teachers and heads of selected schools, and on the other, we interviewed key informants with high levels of expertise in our subject matter. The purpose of the in-depth interviews was to obtain knowledge on the relationship between teacher evaluation and teacher development and performance, as well as to shed light on the mechanism through which the relationships identified in the statistical analysis take place.

The questions revolved around issues of governance and one of the dimensions interviewees were asked to explain is how they are assessed and their perceptions about the mechanisms in place. Data was collected using a theoretical purposeful non-random sample (Miles and Huberman 1994). In this type of sampling one proceeds to selecting 'incidents, slices of life, time periods, or people on the basis of their potential manifestation or representation of important theoretical constructs' (Patton 2002, 238). The educational centers were selected following two main criteria: (a) type of center provision, allowing us to grasp different institutional formats and assess their impact on provision, and (b) educational outcomes (measures of achievement). In the case of Uruguay, where information about outcomes is not readily and publicly available, we resorted to key informants.

## 6. Teacher evaluation, accountability, and learning outcomes

Public accountability of educational results is one of the many ways in which teachers can be evaluated for their performance. By providing information about school performance, accountability has been said to positively support teacher development and, indirectly, lead to better educational results (Corvalán and McMeekin 2008). However, some authors have criticized the excessive emphasis that accountability policies have placed on evaluation: 'although assessment serves as input for systems of accountability, it is not the system itself' (Darling-Hammond 2004). Evaluation systems that promote curricular reforms rather than punish teachers and schools for low performance results have been more successful in states and districts in the United States. For the Latin American case, Poggi points out that the exclusive emphasis on monitoring and external evaluation, coupled with sanctions, fosters the development of 'as-if' procedures (2008, 4).

There is a clear difference between centers' accountability model in Chile, where 32–40% of educational center heads stated that educational results are publicly known, and the social accountability predominant in Uruguay, where 0% of private centers and 10% of public centers reported to publicly know educational results. As we can see, the Uruguayan results are faithful to its social accountability philosophy while Chile has a much longer tradition in the dissemination of assessment data.

Despite the fact that this is related to the educational center evaluation more than to teacher performance, it is undeniable that the notion of accountability that characterizes each educational system is extremely different and that is directly related to the perception of relative responsibility for school success and failure for the different actors involved. As a teacher from a municipal school in Chile explained:

Interviewer: How important are those center rankings for you?

Teacher: They are published on the newspaper, everyone knows about them. The guardians get nervous, especially within private schools, there are claims, there are very happy people, and there are schools which are doing very well. There are other schools that do not.

Regarding teacher evaluation, PISA allows the analysis of the extent to which the information on students' educational results is used to evaluate teachers in each country. As the data supports, the use of achievement data for evaluation of teacher performance is more frequent in private than in public schools, in both countries. Contrary to what the literature indicates, the heads of public schools of Chile and Uruguay report similar mean values of this variable (35% and 37%, respectively), while private Chilean schools show higher values than their Uruguayan counterparts (64%, 70% for Chilean private, private subsidized schools, and 42% for Uruguayan private schools). In spite of the fact that Chile and Uruguay have radically different policies in terms of using student achievement as an indicator of teaching performance, the perception of heads of schools in the public system is quite similar.

With regard to monitoring of teacher lessons by external supervisors, in Chile and Uruguay the school heads and supervisors are usually hired among teachers; this is consistent with a teaching career that rewards antiquity for career advancement. Transition from teacher to principal is a path chosen by many teachers in order to achieve higher pay and recognition (Morduchowicz 2002; Vaillant 2012). TALIS reveals that the average principal in Chile has been in that position for 12 years, and has more than 25 years teaching experience (OECD 2014). As the data shows, the monitoring of teacher lessons by external inspectors as a form of assessment is a much more common practice in Uruguay than in Chile. While 84% of Uruguayan public school heads report the existence of teacher lesson monitoring by external inspectors, this is only the case with 19% of Chilean public school heads. The difference in private schools is also very significant, as almost twice as many Uruguayan private school heads as Chilean public school heads report the existence of external inspection as a way of monitoring teacher lessons (61% in Uruguay and 31% in Chile).

Teacher performance monitoring by the center's head and by supervisors is more common in public institutions in Uruguay than in municipal centers in Chile. While in Uruguay, 90% of centers' heads claim to monitor teacher performance within their center, only 63% do it in Chile. In addition, teacher supervision is, according to centers' heads, a common practice in Uruguayan private schools where 81% of cases are reported. This is similar to the report in private schools in Chile, where 79% of the heads of schools report monitoring teachers' lessons.

As we have seen, while mechanisms of evaluation are reported in both countries, there is a noticeable emphasis on accountability as a form of evaluation of teachers and centers while in Uruguay the figure of the external inspector plays a significant role. Additionally, the use of achievement data to evaluate teacher performance is a practice that is reported as taking place in both contexts, with noticeably higher presence in Chilean private and subsidized schools.

**Table 3.** Effects of teacher performance evaluation on test scores, after controlling by individual characteristics and school inputs, in Chile, Uruguay, and the complete sample.

	Chile			Uruguay			Full sample		
	Science	Mathematics	Reading	Science	Mathematics	Reading	Science	Mathematics	Reading
Assessment data is posted publicly	8.977*	11.08**	7.084**	-1.236	-2.815	-1.757	7.915*	10.33**	6.492*
Monitoring of teacher lessons by external inspectors	-9.632	-5.757	-10.05**	-5.505	-5.559	-5.605	-7.322	-4.531	-8.316*
Monitoring of teacher lessons by head of school	-11.57**	-17.23***	-12.98***	1.723	6.286	-0.741	-9.395*	-13.97***	-11.04***
Achievement data used to evaluate teacher performance	-11.18	-4.970	-14.766**	-12.09***	-12.30***	-12.75***	-11.33*	-5.491	-14.00**

Note: The table summarizes the variables that were of interest for this paper. Several other variables were tested in the regression models conducted for the project but these are not reported as they pertain to other sets of variables.

Data: Self-administered questionnaire by 2009 PISA to heads of schools, in Uruguay and Chile.

\* $P \leq 0.05$ ; \*\* $P \leq 0.01$ ; \*\*\* $P \leq 0.001$ .

Table 3 shows that, after controlling for individual characteristics, grade, and school characteristics, there is a positive association between publishing the results and performance, in all evaluated areas. This association is significant in the full sample and the case of Chile. For Uruguay, the practice of publishing assessment data publicly is inexistent. This is an indirect indicator of teacher assessment as it is mostly the school center that is assessed through this mechanism.

We also found a negative association between the use of achievement data to evaluate teacher performance and student results in Uruguay. It should be noted that this variable can be strongly endogenous, and therefore the association may reflect reverse causality (assessment data is used to assess teacher performance in worst performing schools). Similarly, we find that monitoring teachers in class by the principal is negatively associated with educational outcomes. This contradicts the existing literature in the field but it should be relativized as it is difficult to know what type of monitoring is taking place in the schools. As we will see, the nature and frequency of the monitoring is very heterogeneous. Again, this variable may also reflect reverse causality, i.e. schools in which principals report using achievement data to assess teacher performance are those in which lower results are achieved. As for the monitoring of teachers by external inspectors, we did not find conclusive results. Considering both samples and the full samples, we do not find robust evidence of association between educational performance and the teacher performance evaluation in PISA (see Table 4).

To sum up, the results presented in this section illustrate that there are some differences in teacher evaluation in both countries. This is aligned with the description carried out of both cases in the preceding section. While in Chile, the assessment data is posted publicly to inform the community about education results, in Uruguay this practice is practically inexistent. The data on student results would indicate that this practice might have positive effects on student achievement.

Some differences can also be found in the type of teacher evaluation that is prevalent in each of the countries, as we will see in the section that follows; the use of achievement data to evaluate teacher performance is more common in Chile than in Uruguay because official systems are in place for this to be done. Furthermore, the predominant way of teacher evaluation in Uruguay is related to external evaluations by supervisors which focus more on

**Table 4.** Evaluation variables from PISA, by country.

	Chile						Uruguay			
	Private		Subsidized		Public		Private		Public	
	Mean	Std. D.	Mean	Std. D.	Mean	Std. D.	Mean	Std. D.	Mean	Std. D.
Assessment data is posted publicly	0.32	0.47	0.33	0.47	0.40	0.49	0.00	0.00	0.10	0.30
Achievement data used to evaluate teacher performance	0.64	0.48	0.70	0.46	0.35	0.48	0.42	0.49	0.37	0.48
Monitoring of teacher lessons by external inspectors	0.31	0.46	0.22	0.42	0.19	0.39	0.61	0.49	0.84	0.36
Monitoring of teacher lessons by head of school	0.79	0.41	0.79	0.41	0.63	0.48	0.81	0.39	0.90	0.30

Data: Self-administered questionnaire by 2009 PISA to heads of schools, in Uruguay and Chile.

the teaching practices than on student results. Strikingly, despite the different institutional configurations, PISA variables regarding teacher evaluation do not seem to suggest significant effects of these practices on student results in either country. When we did find significant relationships, they were negative. This takes us to reflect on the key factors of teacher performance effectiveness and their incidence on educational achievements and to analyze some of the issues that could be operating within the educational policies 'black box' to hinder the positive effect of teacher evaluation on student performance.

## 7. From everyday job in the classroom to teachers' evaluation

As explained in previous sections, although Chile and Uruguay had similar educational systems in the past, both countries have adopted different institutional arrangements in the last decades. Chile gave municipalities the administration of schools and consolidated the subsidy system. Uruguay remained loyal to its historical educational matrix centralized in the state. The Chilean model places great importance on the management control through results, while Uruguay makes a limited use of students' evaluation scores for any kind of school control.

The aforementioned differences between Chile and Uruguay are reflected in teacher performance evaluation systems. In Chile the National System of Performance Evaluation (SNED) evaluates every two years the municipal and private subsidized schools taking into account the measurement of students' results according to the Assessment of Educational Quality (SIMCE). Those schools that obtain good results in their evaluation receive additional resources during two years through the Grant for Excellent Performance. These resources are distributed among teachers of schools selected (Peirano, Falck, and Dominguez 2007).

According to UNESCO/OREALC (2012), Chile has a Professional Teacher Performance Evaluation System also in place at the municipal level, in order to evaluate educators every four years based on the criteria contained in the Good Teaching Framework (MBE). Four instruments support this kind of assessment: self-evaluation, a report by the school director and the pedagogical technical director, an evaluation made by a peer from a different school, and a portfolio of written and video files with proof of experience. After passing a general knowledge exam, teachers who get the best results receive an economic incentive (Variable Acknowledgement of Individual Performance, AVDI). Educators who get the lowest scores must undergo training through professional advancement plans, addressing the weaknesses revealed by the evaluation. Teachers showing 'unsatisfactory' performances must leave the classroom and are subject to evaluation the following year, after having received training. If an educator still gets the same score, he or she must abandon the education system.

The qualitative study carried out from in-depth interviews shows that the case of Chile is different from the one of Uruguay, not only because of the existence of a consolidated system of teacher performance evaluation, but also for the existence of three categories of schools (municipal, private subsidized, and private) as opposed to the public-private category in Uruguay that mediates the relationship between teacher evaluation and effectiveness.

The interviews conducted in Chile in municipal schools show that many teachers do not consider the system of evaluation as an input for the improvement of their development. In the first place, there is criticism of the teacher evaluation process and the way it is structured. Thus, one of the interviewees, head of a municipal school, states:

In municipal education, teachers are not in favor of being evaluated since they do not experience changes in their practice in the classroom. It is a process that the Ministry implemented in recent years and it was very difficult to have 100% of teachers in a system of evaluation. This is a mechanism that allows teacher ranking. A class is recorded and this provides a result which is not an input that provides feedback, there is a lot of formality in the process.

Even though the literature has pointed out that it is important to embrace a holistic understanding of teachers effectively, we see here how administrative and bureaucratic structures can get in the way of effective teacher evaluation processes based on evidence beyond student results in standardized tests.

Other interviewees understand that the teacher performance evaluation system has generated resistance; it is positive but there is still a lot to do. One of the teachers told us:

There has been a lot of resistance ... the result is that in the end you get a label of outstanding teacher, basic teacher or below standard teacher. It is a mechanism that the Ministry has. It is a good attempt to ensure some type of quality. From my point of view, it is not very effective. It is a step forward though.

When asked to explain what aspects of the teacher evaluation generate resistance, the consequences appeared as the most problematic aspect. Several interviewees pointed to the fact that the evaluation has a punitive backlash that should be avoided as it discourages many teachers. In the words of one of the interviewees from a municipal school:

The person who gets an unsatisfactory degree after 3 times, has to leave the system. I don't find this very educational; it is rather punitive, instead. The person that gets a basic or below standard grade has to improve their performance. They have to prepare some special classes ... but there are people who try to get the basic qualification, for example: 'I got the basic qualification; I will be sent to study, ah! No! This year I'm leaving this, I won't work in this field; I'll work doing something else.'

As the heads of school explained, teachers who get the basic or below standard grade have to study in order to improve their results. The evaluation identifies the topics upon which professional improvement plans should be built. Those who get the below standard grade have to take the evaluation again the following year after completing the training course. If the same result is obtained, they have to temporarily leave their position in the center where they work and take a full-time professional improvement training course. After completing it, they are evaluated again. In the event of obtaining an unsatisfactory result, they lose their teaching position, something that rarely happens.

The case of Uruguay is quite different from Chile's since it shows a hierarchic evaluation system implemented with detail in which the central authority has a main role. There is no self-assessment or teacher assessment by colleagues from the same establishment, there are no performance standards for teachers, and the association of teacher evaluation with economic incentives is not being considered even in the educational debate. It was interesting to note that, while teacher evaluation is an issue on the agenda in the Chilean case, for the case of Uruguay, performance evaluation of teachers was mostly a non-issue.

Uruguayan teachers are evaluated by their heads and supervisors as shown in the PISA questionnaire to heads (see Table 5). Interestingly, as we saw, even though there is no systematic information on student achievement that can be used by heads of school to assess teacher performance, there is a widespread perception that students' achievements are used in a meaningful way to assess teacher performance. The evaluation mandated by the Uruguayan education authorities by the heads is done annually and mainly refers to the teachers' performance in the educational center and with the community.

**Table 5.** Interview sample across selection variables.

Teacher's evaluation	Chile			Uruguay		Total	
	Schools (three interviews per center)						
	Municipal	Private	Private subsidized	Public	Private		
Good results	xx	xx	x	x	xx		
Bad results	xx		x	xx	x		
	4	2	2	3	3	42	
			Key informant				
<i>N</i>		9			9	18	
Total						60	

Interviews to heads and teachers from public educational centers in Uruguay reinforce the idea that teacher performance evaluation takes the form of heads' or supervisors' visits to the teachers' groups. Nevertheless, there were a lot of criticisms of the way evaluation operates in practice. Heads and teachers say that heads devote a lot of time to management tasks. Within these, the merely administrative tasks are ubiquitous. One of the interviewed heads tells us:

I'd like to spend my day visiting classes and observing how teachers work ... I realize that we are doing something wrong, many students leave school, others repeat ... but I do not have time, parents ask for interviews, forms, requirements from the Administration ... all the things that do not work and I have to find solutions ... time is not enough. (Head, public center)

According to the interviewees, the two aspects in which Uruguayan teachers are objectively evaluated are attendance and punctuality. In many cases, teacher performance in their subjects is evaluated from formal aspects without an effective look at and feedback on their work.

Additionally, the system includes the observation of teachers by supervisors who are experienced educators within the teacher's field of expertise but these evaluations have 'soft' effects on teachers. Supervisors' evaluation focuses on the teachers' work inside the classroom, prioritizing didactic aspects. In relation to this, one of the interviewed teachers tells us:

... the Secondary Education evaluation system is a joke ... we pretend we work and they pretend they evaluate us, the supervisor comes once or twice a year at the most, and sometimes they do not even step inside the school ... then we have the head and everything depends on him, some evaluate from sitting on their desks while others take it more seriously and come to see us in class ... this system should definitely change since, if you are bad, they cannot fire you, the head makes a negative report and in the best scenario he gets rid of the teacher who goes to another high school which means that the incompetence is transferred elsewhere ...

Even though teacher assignation to the different educational centers depends on a hierarchical system that is reliant on years of experience, the system also relies on the teacher evaluation. This means that the best and most experienced teachers have priority when it comes to choosing schools and they tend to select the best educational centers (Labadie et al. 2006). This system promotes a perverse cycle that might reproduce differences in educational results.

Regarding evaluation by supervisors, which clearly differs from the Chilean teachers' evaluation, Uruguayan teachers are also highly critical about the actual capacity of evaluating a teacher through short and sporadic visits to their classes. A teacher from a public center well known for its educational achievements explained:

I think that evaluating what a teacher does in a whole year in 45 minutes when the group can be nervous or, as it happened to a colleague, several students were absent because it was raining hard and just a few students showed up ... They evaluate you for what happened that day but it lasts for the whole year or until the next evaluation. There are teachers who have not been evaluated for five years.



Nonetheless, even though teachers in Chile do not question the importance of an evaluation system of teacher performance, the artificialness of the teacher evaluation system in Chile is also pointed out by the teachers.

I completely agree with the teacher evaluation. But I have to say, it has errors, for example, there are teachers that don't do it; they hire other teachers to do it for them. So it's not very effective ... The class that you film is a small percentage of the evaluation, and since they know they will be observed, teachers prepare wonderful classes, they even rehearse them ... So, it doesn't really measure what it seeks to measure. What it does measure is that teachers are acquainted with theories of good teaching; that they are capable of teaching; but it doesn't measure what teachers do daily, it doesn't measure if they apply what they know in their everyday teaching ...

There is therefore a common demand on the part of educators for the need to implement meaningful teacher evaluation mechanisms that contemplate long-term achievement and student progress by contemplating variables that cannot easily be 'measured' in short observations or in standardized tests.

To sum up, there are mechanisms of teacher performance evaluation in both countries, especially regulated in the case of municipal or public centers. In the case of Uruguay, the system is restricted to formal class visits by heads and supervisors; it is not associated with educational results and has no impact on the teachers' salary, even though it does have a minor impact on their classification. In the case of Chile, the teacher performance evaluation system is more complex and it is related to material incentives. However, when going into this 'black box' that is teacher performance evaluation, it is possible to say that both systems miss out on important dimensions of the teacher profession and do not allow them to work as expected.

## 8. Discussion

While there is ample evidence that good teaching has positive effects on student results (see Section 1), the evidence presented here regarding the effects of the evaluation systems on teacher development and therefore on the students' results is contradictory. Although our research could not establish that a system of evaluation as existing in the case of Chile or Uruguay represents more possibilities to improve students' learning results, by careful observation of the mechanisms within educational policies 'black boxes' it is possible to identify a complex relationship that exists between teacher performance evaluation and educational results. And this is exactly the information policy-makers need to design educational policies.

As we have seen, Chile and Uruguay have developed different teacher evaluation systems and these have produced similar educational results (as measured in PISA). Furthermore, the evaluation systems in the cases presented here do not appear to have a significant direct impact on educational results.

Even though literature points to the positive effect of using student achievement to evaluate the effectiveness of teacher practices (see Section 2), there are all sorts of unintended consequences that are oftentimes overlooked by policy-makers that mediate the relationship between teacher performances and student results.

In this article, we proposed that the value of an evaluation system might not so much be in the results themselves but rather in that they oblige organizations and systems towards an administrative rationalization which has as a consequence better 'accountability' and a



more transparent management of the system in general. Obviously, it is paramount to consider the diversity of contexts and the distance between public policy theory and its implementation. The literature shows that teacher evaluation should make up a very important aspect of the improvement of teacher performance but, as we showed in this study, its efficiency is strongly related to the context in which they are developed and the participation and complicity of actors involved. The case of Chile is useful to see that the implementation of policies that aim at teacher evaluation development is complex and that it is crucial to involve teachers in the design processes as well as permanently revise policies' intended and unintended effects. Furthermore, both cases underscore the importance of putting formative systems in place to support teachers in their ongoing improvement.

Our study provides another important insight in relation to the use of scores of standardized evaluations such as PISA to measure students' learning. There are many problems with the use of scores in this kind of evaluation, designed and validated with the specific aim of taking a 'picture' of educational results in a given moment. As we have explained before, this picture should be complemented by other qualitative studies such as this one, to look at the nuances and contextual factors that might be intervening. And these nuances and contextual factors oftentimes make all the difference. We have looked at snapshots of two national cases in a given time but educational achievements must also be measured with regards to within-country evolution across time. Furthermore, there is a lot of potential for longitudinal studies that look at the issues explored in this paper across the years to see if differences in educational governance regarding teacher evaluation might not be having an impact in long-term tendencies.

The study shows that in Chile and in Uruguay the effective ways to evaluate teacher performance have not yet been found.<sup>2</sup> One of the reasons probably refers to the fact that teaching is a complex job, and so is its evaluation. The qualitative evidence of this study confirms the international literature's (OECD 2011) finding that the evaluation of teacher performance needs to be valid and trustworthy but above all accepted by teachers. Both in Uruguay and in Chile, teachers identified weaknesses in the evaluation systems in place that are mostly seen as top-down. This study confirms that the definition of what is considered a good teacher needs to be defined and shared by key actors (see Section 3); teachers need to buy in to their evaluation process for it to have a significant effect on their practice. It is also important to recognize the diversity of working conditions in Latin American schools to avoid a 'one measure fits all' approach. Furthermore, teacher evaluation should have consequences not only on the teachers' initial and continuous training but also on the working conditions established by governments and on the effectiveness of educational policies.

The empirical evidence of this study does not allow us to identify *what* characteristics of good teaching correlate with positive educational outcomes. This is yet another aspect of a 'policies black box' that needs to be explored by future studies.

## Notes

1. 'Public service delivery in basic education: institutional arrangements, governance and school results in Uruguay and Chile', coordinated by Mancebo and Vaillant with the collaboration of Llambi, Piñeiro, and González.
2. Since research for this paper was carried out, Chile has approved a new educational law (General Law of Education) that sought to remedy some of the shortcomings that were identified by interviewees in this paper.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

The funds for the present study were provided by the Institutional Capacity Strengthening Fund (ICSF), managed by Interâ American Development Bank (IDB) and Global Development Network.

## Notes on contributors

**Denise Vaillant** holds a PhD in Education from the University of Québec in Montreal, Canada, and an MA in Educational Planning and Management from the University of Geneva, Switzerland. She is the academic director of the Institute of Education at Universidad ORT-Uruguay and is the president of the Teaching Profession International Observatory, based at the University of Barcelona (OBIPD).

**Gabriela Gonzalez-Vaillant** holds a PhD in Sociology at Stony Brook (State University of New York). She obtained her MA in Sociology and a Diploma in Gender Studies from this same institution. She has undertaken numerous investigations on diverse issues in the field of political sociology, social movements, memory, and civic participation. Currently she is teaching at the Catholic University of Uruguay.

## ORCID

Denise Vaillant  <http://orcid.org/0000-0001-8110-4634>

Gabriela Gonzalez-Vaillant  <http://orcid.org/0000-0002-7970-4195>

## References

- Burns, B., and J. Luque. 2014. *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*. Latin American Development Forum Series. Washington, DC: The World Bank.
- Corvalán, J., and R. McMeekin, eds. 2008. *Accountability educacional: posibilidades y desafíos para América Latina a partir de la experiencia internacional* [Educational Accountability: Opportunities and Challenges for Latin America from the International Experience]. Santiago: PREAL-CIDE.
- Danielson, C. 2006. *Teacher Leadership that Strengthens Professional Practice*. Alexandria: ASCD.
- Darling-Hammond, L. 2004. "Standards, Accountability and School Reform." *Teachers College Record* 106 (6): 1047–1085.
- Darling-Hammond, L. 2006. "Assessing Teacher Education: The Usefulness of Multiple Measures for Assessing Program Outcomes." *Journal of Teacher Education* 57 (2): 120–138.
- Darling-Hammond, L., D. Holtzman, S. J. Gatlin, and J. V. Heilig. 2005. "Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness." *Education Policy Analysis Archives* 13 (42): 1–48.
- Darling-Hammond, L., and G. Sykes. 2003. "Wanted: A National Teacher Supply Policy for Education: The Right Way to Meet the 'Highly Qualified Teacher' Challenge." *Educational Policy Analysis Archives* 11 (33): 33. doi: 10.14507/epaa.v11n33.2003. <http://epaa.asu.edu/ojs/article/view/261>.
- Day, C., P. Sammons, G. Stobart, A. Kington, and Q. Gu. 2007. *Teachers Matter: Connecting Lives, Work and Effectiveness*. Maidenhead: Open University Press.
- Gershberg, A. I., and B. Meade. 2006. *Accountability of Educational Reform in Latin America: How can Decentralization Increase (or Not) Accountability?* Santiago: PREAL.
- Goldhaber, D. 2006. "National Board Teachers Are More Effective, But Are They in the Classrooms Where They're Needed the Most?" *Education Finance and Policy* 1 (3): 372–382.
- Gordon, R., T. Kane, and D. Staiger. 2006. *Identifying Effective Teachers Using Performance on the Job*. Washington, DC: The Brookings Institution.
- Hunt, B. 2009. *Effectiveness of Teachers Performance. Review of International Literature and its Relevance to Improve Latin American Education*. Document No. 43. Santiago de Chile: PREAL.

- Kaztman, R., A. Aristimuño, and L. Monteiro. 2003. *Uruguay. Empiric Information. How it is Used, and its Impact on National Learning Evaluations in the Improvement of Public Education in Uruguay?* Universidad Católica del Uruguay. Final Reports, Educational Research Fund. Montevideo: PREAL.
- Labadie, G., A. Mizala, A. Retamoso, R. Operetti, and P. Romaguera. 2006. *Gestión Educativa, Remuneraciones e Incentivos: Análisis Comparado de las Experiencias de Chile y Uruguay* [Educational Management, Remunerations and Incentives: Comparative Analysis of the Chilean and Uruguayan Experience]. Documento de Trabajo No. 20. Montevideo: Facultad de Administración y Ciencias Sociales. Universidad ORT-Uruguay.
- Miles, M. B., and A. M. Huberman. 1994. *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks: Sage.
- Morduchowicz, A. 2002. *Carreras, Incentivos y Estructuras Salariales Docentes* [Teacher Career, Incentives and Salary Structure]. Working Paper No. 23. Santiago de Chile: PREAL.
- Morduchowicz, A. 2009. *Supply, Demand, and Teachers' Salary: Model for Working Document No. 45*. Santiago de Chile: PREAL.
- Murillo, F. J. 2006. *Evaluation of Teacher Performance and Professional Career. America and Europe, An Overview*. Santiago de Chile: UNESCO.
- Murillo, F. J., and M. Román. 2010. "Challenges in the Assessment of Teaching Quality in Latin America." *Ibero-American Magazine of Education* 53: 97–120.
- Navarro, G. 2003. "What Do We Understand by Education for Social Responsibility?" In *Educando para la Responsabilidad Social: La Universidad en su función docente (Tomo 3)* [Educating for Social Responsibility 3: The University and its Teaching Role], edited by C. Ardiles, 21–45. Santiago: Funny.
- OECD. 2011. *Building a High-Quality Teaching Profession – Lessons from around the World*. Paris: OECD.
- OECD. 2012. *Teachers for the 21st Century, Using Evaluation to Improve Teaching*. Paris: OECD.
- OECD. 2014. *TALIS 2013 Results: An International Perspective on Teaching and Learning*. TALIS. Paris: OECD.
- Patton, M. Q. 2002. *Qualitative Evaluation and Research Methods*. 3rd ed. Thousand Oaks: Sage.
- Peirano, C., D. Falck, and M. Dominguez. 2007. "Evaluación Docente en América Latina [Teacher Evaluation in Latin America]." Working paper for the working group of teacher development PREAL, Washington, DC: Santiago de Chile..
- Poggi, Margarita. 2008. "El Sistema Educativo: De la rendición de cuentas a las responsabilidades [The Education System: From Accountability to Responsibilities]." *Revista El monitor de la Educación* 5 (17): 1–5.
- Ravela, P. 2002. *What is Being Assessed by This Test?* Abbreviated Version. Document No. 7. Santiago de Chile: PREAL.
- Rivkin, S. G., E. A. Hanushek, and J. F. Kain. 2005. "Teachers, Schools and Academic Achievement." *Econometrica* 73 (2). Accessed October 14, 2007. <http://edpro.stanford.edu/Hanushek/admin/pages/files/uploads/teachers.econometrica.pdf>
- Rocco, T. S., L. A. Bliss, S. Gallagher, and A. Perez-Prado. 2003. "Taking the Next Step: Mixed Methods Research in Organisational Systems." *Information Technology, Learning and Performance Journal* 21 (1): 19–29.
- Tashakkori, A., and C. Teddlie, eds. 2003. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks: Sage.
- Terigi, F. 2009. School Career, from Individual Problems to the Challenge of an Education Policy. Buenos Aires, National Ministry of Education / OAS. <http://www.porlainclusionmercursosur.educ.ar>
- UNESCO/OREALC. 2012. *Background and Criteria for the Making of Teaching Policies in Latin America and the Caribbean: Strategic Regional Project on Teachers*. Santiago de Chile: UNESCO-OREALC.
- Vaillant, D. 2010. "Capacidades docentes para la educación del mañana [Teacher Capacity for Tomorrow's Education]." *Pensamiento Iberoamericano* 7: 113–128.
- Vaillant, D. 2012. "La gobernanza educativa y los incentivos docentes: el caso de Chile y de Uruguay [Educational Governance and Teacher Incentives: Chile and Uruguay]." *Revista Uruguaya de Ciencia Política* 21: 119–139.
- Vaillant, D., and C. Rossel, eds. 2006. *Docentes en Latinoamérica hacia una radiografía de la profesión* [Teachers in Latin America: An X-ray of the Profession]. Santiago de Chile: PREAL.
- Wenglinsky, H. 2002. "How Schools Matter: The Link between Teacher Classroom Practices and Student Academic Performance." *Education Policy Analysis Archives*, 10 (12): 1–30.