

# TESTING OUR TEACHERS

Keys to a successful teacher evaluation system



Yyannú Cruz-Aguayo | Diana Hincapié | Catherine Rodríguez



# TESTING OUR TEACHERS



# TESTING OUR TEACHERS

Keys to a Successful Teacher Evaluation System

Yyannú Cruz-Aguayo | Diana Hincapié | Catherine Rodríguez

**Cataloging-in-Publication data provided by the  
Inter-American Development Bank  
Felipe Herrera Library**

Cruz-Aguayo, Yyannú.

Testing our teachers: keys to a successful teacher evaluation system / Yyannú Cruz-Aguayo,  
Diana Hincapié, Catherine Rodríguez.

p. cm. — (IDB Monograph; 749)

Includes bibliographic references.

978-1-59782-387-6 (PDF)

1. Teachers-Rating of-Latin America. 2. Teacher effectiveness-Latin America-Evaluation. 3. Teaching-Latin America-Evaluation. I. Hincapié, Diana. II. Rodríguez, Catherine. III. Inter-American Development Bank. Education Division. IV. Title. V. Series.  
IDB-MG-749

Keywords: Education quality, Teacher quality, Teacher evaluation

JEL Classification: I21, I28, J24, O57

Copyright © [2020] Inter-American Development Bank. This work is licensed under a Creative Commons IGO 3.0 Attribution-NonCommercial-NoDerivatives (CC-IGO BY-NC-ND 3.0 IGO) license (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) and may be reproduced with attribution to the IDB and for any non-commercial purpose. No derivative work is allowed.

Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the UNCITRAL rules. The use of the IDB's name for any purpose other than for attribution, and the use of IDB's logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this CC-IGO license.

Note that link provided above includes additional terms and conditions of the license.

The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the Inter-American Development Bank, its Board of Directors, or the countries they represent.



# CONTENTS

<b>List of Tables</b> .....	<b>vii</b>
<b>List of Figures</b> .....	<b>ix</b>
<b>Acknowledgments</b> .....	<b>xi</b>
<b>About the Authors</b> .....	<b>xiii</b>
<b>1 Teacher Evaluation: Identifying Effectiveness and Supporting Teacher Development</b> .....	<b>1</b>
<b>2 Why it is Important to Evaluate Teachers and How to Do it Well</b> .....	<b>7</b>
Instruments Available for Teacher Evaluation .....	10
Standardized Student Tests .....	11
Measure of Learning Objectives .....	13
Standardized Classroom Observations .....	13
Teacher Portfolio .....	15
Surveys of Students and other Actors .....	15
Standardized Tests for Teachers .....	16
Studies Analyzing Different Instruments .....	16
Impact Evaluations of Teacher Evaluation Systems .....	19
Impacts on the Composition of the Teaching Force, and Teacher Retention and performance .....	20
Impacts on Students' Academic Achievement .....	22
<b>3 How Teachers are Evaluated in Successful Education Systems</b> .....	<b>27</b>
Instruments Used in Teacher Evaluations .....	30
Weights Assigned to Each Instrument in the Teacher Evaluation .....	35
Consequences Associated to the Evaluation's Results .....	37

<b>4</b>	<b>Teacher Evaluation Systems in Latin America and the Caribbean</b> .....	<b>41</b>
	Teacher Performance Standards in the Region.....	46
	Regulation of the Teaching Profession in the Region.....	49
	Evaluations for Entry Into the Teaching Career.....	53
	Induction Programs for Novice Teachers and Evaluations for Approving the probationary Period .....	58
	Regular Mandatory Evaluations.....	62
	Evaluations for Horizontal Promotions .....	77
<b>5</b>	<b>Five Keys to a Successful Teacher Evaluation</b> .....	<b>85</b>
	<b>References</b> .....	<b>93</b>

# LIST OF TABLES

Table 3.1	Instruments Used in Teacher Evaluations in Selected High-Performing Education Systems	32
Table 3.2	Weights Assigned to Instruments in Teacher Evaluations in Some US States and Districts	36
Table 3.3	Consequences Associated to the Results of the Teacher Evaluation in Selected High-Performing Education Systems	38
Table 4.1	Evaluations Mentioned Explicitly in Teacher Statutes	53
Table 4.2	Number and Type of Instruments Used in Teacher Selection Processes in Latin America and the Caribbean	57
Table 4.3	Effective Implementation and Multiplicity of Instruments in Regular Mandatory Teacher Evaluations by Frequency of Application	64
Table 4.4	Instruments Used in Regular Mandatory Teacher Evaluations with Multiple Instruments	66
Table 4.5	Participants in Regular Mandatory Teacher Evaluations with Multiple Instruments	67
Table 4.6	Weights Assigned to Instruments in Regular Mandatory Teacher Evaluations with Multiple Instruments	69
Table 4.7	Instruments for Horizontal Promotion Evaluations in Countries Where this Possibility Exists and the Legislation Includes Meritocratic Aspects	80



# LIST OF FIGURES

Figure 2.1	Theory of Change in Teacher Evaluations	9
Figure 4.1	Types of Teacher Evaluations Analyzed in the Study	42
Figure 4.2	Progress of Teacher Performance Standards in Latin America and the Caribbean	48
Figure 4.3	Regulations of Teaching Careers in Latin America and the Caribbean over Time	50
Figure 4.4	Mandatory Basis, Centralization and Degree Required in Teacher Selection Processes in Latin America and the Caribbean	55
Figure 4.5	Evaluations in the Probationary Period for Novice Teachers in Latin America and the Caribbean	60
Figure 4.6	Regular Mandatory Evaluations in Latin America and the Caribbean	63
Figure 4.7	Results of Regular Mandatory Teacher Evaluation in Selected Countries	73
Figure 4.8	Consequences of the Results in Regular Mandatory Teacher Evaluations in Latin America and the Caribbean by Frequency of Implementation	74
Figure 4.9	Possibilities of Horizontal Promotions in Teaching Careers and Their Requirements, According to Teacher Statutes	78
Figure 4.10	Instruments Used in Horizontal Promotion Evaluations in Meritocratic Teaching Careers in Latin America and the Caribbean	82



# ACKNOWLEDGMENTS

*Testing our Teachers: Keys to a Successful Teacher Evaluation System* is a publication of the Education Division of the Inter-American Development Bank (IDB).

We especially thank Norbert Schady, economic advisor of the Social Sector, and Gregory Elacqua, principal economist of the Education Division, for their guidance and feedback throughout this project. We also appreciate the support of Emiliana Vegas, former Chief of Education Division; Sabine Rieble-Auborg, acting Chief of the Education Division; Marcelo Cabrol, Social Sector manager; Ana María Ibáñez, principal economic advisor in the Sectors and Knowledge Vice Presidency; and Ana María Rodríguez, vice president of Sectors and Knowledge.

Our colleagues in the Education Division provided us with contacts in the Ministries of Education and other authorities and provided useful feedback at various stages of the production of this study.

Several researchers participated in the preparation of the background documents used for this study: María Camila Rivera helped with the preparation of a study on teacher evaluation systems in Latin America and the Caribbean; James Wyckoff and Verónica Katz prepared a study on teacher evaluation systems in the United States; and Phillippa Codringley and Paul Crisp of the Center for the Use of Research and Evidence in Education (CUREE) produced a report on teacher evaluation in high performing education systems. Diana Paredes assisted in the publication process, particularly in the implementation of peer reviewers' comments and editing.

We thank the education experts and the authorities in Latin America and the Caribbean, who talked with us, clarify our doubts, and validated a large part of the information collected.

The study received comments from anonymous external peer reviewers. We also appreciate the observations of the discussants and participants in the Regional Policy Dialogue of the IDB Education Network, especially Santiago Cueto and Ximena Dueñas.

The comments and opinions expressed in this publication are the authors' and do not reflect in anyway the views of the Inter-American Development Bank or its Executive Board.



## ABOUT THE AUTHORS

**Yyannú Cruz-Aguayo** is senior economist in the IDB Labor Market Division. She is a specialist in impact assessments of social programs and recently co-led a research program on the impact of teacher quality on learning outcomes in early schooling. She has worked at the World Bank, Banco Nacional de Mexico and the Ministry of Finance of that country. She is author and co-author of academic articles (in journals such as *Quarterly Journal of Economics*) and book chapters on the effectiveness of social programs and early childhood development, among other topics. She is currently collaborating on the development of loan instruments for the Bank's member countries. She has a master's degree and a PhD from the University of Maryland, College Park. Her research agenda focuses on the effectiveness of labor interventions for disadvantaged populations.

**Diana Hincapié** is an economist at the Education Division of the Inter-American Development Bank (IDB). She leads research projects related to improving the quality of education, teacher policy, skills development, early childhood development and the extended school day. She works on the design and implementation of impact evaluations of education programs and policies and providing technical assistance for the preparation and implementation of loans and technical cooperations for Latin American and Caribbean countries. She is co-author of the 2017 IDB flagship publication *Learning better: public policies for skills development* and the book *Profession: Teacher in Latin America. Why was the prestige lost and how to recover it?* and is author of academic articles. She has worked in the IDB Research Department, the World Bank's Poverty and Gender Unit, and the Center for Economic Development Studies (CEDE) of the University of Los Andes in Colombia. She graduated in Economics from University of Los Andes (Colombia) and has a master's degree in Economics from the same university. She has a PhD in Public Policy and Public Administration from George Washington University.

**Catherine Rodríguez** is an economist with expertise on impact evaluations of education programs, the design of teacher policy, and on issues associated with the consequences of armed conflict on civilian population. She is the author of books, book chapters and academic articles in national and international journals. She currently works as a researcher at

the Faculty of Economics and the Center for Economic Development Studies (CEDE) of the University of Los Andes. She has been an associate professor in the same faculty and a consultant for the IDB, the Development Bank of Latin America (CAF) and the World Bank, among other institutions. She has an economics degree from the Del Rosario University (Colombia), a master's degree in Economic Policy and a PhD in Economics from Boston University.



## CHAPTER 1

# TEACHER EVALUATION: IDENTIFYING EFFECTIVENESS AND SUPPORTING TEACHER DEVELOPMENT

Quality education is essential for promoting the wellbeing of people, economic growth and the reduction of social inequalities in countries. It is not surprising therefore that the governments of Latin America and the Caribbean have made significant efforts in recent decades to expand the education coverage. Most Latin American countries have achieved universal enrollment in primary education, while almost 77% of children aged 4 to 5 and 80% of young people aged 15 to 17 are attending school (IDB-CIMA, 2019). With the expanded enrollment, average years of education have risen from practically zero at the start of the 20th century to nine today (Busso, Cristiá, Hincapié, Messina and Ripani, 2017).

Despite this undeniable progress, there are still large lags in the quality of the education in the region. For example, only about one third of fourth-grade students who participated in the Third Regional Comparative and Explanatory Study (TERCE) and one third of 15-year-olds who participated in the Program for International Student Assessment (PISA) achieved a minimum level of proficiency in math and reading (Busso et al., 2017). Consequently, it is necessary to continue moving forward with the implementation of policies aimed at improving the quality of the education in our countries.

Efforts to achieve this need to include policies covering all the aspects that are related to quality. On the education supply side, one of the fundamental variables that determines the quality of the education provided by a given system is its teacher quality. Research in the education area has shown that teachers are the most important school input for promoting learning. Some analyses of the literature have concluded that interventions involving teachers are particularly successful in improving educational quality (see, for example: Ganimian and Murnane, 2016; Kremer, Brannen and Glennerster, 2013; Evans and Popova, 2016;

McEwan, 2015). Moreover, there is evidence that the impact of a good teacher goes beyond learning in school and influence other areas, such as increasing graduation rates, the probability of accessing and completing higher education, labor income and even reducing teenage pregnancy (Jackson, 2018; Chetty, Friedman and Rockoff, 2014b).

For this reason, it is not surprising that some of the factors behind the success of education systems in countries such as Finland, Japan, Denmark, Switzerland and the Republic of Korea include having coherent policies for selecting, training and developing the teaching force, which has led to the profession being highly respected and valued. By contrast, as evidenced by Elacqua, Hincapie, Vegas and Alfonso (2018), the value and prestige of the teaching profession in Latin America fell in the last decades of the 20th century. In addition, the pedagogical skills and content knowledge of teachers in the region are below those suggested by international standards. For example, teachers in the region spend 20% less time teaching than recommended. The differences in effective teaching times implies that students in the region receive on average one day less of class a week (Bruns and Luque, 2015). Similarly, with respect to content knowledge, standardized tests applied to teachers from Peru, Chile and Mexico show that their math content knowledge is unsatisfactory (Elacqua et al., 2018).

To reverse this situation and achieve a true transformation of educational quality in the region, governments need to make serious commitments to developing and implementing a systemic policy to improve the profession's prestige and teacher quality. This policy must include actions throughout the teaching career, such as: i) attracting better candidates to teacher training programs; ii) improving the quality and relevance of the pre-service training; iii) merit-based selection of candidates for the teaching profession; iv) a clear, transparent and meritocratic framework to regulate the profession; v) a competitive salary structure that creates incentives for excellence; and vi) a teacher evaluation system that ensures continuous improvement (Cruz-Aguayo, Hincapié, and Rodríguez, 2020; Elacqua et al., 2018; Bruns and Luque, 2015; García et al., 2014).

All these policies must be aimed at developing and enhancing teachers' content knowledge and pedagogical skills. But this is not a simple task. For both developed and developing countries, the evidence suggests that even in the same country (with the same regulatory and administrative framework) there are significant variations in teacher quality. These differences are present even in the same school, where teaching practices and skills to promote student learning are quite different among teachers. This difference has been found in recent studies in the United States (Hanushek and Rivkin, 2010; Chetty, Friedman and Rockoff, 2014a) and in several Latin American countries (Araujo et al., 2016; Bruns and Luque, 2015).

Teacher evaluations can help identify differences in teacher effectiveness. A correct use of their results can provide the information needed to promote their strengths, overcome their weaknesses, and promote teacher excellence. **For teacher evaluations to achieve these objectives and be widely accepted, implemented and used in an education**

**system, they must be *valid and reliable*. This means that the information they produce must be objective, be really associated with the teachers' work, and reflect fundamental aspects of the students' learning processes, so that education authorities and teachers themselves understand and recognize the strengths and weaknesses of teachers' performance.** Moreover, it is only when the evaluation results are used effectively in professional development programs and/or to provide career incentives—including bonus pay, promotions or dismissal depending on the case—that the quality of teaching and student learning can really improve.

Education systems around the world are making great efforts in this respect and transforming teacher evaluation systems. Several factors can explain the evolution of teacher evaluations in recent years. The low impact of the indicators traditionally used to measure the effect of teachers on student learning—such as the teacher's education level or years of experience—has created incentives to search for measures that are more directly related to their productivity and effectiveness. The introduction of labor policies aimed at encouraging greater effort and assigning greater responsibility for student performance to teachers themselves can also explain this change. Finally, the search for academic excellence in education systems and the evidence of the importance of the work that teachers do every day are other factors behind the recent emphasis on teacher evaluation systems.

Latin American and the Caribbean countries are no stranger to this transformation. In recent years several countries have launched initiatives to improve teacher evaluation systems in the public education system. The region also has some examples of teacher evaluation systems that can be considered “second generation” systems, which are at different levels of development and implementation.<sup>1</sup> In a context where it is vital to make a leap forward in teacher effectiveness to improve the quality of education in the region, **this study presents some of the key factors for designing a successful teacher evaluation system.**

The study is organized into five chapters, including this introduction. The second chapter explains why it is important to evaluate teachers, describes different instruments that can be included in evaluation systems and the most relevant aspects to be considered in their design and expected results. The third chapter analyzes the teacher evaluation processes in 12 systems that offer their students high quality education. The fourth chapter details the current status of teacher evaluation in 19 education systems in Latin America. Finally, the fifth chapter presents policy recommendations for designing a successful teacher evaluation system.

Ensuring that teacher evaluation systems are reliable and valid is not trivial. At the time of their design and implementation, education authorities face a multiplicity of possibilities,

---

<sup>1</sup> As will be explained in detail later, these types of evaluations, consistent with the multiple tasks that teachers carry out, are characterized by having a multiplicity of instruments and evaluators with the objective of understanding in detail all aspects of teaching processes, and ensuring the characteristics of validity and reliability mentioned above.

including selection of instruments, deciding on how they can be implemented and the individual weights to be assigned to each instrument in the final score, among other elements. Likewise, once the evaluation process has finalized, the authorities need to decide how to deliver the information collected to teachers and, more importantly, how to use it to ensure improvement of teachers' performance.

It is worth clarifying that an important preliminary step for the design of evaluation systems and the choice of instruments has to do with the definition of what a good teacher needs to know and know how to do. These characteristics, generally defined in teacher performance standards, constitute an essential tool for guiding and laying the basis for all aspects of the teaching profession, including evaluation. Of course, it is not possible to define under a common framework what is expected of a teacher of excellence in all countries of the region. Even within each country, what is expected or required of teachers may vary depending on their context. For example, whether they teach at a rural school, a school with an indigenous population or a multigrade school. Although performance standards are a fundamental aspect of teacher evaluation systems, they are not studied in depth in this book.

However, it is clear that teachers are expected to have content knowledge and pedagogical skills appropriate to the subject and the grade in which they are teaching; understand the students' learning process and their context; and use teaching materials and technological resources in the classroom as a mean of ensuring effective teaching.<sup>2</sup> This is precisely what an appropriate use of the information produced by a valid and reliable evaluation system based on clear standards could help ensure, and what this book analyzes in detail.

**The second chapter summarizes relevant information for policymakers on some of the most common instruments for teacher evaluation and describes the most important aspects about their design and results.** The choice of the instruments to be included in a teacher evaluation is not trivial. The second chapter highlights two extensive and recent research programs on the subject. The first—the Measures of Effective Teaching (MET) project of 2013—led to the conclusion that, through a combination of instruments, including classroom observation, value-added measures and student surveys, it is possible to identify the most effective teachers i.e. who achieve the greatest impact on their students' learning outcomes. However, the conclusions obtained in a second study (Stecher et al., 2018) make it clear that implementing these lessons into practice is difficult and expensive.

---

2 The technological transformation is changing learning and teaching practices in the world and is an aspect that must be considered in the region. Technology can complement the work of teachers to strengthen learning in different ways (Muralidharan, Singh and Ganimian, 2019; Lai et al., 2012; Banerjee et al., 2007). For example, dynamic learning software can help students learn at their own pace and dynamically adapt to their knowledge. With this type of software, teachers can identify the strengths and weaknesses of their students and provide a personalized service.

It is only possible to obtain positive results when there is a careful implementation of multiple evaluation instruments.

**The third chapter analyzes teacher evaluation processes in 12 education systems characterized by offering their students high quality education.** Although it is not useful or desirable to replicate evaluation systems that work in different contexts, a review of international experiences can offer valuable ideas and lessons for understanding how these countries are handling difficult decisions on teacher evaluation processes. The systems included in the review are Estonia, Ontario (Canada), New Zealand, United Kingdom, Republic of Korea, and Singapore; and in the United States the analysis covers the systems of Denver Public Schools (DPS), District of Columbia Public Schools (DCPS), New Haven Public Schools (NHPS), New York City Public Schools (NYCPS) and the states of New Mexico and Tennessee. Although there are multiple lessons that can be obtained from the study of these cases, the review focused on three particular aspects that, in addition to being fundamental in the evaluation processes, allow a detailed comparison throughout the 12 systems: i) the type of evaluation instruments chosen in each system; ii) the weights assigned to each instrument in the total score of the teacher evaluation; and iii) the use of the evaluation results.

This analysis of international systems leads to three key conclusions. First, the teacher evaluations implemented in these education systems—except in three cases—encompass a multiplicity of instruments aimed at addressing the complexity of the teaching work. In all cases classroom observation is included, the crucial difference between them is in the details and the rigor of its implementation. For the states and districts analyzed in the United States, it is also common to include measures of student performance both objective—based on value-added measures or percentiles of student academic progress—and subjective, based on learning objectives. Some systems have also added surveys of students or parents, and to a lesser extent use of instruments such as self-assessment or portfolios of teaching work. Second, based on the analysis of United States cases, the instrument that receives the highest weight in the teacher evaluation is the classroom observation, followed by the instruments that evaluate student learning. Third, most systems link evaluation results to bonus pay or promotions in teaching careers. Strikingly few systems have explicitly defined aspects of in-service training and possibilities of tenure. Two exceptions are Singapore and the Republic of Korea, where in-service training is an essential part of the teacher evaluation processes.

**The fourth chapter details the current state of teacher evaluations in 19 education systems in the region: Buenos Aires province (Argentina), Belize, Bolivia, Santa Catarina (Brazil), Chile, Colombia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Mexico, Panama, Paraguay, Peru, Trinidad and Tobago, and Uruguay.**

The review begins with the diagnosis of two topics closely linked to evaluation: the existence of teacher performance standards and the level of meritocracy defined by the teacher

statutes in each system. The former, as already emphasized, is a crucial link to design and implement any teacher policy, particularly in the areas of training and evaluation. The latter aspect identifies what the explicit policy objectives are, focusing on evaluation and career promotion. After understanding these two aspects, the chapter systematizes the different evaluations currently implemented in the region at three points in the teaching career: i) entry evaluations; ii) evaluations during/after the probationary period and iii) evaluations for in-service teachers. This last stage analyzes two types of evaluations: mandatory, conducted with a defined frequency and with the possibility of dismissal in case of repeated low performance; and evaluations to access career promotions. For each of these evaluations, the chapter summarizes their main characteristics, the type of instruments used, the weight assigned to each instrument, the people responsible for their implementation, and the use of results by the education authorities and the teachers themselves.

Although progress has been made in the evaluation systems of the region, the analysis in Chapter 4 highlights three fundamental conclusions about teacher evaluations in Latin America. **First**, there are notable differences regarding the levels of meritocracy defined in the statutes that govern the teaching profession in the region, including aspects related to entry processes, promotions and leaving the career, among others. An analysis of teachers' statutes found that there are countries where teacher evaluations are not even mentioned, while in other countries the statutes mention explicitly and in detail the four types of evaluations considered in the study. **Second**, the *de facto* situation does not necessarily reflect the *de jure* situation in terms of implementation and use of teacher evaluation results. Although the legislation defines in detail the mandatory nature and labor-related consequences of teacher evaluations at different stages of the profession, these evaluations are not always applied, or if they are, they are not implemented according to the legislation or not all the consequences that arise from them actually occur. **Third**, despite all these shortcomings, many countries in the region are making valuable efforts to improve the attraction, retention and professionalization of the teaching force through evaluation processes. In fact, in some countries, evaluations that can be classified as *second generation* are being applied or have started. In addition, in a few countries evaluation has been accompanied by research into the processes and the impacts of implementing these evaluation systems, which has improved the instruments and protocols used. It is precisely this type of information that will allow principals and authorities to make informed decisions on how to improve the quality of education.

Finally, **the fifth chapter concludes with some policy recommendations and five keys to a successful teacher evaluation.** We hope that these suggestions, together with the analyses presented, will be useful for guiding and strengthening the efforts to design and implement teacher evaluation systems whose results can be used to improve teacher effectiveness. This is a necessary path to ensure that all children and youth in the region have access to effective teachers and through them to a high-quality education.



## CHAPTER 2

# WHY IT IS IMPORTANT TO EVALUATE TEACHERS AND HOW TO DO IT WELL

Research in the area of education has increased and has been greatly enriched in recent decades. The tools and statistical techniques for data analysis have improved. Likewise, the quantity and quality of the data and the information available have also made significant advances. Currently, not only there are better measures of the level of student learning, but researchers in the education area are collecting much more accurate information about teachers' skills and content knowledge.

These advances have demonstrated that teachers are the most important factor for student learning in school. It is also clear that there are significant differences in teacher quality, and that these differences are often even greater within a school than between schools. For example, in an analysis of the use of classroom time by teachers in Latin America, Bruns and Luque (2015) found that while one teacher in a school in Mexico spent 80% of the time on effective instruction, another teacher in the same school did not spend a single minute of class on it. In fact, according to the authors, these differences are so common that dispersion in the use of time between the “best” and “worst” teacher at the same school accounts for an average of two thirds of the total variation in learning among students at that school.

In this context, it is not surprising that students in the same school that are assigned to teachers with different levels of effectiveness achieve very different levels of learning. The analysis of Hanushek and Rivkin (2010) for the United States, and more recently demonstrated by Araujo et al. (2016) for Ecuador, suggest that students assigned to high-performance teachers master the content taught about 50% more than students assigned to low-performance teachers, which is equivalent to just over a school year. Recent studies show that teachers also have an impact on the socioemotional skills of their

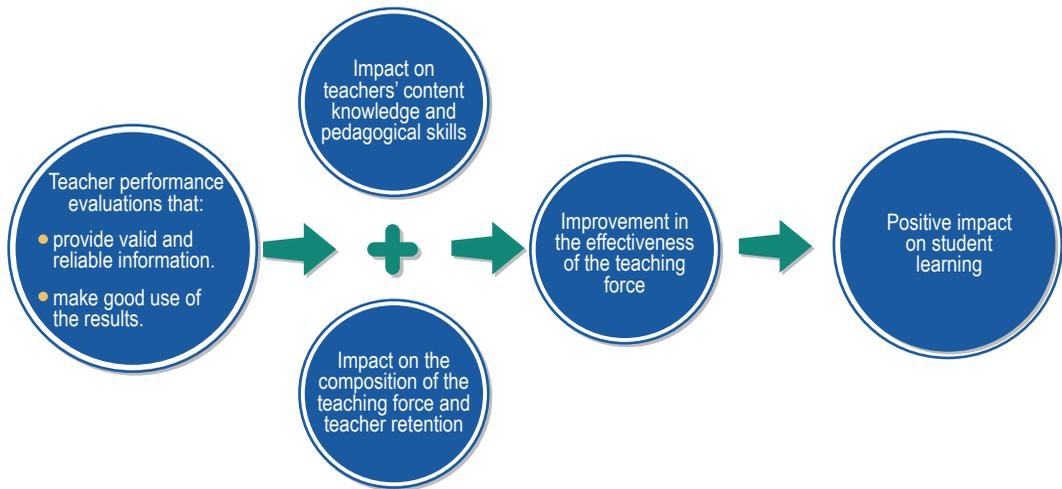
students measured by observable proxies in the behavior of the young people, such as absences, suspensions and high school graduation (Jackson, 2018), and also on direct measurements of these skills, including determination, the way of thinking and effort (Kraft, 2019), although the difficulty in measuring these skills has influenced the fact that some studies do not find significant impacts. Moreover, it is known with certainty that the impacts of good teachers not only affect the learning of their students in school, but also have an impact on longer-term results such as higher education decisions, work aspects and even issues such as teenage pregnancy (Chetty, Friedman and Rockoff, 2014b). It is important to keep in mind that, although we would ideally like to have teachers who will have a positive impact on all these skills and variables, the literature suggests that teachers who have impacts on learning are not necessarily going to have an impact on socioemotional skills and vice versa. This is an important aspect to consider when measuring and assessing teacher effectiveness.

All this has put policies aimed at attracting and training effective teachers and keeping them in the classroom at the center of the education debate. One of these policies is teacher evaluation. As will be shown in the following chapters, several governments in different countries of the world—including Latin America and the Caribbean—have made significant efforts in time and money to improve this specific aspect of the teaching career. These efforts are relatively recent, so research on this subject is scarce and concentrated in a few cases. This does not mean, however, that there are no important lessons to be learned by policymakers, academics and, in general, those interested in the subject. This chapter summarizes the main conclusions that can be drawn from causal studies that analyze aspects of the design, implementation and impacts of teacher evaluation on education systems.

The first question raised when analyzing the subject is how teacher evaluation can attract, train and retain the best teachers in the classrooms, and thus improve student learning. Figure 2.1 presents the theory of change, summarizing the mechanisms through which this can happen.<sup>1</sup> The process starts with the design and implementation of teacher performance evaluations based on instruments that provide valid and reliable information. An instrument (or set of instruments) of teacher evaluation provides **valid** information if the results of its implementation can identify, through rigorous research, teachers who are highly effective or have an impact on their students' learning. An instrument (or a set of instruments) provides **reliable** information when the results obtained from each teacher reflect their typical classroom performance and do not depend on the day on which the information was collected or on the person responsible for the collection. These two factors are fundamental for helping the education authority understand what the strengths

---

<sup>1</sup> This theory of change assumes that teachers with greater pedagogical skills and content knowledge effectively deploy them in the classroom and so improve the learning of their students. It is also assumed that teachers have the teaching resources and material support needed to be effective in the classroom.

**FIGURE 2.1 | Theory of Change in Teacher Evaluations**

Source: prepared by the authors.

and weaknesses of its teachers are and take the necessary steps to strengthen or overcome them, respectively. These factors are also essential if the teaching body is to accept the evaluation, trust it and act on it in line with its results. For this reason, when conducting the evaluations, it is essential to produce information and analyses that back up their reliability and validity, as recommended by international standards on education evaluations (AERA, APA and NCME, 2014).

The second necessary characteristic in the theory of change (reflected in the second point of the first circle from left to right) is the use of the evaluation results. Although some of the research detailed below demonstrates that mere participation in the evaluation process can have positive impacts on teaching practices (Taylor and Tyler, 2012), experts argue that the mechanism through which carefully designed and implemented evaluations can improve teacher quality depends on the use made of the results. According to the Organization for Economic Cooperation and Development (OECD) this use can have two objectives: i) improve teaching practices and content knowledge based on the diagnosis and linking to professional development programs designed to improve the results; ii) improve the composition and motivation of the teaching force by granting bonuses, special awards or promotions for teachers with excellent results or dismissing teachers—or at least not allowing them to teach in a classroom—in cases where they consistently show that they do not meet the minimum required conditions required by the profession (OECD, 2009 and 2013).

The problem raised in relation to these two objectives (reflected in the second link of Figure 2.1) is that they are difficult to achieve with a single evaluation tool. To be able to

detect aspects for improving teaching practices and content knowledge, and to recognize teaching excellence, teachers need to be completely open about revealing their practices and achievements and willing to share them with the authorities. Such openness would help to understand the training needs of the teaching force and to provide for them. However, if the evaluation results are also linked to aspects of the career such as salary levels, promotions and even job stability, acceptance of implementation of evaluation processes and the positive and open attitude of teachers and evaluators to them could be compromised.

The solution to this problem is not simple. In fact, as previous analyses show (OECD, 2013; Steinberg and Donaldson, 2016; Wyckoff and Katz, 2018; CUREE, 2018) and—as will become clear in the review of international experiences and of Latin America and the Caribbean reviewed in the following chapters—countries have chosen different solutions. Some education systems focus on using the results to achieve only one of the two objectives; others implement multiple types of evaluations to achieve a different objective with each evaluation, and others, in contrast, avoid this complexity and aim to achieve both objectives through a single evaluation. Obviously, there is no single and correct answer to this dilemma and the appropriate solution for each education system will depend on the context .

The last two links in Figure 2.1 close the theory of change underlying teacher evaluations. As can be seen, it is assumed that the proper use of the results obtained in teacher evaluations impacts the levels of effectiveness of the teaching force in both their practices and content knowledge (depending on the type of instruments used). Finally, this will end up by impacting the knowledge that students acquire in the classroom and improve the educational quality of the system.

Irrespective of the decisions on the use of the results, it is crucial for the system to ensure that the information provided is valid and reliable. In this respect policymakers have to make multiple decisions, since a wide range of instruments is available for teacher evaluation. The next section summarizes some of the most common instruments and what the research says about their effectiveness. The last section of this chapter summarizes what rigorous research has found about the impacts of the design and implementation of teacher evaluations in some education systems.

## **INSTRUMENTS AVAILABLE FOR TEACHER EVALUATION**

Since teaching is a complex activity, a holistic evaluation of teaching work requires a combination of different instruments. As a result, when designing teacher evaluation systems, education authorities must select not only the instruments to be used but also the weights to be assigned to each one. These decisions will have repercussions on the costs of the system, the level of acceptance by the different actors and the success of its implementation. For example, assigning greater weight to a given quantitative indicator can increase the pressure on the results obtained and increase the likelihood of distortion of processes that, it is supposed to monitor (Campbell, 1979).

Although a detailed review of available instruments, their characteristics and the requirements for their correct application is beyond the scope of this work, it is worth mentioning the most used instruments and what the most important research on them has found.

## Standardized Student Tests

The ultimate objective of an adequate design, implementation, and use of the results of teacher performance evaluations is to improve student learning in the system. It is not surprising, therefore, that various education systems use instruments based on standardized student exams in teacher evaluation. In fact, 97% of teachers included in the sample of the 2013 International Study on Teaching and Learning (TALIS) report that some measure of student performance is used as part of the teacher evaluations conducted in their schools (Smith and Kubacka, 2017). Consistent with this evidence, the review by Steinberg and Donaldson (2016) for the United States shows that this information is also widely used by that country's education authorities in their evaluation systems.

Among the instruments that use this type of information, the two most common are value-added measures of teacher effectiveness and student academic progress percentiles. In the first case, the previous results of tests performed on students are used to estimate, controlling for a set of personal and school characteristics, the expected result for each one. The differences between the expected value in each student's test and the actual value are averaged among all the students of each teacher, which results in the teachers' value added. In the second case, the percentile growth measures compare the growth in a student's achievement with the growth of other students who in previous years had similar performance measures. The teacher's effectiveness is then measured by estimating the growth in the average percentile of the students.

Proponents of the inclusion of standardized student tests as an instrument in teacher evaluations attribute three positive aspects to them:

- i. Although they are aware that student knowledge tests do not capture all the objectives schools seek to achieve, they recognize that these tests do measure student learning, which is one of the most important objectives.
- ii. The literature on impact has shown that teachers' present value-added measures are good predictors of value-added measures in the future (Kane et al., 2013; Araujo et al., 2016). For example, in the case of Ecuador, Araujo et al. (2016) find that, in addition to the experience and quality of teacher interactions with their students in the classroom, value added also predicts student learning.
- iii. Teacher value-added measures are correlated with aspects of individual development of their students in the long term, such as participation in higher education, earned salary at age 28 and teenage pregnancy (Chetty et al., 2014a and 2014b).

In contrast, there are arguments *against* the use of standardized student tests as part of teacher evaluations, as detailed below.

1. The first of these points out that standardized tests cannot identify the whole set of results desired from the teaching-learning process.
2. The second stipulates that, in part, student performance is outside the absolute scope of the teacher and depends on the students themselves and their motivation and interest in taking the tests. Since teachers cannot have an effective impact on student test results, it is not useful to assign a significant weight to their results.
3. The third argument states that, even if learning in traditional subjects such as math and reading is commonly accepted as a basic outcome that should be achieved in the education system, it is necessary to take into account that value-added measures are not usually available to all teachers in education systems but only to those whose students take the standardized tests on which they are based. This is not a minor point since, for example, in developed countries the information is only available for about 20% of teachers and in Latin America the percentage is even lower (Wyckoff and Katz, 2018).<sup>2</sup>
4. The fourth argument points out that, like any measure, its reliability depends on the validity of the student knowledge exams, a factor that is not always guaranteed (Lockwood et al., 2007; Papay, 2011).
5. The fifth establishes that when there are incentives linked to the results of the student learning evaluation, they must be carefully designed to ensure that they do not become “teaching to the test” strategies or even encourage copying or cheating<sup>3</sup> to increase scores and thus obtain incentives (Koretz, 2015; Stecher, 2002).
6. The sixth criticism is that calculation of value-added measures is complex, including methodologically, and requires decisions that can alter the results, such as selection of the type of controls of students’ and schools’ socioeconomic characteristics which are included in their estimates (Donaldson and Papay, 2015). Moreover, there are doubts about the degree of reliability of the use of these measures, since their annual correlation is relatively low and this only increases to acceptable levels with averages over five years (Kane, Kerr and Pianta, 2014).
7. Lastly, the seventh criticism is linked to the point when teacher effectiveness measures based on these results can be estimated and what they may be used for. In general, these measures are calculated and disseminated at the end of the academic year, or even later. More importantly, they provide little information for teachers and

---

2 To partly overcome the criticism that these measures are available only for a small percentage of teachers, some education systems have opted to use average value-added measures at school level. However, in these cases the distance between the teacher’s effectiveness and aggregate measures of outcome is even greater, which detracts from the use of such strategies.

3 For example, teachers may exclude some students with lower academic performance from the tests to try to get a better score.

the education authority to understand how each teacher can and/or must improve. According to Smith and Kubacka (2017), teachers consider that evaluation systems based significantly on this instrument have a limited impact on their teaching strategies and see them as a purely administrative and inadequate training process.

While many of these criticisms may be valid in certain contexts, the key to a successful teacher evaluation is finding ways to use the advantages they can offer, and incorporating into the actual design of the evaluation ways of mitigating the negative elements usually attributed to them.

### Measure of Learning Objectives

The instrument based on the measure of learning objectives is based on the objectives that teachers set at the start of the school year for their students. The teacher's supervisor reviews and approves these objectives and at the end of the academic year evaluates, together with the teacher, the progress achieved by the students. Based on this progress, teachers receive a performance rating. The most obvious advantage of using this type of instrument compared to the ones based on standardized tests is that it can be applied to all teachers in the education system, irrespective of their subject or. However, its level of standardization and objectivity is low. Although, in general, they are subject to objectives or competencies previously defined by each system for each knowledge area and grade, they ultimately depend on the autonomy each teacher has in their definition and on the rigor of the supervisors in evaluating the level at which these objectives have been met.<sup>4</sup> In fact, Lachlan-Haché et al. (2015), in one of the most complete studies on the topic, mention that the specialized literature concludes that implementation of this instrument is not only subjective but also requires a lot of the time of teachers and supervisors and has a very weak correlation with student results on standardized tests. Despite this, as will be clear in the reviews of international experiences, it is one of the most widely used instruments in most of the education systems of the United States and has a significant weight assigned to it in the final rating given to teachers.<sup>5</sup>

### Standardized Classroom Observations

Classroom observations of the teacher's work are the oldest teacher evaluation instrument of all. Even before the implementation of evaluation systems with multiple instruments and

---

<sup>4</sup> Although student progress can be evaluated through specific tests, the use of measures of learning objectives is based precisely on the difficulty of evaluating all students, in all subjects, every year through standardized tests.

<sup>5</sup> According to Steinberg and Donaldson (2016), 52% of states and 39% of the largest districts in the United States use this instrument. The relative weight assigned to the instrument varies between 8% and 25% in the total evaluation score.

with different consequences around the world, teachers have always been observed in one way or another by supervisors. What has changed in recent years is the structure in which these observations take place, the instruments used, and the consequences this action can have. In some education systems they are based on classroom observation rubrics or measures designed specifically for each system. Ideally, these rubrics should be based on teacher performance standards that in turn define what a good teacher is expected to know and know how to do in each country. Although there are rubrics already predefined and widely used in education research, such as the Framework for Teaching (FFT) (Danielson, 1996, 2007); Classroom Assessment Scoring System (CLASS) (Pianta et al., 2008); Stallings Classroom Observation System (Stallings and Knight, 2003); or TEACH (Molina et al., 2018), very often the systems must find the most appropriate rubric based on the objectives of the evaluation, the context in which it is framed and the time and resources available for it.

In that respect, Donaldson and Papay (2015) and Steinberg and Donaldson (2016) point out that this design requires making important decisions, for example, the number of observations for each teacher, the length of each observation, if they will be scheduled or surprise observations, who will be responsible for them (a supervisor, a peer in the school, a peer from another school or an outside expert) and the type of training that will be given to each observer or coder, among other aspects. These decisions also have direct consequences on the validity and reliability of the information collected through this instrument (Cohen and Goldhaber, 2016) and on the time and money required for implementation. For example, the CLASS observation instrument generates data on the quality of teachers' interaction during teaching, emotional support to the students and how teachers manage the class, while the Stallings instrument mainly captures the use teachers make of instructional time and some teaching practices. Because the CLASS instrument aims to capture a wider spectrum of teacher behavior, its implementation, although more expensive than the Stallings instrument, allows for more feedback which improves the teaching practices of the teachers evaluated. However, as Bruns, De Gregorio and Taut (2016) argue, in Latin American education systems, where a high percentage of teachers have not yet reached the minimum effective teaching times suggested by the OECD, it may be beneficial to start with implementation of simpler instruments.

Despite these complications, the use of classroom observation as an instrument has several advantages. Unlike the value-added measures, classroom observations encounter lower levels of controversy from teachers, for whom the results obtained from standardized tests restrict the objectives of teaching work to a single measure and do not give them relevant information about what to improve in their daily work or how to do it. Classroom observation, on the other hand, is rooted in the profession and if implemented correctly can provide teachers with specific information about their weaknesses and strengths. Studies such as the Measures of Effective Teaching (MET) Project or the study by Rockoff and Speroni (2010), among others, suggest that these measures are valid—are correlated with student performance on standardized tests—and reliable when designed and implemented correctly.

Although it is the oldest and possibly one of the most used instruments in the world, there is not yet enough rigorous evidence about its results (Cohen and Goldhaber, 2016). For example, it is not known if its results are also related to long-term impacts on students, something that is certainly the case with value-added measures (Chetty et al., 2014b). In addition, more research is required on what the crucial practices are that need to be measured and therefore corrected through training programs. Similarly, although the MET project conclusions outlined below are valuable, further research is needed on what is the most appropriate and cost-effective design to be applied depending on the specific objective to be achieved.

## Teacher Portfolio

A teacher portfolio is the set of materials compiled by teachers which makes it possible to understand their teaching process during the school year. The advantage of this tool is that it opens the possibility of a holistic view of teaching work by including the components of planning, evaluation and reflection and by collecting evidence of the quality of their practice and how this is adapted to the needs and progress of their students and depends on these elements. For example, in Chile the portfolio consists of three modules: i) description of the teaching practice and reflection on the teaching work, ii) a recorded class and iii) description of the collaborative work between teachers and reflection on it.

The portfolio is an instrument that seems to be particularly useful for self-reflection. However, little research has been done on its effectiveness. An exception is the study for the United States by Goldhaber and Anthony (2007) who evaluate the impact of the National Board for Professional Teacher Standards (NBPTS), the most recognized voluntary teacher certification system in the country which is based on development of a portfolio of teaching evidence. The results obtained from data from North Carolina state suggest that the system can identify more effective teachers both in the group who apply to the program and among those who never do. There is also evidence of the advantages of the portfolio for Latin America and the Caribbean: in an analysis of the correspondence between the different teacher evaluation tools in Chile, Taut et al. (2016) find a significant positive correlation of the results obtained by teachers in the portfolio score and the value-added measures of their students. However, since the portfolio in Chile includes a recorded class, as well as evidence and reflections on teaching practice and collaborative work, it is not clear whether the positive correlation comes mainly from the components of evidence and reflections or if it is due to the classroom observation component.

## Surveys of Students and other Actors

Recent evidence from the MET project and subsequent studies using its data (Polikoff, 2015) suggest that student surveys may be an important complement to teacher evaluations. As with the other instruments, if they are well designed, surveys can have an

advantage in terms of reliability since they depend on information collected by various actors, provide information from another perspective and are much less expensive than other instruments. Although there is a possibility that they face some bias, studies on students and teachers in higher education indicate that surveys are a valid instrument (Benton and Cashin, 2012). Other instruments used by different education systems are interviews with principals or supervisors, interviews with peers and parent surveys, among others.

### Standardized Tests for Teachers

Standardized tests applied to teachers are a useful tool that, depending on their design, allows to identify their pedagogical or content knowledge. As will be clear in later chapters, the use of this type of instrument is common within the educational systems of Latin America and the Caribbean, particularly in the evaluations associated with entering the career and in those required to ascend. It is a common instrument also in educational systems within the United States where it is a requirement for teacher certification. Probably, the greatest advantage of this instrument is that, once its design ensures that the questions included effectively assess whether the teacher meets the required performance standards, its implementation is much simpler and less expensive than other instruments. In addition, it is the only instrument for which there is positive causal evidence in the region that shows its implementation can effectively improve student performance (Ome, 2012; Brutti and Sánchez, 2017; Estrada, 2019).

### Studies Analyzing Different Instruments

The abundance of instruments offers the education authorities a disjunctive when choosing how many and which to include in their teacher evaluation systems. In addition, after answering these two questions, policymakers must decide what weight to assign to each one in the final rating. These decisions will have an impact on the validity and reliability of the evaluations, so they need to be based on the available evidence in each context and, if not available, on the international evidence.

The MET project is one of the most influential and extensive studies on the subject, created with the objective of constructing and testing measures of teacher effectiveness for use as inputs in teacher evaluation processes to provide feedback for teacher classroom practices and increase their effectiveness.<sup>6</sup> The project worked with about 3,000 teachers from seven school districts in the United States from 2009 to 2013. Three types of instruments were analyzed: classroom observations (with both generic and

---

<sup>6</sup> The results are available at: <http://k12education.gatesfoundation.org/blog/measures-of-effective-teaching-project-faqs/>.

specific observation protocols for science and math), student perception surveys and measures of gains in student learning based on standardized tests.<sup>7</sup> One of the conclusions from the study is the validity of student surveys and classroom observation protocols which, carefully designed and applied, are able to identify and provide feedback on teaching practices which in turn are correlated with the level of student learning. In that context, three major questions that researchers in the MET project (2013) sought to answer are outlined below.

The first question revolves around whether effective teaching measures can identify teachers who promote greater learning in their students. To find out, the analysts collected teacher efficiency measures based on classroom observations, student surveys and value-added measures adjusted for the socioeconomic characteristics of the students for the 2009–10 academic year and constructed a composite measure. The following year they conducted an experiment where they assigned students to each teacher randomly and then determined whether those assigned to teachers identified as more effective according to the composite measure learned more. The results suggest that these students in fact performed better on traditional standardized tests and even on tests with higher levels of analytical requirements than students assigned to less effective teachers.

The second big question they attempted to answer in the MET project relates to the optimal weighting each instrument should have in a performance evaluation. Due to the complexity of teaching work, conducting an evaluation with a multitude of quality instruments is better than making an evaluation with a single instrument. However, there are not many studies that identify which combination is most effective. The researchers opted for four different weightings that assigned alternative weights to value-added measures, classroom observations and student surveys. According to the authors, an equal weight between the three measures achieves the highest correlation with results in different standardized tests on students and greater reliability over the years, so they consider that it is one of the best options.<sup>8</sup> A weighting of 50% to classroom observation and 50% divided equally between student surveys and value-added measures gives similar results. Thus, the

---

<sup>7</sup> Classroom observation protocols included the Danielson Framework for Teaching; Classroom Assessment Scoring System (CLASS) of the University of Virginia; Mathematical Quality of Instruction (MQI) of the University of Michigan; Protocol for Language Arts Teaching Observations (PLATO) of Stanford University. In addition, specific protocols were used from certain areas of knowledge, such as the Quality of Science Teaching (QST) of Stanford University, and the UTeach Teacher Observation Protocol (UTOP) of University of Texas-Austin, for mathematics and science. The student surveys were the Tripod Student Perception Survey, developed by Ron Ferguson. The knowledge exams included state exams and the Stanford 9 Open-Ended Reading Assessment, the Balanced Assessment of Mathematics (BAM) and the ACT QualityCare for algebra, English and biology.

<sup>8</sup> As expected, the authors found that a weight that assigns 80% to value-added measures based on state standardized tests is the one that best predicts of students' future performance in these same tests. However, it is not necessarily the best option for capturing other aspects of teaching. In fact, it is the combination that obtains the lowest correlation with other types of standardized tests applied to students.

MET study concluded that a combination of instruments such as classroom observation, teacher value-added measures and student surveys is valid.<sup>9</sup>

Lastly, the MET research project set out to determine the best way to implement classroom observation instruments to ensure its reliability, so that the results consistently reflect aspects related to teaching practices and are not subject to other factors such as who the observer is or if it is a particularly bad day for the teacher. The research concluded that increasing from one to two observations by the same person increases the reliability of the instrument from 0.51 to 0.58 on a scale of 0 to 1. If the second observation is made by a second person, the reliability increases further (to 0.67). Based on this result, the MET project investigated whether it was necessary for the second observation to be made during a complete class (45 minutes) and found that if the second class is observed by three different people in 15-minute blocks, the same level of reliability is obtained as if the observation of the whole class is done by one person.

Although the impacts reflected by the MET project and the particular studies referred to above are positive, it is important to highlight cases in which, despite the efforts in time and money made by education systems to design and implement teacher evaluation processes, they may not lead to the expected impacts. A case is the RAND Corporation and the American Institute for Research project, also funded by the Gates Foundation. This project aimed to improve students' learning and their admission to higher education, particularly of young people from minority groups, by increasing access to excellent teachers. Between 2009 and 2016, three education districts and four charter management organizations agreed to participate in the study. All were committed to developing in their systems a detailed teacher evaluation process and reflecting its results in the professional development and teaching career of their teachers.

The final project report contains valuable lessons for all interested in the subject (Stecher et al., 2018).<sup>10</sup> The first lesson relates to implementation of these evaluation systems. Although in all participating schools, teacher evaluation systems were designed and applied based on classroom observation rubrics and, where possible, on students' academic results, their implementation was not simple. For example, in most schools' observations were made by principals, which significantly increased their workload. This situation led to a reduction in the number or length of the observations in some systems, or designation of another internal actor, without the same level of preparation, as an evaluating agent to do the work. Also, as is the case in many of the systems analyzed in later chapters, most teachers were rated in the most efficient performance categories, a fact that diminished the power of differentiation of the instrument in terms of teacher quality.

---

<sup>9</sup> The study also found that the score obtained by teachers in tests of their content knowledge and pedagogical skills, was not correlated with student learning.

<sup>10</sup> The results of the study are available at: <https://www.rand.org/education-and-labor/projects/evaluating-teaching-effectiveness.html>.

Consistent with other studies, the project found that implementation of teacher evaluation increased the exit rate of ineffective teachers while failing to retain or hire more effective teachers. The explanation for this is that the results were not used as expected. For example, although some policies that associated the results with career promotions or bonus pay were implemented, it was not done to the extent expected, their design differed from that suggested by the researchers at the start of the project and in some cases they were associated with aspects that were not important for teachers.<sup>11</sup> Moreover, the systems were designed to integrate professional development programs based on the results, but the research showed that the schools did not know how to do this and their implementation was not complete. As explained by the authorities of the institutions that participated in the project, the results and the feedback were not detailed enough for designing suitable professional development programs, there were no programs to be used as models and the education systems did not have the time or monetary resources to design them. In addition, when general professional development and in-service training programs existed, teachers were under no obligation to participate, there was no supervision of which teachers took them, and there was no control over whether teaching practices of teachers participating in the programs improved as result.

Given these results, it is not surprising that the research found no impact on access to more effective teachers, on student performance in standardized tests and graduation rates. However, according to the researchers, the analysis of these results should consider that, in three of the four states where the study was conducted, evaluation systems were implemented with wide-ranging consequences, which contaminated the control groups. The authors also suggest that there may not have been enough time to observe results. Even so, the study leaves important lessons for policymakers and academics interested in the subject: i) if implementation of evaluation systems is not accompanied by strong mechanisms to make effective use of the results, the process is basically incomplete; ii) as discussed above, the authors emphasize that the systems reported difficulties in trying to use the same results to achieve the objectives of development and of implications for the professional career.

## **IMPACT EVALUATIONS OF TEACHER EVALUATION SYSTEMS**

Impact evaluations are intended to respond to a specific cause and effect question: what is the impact (or causal effect) of a practice or program on a result of interest? (Gertler et al., 2016). Although many countries evaluate teachers in one way or another, there is limited

---

<sup>11</sup> For example, the study mentions that bonuses for obtaining exceptional results were not significant and that, although some specific roles were created for teachers with outstanding results in evaluations, a system of career advancement as imagined by researchers at the start of the project was not implemented.

knowledge about the causal effects of evaluation practices on teacher performance and, ultimately, on student learning. As mentioned before, the aim of teacher evaluation and the appropriate use of results through training programs and career incentives is to improve student learning. This section summarizes the main conclusions drawn from the impact assessments that exist so far, based on a review of the three steps of the theory of change (Figure 2.1). First, the impacts on the teaching force are summarized through changes in the composition of the teaching force and improvement in teaching practices, followed by the impacts on student learning.<sup>12</sup>

### Impacts on the Composition of the Teaching Force, and Teacher Retention and Performance

Two studies for the United States suggest that holistic teacher evaluation systems can have significant impacts on the composition and retention of the teaching force. The first study by Dee and Wyckoff (2015) analyzes the impact the incentive system had on the teacher evaluation implemented in the public schools of the District of Columbia (known as IMPACT), on the differential rotation of teachers rated as effective and ineffective by the system itself, and on the performance of teachers who remain after evaluation. The IMPACT evaluation system was introduced on a general basis in academic year 2009–10 with the explicit objectives of improving teaching quality and student learning. Under IMPACT, teachers received a final rating that depended heavily on results obtained from rigorous classroom observations linked to the district’s performance standards (teaching and learning framework). The rating ranged from 100 to 400 points and the points obtained defined each teacher’s position in one of four possible performance categories: ineffective (100–175), minimally effective (176–250), effective (251–350) and highly effective (351–400).

The system considered five observations during the year: three were conducted by a school administrator (principal or assistant principal) and two by an external expert. The only announced observation was the first one, which was conducted by the school administrator. For a small group of teachers (17%) whose students took part in standardized tests, value-added measures based on those tests were also included. For teachers who did not have the possibility of calculating value added, a measure of learning objectives was incorporated for the students set by the teacher and the principal. At the end of the year the school administrators evaluated to what extent these objectives had been achieved. There were also two additional measures: commitment to the school and the community and a

---

<sup>12</sup> It is also interesting to analyze the literature that assesses the impact of the individual or group monetary incentives that teachers receive based on the performance of their students. Since it does not focus on evaluating teachers’ pedagogical skills and content knowledge, this analysis has not been included in this study. However, it is worth mentioning studies such as Springer (2009), Fryer (2013), Bruns and Luque (2015) and Barrera-Osorio and Raju (2017).

value-added measure for the school. Finally, the design included improvement mechanisms through tutors for teachers who required help based on the results obtained.

The system consisted of two types of incentives based on the results obtained in each teacher evaluation. The first consisted of a monetary incentive with a bonus for teachers rated as highly effective. Likewise, the system ensured a permanent increase in the base salary (equivalent to three to five years' service) for teachers rated as highly effective on two consecutive occasions, and immediately dismissed teachers rated as ineffective. The second type of incentive is associated to job stability. In that framework, teachers rated in the minimally effective category for two consecutive years were dismissed. Thus, both the IMPACT incentives were not only linked to multidimensional measures of academic performance associated with higher levels of reliability and transparency—as explained above—but were also based on actions that teachers themselves directly control.

To estimate the impact of the evaluation, the researchers compared the tenure and performance of teachers whose scores in the previous year's evaluation were close to the limits that separated an effective from a minimally effective teacher (about 250 points) and in the limit between teachers rated as effective and highly effective (350 points).<sup>13</sup> The results show that the threat of dismissal increased the voluntary withdrawal of teachers rated as minimally effective in the first evaluation. In addition, the performance of teachers rated in this category who remained in the system improved in the following evaluation, measured by the evaluation score in later years. Lastly, the authors find that the monetary incentives given to those rated in the upper tail were also effective in improving their results in later evaluations. The study also reveals that impacts grow over time, as bonus pay and threats become more credible. However, the authors cannot differentiate what part of the impacts came exclusively from use of the evaluation and what part from the incentives.

Sartain and Steinberg (2016) answer a similar question based on analysis of the impacts of a pilot teacher evaluation system, Excellence in Teaching Project (EITP). This program, applied in a group of Chicago public schools, was aimed at improving the quality of student learning by improving the feedback provided by the principals to their teachers with a view to improving classroom teaching practices. The program used the Danielson rubric (Framework for Teaching, FFT) and implemented a deep program to train the principals in its correct use. The teachers were not subject to a specific benefit or punishment based on the results obtained. The authors analyze two aspects of the impact of the evaluation system: i) exit of teachers from the school where they work, but not from the Chicago public school system (rotation within the district) and ii) the definitive final dismissal of teachers from the city's public school system. In this case, to find the impacts, a randomized experiment was conducted in 92 selected public schools: 44 of the schools received the treatment in the 2008–09 academic period and the remaining 48 schools in the next academic year.

---

<sup>13</sup> This is a quasi-experimental methodology, widely used in specialized literature, known as the Discontinuous Regression methodology.

Consequently, it is a very short-term evaluation in which the indicators of interest were compared one year after the program was implemented in the first 44 schools. The results suggest that the EITP had no impact on the average teacher turnover rate in the district. However, by separating teachers by their level of effectiveness (according to evaluations prior to 2009), the authors find that the probability of leaving the Chicago public school system increases for teachers rated as unsatisfactory or satisfactory, very similar to the recent result found by Stecher et al. (2018). They also determine that the probability of leaving the system is greater for teachers with few years of experience. Thus, the authors conclude that this type of evaluation system has the potential to exit teachers with low levels of effectiveness from the education system. The authors also point out that the new teachers hired in the schools, as well as the teachers who remained after implementing the system of observation and feedback on teaching practices, have a higher score in the FFT than the teachers who left.

Both studies suggest, therefore, that classroom observation with rubrics linked to clear standards, careful and intensely trained observers and continuous feedback for teachers on their practices can achieve improvements in teaching quality and promote the exit of teachers with deficient practices. Moreover, in the case of Chicago, the training of the principals themselves on how to evaluate their teachers produced positive impacts on the quality of teachers hired in the system. These results are consistent with previous evidence also available for the United States in which delivery of information to school principals on the relative effectiveness of teachers, measured from value added estimates, has an impact on retention of ineffective teachers and, through this channel, on student results (Rockoff et al., 2012).

## Impacts on Students' Academic Achievement

As Figure 2.1 shows, the ultimate objective of teacher evaluation is to ensure that the system has excellent teachers that effectively promote student learning. The study by Taylor and Tyler (2012) analyzes this possibility with data from teacher evaluations conducted in Cincinnati, United States. The authors use the progressive application of the evaluations to compare a group of teachers who participated in the program first, with another group that began to participate later. The state evaluation program made four methodological classroom observations during one academic year. Three of them were conducted by two experts external to the school—chosen because of their own levels of effectiveness and expressly designated for that purpose. The fourth was conducted by the principal or administrator of the school itself. The observations were based on Danielson's rubric; while all the observers had rigorous training on how to conduct them. The first observation was announced during the week it was to take place, but the following observations were not notified in advance. At the end of the annual process, each teacher received a rating as part of a final performance report on one of four levels: distinguished, effective, basic and

unsatisfactory. The evaluation results had few implications for career stability: excellent levels could be linked to promotions, low levels led to assignment of peers for an improvement program, and in very few cases they led to termination of the employment contract. About 90% of teachers were placed in the first two categories, which confirms the low impact that the evaluations per se could have on career.

The authors use a methodology to analyze how each teacher's effectiveness changes before during and after going through the evaluation process.<sup>14</sup> They found that teacher productivity, measured as learning based on student score in their math class, increases during the year when they are evaluated and even increases in greater proportion in later years at end of the evaluation process. Students raise their standardized math test score by 0.1 standard deviations more than a student with similar socioeconomic characteristics and similar prior levels of knowledge who was assigned to the same teacher but before he or she went through the evaluation process. Although the data cannot identify the channels which could explain the impacts, the authors suggest that they are based on the learning teachers obtain during the evaluation process about their performance and teaching practices, and the subsequent development of additional skills, as well as greater future effort.<sup>15</sup>

Steinberg and Sartain (2015) analyze the impact of the EITP pilot teacher evaluation system in Chicago (described above) on student learning. As mentioned before, the pilot began in 2008-09 in 44 primary schools (first cohort) and in 2009-10 48 additional schools (second cohort) were included. The authors find that students attending schools whose teachers were treated in the first cohort get better results in reading and math than students attending schools whose teachers were treated in the second cohort, although only the language results are statistically significant. They also find that this advantage is maintained even after students in the second cohort have been treated, and that the impacts were greater for schools attended by students with higher average income levels.

In their conclusions the authors highlight three important points. The first is that immediate results could be obtained in student learning because the system was based on classroom observations made by the school principals, who then discussed in detail the results and possible improvement strategies with teachers. However, this scheme had two implications: i) training the principals, and in general the person responsible for observation and feedback, must be intense and of the highest quality; ii) implementation of the evaluation and the subsequent discussions put intense pressure on principals in time and

---

<sup>14</sup> In technical terms, the authors use a fixed effects strategy by teacher. In their estimates they also consider the students' previous scores, which did not enter the evaluation process of the Cincinnati public schools they were analyzing.

<sup>15</sup> This same analysis was carried out with teachers in the language area through reading tests. In this case the authors do not find any positive impact and do not offer explanations in this respect, apart from suggesting that students learn and practice reading in contexts that are outside the school much more than math and that, therefore, the effectiveness or differences between teachers in this area of education are per se less.

work which policymakers need to take into account. Finally, the authors point out that, in general, implementation of the evaluation was significantly weakened from the second year, to the point where the program was in general never applied after the pilot. According to them, this explains the lower impacts observed for cohort two and underlines the need for a strong system in order to obtain positive results. In that respect, they stress the importance of proper implementation faithful to the original design.

Adnot et al. (2017) also found positive effects of teacher turnover on student learning as a result of the IMPACT program.<sup>16</sup> The authors find that DCPS replaced teachers who left with teachers who increased student achievement by 0.08 standard deviations in math. When only considering the effects of lower-performing teachers who were induced to leave the system for poor performance, the student achievement improves by 0.14 standard deviations in reading and 0.21 standard deviations in math. The authors also mention that the effect of the restructuring on the effectiveness of the group of teachers in the system was greater in schools with high levels of vulnerable population, which means the improvements in learning are concentrated in these students. In turn, this implies that the impacts on these types of students are even greater and are equivalent to about 35%–65% of their average learning in a year.

For Latin America and the Caribbean, the evidence on the impacts generated by teacher evaluations is scarce and focuses on a few cases. However, it suggests that teacher evaluations can have positive impacts on student learning. For example, Ome (2012) and Brutti and Sánchez (2017) in Colombia, and Estrada (2015) in Mexico find that evaluations at the start of the teaching career have positive marginal impacts on indicators of efficiency and quality of the education received by students. For Chile, although it is not possible to classify it as an impact assessment of the teacher evaluation system, the study by Taut et al. (2016) finds evidence suggesting that learning by students whose teachers scored higher in the portfolio instrument is greater than learning by students assigned to other teachers.

While not analyzing a teacher evaluation system as such, a related study (Santibáñez et al. 2006) estimates the impact generated by the introduction of the Teachers' Career ("Carrera Magisterial") program in Mexico. This competency-based pay program, implemented in the country between 1992 and 2015, gave teachers the possibility of achieving a promotion or horizontal promotion. In horizontal promotions, the teacher's prestige, functions, responsibilities and compensation can increase without the need for them to abandon the classroom or the focus on teaching. To access a promotion, interested teachers were evaluated annually in six dimensions: i) maximum grade of studies, ii) years of service, iii) professional performance according to the ratings given by their peers, iv) score in the

---

<sup>16</sup> The authors use information at grade-school (primary) level and a methodology of differences in differences which assess the impact of teacher turnover from that particular grade-school, including those who left the system due to IMPACT, has on student performance.

state and federal evaluations of professional development courses, v) scores in teacher knowledge tests and vi) average scores on standardized tests of their students. According to the result obtained, teachers were rated in six possible ranges which then determined the amount of the bonus.

The authors analyze both the reliability and internal consistency of the tests applied and the impacts of the program on the quality of the education the students received. In this respect, they point out that, although the teacher tests were aligned with the curriculum and had adequate levels of reliability and internal consistency, the questions had a low level of complexity and there was room for improvement. On student exams, the authors mention that reliability and internal consistency were lower than in the teacher tests, and the required cognitive ability was also low. Taking this into account, the impacts of the Teaching Career on student learning are mixed. Although at secondary level there are improvements of between 3% and 15% of standard deviation in the standardized tests of students whose teachers were classified in the higher incentives group. No impact is found for primary students and there are even some negative impacts for primary students whose teachers had already been classified in the higher incentives group in previous years.

Overall, the evidence on teacher evaluations shows that they have the potential to improve educational quality in the systems where they are applied. However, for this to happen, policymakers must understand that design and implementation are fundamental factors. If design and implementation of teacher evaluations do not ensure that the instruments and information that result from them are valid and reliable, the resources invested in them are unlikely to produce real fruits. Even if this first condition is met, the way the results produced by the evaluations are used is also important. Although there is evidence to suggest that teacher effectiveness improves through participation in evaluation processes, the theory of change clearly indicates that it is necessary to include formative processes or career incentives linked to the evaluations. This means that the decisions that need to be made on the issue are not only complex but also expensive. Understanding how other education systems have dealt with this dilemma can be informative and enriching for future decision making. This is precisely the aim of the next two chapters, which analyze how the high performing education systems in other regions and 19 Latin American education systems have addressed these dilemmas of public education and teacher policy.





## CHAPTER 3

# HOW TEACHERS ARE EVALUATED IN SUCCESSFUL EDUCATION SYSTEMS

The previous chapter set out the multiplicity of decisions that policymakers must make when designing, implementing and using information from teacher evaluation systems. This multiplicity also suggests that there cannot and should not be a single solution to these dilemmas for all education systems. Each country is immersed in different contexts and the teacher evaluation systems finally implemented must match the objectives, priorities and possibilities that each government wants and can achieve with them. What works in education systems with good education performance will not necessarily work in the education systems of Latin America and the Caribbean, either because of budgetary reasons, experience in conducting evaluations, or the characteristics of the teaching staff, among other causes. However, an analysis of the main characteristics of teacher evaluation in education systems that are recognized for their educational quality can offer valuable insights into their design, implementation and the consequences.

This is precisely what this chapter does, through a general review of the recent history and main characteristics of teacher evaluation in education systems considered to have a good education performance.<sup>1</sup> Specifically, the analysis covers teacher evaluation systems for Ontario (Canada), United Kingdom, New Zealand, Singapore, Republic of Korea and Estonia. These education systems were selected for their high performance in standardized student tests and/or because they have innovative designs of teacher evaluation. In the United States, the systems analyzed were Denver Public Schools (DPS), District of

---

<sup>1</sup> This analysis is based largely on the background documents produced by Wyckoff and Katz (2018) and CUREE (2018) for the “Measuring teacher effectiveness” project, and on the recent study by Steinberg and Donaldson (2016) which summarizes the teacher evaluation systems of 50 states and 25 school districts in the United States.

Columbia Public Schools (DCPS), New Haven Public Schools (NHPS) and New York City Public Schools (NYCPS) and the states of New Mexico and Tennessee. They were chosen because they incorporate value-added measures in their teacher evaluations.

To understand the final decisions on the design and implementation of teacher evaluation systems, it is important to start by learning the history behind each case analyzed. In the United States, development of teacher performance evaluation systems, which use a combination of instruments and aim to understand and differentiate teaching practices in schools, is a relatively recent phenomenon. According to Wyckoff and Katz (2018), a combination of at least three national and international factors were influential in promoting the introduction of these systems. First, the evidence of national and international standardized knowledge tests, existing from the 1980s onwards showed that the quality of education provided by the system was not ideal and that concrete actions were needed to improve it. Second, some studies showed that: i) teachers were the most important of all school inputs for student academic performance (Aaronson, Barrow and Sander, 2007); ii) teacher efficiency/quality in the country varied significantly even at school level (Aaronson, Barrow and Sander, 2007; Rivkin, Hanushek and Kain, 2005; Rockoff, 2004), and iii) teacher evaluation in the country's schools was simple, rating the majority of the teaching body with satisfactory performance, without providing valuable feedback to improve teaching quality (Weisberg et al., 2009). Third, the clear evidence provided by the two previous points created the space for specific programs to be promoted and financed with the objective of improving the educational quality of schools and developing detailed teacher evaluation systems.<sup>2</sup>

The flow of new federal funds since the beginning of the millennium, coupled with pressure from various sectors, influenced many states and school districts in the United States to radically reform teacher evaluation systems. In addition to including multiple instruments to identify teacher performance, in many cases they also linked the results to consequences in monetary or employment terms. The idea was, of course, to base the design and implementation of these systems on the available evidence, which laid the bases for ensuring reliable and valid evaluation systems that would lead to a real improvement in educational quality. However, the short period given by the federal programs to implement the new systems, the high monetary cost involved and opposition from teachers' unions weakened their implementation in many cases (Stecher et al., 2018).

The review of the other education systems included in the chapter also shows the diversity in terms of history, maturity and objectives of the teacher evaluation implemented in each education system. For example, the Republic of Korea system, created in 1964 and reformed over the years and for the last time in 2010, establishes a mandatory evaluation

---

<sup>2</sup> Three well known programs that appeared in response to academic evidence and a sense of urgency about the need to improve educational quality in the country include Race to the Top (2009), the Teacher Incentive Fund (2006) and No Child Left Behind (2001).

model for all teachers working in a school for more than two months, based on three components.<sup>3</sup> The first component in the system is “Performance Evaluation” which checks the characteristics, attitudes and work of all teachers in the country’s public schools through surveys of students and peers, and classroom observation, as will be detailed later. The second component is the “Performance-based incentive system” whose function is to determine teacher salary based on performance, and to raise their levels of motivation and educational quality. Finally, the third component is “Evaluation for professional development,” introduced in 2010 and designed to provide feedback to teachers and develop their skills. This professional development component is based on the evaluation results of the first component but adds a detailed self-evaluation report in which teachers reflect on their goals, strengths and weaknesses. The three components are complementary and have been introduced gradually since the education authorities realized that a single component was failing to achieve the evaluation objectives (detailed in the theory of change of the previous chapter, figure 2.1).

Although Singapore does not have three different evaluation systems, the country’s education authorities developed a central system known as the “Education Service Professional Development Plan,” which has three complementary components: the Enhanced Performance Management System, Recognition through Monetary Bonuses System and a Career System with three different options (classroom teacher career, education leader career, experienced specialist career). The teacher evaluation system, linked to the first component, was announced in 2001 and finally implemented in 2005. It is a competency-based system that determines the knowledge, skills and abilities that teachers must demonstrate at each stage of their career, with a double objective:

- i. Promote the professional development of teachers by reflecting on their achievements and potential and, through this process, improve the educational quality offered to students.
- ii. Link the teacher evaluation system with decisions on salary levels, bonus pay and career progression.

Unlike these two modalities, the remaining four cases analyzed have much less centralized and detailed evaluation systems. The teacher evaluation system in Ontario province (Canada) dates from 2002 and was last reformed in 2010. This is the third education system in terms of a high level of centralization in the decisions and requirements that must be included in an evaluation. However, unlike the other two cases mentioned above, since its inception, evaluation in Ontario has aimed above all for a single objective: professional

---

<sup>3</sup> To some extent this is the case most similar to what is happening in the Latin American region where different types of evaluations in use in each country coexist with different purposes, as will become clear in the next chapter.

development through a diagnostic evaluation that provides feedback to teachers on their classroom practices. This high level of centralization is followed by the United Kingdom, whose system, developed in 2000 and reformed in 2006 and 2012, was initially aimed at ensuring job satisfaction and high levels of excellence in the teaching staff. Although the country's regulations detail some basic principles for the evaluation, the responsibility for the system relies in collegiate bodies at the level of each institution.

Of all the systems analyzed, the New Zealand system is the most recent, with implementation beginning in 2018. Both its design and implementation may vary in different schools, but the system must be based on guidelines specially designed by the central authority. Finally, in Estonia the evaluation system required by the legislation is at school not central level. This is consistent with the fact that teachers in the country enjoy the highest level of autonomy of all member countries of the Organization for Economic Cooperation and Development (OECD), which means that little is known about the criteria used in the evaluations of schools and even less of teachers. However, the government issues guidelines for decision-making by principals with inclusion of observation of their teachers' classes as part of the internal evaluation of the schools.

Despite the clear differences in their history, it is possible to observe three fundamental aspects of teacher evaluation systems for the 12 cases presented here. Specifically, the focus of the analysis of these education systems is centered on three specific aspects: i) the instruments used in each case; ii) the weights assigned by the education authorities to each of these instruments in the final score and, lastly, iii) the consequences linked to the evaluation results. This analysis will enable the reader to understand how in these successful cases fundamental decisions were made at the time of the design and implementation of the teacher evaluation systems, and also make possible an analysis of the chief similarities and differences in the teacher evaluation systems of the region.

## **INSTRUMENTS USED IN TEACHER EVALUATIONS**

Deepening the comparative analysis of the characteristics of the evaluation systems, it is worth noting as the first aspect that, under different designs, the high-performance systems studied have as their fundamental basis certain teacher performance standards. These standards explicitly describe what teachers should know and know how to do in the process of building the knowledge of their students and their daily work. Therefore, the design and implementation of this roadmap of the profession constitute an essential step for guiding and laying the bases for all aspects of the career and strengthening the growth and continuous improvement of the teaching force. Obviously, standards are also the necessary previous step to understanding the scope and effectiveness of teacher evaluation systems.

Perhaps the most detailed system of standards is found in Singapore, where each stage of the career is aligned with its own set of duly specified and evaluated competencies.

In Ontario, 16 competencies are established, including teachers' commitment to student learning and wellbeing, and their content knowledge. As in Singapore, in Ontario the level of requirement varies depending on the point in the teacher's career. For example, novice teachers are evaluated in only eight of the 16 competencies, and promotion to the next level in the career ladder depends on achieving these competencies.

Similarly, New Zealand has central performance frameworks that also vary according to the specific context of schools or groups (for example, schools where the curriculum is taught in the Maori language). Performance standards have also been developed in Estonia and, although their use is not mandatory in the evaluation process, they are found in the base guidelines conceived by the authorities. Definition of these standards and their inclusion in the evaluations are important since, as well as guiding the teachers themselves on what is expected of them, they contribute to making the instruments, the actors responsible for the evaluation and the results clear and impartial and correspond with the objectives of educational excellence unique to each country (OECD, 2013).

The second aspect that can be analyzed for each of the 12 systems covers the frequency of evaluations, whether they are mandatory and who is responsible for their implementation. In most of the cases analyzed for the United States, teacher evaluations are mandatory, annual and implemented by the authorities of each school. However, in New Haven and New Mexico external evaluators also participate in the application of some instruments. This is a prominent feature that adds a higher level of impartiality and consistency to the results, provided these actors are properly trained for such work.

The variation around these two aspects is slightly greater in the cases analyzed in the other high-performance systems. For example, in Ontario teacher evaluation is mandatory at least once every five years—although they must prepare an annual learning plan—based on the decision of the principal of the education institution where they work. In both the United Kingdom and New Zealand, teacher evaluations must be annual and, as explained above, current legislation gives principals a flexible space to accommodate them, according to what they consider most convenient for their teachers and schools.

In Singapore and the Republic of Korea evaluations are annual and their structure is more closely coordinated by the central authorities. In Singapore, although teachers are regularly evaluated by their direct supervisor, the formal evaluation is conducted by a panel of supervisors composed of the principal, deputy principal and department heads, who collectively rate them. In the Republic of Korea, principals are responsible for ensuring that the guidelines issued by the Ministry of Education for evaluations are complied with. Finally, in Estonia internal evaluation of the institution takes place every three years, although an internal progress report must be prepared every year. In addition, teachers can apply voluntarily for processes for career promotion, and for certification whose characteristics are defined by the schools and the country's teachers' association, respectively.

Although the frequency and mandatory nature of evaluation systems are essential aspects that provide a general understanding of their functioning in each case, one of the

**TABLE 3.1 | Instruments Used in Teacher Evaluations in Selected High-Performing Education Systems**

Education system	Classroom observation	Portfolio	Results of student tests	Student learning objectives	Survey of students and/or parents	Self-evaluation	Others
Ontario (Canada)							
United Kingdom							
New Zealand							
Singapore							
Rep. of Korea							
Estonia							
Denver							
DCPS							
NHPS							
New Mexico							
NYCPS							
Tennessee							

**Source:** Prepared by the authors based on information from Wyckoff and Katz (2018) and CUREE (2018). DCPS = District of Columbia Public Schools; NHPS = New Haven Public Schools; NYCPS = New York City Public Schools.

fundamental decisions that the education authorities in each system have to make has to do with the types of instruments to be used, as has been clear from the beginning from the previous chapter. Table 3.1 summarizes the instruments used for teacher evaluations in each of the 12 systems analyzed here. Due to the history behind each system, which has already been reviewed, it is no surprise that important differences are apparent in the selection of instruments in each case.

The first obvious conclusion from analysis of the information is that in the US systems studied, teacher evaluations usually include a multiplicity of instruments, while this is not the case in half of the other high performing education systems discussed in this chapter. The first exception is New Zealand, where teachers, in addition to being subject to classroom observation, must also prepare a self-assessment that includes a portfolio with evidence demonstrating that the teacher is effectively meeting the professional standards required.<sup>4</sup> Unlike the other four systems, the structured evaluations of Singapore and the Republic of Korea cover the use of at least four instruments. For these two countries, the current evaluation system includes, in addition to classroom observation, analysis of teacher portfolios, self-evaluation and the results of student exams (Singapore) and student surveys (Republic of Korea).

This multiplicity of instruments is much more common in the systems analyzed in the United States, which use between three and four instruments in their evaluations (except

<sup>4</sup> In fact, guides for building this portfolio were developed in the country.

for the NYCPS which only includes two). In fact, as detailed in Figure A1 in the appendix, the use of various types of instruments is common throughout the country and is not a decision taken only by the authorities of the six cases examined. Clearly classroom observation is the most used instrument in teacher evaluation systems in the United States. In fact, it is the only instrument included in 100% of the systems (Table 3.1 and Figure A1). Instruments based on student learning are also widely used. Indeed, as shown in Figure A1 in the appendix, the use of this type of information in teacher evaluations is common both in the states and in general in the large districts of the country. The variation is in the type and construction of the indicator chosen. In 60% of the districts considered by Steinberg and Donaldson (2016), value-added measures are used as indicators of teacher performance. These are complemented by measures of growth in student percentiles and learning objectives which, according to the authors, are included in 35% and 39% of districts respectively. In the case of the states, Steinberg and Donaldson find that use of growth measures in percentiles (59%) and learning objectives (52%) is much more common than the use of value-added measures (30%). Therefore, it is not surprising that in four of the six cases analyzed here, student performance measures in standardized exams are included as evaluation instruments.<sup>5</sup> The exceptions are New Mexico and New York City where the instrument associated with the learning objectives is used as a student learning measure.

As Table 3.1 also shows, of the high-performance education systems, the only one that incorporates student performance measures is Singapore, although the Republic of Korea takes students into account by including surveys applied to them. The latter instrument is also used in about 17% of the states and districts of the United States. In fact, student surveys are not used in teacher evaluation in the cases of New Haven and New York City only. It is also noteworthy that all cases analyzed for the United States, except New Haven, include as instrument a measure with few levels of standardization, as are achievement of learning objectives. Finally, Denver, DCPS, New Haven and New Mexico use as an instrument forms that ask about teachers' level of professionalism measured in aspects such as punctuality, compliance with responsibilities, teamwork, and even care for their appearance. It is possible that this type of instrument is very similar to some instruments used by several education systems in the region which, although they certainly provide some information about the teacher, are unlikely to lead to identifying what the strengths and weaknesses in their content knowledge and teaching practices are.

Although in the case of Ontario, the United Kingdom and Estonia, evaluation systems also include meetings with supervisors, rather than specific instruments, these are used to obtain basic teacher information, coordinate observation activities and deliver the final evaluation

---

<sup>5</sup> It is important to clarify that the information required for inclusion of measures of value added or of growth of student percentiles in evaluation systems in the United States is only available for about 20% of teachers in the country. Consequently, although some of these instruments are mentioned as used in a particular state or district, this does not imply that all teachers in these systems are evaluated with them.

results. For example, in Ontario (Canada) the school principal must hold a pre-observation meeting to explain its details to the teacher, and a later meeting to analyze the results, general ratings and strategies to be followed. In the case of the United Kingdom, the regulations only require observation and an interview with the teacher to define work objectives. Finally, as already mentioned, in Estonia the evaluation system focuses on the general functioning of the school and not specifically on each teacher. Despite this, the review reached the conclusion that perhaps observation is the most common instrument used in that country.

Given the preponderance of the use of classroom observation as an instrument in all the cases analyzed here, it is worth detailing aspects of its design and implementation. As noted above, for this tool to provide reliable and valid information, understanding the details of these two aspects is fundamental. In the design, the observation rubrics must be detailed and based on clear performance standards, as well as defining with transparency the actors responsible for conducting them, the way in which the results will be delivered to teachers and how they will be given feedback on their practices. For its part, implementation requires detailed and continuous training of observers, along with well-defined protocols on the number and type of observations to be made (formal/informal; scheduled/surprise), and on the methodologies and coding criteria, which have an impact on the final scores and results.

The international studies clearly show that there is a great variety of levels of rigor in the design and implementation of classroom observation instruments, even when comparing different cases in the United States. An analysis of the education systems with high performance clearly shows that Estonia is the nation with the lowest level of centralization and standardization. The education authorities have not established mandatory frameworks or procedures for conducting the observations. In such a decentralized system, the internal evaluation process and activities differ from one school and another. This case is followed by New Zealand, the United Kingdom and Ontario, in that order, where, although there are no mandatory common processes for classroom observation, the authorities have developed guidelines to facilitate its implementation. In the specific case of New Zealand, the regulations require at least two observations, and detail some suggestions that must be applied.<sup>6</sup> Finally, in the systems of Singapore and Republic of Korea, classroom observation is much more detailed and standardized. In the former, for example, evaluations are always conducted at the end of the school year (October) and are in the hands of the supervisor, together with a panel of superiors who have worked with the teacher (such as the principal, deputy principal or department heads) who under different protocols evaluate the teacher's potential with the instruments designed for this purpose.

As noted above, the variation in the design and implementation of classroom observations is also important in the United States. Where there is evidently greater attention to the

---

<sup>6</sup> See <https://educationcouncil.org.nz/content/observations>.

design and standardization of teacher observations is in the New Haven district. For example, this district funds and trains external observers who validate the observations from teachers that received either the worst or the best performance ratings. As a result, these teachers who, because of their rating might be eligible for a leadership position or possible dismissal, can receive unbiased and more accurate ratings. The case of DCPS is also interesting, since at the start all teachers had to be rated by external observers twice a year, but this desirable modality was eliminated from the process in 2016 due to the high monetary costs involved.

Analysis of this instrument must also include a review of the number of mandatory observations in each process. As the MET study (2013) demonstrated, variations in this aspect significantly affect the reliability of the information collected. For the United States it was found that Denver is at one extreme, where a single formal and informal observation are required, and Tennessee at the other extreme, where six observations are required each year. In the middle between these two cases are the city of New York, which requires one formal and three informal observations; the DCPS and New Mexico systems where a maximum of three formal observations are required, and New Haven which requires two formal and three informal observations.

As already noted, almost all the cases analyzed in the United States use some measure of student performance as a second instrument. As with classroom observations, there is great variation in the use of this type of student information in teacher evaluations in the United States. Although DCPS, Tennessee and New Mexico use teacher value-added measures, their design is different in each case. For example, while DCPS estimates the value added for math and English teachers in grades 4 to 10, in Tennessee and New Mexico, student performance is considered in a greater variety of subjects. Similarly, in calculating the value-added estimator DCPS includes some socioeconomic characteristics of the students as controls, whereas Tennessee and New Mexico do not take them into account. The differences in these three cases compared to Denver are even greater, since the instrument used by this district, although it uses student information, is based on growth percentiles instead of value-added measures.

## **WEIGHTS ASSIGNED TO EACH INSTRUMENT IN THE TEACHER EVALUATION**

For education systems that use a multiplicity of instruments in their observations, it is important to understand the weight assigned to each of them, since, as MET research (2013) has shown, this decision has important impacts on the validity of the system. The information available only permits this analysis in the case of teacher evaluation systems in the United States. As shown in Table A1 of the appendix for the case of the United States and in Table 3.2 for the six particular cases analyzed here, classroom observation is the instrument that receives the highest weight on average in the teacher evaluation, reaching more than 50% and 37% respectively. Thus, classroom observation is not only the most used instrument but also the one with the highest weight in the final evaluation ratings in

**TABLE 3.2 | Weights Assigned to Instruments in Teacher Evaluations in Some US States and Districts**

District/ State	Classroom observation	Student test results	Learning objectives	Surveys of students and/ or parents	Professionalism	Other
Denver	30%	10%	30%	10%	10%	10%
DCPS	30%	35%	15%	10%	10%	
New Haven	40%		50%		10%	
New Mexico	25%	50%		5%	15%	5%
NYCPS	50%		50%			
Tennessee	45%	35%	15%	5%		
<b>Average</b>	<b>37%</b>	<b>33%</b>	<b>32%</b>	<b>8%</b>	<b>11%</b>	<b>8%</b>

**Source:** Prepared by the authors based on information from Wyckoff and Katz (2018).

**Note:** The level of teacher professionalism is measured in aspects such as punctuality, fulfillment of their responsibilities, teamwork and even care of their appearance.

that nation. As a result, even more emphasis must be given to the previous discussion on the need for careful design and implementation of this instrument to ensure that it provides accurate, valid and consistent information on the teaching practices.

As detailed in the previous chapter, the MET study (2013) suggests that the instruments based on the results of standardized tests of students' knowledge should be included in the evaluations and receive a third of the weight in the final rating when it is possible to complement them with classroom observations and student surveys. With this weighting, it is possible to obtain high correlations with later results of students in different standardized knowledge tests, and it is also the weight that gives the highest confidence of all. Although, as noted, most US states and districts do in fact include this instrument in their evaluations, the weighting assigned is not high. On average, Steinberg and Donaldson (2016) find that the measures of value-added and of growth in percentiles have a weight of 16% and 10% in the evaluation systems, respectively (Table A1 in the appendix). In the six US cases detailed in this chapter, this average rises to 37% which is explained by New Mexico state which grants 50% weighting in its evaluations to instruments based on the results of standardized tests of student knowledge, and by Tennessee and DCPS which grant them 35% of the final score. In contrast, Denver only assigns 10% of the final score to student performance measures.

In fact, a comparison of the weighting assigned to the different instruments in the six cases examined for the United States reveals that the instrument with the second highest weight is related to analysis of achievement of the learning goals set by the teacher at the start of the year. For example, New Haven and New York City assign 50% of the final rating to this instrument and, of the six cases considered, only New Mexico does not include it in teacher evaluation. The predominance granted to this instrument goes against the expert recommendations in the field since, because of its design and implementation, it is a much more subjective measure that is difficult to standardize. However, it should made clear

that these six cases would seem to be atypical in relation to what happens in the United States since, as Table A1 of the appendix shows, the national average in terms of the weight assigned to this instrument falls to 10%.

The last aspect to be decided by the education authorities on the teacher evaluation score is the type of rating to be assigned to each teacher and its consequences. On the type of rating, according to the study by Kraft and Gilmour (2017), most states chose to have four categories: Very Effective, Effective, Developing and Ineffective. Only three of the 24 states considered in the study included a fifth category of “Exceptional.” However, despite progress in the use of multiple instruments for teacher evaluation, accordingly with the complex work and the multiple dimensions they wish to analyze, and the efforts in time and money involved, the final rating granted to teachers in the United States is still concentrated in the upper distribution tail. In most of the states analyzed, less than 10% of teachers were rated in the last two categories (see Figure A2 in the appendix).<sup>7</sup> As Wyckoff and Katz (2018) mention, this concentration in the upper tail is not consistent with other educational quality measures, such as the student results in standardized tests. According to the authors, teacher ratings could be expected to partly follow the student distribution. The differences between the two lead to the conclusion that, although there is undoubtedly an improvement in the current evaluation systems when compared with the previous ones, greater efforts are needed to continue this improvement.

## CONSEQUENCES ASSOCIATED TO THE EVALUATION’S RESULTS

As was made clear in the theory of change (Figure 2.1), the use made of the results obtained in teacher evaluations is crucial for achieving effective improvements in teaching staff and the quality of the education received by students. Table 3.3 shows the consequences of the final rating obtained by teachers in evaluation systems in terms of their professional career, salary levels or bonus pay and the possibilities of continuous training. As detailed in the previous chapter, although there is evidence that the mere implementation of evaluation can have positive impacts (see the work of Taylor and Tyler, 2012), it is crucial to understand how the results produced by the evaluation systems are used. The most common consequences linked to the evaluation results in the 12 cases analyzed have to do with salary increases or bonus pay and possibilities of promotion. There are only three cases where evaluation results are linked either to aspects of professional development or job security.

---

<sup>7</sup> The CUREE review (2018) does not detail these aspects and only mentions the final classification granted to three countries in its analysis. In Ontario there are only two categories (satisfactory and unsatisfactory); in Singapore the system grants a numerical score of 1 to 100. Based on this, teachers are rated with an excellent to poor performance; In the Republic of Korea, teachers are rated in one of five categories, ranging from A (highest) to E (lowest). There is no information on the distribution of the final rating obtained by teachers for any of the education systems analyzed by the authors.

**TABLE 3.3 | Consequences Associated to the Results of the Teacher Evaluation in Selected High-Performing Education Systems**

Education systems	Diagnostic	Bonus pay/ salaries	Promotion	Training	Tenure
Ontario (Canada)					
United Kingdom					
New Zealand					
Singapore					
Rep. Of Korea					
Estonia					
Denver					
DCPS					
New Haven					
New Mexico					
NYCPS					
Tennessee					
<b>Percentage of education systems</b>	<b>17%</b>	<b>75%</b>	<b>67%</b>	<b>25%</b>	<b>25%</b>

Source: Prepared by the authors based on information from Wyckoff and Katz (2018) and CUREE (2018).

This means, respectively, that teachers who obtain low results in their evaluations are subject to training programs or even dismissal when such results are obtained consecutively.

Of the high-performance education systems, only in Ontario does the evaluation, as its sole function, play an exclusively diagnostic role by identifying the scope of teachers' professional development objectives. In the United Kingdom and Estonia, for example, evaluation results are the main input used for discussion of professional development needs, career progression and of course salary levels. Similarly, in New Zealand final rating is linked to salary progression and an annual review of professional objectives. In the Republic of Korea, evaluation has both additive and training consequences, since the results are used to identify teachers with high and low performance to provide them with access to research opportunities or professional development programs, respectively. Lastly, the results of the three Singapore evaluation systems define aspects related to career promotions, annual performance bonuses and even the possibility of dismissal for teachers who do not meet the professional standards required. It is worth noting the importance that Korea gives to in-service training systems, reflected by the fact that the Ministry of Education finances up to 100 hours of professional development. Although these training hours are not totally tied to the evaluation results, the results are directly linked to these activities, since department heads of each school use them to define each teacher's needs and the organization of workshops in the institutions.

On the systems analyzed for the United States, the only case that does not tie evaluation results directly to aspects of salary is the New York City district. Denver, DCPS and New Haven link the results to bonus pay and salaries, both positively (increases) and negatively (freezing); New Mexico does so only positively (bonuses) and Tennessee only negatively (freezing). The results obtained by teachers in Denver, DCPS, New Haven and NYCPS define their chances of promotion, and for DCPS and New Haven even their tenure in the teaching profession.

The only case in which the results are associated with improvement and training plans is the New Haven district. As established by law, in New Mexico and Tennessee, as in New York City, the results of teacher evaluations should provide information for professional development plans, but there is no evidence of the existence of holistic plans for teacher professional development in these systems. On the contrary, in New Haven there is evidence that all teachers undertake a development plan with specific goals. In particular, during the year, teachers can enter a “Structured Support Plan,” while teachers who have received a low performance rating in their evaluations enter a more comprehensive program known as “Intensive Improvement Plan.”<sup>8</sup> A comparison of these consequences with the much larger group of states and districts analyzed by Steinberg and Donaldson (2016), as summarized in Table A2 of the appendix, reveals notable differences. According to these researchers, professional development is the aspect most linked to the results of performance evaluations. The most common consequence is termination of contracts for teachers who receive a poor performance rating; finally, in less than half the systems analyzed, the results are tied to promotions or bonus pay. The only exception is for home states where, in 75% of the cases analyzed, the results of teacher evaluations are in fact linked to aspects of career progression.

Several key conclusions can be drawn from the analysis of the main characteristics of teacher evaluation systems in the 12 education systems considered here.

The first is that, although there is no single ideal design, the great majority of systems have moved toward use of evaluations with multiple instruments aimed at understanding the dimensions of teacher performance that education systems consider important.

Second, although they all include classroom observation as a core tool in evaluation processes, its design and implementation vary considerably. As research on the subject

---

8 As detailed by Wyckoff and Katz (2018), some of the systems analyzed have plans or programs to support teachers’ professional development. However, these are not necessarily linked to teacher evaluation systems. For example, Tennessee conducted a pilot program of a peer collaboration model. Based on classroom observational data, low-performance teachers were partnered with high-performance teachers with the objective of working together during the 2013-14 school year. An impact assessment of the pilot suggests that the strategy increased student achievement (Papay et al., 2016) and the state may decide to expand the program. DCPS also has a program known as LEAP, designed to strengthen teachers’ skills in small groups where aspects related to planning, teaching strategies, student progress and event content knowledge are discussed. However, none of these programs, for now, are directly linked to teacher evaluation systems.

has shown, this produces impacts on the validity and reliability of the instruments. Consequently, implementation of a system such as New Haven's, where trained external observers are used to make at least two observations per teacher, is desirable but clearly expensive.

Third, use of instruments based of the results obtained by the students themselves in standardized exams seems to be a much more common tool in US education systems than in the other countries studied. This phenomenon can be explained in part by the recent funding granted by the federal government to education systems for implementation of these exams in multiple grades and subjects. The evidence suggests that this type of tool may have most validity for identifying the effectiveness of teachers in the content knowledge acquired by students as measured by these specific tests, but it clearly requires large initial investments that do not currently exist in Latin American and Caribbean countries and which would leave most teachers without the possibility of being evaluated.

Lastly, use of the results produced by the evaluations is essential. Interestingly, the cases analyzed in this chapter suggest that most systems link the consequences with respect to training and monetary and career incentives to the same type of evaluation. This is in marked contrast to some cases in Latin America where systems have evolved toward implementation of different types of evaluations to achieve different purposes, as in the case of the Republic of Korea.



## CHAPTER 4

# TEACHER EVALUATION SYSTEMS IN LATIN AMERICA AND THE CARIBBEAN

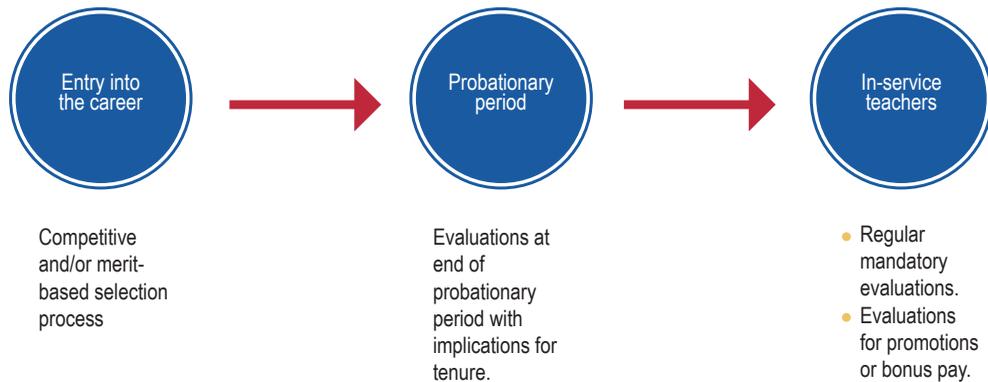
Several countries in Latin America and the Caribbean are making significant efforts to develop their teacher evaluation systems. This chapter analyzes and systematizes the information collected on these systems, providing a holistic look at the state of evaluation systems in the region and an analysis of the similarities and differences in the type of teacher evaluations being implemented. The information from each country was gathered from a review of official documents and from meetings with officials responsible for teacher evaluation in the respective ministries and agencies.<sup>1</sup> This chapter, together with the evidence presented in previous chapters, makes it possible to understand and contextualize the progress made by the countries of the region and serves as a guide to relevant aspects of the policies associated with this fundamental aspect of the teaching career for the future.

The 19 education systems analyzed, as part of this systematization, are: Buenos Aires province (Argentina), Belize, Bolivia, Santa Catarina (Brazil), Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Mexico, Panama, Paraguay, Peru, Dominican Republic, Trinidad and Tobago and Uruguay. For all of them, the information relates to teacher evaluation practices in use in recent years.<sup>2</sup>

The chapter analyzes four types of teacher evaluations (see Figure 4.1) implemented at three different points of the profession: i) evaluations for entry into the career, ii) evaluations during or at the end of the probationary period and iii) evaluations for in-service teachers. While in the first two stages there is only one specific type of evaluation, in the third stage two different evaluations are considered: regular mandatory evaluations and

<sup>1</sup> This information is also based on the background document produced by Rodríguez and Rivera (2018).

<sup>2</sup> The information was collected during 2017 with updates to 2019 as far as possible.

**FIGURE 4.1 | Types of Teacher Evaluations Analyzed in the Study**

**Source:** Prepared by the authors based on Rodríguez and Rivera (2018).

evaluations for promotions or bonus pay. For each type of evaluation, the authors checked to see if it was mentioned in the statutes, investigated if it was effectively implemented, compared the instruments used and how the results obtained were linked (or not) to continuous training processes and if they had implications for the entire teaching career.

Given that evaluations are applied at different stages of the career, their objectives, design and implementation have different attributes and differ even within the same country. The first evaluation analyzed, which is implemented in almost all the education systems of the region covered by this study, except in Belize and Jamaica, defines whether an individual can join the public system teaching body. In some countries, this type of evaluation is known as a “teacher selection process” (*concurso docente*). The second evaluation, which is the least common of all in the systems analyzed, determines whether novice teachers who enter the teaching profession have the content knowledge and pedagogical skills required to promote their students’ learning. Finally, in the last stage, two types of evaluations are administered to teachers who have successfully passed the probationary period, have received a stable teaching position and are therefore governed by each country’s teacher statutes. The first type (which in this study is called *regular mandatory evaluations*) is applied to teachers during specific set periods. According to the teacher statutes, in most cases, the results of these evaluations determine tenure or job stability and, occasionally, also the training that teachers need. The second type (which in this study is called *evaluations for horizontal promotions*) are implemented to establish whether teachers can progress in their professional career.

The systematization makes it possible to draw important conclusions from each of the four types of evaluations analyzed in this study. On the evaluations for entry into the career, the legislation of 17 of the education systems included in this work mentions them explicitly and makes them mandatory for entering the teaching profession. However, it was

found that in five of these systems the evaluations are not implemented. In these cases, three situations could occur: i) the selection is based solely on verification of academic qualifications, ii) the process can be corrupt, iii) there is no information that validates if the evaluations are being implemented and under what conditions (FEREMA, 2015). In most of the 12 remaining systems, selection is based on evaluation of the applicants' pedagogical skills and content knowledge with exams being the main instrument. In almost half, exams are complemented by interviews. However, it is not possible to know their structure and level of depth.

Deficiencies in the quality of pre-service training and the characteristics of the people interested in entering the teaching profession in the region (reviewed by Elacqua et al., 2018), along with the difficulty faced by ministries of education when dismissing teachers with low levels of performance, mean that career entry evaluations become an essential element for better identifying the characteristics and capabilities of future teachers. In that sense, continuing the processes of improving and correctly implementing this type of evaluation could bring benefits in the quality of the teaching force in the region. Although the evidence is still scarce, studies for Colombia (Ome, 2012; Brutti and Sánchez, 2017) and Mexico (Estrada, 2015) suggest that these selection systems (with career entry evaluations) are having some positive impacts on the educational quality of the countries that implement them. Also, the available research indicates that motivational level is also essential for adequate performance in the public sector (Banuri, Keefer and De Walque, 2018), and teaching should be no exception. Selection processes for the teaching career should also take this aspect into consideration in order to attract motivated candidates (Elacqua et al., 2018).

Similarly, the early years of practice of the teaching career play a significant role in development of teacher skills (Jensen et al., 2012; Ávalos, 2016). Currently, most of the systems analyzed include some type of teacher induction program in the legislation which they implement. Although the quality of the programs is not clear, their existence is, in itself, an important step. There is evidence that the correct entry of novice teachers and support for new entrants in the career are part of the political discussion in just over half the countries in the region.<sup>3</sup> Even so, the review of teacher evaluation systems at this stage of the career shows that evaluation at the end of the probationary period is the least common of all.

In 13 of the 19 education systems analyzed in this study no evaluation takes place at this stage of the profession, even though the evaluation is explicitly mentioned in the legislation in several of them. Although three systems include an evaluation, it is based on a single instrument, consisting of a form completed by the principal and the teacher. Due to

---

<sup>3</sup> The design, implementation and impacts of these programs are beyond the scope of this study. Readers can access important information in documents such as Avalos (2016) and Marcelo and Vaillant (2017), among others.

its structure, this evaluation does not detect aspects of teaching practice that could provide useful information to ensure a correct entry of teachers into the system. Only in three systems (Santa Catarina in Brazil, Chile and Mexico) is the evaluation of novice teachers based on multiple instruments that really test their performance. In the case of Santa Catarina (Brazil), the tests are designed specifically for this purpose, while in Chile and Mexico teachers who complete two years of practice must take the same regular mandatory tests as other members of the teaching staff (the characteristics of these tests are described in detail below). Moreover, even though in the systems where these tests are applied, the results are tied to the possibility of continuing or leaving the teaching career, the reality is that few teachers are dismissed because of poor performance.

In 17 of the 19 education systems analyzed, the legislation establishes that after taking up their position and passing the probationary period, teachers must undertake a regular mandatory evaluation. In 11 of these systems the frequency of this evaluation is annual while in the other six systems frequency varies between three and four years depending on the country and the teachers' results in the evaluation. This division, although it seems insignificant, leads to immense differences in the type of evaluation regulated in the legislation and implemented in the system. The great majority of the statutes that oblige teachers to take an annual evaluation define them as based on a single instrument composed of pre-established forms to be completed by the principal or supervisor.<sup>4</sup> Only five of the systems analyzed in this group provided evidence that this evaluation actually takes place. Moreover, being based on a simple instrument that does not collect evidence of teacher performance, most teachers are rated with outstanding performances and the results have few or no consequences for teachers' professional development. This high number of teachers rated with outstanding or good performance implies that dismissal of ineffective teachers from the profession is almost non-existent. Perhaps, the greatest use given to the results produced by the evaluations has to do with the promotion in the teacher hierarchy which, due to the structure of the evaluations, depend more on years of experience and qualifications than on teacher performance.

In the education systems that establish mandatory evaluations with a higher frequency the situation is different. First, in all these systems evaluations are implemented, although not in all cases established by the legislation. Second, in general systems include multiple instruments aimed at effectively evaluating teacher performance. Although the final choice of instruments varies, classroom observation and standardized exams are the most used and the ones that receive the highest weighting in the evaluation by the education authorities. Third, most of the systems include use of evaluators external to the education

---

<sup>4</sup> As detailed in the respective section, the only exception to this trend is Paraguay, which introduced an evaluation with multiple instruments in 2017. As the system is still in the initial phase and its implementation is expected to be gradual, the frequency will have to be modified in the future to make its application viable.

institution to implement the evaluations. In several of them, there are even institutions dedicated to this work, which makes the system more independent than when the evaluation is implemented only by the principal. Finally, although in most of these systems the legislation links the evaluation results to decisions on tenure and/or in-service training programs, the systems have not yet been able to fully implement this in practice. From what has been observed in this review, few teachers are actually dismissed as a result of these evaluations, which is also the case in systems with annual mandatory evaluation. In addition, although Chile, Mexico and Peru link evaluation results to these types of programs, they are only being effectively applied in the first country.

The last type analyzed is the evaluation that the systems implement for promotion in the teaching career. In particular, the chapter studies the possibilities and requirements required by the legislation to advance in the career and access higher salary levels, prestige and, perhaps, responsibilities, without need to leave the classroom to take up more administrative positions. These promotions, which as explained above are known as horizontal, are different from those in which teachers are promoted to administrative positions as principals or coordinators, known as vertical promotions. The review divided the education systems into three clear groups according to the possibilities and requirements of horizontal promotions: i) systems where there is no such promotion, ii) systems where there is the possibility of horizontal promotions depending exclusively on years of experience and/or academic qualifications and iii) systems in which horizontal promotions are meritocratic and specific tests must be taken to access them. For this last group it is important to emphasize three points. First, the type of evaluation used for horizontal promotions and its associated characteristics—such as the frequency, instruments and those responsible—vary significantly between systems. Second, the review found that in some systems teachers do not need to undergo another evaluation in addition to the mandatory one, since the scores obtained in the mandatory evaluation are considered in decisions on horizontal promotion. In general, systems that implement mandatory evaluations with a multiplicity of instruments use the information provided by these tests as an essential input when making promotion decisions. In other systems, even though the regulations are clear about the requirement of a specific evaluation for horizontal promotions, there is no evidence that what is written in the law is being applied. Lastly, as seen in other evaluations, there are important differences between systems in relation to the level of implementation of the requirements, even within the same system, and the *de facto* situation is sometimes the opposite of the *de jure* position.

The next section of the chapter analyses two issues closely linked to teacher evaluation and the career in general: i) the existence of teacher performance standards that guide not only the structure of implemented teacher evaluations but also vitally important issues such as pre-service training, in-service training and the actual expectations of the teaching force in the country; ii) regulation of the teaching career, especially teacher evaluation and how it determines, one way or another, the possibilities of tenure, training, bonus pay and promotions in each system. In several of the education systems analyzed,

teacher evaluations are applied in the framework of recent laws on the teaching career, aim at making the career more meritocratic and enhancing the prestige of the profession. The new careers give teachers access to higher salaries and better opportunities for professional growth. Next the chapter continues with all the information collected and studied on teacher evaluation and a detailed description of essential aspects, also relevant for the review of experiences outside the region, such as instruments used, assigned weights, consequences of the results and the actors involved at the different stages, among other characteristics that the specialized literature considers fundamental.

Finally, although it is not a topic that is analyzed in detail in this book, it is important to note that, in most countries of the region, initial implementation of the new evaluations was rejected by teacher unions, and many governments had to deal with (and in some cases are still doing so) major national strikes by the teaching force opposed to application of evaluations. To overcome this resistance, governments have used a variety of strategies, including linking the results only to positive consequences for teachers, publishing the low learning results of students to give a greater sense of urgency to implementing the reforms, mobilizing the support of public opinion for the evaluation processes, and, in some cases, introducing penalties for teachers who do not participate in evaluations, among others. The political compromises needed to implement these evaluations, along with the large number of public sector teachers in many countries of the region, have resulted in a gradual implementation (Elacqua et al., 2018). Some studies that have researched aspects related to the agreements and final implementation of education programs and policies in the region include Mizala and Schneider (2014 and 2019) who analyze the political economy around the reforms implemented in the teacher evaluation system in Chile and regulation of the teaching career in Chile, respectively. While Bruns, De Gregorio and Taut (2016) study the issue more generally with emphasis on the Latin American region.

Despite the difficulties and although more actions are still needed in this area of the profession, the progress made in several countries of the region in terms of teacher evaluation has been positive, as described below.

## **TEACHER PERFORMANCE STANDARDS IN THE REGION**

The existence and implementation of teacher performance standards in each country are necessary first steps to understanding the scope and effectiveness of teacher evaluation systems. The standards are a cross-cutting tool for the three types of teacher evaluations and an essential instrument for guiding and laying the foundations for all the main aspects of the teaching profession. Each country's standards are different, but they generally include dimensions to ensure that teachers have content knowledge. In each country, what is expected of teachers may vary according to the context and reality of the students, so each system must define specific indicators by level, modality, language of teaching, among others.

As the literature mentioned above emphasizes, a teacher evaluation system must be valid, reliable and transparent if teachers are to accept it and if the efforts required for its design and implementation are to generate the expected results. To achieve this objective, all actors in the education system need to be clear about what a teacher must *know* and *know how to do*. Pre-service training and in-service training should be designed to ensure that future and in-service teachers achieve these standards in their daily performance. In the same way, a logic must be followed in the design and implementation of the evaluation instruments to identify if an individual who practices the profession adequately complies with the performance standards. Lastly, the teachers themselves must know what the system expects of them and the characteristics that will allow them to be better professionals. All this can only be achieved when there are defined performance standards, disseminated and used by all.

Despite its importance, few countries in the region have designed this tool and incorporated it into their education systems, in contrast to what happens in developed countries or in successful education systems. Figure 4.2 shows that performance standards are used in only five countries with different intensity: Chile, Ecuador, Mexico, Peru, and Dominican Republic. In these five countries, performance standards have been discussed and agreed between various stakeholders, such as ministries, teachers and in some cases non-governmental organizations in the education sector, and guide the design and application of education policy. Currently, Guatemala and Panama are in the process of developing standards; while Colombia, Honduras, Jamaica and Paraguay have incorporated certain specific standards into their recently developed teacher evaluation processes. The other education systems do not have teacher performance standards. Some, such as Bolivia and El Salvador, have defined profiles for graduation from teacher training careers and others have teacher statutes that generally describe the minimum obligations that teachers who are governed by them must comply with.

In the group of countries that have carried out the process of discussion, design, approval and implementation of teacher standards in the region, Chile is the pioneer. In 2003, this country created the Framework for Good Teaching, based on four domains: preparation for teaching, creation of a favorable classroom environment, teaching for learning by all students and professional responsibilities. Since its creation, the Framework has guided policies aimed at strengthening the teaching profession, has laid the foundations for the design of pre-service and in-service training programs and has been used in performance evaluation processes. More importantly, it has given teachers in the country a guide to reflect and evaluate their own teaching and education practices. In 2016 and 2017 consultations were held to review the Framework based on new discussions and participatory processes with academia and the teachers themselves. The proposal was submitted to the National Education Council which will review and approve it before the new version comes into force.

In the other countries in this group, performance standards are more recent and have therefore been used less. However, the reflective processes that took place made

**FIGURE 4.2 | Progress of Teacher Performance Standards in Latin America and the Caribbean**



Source: Prepared by the authors based on Rodríguez and Rivera (2018).

it possible to clearly establish what is expected of teaching work. In the case of Ecuador, the 2012 standards consist of four dimensions related to the content knowledge and curricular domain, learning management, professional development and ethical commitment. In addition, each of the four dimensions has general standards, specific standards and indicators. In Mexico the standards define what a good teacher should be: i) know their students; ii) organize, evaluate their work and carry out a teaching intervention; iii) continually improve their training and teaching practices; iv) assume their legal and ethical responsibilities, and v) participate actively in the school and in the community. In Peru, the Good Teaching Performance Framework approved in 2012 consists of four domains: preparation for student learning, teaching for student learning, participation in school management in coordination with the community and development of professionalism and teacher identity. Finally, the standards of the Dominican Republic approved in 2014 also include four dimensions: i) recognition of the student and their individualities as center of attention, ii) content and curricular knowledge, iii) implementation of effective teaching and learning processes and iv) personal and professional commitment.

The differences between performance standards in the region, evident in the number of dimensions included in each one, confirm the complexity of teaching work. Likewise, these differences show that it is not even possible to establish what is expected of a teacher of excellence in a common framework, so that it can be applied in general in countries of the region that do not yet have it. Education systems that do not have this tool need to define and agree on what they expect from their teaching staff in accordance with their context and educational objectives. In addition to setting performance standards, it is essential that once designed they guide the different aspects of the teaching career, including teacher evaluation. To do this, it is crucial for teachers to participate actively in design discussions and build consensus on what is relevant in teaching practice in each system, as in the case of the five countries that already have defined standards. It is only through these agreements that a favorable environment can be created to facilitate their effective incorporation into each stage of the teaching career and the education system in general. It is also important to mention that the vagueness of some of the standards creates a challenge when measuring them effectively through specific indicators and teacher evaluation.

## **REGULATION OF THE TEACHING PROFESSION IN THE REGION**

The second aspect closely related to the structure of teacher evaluations in each country is the teacher statutes that govern the career and regulate the evaluations. As detailed by the UNESCO study (2015), regulation of teaching careers in the region is very diverse with aspects that include, the legislative level of the regulation, when the regulation was established, but especially the type of career that it produces. On the legislative level, in some countries regulation of the career is linked to the Constitution, in others it takes the form of specific statutes that govern the profession, while in a third group the regulation is included

### FIGURE 4.3 | Regulations of Teaching Careers in Latin America and the Caribbean over Time



**Source:** Prepared by the authors based on the teacher statutes of each country.

in the general education laws. Similarly, in federal countries, such as Argentina and Brazil, there are specific statutes for each administrative entity; while in centralized countries there is a single legal framework that everyone must comply with.

The career has been regulated in the countries of the region over a long period of time which partly explains the differences in the view of the profession expressed by the regulations in each country. For example, as shown in Figure 4.3, the region has teacher statutes developed in the 1950s, such as that of Bolivia of 1957, which coexist with many more recent statutes such Chile's, reformulated in 2016. The figure divides the education systems included in this review into three broad groups by the date on which the career regulation was established: i) before the 1980s, ii) from 1980s to the new millennium and iii) in the last twenty years. According to UNESCO (2015), teaching careers regulated by statutes or laws before 2001 are, in general, more bureaucratic and contain fewer meritocratic aspects than the ones reformed more recently.

Although the processes that have taken place around these recent reforms are beyond the scope of this review, it is worth mentioning some of the key aspects that marked these processes. For example, Colombia is the only country whose reform led to the coexistence of two teacher statutes: Decree Law 2277 of 1979 (or old statute) and Decree Law 1278 of 2002 (or new statute). Consequently, the teaching force is governed by two different career rules, the first of which lacks mechanisms for strengthening teacher excellence. Only teachers hired by the teaching authority after 2002, and therefore governed by the statute of that year, are selected through competitive processes and are subject to teacher evaluations with implications for their career and continuing education. For this country,

the analyses presented in this study are based on the regulations derived from the most recent statute.<sup>5</sup>

In the case of Ecuador, as detailed by Cevallos-Estarellas (2016), between 2007 and 2013 the country implemented ambitious education reforms, many of them focused on teachers. Among the changes introduced were reforms in pre-service training programs, selection, regulatory framework, teacher evaluations, continuous training and salary structure. All these changes, which sought to increase the prestige and quality of teaching in the country, modernized the career and included aspects of accountability that did not exist before. Similarly, in Honduras the regulation of the Basic Law of Education (2012) passed during 2014 introduced important changes in the legislation on the teaching force, including the Regulation on Pre-Service Teacher Training, Regulation on Permanent Teacher Training, Regulation on Teacher Evaluation, and the new Regulation on the Teaching Career. Although these new instruments are not yet being applied in the country, the fact that the legislation provides for them is undoubtedly an advance.

In Peru the Teacher Reform Law (Law No. 29944) was passed in 2012, which modifies the rules for selection of teachers, their progression in the career and development as teachers. The fundamental idea of the new regulation is that the career must be governed by excellence and the principle of merit and ability. This means that the country is making significant efforts to set up an evaluation process capable of guaranteeing that the most suitable teachers enter and remain in the profession, strengthen training and continuous education programs and increase teacher salaries.

In Mexico 2012 was also a turning point for teacher policies. Although the country had already made important reforms, such as introducing the Teaching Career Program in 1993, it was only in 2012 that reform of various areas of the education system was initiated. One of the most important reforms was creation of the National Institute for Evaluation of Education (INEE), an entity with responsibility for regulating and supervising teacher evaluations in the country, and the General Law of the Professional Teaching Service, which regulated a new type of meritocratic career for teachers. However, these reforms were repealed in 2019 (in practice eliminating INEE and teacher evaluation).

Finally, Chile is the country where reforms have most recently been implemented to regulate the teaching career. Although the first statute dates from 1993, in 2016 Law No. 20903 created the National System for Professional Teacher Development (SNDPD). The new system aims to enhance the prestige of the career through high quality pre-service training, ensuring an entry to the career through a monitoring process in the first years of practice, increasing continuous training and improving teacher salaries in line with the level of professional development. Thus, the SNDPD Law, the Teacher

---

<sup>5</sup> At present, about 50% of teachers are governed by the old statute, 30% by the new statute and the remaining percentage relates to provisional teachers.

Evaluation Law of 2004 and the 20,501 Law of 2011 on quality and equity in education are promoting a meritocratic career for all teachers working in public or subsidized schools in the country.

The differences in the statutes that define the rules governing the teaching profession in the region are immense. However, due to the scope of this study, two crucial aspects related to teacher evaluations are highlighted. The first is related to the type and number of evaluations established in each teacher statute. The second aspect, analyzed in detail below, has to do with the type of career determined by each statute and each teacher's possibilities of promotion.

Analysis of the statutes shows that, with few exceptions, the systems that implemented recent reforms have defined more meritocratic and less bureaucratic careers than those in systems with statutes that have not been modified in more than 30 years. Also, in the systems mentioned in first place, teacher evaluation is more complete in terms of design (multiplicity of instruments and evaluators used) and implementation (understood as the effective application of the legal provisions), and in the use made of the results of the evaluations when they actually take place. Details of each type of evaluation are presented in the corresponding sections. However, Table 4.2 indicates whether the statutes explicitly refer to each of the four evaluations analyzed in this study: selection process to enter the teaching career, evaluation of probation, regular mandatory evaluations and evaluations that define horizontal promotions in the career. In other words, it establishes the *de jure* situation of each type of evaluation for each education system.

At least four clear messages can be extracted from Table 4.1. The first is that the education systems of the region have included in the teacher policy agenda discussion of evaluation at different stages of the career and its mandatory nature. The second message is that these efforts have focused on two evaluations: evaluations for career entry and regular mandatory evaluations. All the education systems analyzed now require that at least some of their teachers meet minimum requirements and undergo an evaluation to enter the career. Likewise, except for the Belize and Santa Catarina (Brazil) systems, all aim to implement regular mandatory evaluation processes for their teachers. Third, evaluations to determine horizontal promotions have been used to allow teachers to progress in their career in terms of prestige, economic resources and responsibilities without having to leave the classroom. Although this possibility and the related meritocratic requirements are less common, the evidence suggests that it is an aspect being targeted by some education authorities in the region. Lastly, perhaps the aspect where education systems have made less effort in terms of teacher evaluation is after the probationary period.

The mere mention of evaluations and of being mandatory is an important step forward and lays the legal basis for their implementation. However, their effective application and quality vary significantly in the region. The following sections detail the main characteristics of each of the four evaluations.

**TABLE 4.1 | Evaluations Mentioned Explicitly in Teacher Statutes**

Education system	Types of teacher evaluations			
	Selection process	Probationary period	Regular mandatory	Horizontal promotions
Buenos Aires Prov (Argentina)	■	■	■	□
Belize	■	■	□	□
Bolivia	■	■	■	■
Santa Catarina (Brazil)	■	■	□	□
Chile	■	□	■	■
Colombia	■	■	■	■
Costa Rica	■	■	■	□
Ecuador	■	□	■	■
El Salvador	■	□	■	■
Guatemala	■	□	■	■
Honduras	■	□	■	□
Jamaica	■	■	■	■
Mexico	■	■	■	■
Panama	■	■	■	□
Paraguay	■	■	■	■
Peru	■	□	■	■
Dominican Republic	■	■	■	□
Trinidad and Tobago	■	■	■	■
Uruguay	■	□	■	■

■ Explicitly mentioned in the statutes of each country.

■ These education systems grant a license instead of a competitive selection process.

□ No mention of this in the statutes.

Source: Prepared by the authors based on Rodríguez and Rivera (2018).

## EVALUATIONS FOR ENTRY INTO THE TEACHING CAREER

As described by Elacqua et al. (2018), during the twentieth century Latin American countries made immense efforts to ensure greater education system's coverage. Naturally, to respond to the growing number of students entering the system, it was also necessary to rapidly increase the number of teachers hired by governments. Consequently, most countries relaxed the requirements for entry to the teaching profession and expanded training programs without considering quality criteria. These two decisions, together with a fall in average relative salary of teachers and the opening of new job opportunities for women, largely explain the deterioration of the prestige of the profession in the region.

As a direct consequence, the average training of in-service teachers has fallen significantly and the number of students who hope to enter the profession has fallen significantly. The low content knowledge among teachers is shown by the results of the TEDS-M test, which evaluated the math knowledge of teachers in 17 countries. In Chile, teachers came last. Evidence for countries like Belize (Näslund-Hadley, Alonzo and Martin, 2013), Costa Rica (OECD, 2017) and Dominican Republic (Caraballo et al., 2014) suggest that teachers do not even master the content of what they teach. Similarly, in standardized tests conducted in Colombia, student teachers obtain lower scores than the next set of higher education graduates in the country (Figueroa et al., 2018). With respect to the skills of young people who are close to graduating in secondary education and are interested in the teaching career, the existing information is not encouraging either. The results of PISA tests (Elacqua et al., 2018) and national standardized tests in Chile and Colombia (Alvarado et al., 2012, and García et al., 2014) indicate that the skill level of young people interested in teaching is below interest in careers such as engineering, science, medicine or law.

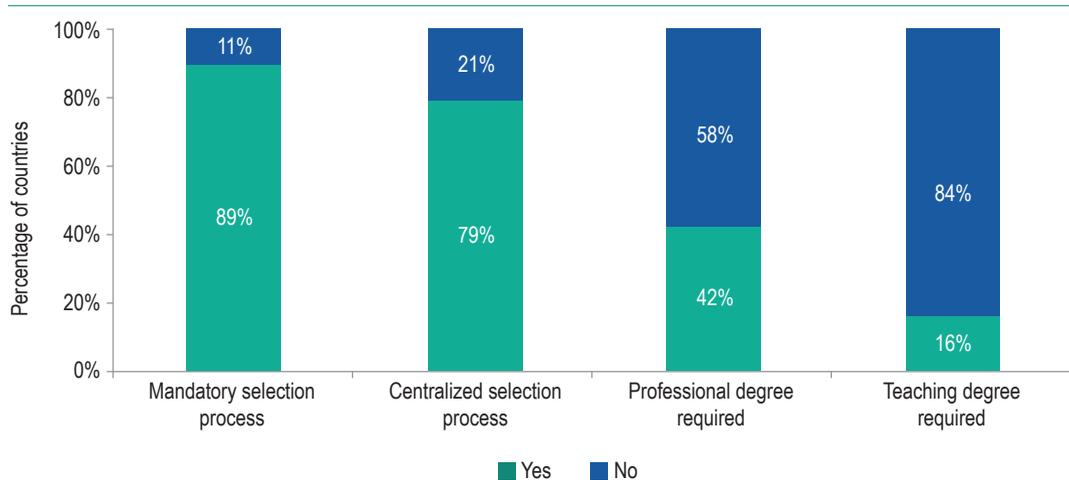
In these conditions, the selection criteria for candidates who are finally hired by Latin American and the Caribbean governments to practice the teaching profession are fundamental. It is only by selecting the best prepared and motivated individuals for this work that the quality of education provided in the region can be improved.

The review of this aspect in the 19 education systems analyzed indicates that all of them currently have some type of candidate selection process (see table 4.1). In two of them, Belize and Jamaica, the state grants a renewable license to practice the profession<sup>6</sup>. In the other 17 systems, current regulations require most candidates to participate and approve a competitive or merit-based selection process to enter the teaching profession. It is worth mentioning that the type of selection process, the instruments used, and the level of implementation are very heterogeneous.

Figure 4.4 summarizes four basic characteristics of teacher selection processes in the education systems included in the study (details of each of the four characteristics for each education system are given in Table A3 in the appendix). The first thing worth noting is that despite all the statutes explicitly mentioning the need for entry into the profession to be through selection processes (or licenses in the case of Belize and Jamaica) in reality this

---

<sup>6</sup> Any person interested in teaching in an education institution in Belize must meet specific requirements in order to apply to the Chief Education Officer for a license to practice. For some time, the country has been having a discussion—now in an advanced stage—on homogenizing and standardizing the entry requirements for the professional teaching career. In 2018 a national entrance exam is expected to be introduced which would identify which candidates had the necessary qualifications and preparation to undertake teaching work in the country. Although the country conducted the first of these evaluations, its results, possible modifications and final application are still under discussion. In Jamaica, every graduate of a teaching career accredited by the government must register with the Ministry of Education to be certified as an academically prepared person to pursue a teaching career. Once this process is completed, the person is registered and has government permission to practice as a teacher in the country.

**FIGURE 4.4 | Mandatory Basis, Centralization and Degree Required in Teacher Selection Processes in Latin America and the Caribbean**

Source: Prepared by the authors based on data from Rodríguez and Rivera (2018).

does not happen. In some countries, competitive selection is not applied to all teachers in the system, which leaves room for possible corruption and loses an opportunity to select the best candidates. Specifically, selection processes are not mandatory for all teachers in two education systems: Bolivia and Guatemala. In Bolivia, according to article 13 of the National Career Ladder of the Education Service, applicants who are studying a teacher career enter the career without having to take an exam or go through competitive selection process. The selection process is mandatory for interim teachers, that is, for teachers with another type of profession in their undergraduate studies. In Guatemala, a competitive selection process is mandatory for teachers interested in teaching at pre-primary or primary level. However, those interested in teaching in middle school are selected by the principals and do not need to go through a competitive selection process.

The second characteristic analyzed is the level of centralization of teacher selection processes. As shown in Figure 4.4, selection processes are centralized in almost all the countries analyzed: irrespective of the region or education level at which the candidate wants to teach, all selection is subject to the same process. Although in Belize and Jamaica the government grants licenses to practice as a teacher and there is no selection process as such, they were classified as countries where the process is centralized because effectively only one institution in the country can grant these licenses and the requirements are the same for all. Argentina and Brazil, while having a federal government organization, have decentralized processes with characteristics, instruments and requirements that depend on what each province, state or municipality establishes. Similarly, due to the organizational level of government that characterizes them, Bolivia, Colombia, Costa Rica, Ecuador,

El Salvador, Guatemala, Honduras, Panama, Paraguay, Peru, Dominican Republic, Trinidad and Tobago and Uruguay have centralized processes under each country's Ministry of Education. Four countries require a special clarification about centralization of selection processes. The first is Chile, where although the teaching career is regulated in detail in the Law of the National System of Teacher Professional Development (SNDPD), entry of new teachers into the system is decentralized. In contrast, in Mexico, which has a federal system, the process of teacher selection can be classified as centralized because the National Institute for Evaluation of Education (INEE) defines the technical guidelines and criteria for evaluation which are implemented by the Ministry of Public Education (SEP), the education authorities and the corresponding decentralized agencies. It is also important to highlight the case of Uruguay, where the characteristics of teacher evaluation vary significantly according to the education level at which the teacher teaches. In Uruguay there are five education councils associated with different levels which direct education policies independently in each one.<sup>7</sup> This division makes understanding and managing teaching in the country more difficult (INEEd, 2017). In the case of the teacher selection process, for example, while annual processes are held at primary level, at secondary level they are much less frequent. In addition, the two levels have different requirements and instruments. Lastly, in Peru the Selection Process for Entry into the Public Teaching Career is held every two years and consists of two stages. The first stage is executed by the Ministry of Education and consists of one exam, known as the National Single Test, which evaluates general skills, content knowledge and pedagogical skills. Candidates who pass this exam can move onto the second stage of the process. This new stage takes place in every education institution with available places under the supervision of an evaluation committee whose members include the principal and assistant principal or academic coordinator. In this second stage, the teacher's experience, education, training and pedagogical skills are evaluated.

Among the requirements for participating in the selection process, most countries require candidates to be nationals or authorized residents in the country, have good health and no criminal record that prevents them from working in a school. The statutes also detail the level of pre-service training necessary for participating. As shown in Figure 4.4 (the third and fourth bars summarize the education requirements imposed by each country for entering the teaching profession), only 42% of the countries require a professional university degree for entering the public teaching profession. In some cases, the education level requirement is not explicit and only mentions, for example, the need for an "enabling" qualification without giving more details. In fact, in many countries—such as Brazil, Colombia, Honduras, Trinidad and Tobago and Uruguay—the teacher statutes clearly establish that candidates without an university degree can enter the selection process, generally, for a

---

<sup>7</sup> The councils in Uruguay are: Central Directive Council, Council of Pre-Primary and Primary Education, Council of Secondary Education, Council of Technical-Professional Education and Council of Training in Education.

**TABLE 4.2 | Number and Type of Instruments Used in Teacher Selection Processes in Latin America and the Caribbean**

Number of instruments included	Education system	Percentage of education systems
<b>No instrument/no information</b>	Chile, Costa Rica, El Salvador, Honduras, Panama	29%
<b>One instrument</b>		
<i>Knowledge test</i>	Bolivia, Santa Catarina (Brazil), Guatemala, Mexico, Uruguay	29%
<i>Interview</i>	Trinidad and Tobago	6%
<b>Two instruments</b>		
<i>Knowledge test and interview</i>	Buenos Aires Province (Argentina), Colombia, Paraguay, Dominican Republic	24%
<i>Knowledge test and demonstration class</i>	Ecuador	6%
<b>Three instruments</b>		
<i>Knowledge test, interview and demonstration class</i>	Peru	6%

Source: Prepared by the authors based on Rodríguez and Rivera (2018).

teaching position at primary level. For example, some of these countries authorize entry to the profession by graduates from teacher training schools (*escuelas normales*). The important thing is that the training given in these schools should provide them with the content knowledge and pedagogical skills required to become excellent teachers.<sup>8</sup>

Finally, an analysis of whether the statutes require a teaching degree found that this was mandatory in only three systems: Buenos Aires province (Argentina), Jamaica and Peru. In the other systems (85%) professionals from other areas can participate in the selection process, and if approved, in most cases they have a maximum period to complete pedagogical studies or specific teaching programs while they are in-service.

To improve teacher selection in the region, it is necessary to have mandatory processes for anyone interested in the teaching career. However, the crucial point that will help in the selection of suitable candidates depends on the evaluation instruments available.<sup>9</sup> Table 4.2 summarizes the main components of the selection processes in each of the 17 systems analyzed where they exist and are regulated (so Belize and Jamaica are

<sup>8</sup> Several countries still have teachers without a professional qualification working in secondary schools. For example, the review found that, in countries such as Belize, only 30% of teachers at secondary level are trained to be teachers. Similarly, in Uruguay, where the proportion of teachers with a university teaching degree reaches almost 100% in primary education and is among the highest in the region, the levels of teacher qualifications in secondary education are much lower. According to the OECD report (2016) about 50% of teachers at this education level had not completed their university preparation according to the 2007 teachers' census.

<sup>9</sup> Table A4 of the appendix explains the components of the teacher selection processes for each country in accordance with the provisions of the legislation.

excluded from this list). Among them, some peculiarities stand out. In Buenos Aires province (Argentina) and Bolivia, the regulations are clear that the process must include exams on knowledge and colloquiums. However, there is no information about whether these exams are applied and what kind of knowledge they evaluate when applied. Since in Chile the organization of the selection process is decentralized and the types of instruments used are not detailed, it was decided to classify the case as “no information available.” In Costa Rica, El Salvador, Honduras and Panama it was found that in practice no instrument is used for recruitment of new teachers in the profession. In Costa Rica, El Salvador and Panama, the review determined that candidates are selected based on their academic qualifications, courses taken and seniority. In Honduras the evidence suggests that the system does not take meritocratic aspects into account (FEREMA, 2015).

With these clarifications, the analysis of the table divides the education systems into three large groups by number and type of instrument used in the selection process to enter the teaching force. The first group lists the systems that have no information on the process or those that do not use an instrument for the selection of teachers hired by the government. In most cases, only academic qualifications are reviewed. Since the quality of many of the pre-service training programs in the region is questionable (Elacqua et al., 2018), there is no guarantee that candidates selected in these systems are properly trained and have the appropriate levels of knowledge.

In the second group the education systems establish a single selection instrument to define if candidates are suitable to enter the teaching force, generally based on a knowledge test. Inclusion of areas evaluated in these exams varies by country and include content mastery, pedagogical knowledge and/or psychotechnical aspects. Only in the case of Trinidad and Tobago is admission based solely on an interview. The third group is composed of education systems that include two or more types of instruments in the process; most use an interview as an additional instrument to knowledge exams. The selection processes in Ecuador and Peru are unique in the region. In Ecuador, a knowledge test and a demonstration class are included. Peru is the only case studied which includes three instruments: knowledge test, interview and demonstration class.

## **INDUCTION PROGRAMS FOR NOVICE TEACHERS AND EVALUATIONS FOR APPROVING THE PROBATIONARY PERIOD**

Experts argue that the early years of teaching are critical (Ávalos, 2016). For example, Jensen et al. (2012) indicate that novice teachers devote less effective classroom time to learning-related activities, spend more time maintaining order and report lower levels of self-efficacy in teaching compared to teachers who have more years of experience. It is not surprising, therefore, that these early years of experience also have an impact on teachers' effectiveness, measured by the value added in learning. Various studies in the United States find that improvements in teaching ability are concentrated in the first three years

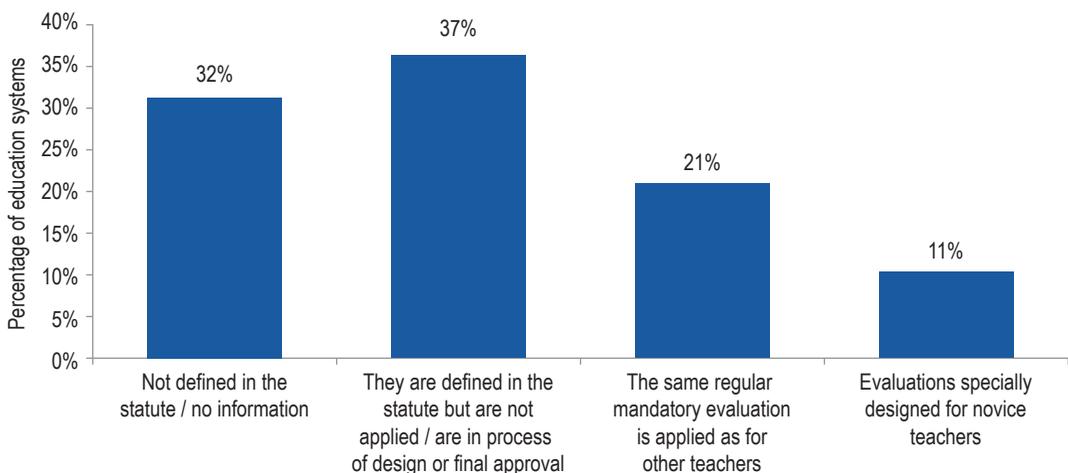
(Hanushek et al., 2005; Rivkin, Hanushek and Kain, 2005), after which there are no positive returns (Chingos and Peterson, 2011). For the Latin American case, Araujo et al. (2016) find similar results in Ecuador where returns on teaching experience increase rapidly in the first three years, and then stabilize.

If this is added to the fact that in several countries of the region novice teachers begin their work experience in schools that, on average, are vulnerable or in vulnerable contexts due to their geographical location and the socioeconomic characteristics of their students (García et al. 2014), programs to promote appropriate entry into the system are critical. Several studies in the United States suggest that the most effective teachers are concentrated in schools with students of higher socioeconomic level or with lower participation by minority students (Steele et al., 2015). In Latin America there is also evidence of an unequal distribution of teachers. Although there are no studies that directly analyze teacher effectiveness, some research reveals that the less prepared and worst evaluated teachers often work in schools with students of lower socioeconomic status (SES) and worse results in learning tests. Bertoni et al. (2018) analyze teacher distribution in public schools in Chile, Colombia and Peru and find that less qualified teachers (with less experience and with temporary contracts) are usually concentrated in schools with students of lower SEL, particularly in rural areas. Similarly, in Mexico teachers with little experience and lower education level tend to concentrate in rural schools with students whose parents have a low education level (Luschei, Chudgar and Rew, 2013).

The review of the type of induction programs, their structure and organization, is beyond the scope of this study. However, research on the evaluations conducted in the early years of the profession identified which countries have teacher induction programs. It was found that about 53% of the education systems analyzed have an induction program formally promoted by the ministry of the area, which, at least in theory, should benefit all the novice teachers hired (the main characteristics of the programs are summarized in Table A5 of the appendix). In many cases the program is based on tutors or mentors who guide and counsel novice teachers. Since in general the programs are relatively recent, very often introduced by the statute reforms, the evidence on their impact in the region is almost nonexistent.

The existence of an induction program is not perfectly correlated with the existence of a probationary period and specific evaluations to determine whether novice teachers remain in the system, which is the primary interest of this section and a topic of great importance for teacher policy in general. As argued in previous paragraphs, the first years of the profession are very important for the teachers' career. It is possible that the teachers themselves, their peers and the school principals can perceive if the novice teachers have the motivation, knowledge and aptitude required by the profession; what weaknesses they should work on, and what their greatest strengths are. Evaluations administered in the early years of the profession could provide objective measure to identify all of these.

The main characteristics of the evaluations in this first stage of the teaching career which came out of the review of the regulations in the 19 education systems analyzed are presented

**FIGURE 4.5 | Evaluations in the Probationary Period for Novice Teachers in Latin America and the Caribbean**

Source: Prepared by the authors based on data from Rodríguez and Rivera (2018).

in Table A6 of the appendix. The table details if there is a probationary period for newly hired teachers and its duration; if specific evaluations are established during the period or at the end of it, and if these evaluations determine whether teachers remain in the public system. Finally, the main characteristics of the evaluations of the probationary period are described.

The analysis of the information contained in Table A6 of the appendix is summarized in Figure 4.5. It can be inferred from this that in one third of the education systems in the study a probationary period is not defined and so there is no specific evaluation in this first stage of the teaching career. It is worth clarifying a point in this regard. Ecuador has an induction program for novice teachers which includes permanent evaluation. However, these evaluations—which are not defined in the statute—do not follow standards and have no impact on the progress or stability of the teacher’s career.

Although in the other 13 education systems the teacher statutes clearly define a probationary period of a set time at the end of which the teacher must take and pass some type of evaluation, the fact is that the *de jure* situation is different from the *de facto* situation. In seven of these education systems evaluations for novice teachers are not applied or no evidence was found that they are being implemented. The latter occurs in Bolivia, where there is no evidence of application of any probationary period or implementation of an evaluation for novice teachers after five years of practicing the profession. In Belize, Costa Rica and the Dominican Republic these evaluations are simply not applied. Lastly, in Jamaica, Paraguay, and Trinidad and Tobago, the information collected suggests that evaluations specially designed for this stage are in process of approval and final implementation or in design stages, so there is no evidence that they are being applied as a

whole.<sup>10</sup> The addition of the first two bars of Figure 4.5 shows that about two thirds of the education systems in the region did not have or did not implement any type of evaluation for novice teachers.

According to the review, only six of the education systems in the sample applied some type of specific test to novice teachers: Buenos Aires province (Argentina), Santa Catarina (Brazil), Chile, Colombia, Mexico and Panama. Table A7 in the appendix describes who is directly responsible for implementing the evaluation and the main instruments used in these six systems. In four of these systems novice teachers are subject to the same regular evaluation that applies to all the teachers.<sup>11</sup>

Only two cases were found where specific evaluations for novice teachers are implemented: Santa Catarina state in Brazil and Colombia. In Buenos Aires province (Argentina), Colombia and Panama, the teacher evaluation test is conducted by the principal or supervisor based on pre-established formats which, therefore, can hardly make a real differentiation of teachers' skills and weaknesses. In Santa Catarina (Brazil), Chile and Mexico, on the contrary, the evaluation is based on multiple instruments and multiple observers. In Santa Catarina, evaluations are in the hands of a commission, a central body of the State Secretariat of Education with representation in each school. This body is composed of an area manager, a coordinator and an area server, while the school commission includes the principal, a management officer, an education assistant, a specialist in education matters and a representative of the teachers. In Chile, which applies regular evaluations, tests for teachers who complete two years in the system are carried out by evaluators using a multiplicity of instruments. In Mexico the second-year evaluation is in the hands of the principal, the local education authorities, the SEP (The Secretary of Public Education) and the INEE. As the INEE report (2018) mentions, the evaluation taken by teachers in Mexico at the end of their second year, which according to the regulations determines if they remain in the system, was the same as that applied to teachers in service in 2015.

Except in Ecuador and Jamaica, the legislation regulating these evaluations in the education systems establishes that novice teachers have, on average, two opportunities to receive approval; otherwise they should be dismissed. Only Jamaica and Uruguay undertake the evaluations at this stage of the career as pure training. Although in practice teachers are hardly ever dismissed, even in countries like Mexico where the tests comprised a multiplicity of instruments and evaluators. For example, 99% of teachers who entered in 2015 and underwent the performance evaluation at the end of their second year passed the exam with a "Complies with the function" grade (INEE, 2018).

---

<sup>10</sup> In Jamaica, the new teacher evaluation formats, including novice teachers, had to be approved in September 2017, the date on which their implementation was also expected to start.

<sup>11</sup> In Chile, the statute does not establish a probationary period or a special evaluation for novice teachers. However, after two years of practice, all of them must offer evidence of the performance evaluation detailed in the next section.

## REGULAR MANDATORY EVALUATIONS

After individuals interested in teaching pass the selection processes, they are formally hired by the ministry and enter the career ladder. After completing the probationary period—in cases where this is a requirement—they then begin their professional careers. Table A8 of the appendix is a summary of an in-depth review of the legislation that defines and structures mandatory evaluations for teachers who have already been appointed to their positions. The summary highlights the most important aspects of mandatory evaluations; for example, if they are in the legislation, their frequency, the instruments used, the implications of their results and, of course, if they are implemented.

A study of the legislation shows that in 17 of the education systems analyzed, teachers are subject to regular mandatory evaluations. Figure 4.6 summarizes presence in the legislation and implementation of this type of teacher evaluation, the third one analyzed in the chapter. As can be seen, in most of the education systems reviewed in the region the legislation explicitly mentions the regular mandatory evaluations to which the teachers appointed in their positions are subject and these are, to a greater or lesser extent, implemented. In Costa Rica and El Salvador, although mandatory evaluations are defined in the statutes, they are not implemented. In Jamaica, the regulation and design of the regular evaluations were in a deep reform process in 2017, so they were classified as such. Finally, the only cases where the statutes do not mention the mandatory aspect of some type of teacher evaluation are in Belize and in Santa Catarina state, Brazil. For this reason, the analyses in this subsection presented below intentionally exclude these two systems.

The differences in the type of mandatory evaluation applied in the countries of the region are profound. The first characteristic to show this is the frequency of implementation. According to the legislation, in 11 of the 19 education systems in the study, teachers are subject to annual mandatory evaluations: Buenos Aires province (Argentina), Bolivia, Colombia, Costa Rica, El Salvador, Guatemala, Jamaica, Panama, Paraguay, Trinidad and Tobago, Uruguay. In the remaining six education systems (Chile, Ecuador, Honduras, Mexico, Peru and the Dominican Republic), the legislation establishes that mandatory evaluations to be carried out every three or four years, depending on the country in question.<sup>12</sup> Although the frequency of mandatory evaluations might seem a minor point, the fact is that it leads to

---

<sup>12</sup> Three points need to be clarified in this respect. The first is that in Honduras evaluation has two parts: an internal one in the school every year, and an external one every three years. Based on this last characteristic, the country has been classified in the second group. The second point is that in 2017 Paraguay began a pilot test of a new teacher evaluation model. However, there is no evidence that a detailed evaluation is being systematically implemented at this time. Although the current regulation establishes that the evaluation must be annual, if the new model begins to be implemented on a census basis for teachers in the country, its frequency may change. The third point is that the regulation in the Dominican Republic states that evaluations have to be at least every three years, but does not set an exact frequency.

**FIGURE 4.6 | Regular Mandatory Evaluations in Latin America and the Caribbean**

**Source:** Prepared by the authors based on data from Rodríguez and Rivera (2018).

**Notes:** i) In Paraguay in 2017 a pilot test of the new teacher evaluation was conducted, which is detailed and includes a multiplicity of instruments. ii) In Buenos Aires province (Argentina), Bolivia and Trinidad and Tobago, although the legislation mentions that these tests are mandatory, there is no concrete evidence that they are regularly implemented.

**TABLE 4.3 | Effective Implementation and Multiplicity of Instruments in Regular Mandatory Teacher Evaluations by Frequency of Application**

Frequency	Education system	Percentage of education systems that have implemented evaluations	Percentage of education systems that use multiple instruments
Annual	Buenos Aires Province (Argentina), Bolivia, Colombia, Costa Rica, El Salvador, Guatemala, Jamaica, Panama, Paraguay, Trinidad and Tobago, Uruguay.	46%	27%
3–5 years	Chile, Ecuador, Honduras, Mexico, Peru, Dominican Republic.	100%	100%

**Source:** Prepared by the authors on data from Rodríguez and Rivera (2018).

**Note:** Belize and Santa Catarina (Brazil) are not included in the table because their legislation does not mention regular mandatory evaluation.

countless differences including the probability that the evaluations are in fact implemented, the instruments used and the repercussions arising from the results, among others.

Table 4.3 divides the 17 education systems analyzed which mention regular mandatory evaluations in their legislation into two groups according to the defined frequency: annual, and every three to five years. The analysis shows two clear differences in mandatory evaluations according to this division. The first is that while in all the systems that establish a mandatory evaluation period of three to five years the evaluations are effectively implemented, in the systems where frequency is annual this percentage falls to less than half.<sup>13</sup>

The second difference has to do with the type of evaluation that is implemented. According to the review of the specialized literature, for teacher evaluations to have a real positive impact on teaching practice and thus on the quality of the education received by students, the evaluations need to have multiple instruments (MET, 2013). Due to the complexity of teaching work, different instruments must be used to observe the many facets an education professional has to achieve. As is clear from Table 4.3, while all the systems that implement these mandatory evaluations with a frequency of three to five years use multiple instruments, this percentage only reaches 27% in systems where frequency is annual. It is important to clarify which are the three systems that form the group with annual evaluations and were classified as systems that use multiple instruments. The first

<sup>13</sup> Education systems were classified in this group when it was possible to verify implementation of a standardized evaluation for all teachers. For this reason, only five of the 11 education systems that have annual mandatory evaluation are in the group: Colombia, Guatemala, Panama, Paraguay and Uruguay. The review concluded that evaluations do not apply in Costa Rica and El Salvador. In the case of Jamaica, the final formats of the new evaluation should have been approved by September 2017, so until that time they had not been implemented. For Buenos Aires province (Argentina), Bolivia and Trinidad and Tobago, it was not possible to find evidence that evaluations are effectively implemented in a continuous and standardized way. Likewise, in the case of Ecuador, as of August 2018, implementation of all the instruments established in the legislation had not been possible.

is Jamaica, which, as explained above, in 2017 was in the process of transforming teacher evaluation. Although the formats and legislation on the requirements of the annual mandatory teacher evaluation were already defined and included use of different instruments, they have not yet been implemented in the country. The second system is Paraguay, which also introduced in 2017 a new teacher evaluation scheme with multiple instruments whose pilot should begin at the end of the year. The third country is Uruguay of which it could be said that, due to the administrative division of the various education levels through the five education councils, the evaluation only applies to teachers at primary level. If these aspects are considered, the percentage could be even less than 27%.

Evaluations that include multiple instruments, and are properly designed and implemented, are costly in time and money. Applying them in the form of an annual census of all teachers in a country is not a practical education policy objective. Therefore, it is not surprising that there are enormous differences in the use of multiple instruments between both groups of systems. Even more important is the fact that this gives a clear lesson for education policy: detailed mandatory census-based evaluations, which provide valuable information and are effectively implemented, cannot be implemented annually.

The next question in the design of an evaluation policy is, of course, what kind of instruments to include in the evaluations. As mentioned in the second chapter, there is a multiplicity of instruments, each of which has its advantages and disadvantages. All education systems that implement annual mandatory evaluations and have a single instrument use a report from the school principal. These reports consider aspects related to academic, administrative and community management, as well as teacher behavior. However, as a basic report, which is not complemented with specific evidence for each aspect, it ends up being an instrument that provides little relevant information on teacher performance that can be used to improve teachers professional development and the education quality. In this respect, it is interesting to analyze the experience of countries in the region which implement mandatory evaluations with multiple instruments. A summary of these decisions and the total number of instruments used in each country according to the legislation is given in Table 4.4.

As can be seen, Ecuador is the country with the highest number of types of instruments mentioned in the legislation (six types), followed by Honduras and Chile (five types).<sup>14</sup> Most countries have established in their legislation the use of four types of instruments.

---

<sup>14</sup> It is important to make a qualification about the number of instruments used in the performance evaluation of three systems. In Chile, according to articles 12 to 17 of Decree No. 192 of the Regulation on teacher evaluation, the instruments to be used in each evaluation are: i) self-evaluation, ii) interview with a peer evaluator, iii) reference reports by third parties, iv) teacher's portfolio. This last instrument has three modules: i) teaching unit, ii) classroom video and iii) collaborative work. As observed in this review, classroom observation is a very important instrument in the region, which is why it was decided to treat it as an additional instrument. As a result, Chile has five instruments instead of four. In Ecuador and Honduras, evaluation systems include both parent and student surveys; if these two instruments are considered independently, it could be said that they use a total of seven and six, respectively.

**TABLE 4.4 | Instruments Used in Regular Mandatory Teacher Evaluations with Multiple Instruments**

Country	Report of principal or supervisor	Peer evaluation	Self-evaluation	Portfolio	Classroom observation	Exam	Student and/or parent surveys	Others
<b>Annual</b>								
Jamaica	■		■		■			
Paraguay			■		■			
Uruguay	■				■			
<b>3–5 years</b>								
Chile	■	■	■		■			
Ecuador	■		■		■	■	■	
Honduras	■	■					■	
Mexico	■		■			■		
Peru					■		■	■
Dominican Rep.	■		■		■			■

**Source:** Prepared by the authors based on Rodríguez and Rivera (2018) using evaluation information for each country.

**Note:** In Honduras, the peer evaluation consists on classroom observation using the Stallings rubric.

This is the case in Jamaica, Mexico and the Dominican Republic. Only three laws specify a smaller number: Paraguay and Peru, three types, and Uruguay, two types. According to the review of the legislation, the two instruments most used in this group of countries are reports by the principal or supervisor and classroom observation. Only Paraguay and Peru do not involve the opinions of principals at this stage, while only Honduras and Mexico do not include an observation of teaching practices as an evaluation instrument. In almost 67% of countries the legislation includes self-evaluation and analysis of a teacher portfolio as instruments. The countries that do not include the first instrument are Honduras, Peru and Uruguay; while countries that do not include analysis of a portfolio are Peru, Dominican Republic and Uruguay. The instruments mentioned less frequently in the legislation of the countries are knowledge tests—which exist in the legislation of Ecuador, Honduras and Mexico—and student and/or parent surveys—included by Ecuador, Honduras and Peru. Finally, only Chile and Honduras include peer reports, while Peru and Dominican Republic use specific instruments to evaluate space management and responsibility (in Peru both are used and in Dominican Republic only the latter is used).

Besides using a multiplicity of instruments, there are also aspects related to its design—such as who is responsible for implementation and which actors participate at each stage—that have an important impact on the results. The group of countries that use a single instrument all use the report of the principal or supervisor, so they are the main actors responsible for the evaluation. The interesting point, therefore, is to analyze which actors are involved in countries that have mandatory evaluation with multiple instruments. A first

**TABLE 4.5 | Participants in Regular Mandatory Teacher Evaluations with Multiple Instruments**

Country	Principal/ immediate superior	Peer	Teachers themselves	Rating experts	Students/ parents
<b>Two evaluators</b>					
Uruguay					
<b>Three evaluators</b>					
Jamaica					
Mexico					
Paraguay					
Peru					
Dominican Republic					
<b>Four or more evaluators</b>					
Chile					
Ecuador					
Honduras					
<b>Percentage of countries</b>	100,0%	55,6%	77,8%	77,8%	33,3%

Source: Prepared by the authors based on Rodríguez and Rivera (2018).

approach to this is presented in Table 4.5 that shows what type of evaluators are involved in the evaluation processes in each of these countries.

It is noteworthy that, although not all the countries include a report from the principal or supervisor as an instrument, this actor participates in one way or another in the mandatory evaluation processes in all these countries. Specifically, in Peru the principal is part of the Teacher Performance Evaluation Committee, along with the assistant principal or coordinator of the level and a teacher of the same education level as the evaluated teacher (or a higher level) from another education institution. In Paraguay, the principal and a peer teacher are responsible for conducting the classroom observation.

The second actors most involved in these multi-instrument evaluations are an expert external to the institution or school and the teacher her/himself. The inclusion of external evaluators is undoubtedly a positive aspect worth highlighting. The advantage of having this type of evaluator is that they can give teachers peace of mind because, in principle, their results will be without the subjectivities of internal evaluators who rate their work on a day-to-day basis. The countries in which the education authority (ministries or councils in the case of Uruguay) designs the evaluations but assigns external personnel to implement part of them have been classified as a first level of independence. For example, in Uruguay the external evaluation, corresponding to the classroom observation, is carried out by inspectors, who report directly to the respective councils. In Paraguay, the role of external evaluator is taken by an academic peer from another education institution.

Honduras and the Dominican Republic are at a second level of independence. In Honduras there is the General Directorate of Curriculum and Evaluation, part of the Ministry of Education, responsible for coordinating the entire process of teacher evaluations, from design of the knowledge exams to training and designation of external observers. In the Dominican Republic, the person responsible for the design and implementation of the latest teacher evaluation is the Dominican Institute of Evaluation and Research for Educational Quality, a technical body attached to the Ministry of Education.<sup>15</sup> Chile, and Ecuador have independent institutions that deal with design of the evaluation and its execution, which gives the process even greater independence: Measurement Center MIDE of the School of Psychology of the Pontifical Catholic University of Chile; National Institute of Education Evaluation in Ecuador. The existence of these institutions means that, at different levels, the design and implementation of mandatory evaluation has enough technical autonomy to prevent decisions on evaluation from depending on preferences or pressures from the government or minister of the day.

Of the countries analyzed, 77.8% also include the teachers themselves as actors (see Table 4.5). This category includes, in addition to the teacher's obvious participation as an evaluated subject, a specific role as instrument of self-evaluation of their own teaching practice (Chile, Ecuador, Jamaica, Mexico, Paraguay and the Dominican Republic) or taking knowledge test (Ecuador, Honduras, Mexico).

The second column of Table 4.5 indicates that half the countries that use multiple instruments in regular evaluations include a teacher's peer as one of the evaluators. In Jamaica and Paraguay (according to the legislation in process of approval at the end of 2017) the peer participates in the classroom observation. In Ecuador, the original plans gave the peer the responsibility of contributing to evaluation of the teacher portfolio. In Honduras, peers are involved both in classroom observation and in the planning measure, class development, time management and evaluation of students. Finally, in Chile the peer is responsible for conducting an interview with the teacher to learn a little more about their teaching practice and the context of their professional work.

The actors included less frequently in the evaluation process are students and parents by means of responding to a questionnaire designed exclusively for them. This is not surprising since arranging for them (especially the parents) to answer the questionnaire is very difficult logistically. Only Ecuador (where this is included in the legislation), Honduras and Peru specifically include them in the process.

---

<sup>15</sup> As noted above, in Colombia the principal is the only actor responsible for carrying out the regular mandatory evaluation. However, this country also has the Colombian Institute for Evaluation of Education, a specialized entity of the national Ministry of Education responsible for conducting the student knowledge exams (state exams at 11th grade and knowledge tests (*Pruebas Saber*) at the 3rd, 5th and 9th) and the teacher evaluations associated with career entry and professional promotion possibilities, the latter described below.

**TABLE 4.6 | Weights Assigned to Instruments in Regular Mandatory Teacher Evaluations with Multiple Instruments**

Country	Report of principal or supervisor	Peer evaluation	Self-evaluation	Portfolio	Classroom observation	Exam	Parents/Students	Others
Paraguay			10%		90%			
Chile	10%	20%	10%	17%	34%			10%
Ecuador	5%		18%	8%	17%	48%	4%	
Honduras	15%	15%		30%		30%	10%	
Peru					50%		17%	33%
Dominican Republic	15%		10%		45%	20%		10%
<b>Average</b>	<b>11%</b>	<b>17%</b>	<b>13%</b>	<b>19%</b>	<b>43%</b>	<b>33%</b>	<b>9%</b>	<b>16%</b>

Source: Prepared by the authors based on Rodríguez and Rivera (2018).

Note: In Honduras, the percentage assigned to peer evaluation (15%) includes class observation.

In addition to choosing the type of instruments to be used, the design of teacher evaluation processes involves making another important decision in systems with a multiplicity of instruments: determining the weighting to be assigned to each instrument. The analyses of the results of the MET project (2013) indicate that this is a fundamental choice which has implications for the validity and reliability of the process. Table 4.6 presents the weight assigned by legislation to each instrument used in six countries that assign weights to each instrument. As can be seen, the instrument given the highest weight on average (43%) is classroom observation. Although Honduras appears as the only country that does not assign weighting to this instrument, the reality is that peer evaluation includes classroom observation as well as measurement of planning (see Table A8 in the appendix). The second instrument to which the countries of the region give most weight is the knowledge test, with an average of 33%. This is followed in order of weight by the portfolio and peer evaluation. Lastly, self-assessment, the principal's report and parent and student surveys receive a similar weighting in the region taken as a whole.

Three countries are not referenced in Table 4.6: Jamaica, Mexico and Uruguay. In Mexico, the weighting in the country's official documents leads to the conclusion that a 60% weighting is assigned to aspects related to evidence and reflection on teaching practice, while the remaining 40% goes to the score obtained in knowledge exams. In Jamaica, according to the legislation, the weighting is assigned in line with the dimensions of the guiding principles that govern the teaching profession in the country and not directly to a particular instrument.<sup>16</sup> The case of Uruguay is similar, where the instructor gives the

<sup>16</sup> The weights for each guiding principle are: i) content knowledge, 23%; ii) pedagogical knowledge, 23%; iii) professional leadership, 23%; iv) management of student diversity, 14%; v) continuous improvement of professional skills, 9%; vi) relations with parents, 8%.

teacher a final rating ranging from 1 to 100 divided between seniority level (up to 20 points), attendance (up to 20 points) and teaching aptitude (up to 60 points). This last category is made up of the rating given by the inspector and the principal of the education institution.<sup>17</sup>

A comparison of the instruments and weightings assigned by the countries of the region that implement regular mandatory teacher evaluations with the United States and the other systems analyzed in the previous chapter reveals significant differences. The first one has to do with the classroom observation instrument. While all the education systems discussed in Chapter 3 include this instrument in teacher evaluations, in the region it is only implemented in seven of the 19 education systems studied. As Bruns and Luque (2015) show, teachers in the region reveal teaching practices that are far from the standards considered adequate by specialized international literature, consequently education policies need to include priority actions in this respect. However, the lack of application of instruments, such as classroom observation, inevitably means that the education authorities do not even have the relevant information to begin to improve this problem.

The second clear difference is that while in the United States most states and districts include performance measures in students standardized tests and assign them, on average, the second highest relative weight, no country in the region uses this instrument. As the specialized literature points out, dispensing with this type of instrument is not necessarily detrimental. Measures of student achievement, although useful, also have disadvantages, because they are generally not available to all teachers, they can encourage inappropriate behaviors by teachers—reducing their work to “teaching to the test” or copying—are subject to measurement problems if not properly designed and their use requires evaluators to be diligent and subject to additional trade-offs when defining which methodologies to use to calculate the value added estimates.

The crucial issue is that evaluation, with the set of instruments selected, can effectively reveal valuable information about teacher practices. Even more important is to use this information proactively to improve the quality of education through the objectives that evaluations are designed to achieve. As mentioned before, three key aspects have to be considered to achieve this: i) adequate design and technical implementation, ii) results that differentiate and rate teachers into different groups based on their content knowledge and pedagogical skills, and iii) appropriate use of the instruments with repercussions on the teaching career or implementation of relevant high-quality professional development programs. The following paragraphs highlight specific points of each of these topics.

Reviewing the design and implementation processes of each instrument in each country is an objective beyond the scope of this study. However, the review did identify important issues. The first is that not necessarily everything established in the law can be

---

<sup>17</sup> No information was found on the weight assigned to the score given by each of these evaluators. However, according to the OECD report (2016), the final rating must be approved at an annual meeting with the participation of both inspectors and principals, where the latter have voice but no vote.

carried out in practice. Table 4.3 on mandatory evaluations clearly shows that less than half the education systems implement annual mandatory evaluations even though they are in the legislation. Although in systems that carry out evaluations with multiple instruments the percentage of implementation reaches 100%, it does not mean that all the provisions of the law are fully enforced. For example, in Uruguay, although the law includes the process of classroom observation and sets the minimum number of observations, and how to treat the results, the real situation is far from full compliance with the legislation. For example, in high school there are so few inspectors that many of the teachers who teach at that level have said they have never been evaluated. The differences are clearly set out in the report by the teachers themselves in the National Teacher Survey held in Uruguay in 2015, where 65% of teachers in public primary schools report having received visits from an inspector in 2015, while this percentage drops to 14% in secondary school (INEEd, 2017). Similarly, although Ecuadorian legislation states that teacher evaluation must include seven instruments, the reality is that, as of September 2017, only one of them had been implemented on a census basis: the content knowledge test.

Moreover, even though all the instruments mentioned in the law are implemented, their form of implementation and their quality can vary significantly from one country to another, as noted above in the case of high-performance education systems. For example, in Chile to conduct classroom observation (included in module 2 of the portfolio) teachers are invited to show their work in the classroom by recording a 40-minute class. This recording, which must be made by a person authorized by Docentemás (responsible for developing the various actions and materials required for the evaluation process) and free of charge for the teacher, is then evaluated by teams of practicing classroom teachers in universities (they are not university professors or students, as is often wrongly claimed), under the supervision and training of Docentemás (Docentemás, 2017). These classroom teachers have training and experience at the same level, in the same subject or modality as the teachers whose portfolios they are correcting. In addition, although these teachers receive detailed rating rubrics aimed at ensuring the quality of work, 20% of the portfolios are rated by two different teams.

The Dominican Republic has also made an interesting effort. With the help of the Organization of Ibero-American States, the government launched a census-based classroom observation process covering almost 60,000 teachers, with specially trained observers. Unlike Chile, in this country the observation takes place in the schools themselves. Precisely because of the logistics, it was not possible to review random or specific cases. In Ecuador, due to budgetary and technical restrictions, the National Institute of Education Evaluation made the decision not to evaluate teaching practices through classroom observation. The option for a viable evaluation of classroom practice is the use of a digital instrument consisting of presenting teachers with a series of videos with specific examples of everyday situations in the classroom. Afterwards teachers answer multiple-choice questions choosing the alternatives that seem most appropriate to them for use in the classroom. At the end of 2018, however, this stage of the evaluation had not yet been applied.

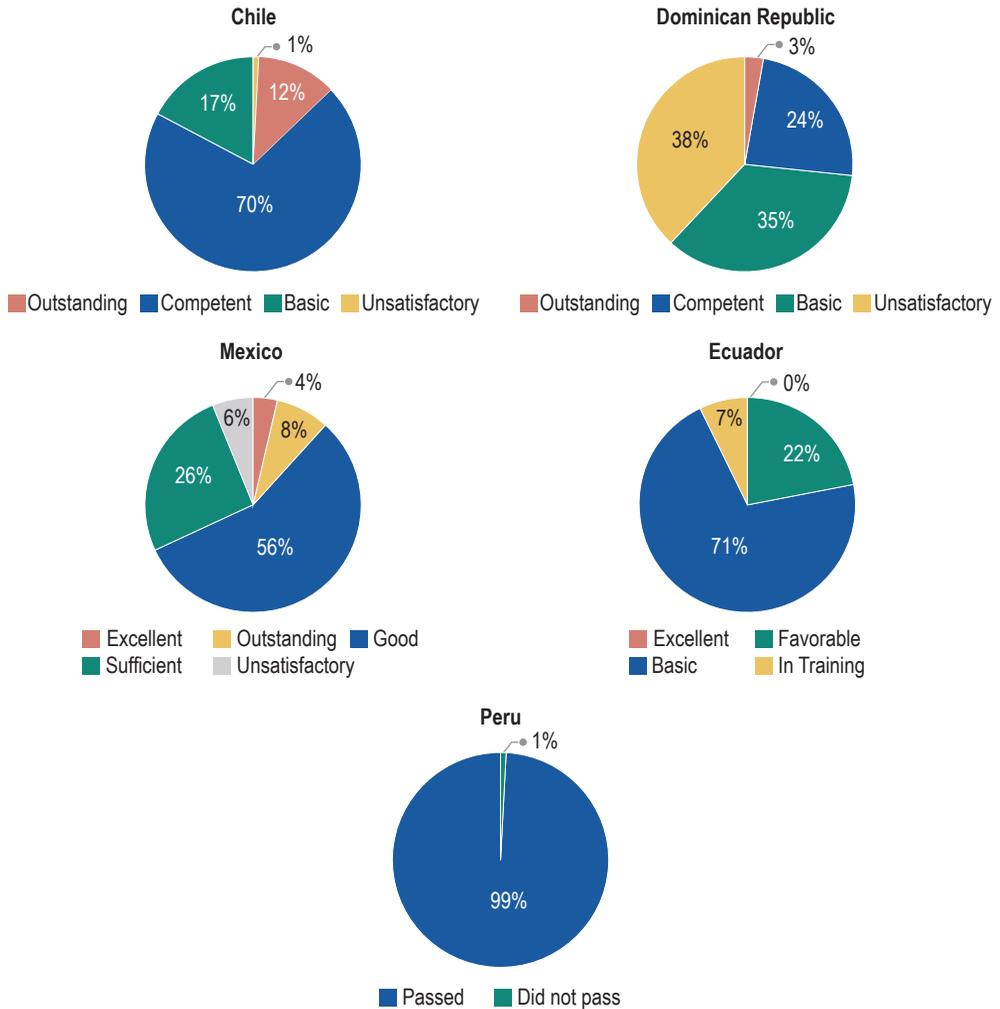
Another important issue identified by the review indicates that differentiation of teachers by pedagogical skills and content knowledge depends, in addition to the instruments and their implementation, on the categories or ratings derived from the evaluations. Table A9 of the appendix presents the analysis for the systems where their legislation mentions mandatory evaluation and shows the ratings that teachers can obtain.<sup>18</sup> Six systems chose to assign a continuous score with specific levels only some of which have an impact on teachers' careers. In the other nine countries, the legislation establishes that the rating given in the evaluation process is categorical. More than half these systems define four categories in which a teacher can be rated on their results. Only three countries opted for a special number of categories. For Peru, evaluation determines if teachers approved or not (two categories); for Colombia it consists of three categories, and finally in Costa Rica five were chosen. Figure 4.7 shows the most recent available results of regular mandatory teacher evaluation for Chile, Ecuador, Honduras, Mexico, Peru and the Dominican Republic.

The number of categories is important for teacher differentiation, but more categories do not necessarily imply better differentiation. Costa Rica, which defined the largest number of categories, does not even implement the evaluation. Similarly, for most of the systems that use annual teacher evaluations, the possible scores—based on the principal's report—are numerical and continuous, giving a wide possibility of differentiation. However, it is estimated that a significant part of the teachers in these systems are rated with scores concentrated in the good or outstanding categories. In addition, in these cases a wide differentiation in scores between teachers would not solve the poor quality of the principals' reports, which prevents valuable information reaching the teacher and the education authorities on their work performance.

At the other extreme, in Peru the evaluation defines only two categories: approved or not approved. However, with multiple instruments including classroom observation, the information collected in the process can be very valuable if used properly. For example, evaluation of kindergarten teachers implemented in 2017 in that country reported that 99.3% of teachers who participated in the process approved the evaluation. However, analysis of the results of the classroom observation rubric for these teachers shows a significant variation in aspects related to teachers' treatment of students, their sensitivity to them and promotion of autonomy and language skills; and identifies teachers who require specific improvements in these aspects. The situation is similar in Chile where in 2015 82% of teachers were rated as competent or outstanding, but an analysis of the scores obtained in the portfolio instrument lowered the percentage to 37%. The information contained in this tool

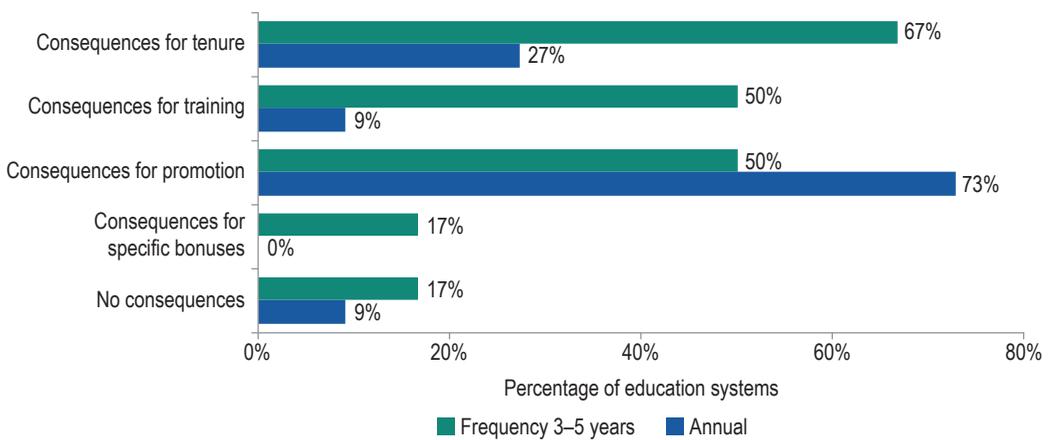
---

<sup>18</sup> The table does not include the observation of four systems. As mentioned, in Belize and Santa Catarina (Brazil) no such evaluation is mentioned in the legislation, so they are naturally excluded. In El Salvador and Trinidad and Tobago, despite being mentioned in the legislation, either they are not applied or there is no information on them.

**FIGURE 4.7 | Results of Regular Mandatory Teacher Evaluation in Selected Countries**

Source: Prepared by the authors based on Rodríguez and Rivera (2018).

is precisely what makes it possible to understand each teacher's strengths and what specific aspect needs to be improved. Lastly, it is useful to mention the evidence collected in the Dominican Republic during the most recent teacher evaluation. According to reports from the ministry, the rating related to self-assessment and the report of the immediate superior suggest that more than 80% of teachers receive a high rating, but the results of the instruments related to classroom observation and planning analysis and professional performance are very different: according to these three instruments, only 40% of teachers are in the high category. There are also specific deficiencies in the probability of teachers

**FIGURE 4.8 | Consequences of the Results in Regular Mandatory Teacher Evaluations in Latin America and the Caribbean by Frequency of Implementation**

**Source:** Prepared by the authors based on data from Rodríguez and Rivera (2018).

activating prior knowledge in their students, guiding their students in activities aimed at research or even making optimal use of the time allocated to the class.

The study of mandatory evaluations implemented in the region also analyzes the consequences arising from their results, something that is naturally linked to the objectives they aim to achieve. Figure 4.8 summarizes the information collected in the review, and again divides the countries by the frequency of evaluations (the consequences of mandatory evaluation for each country are detailed in Table A10 of the appendix). In education systems with mandatory evaluations with a frequency of three to five years the most common consequences are for teacher tenure (67% of cases) and the possibility of training and promotion (50% of cases). In education systems with annual evaluations, these percentages fall to 27% and 9% for tenure and training, respectively, and the figure rises to 72% for promotions.

With respect to the consequence of tenure, generally in both groups the legislation establishes that teachers must obtain a minimum score in the regular mandatory evaluation to guarantee their tenure in the teaching profession. If teachers do not attain this score, they have one or two additional opportunities to do so. In systems with annual evaluation this takes place during the following year and in systems with more spaced evaluations, the regulations generally set the same conditions. If, in these additional evaluations, teachers do not attain the minimum score, the legislation stipulates that they are to be dismissed. However, in the countries of the region, irrespective of the type of regular evaluation implemented, teachers are almost never dismissed due to poor performance in regular mandatory tests.

In the case of the implications of the evaluation results for teacher training processes, once again what the law says is not necessarily what applies. Among the systems with

annual evaluations only Paraguay ties the results to in-service training programs. However, in practice, in 2017 Paraguay had not implemented this aspect. Among the systems with multi-instrument evaluations, Chile, Mexico and Peru include this consequence specifically in the legislation. However, although progress has been made in this area, the small amount of evidence available suggests there is also a long way to go in these systems in this area. Chile has been continuously implementing training processes linked to the results obtained in evaluations. Cortés et al. (2011) analyze some aspects related to the Professional Improvement Plans (PSP) in Chile and conclude that valuable efforts have been made, but there is still room for improvement. In particular, the authors suggest the need to assure the quality of the training, which at the time of the study was a decentralized decision and therefore led to very different qualities of training depending on the municipality. They also point out the importance of implementing training during teachers' working hours. The authors also emphasize that their work is not an impact assessment, and that due to the importance of the subject it is important to make a study of this type.

Mexico has conducted several rounds of evaluations but, according to the education authorities themselves, they have not yet succeeded in integrating the training programs into the process (Schmelkes, 2017). Santibáñez, Rubio and Vázquez (2017) analyze the National Strategy for Continuing Training and Professional Development in that country and conclude that the efforts of the federal and state education authorities have improved coordination between the evaluation and training processes. For example, because it is distance training—the only possible alternative at present according to the authorities—a high value is placed on the involvement of prestigious higher education institutions with responsibility for the training. The study also found that there is ample space for developing more prior teacher training in information technology tools, considering that it is a distance modality; and ensuring the quality of the training provided. Likewise, the study recommends that more resources need to be invested in this fundamental aspect. As in the case of Cortés et al. (2011), the study for Mexico does not constitute an impact assessment, which clearly reflects the vacuum in the region with respect to this essential factor of the theory of change. Finally, in the case of Peru, although the legislation establishes that teachers who fail the evaluation must receive training, it was not possible to find evidence that this is happening in practice.

Figure 4.8 also shows that in systems with annual mandatory evaluations it is much more common to use their results as input or requirement when analyzing career advancements than it is in systems with evaluations with a frequency of three to five years. Among the education systems with annual mandatory evaluations, Bolivia, Colombia, Guatemala, Jamaica, Panama, Paraguay, Trinidad and Tobago and Uruguay have legislation that grants the results the capacity to define the possibility of promotions or to function as criteria for such processes. Among the education systems that implement evaluations with multiple spaced instruments, the only ones that do not use their results as criteria for possible

promotion are Ecuador, Honduras and the Dominican Republic.<sup>19</sup> The final bars of the figure indicate that: i) only in the Dominican Republic does the legislation links evaluation results to specific monetary bonuses and ii) and Honduras and El Salvador are the only cases where the legislation does not link the results of mandatory evaluation to any type of consequence.<sup>20</sup>

Finally, it is worth mentioning the important efforts made by the ministries in research into the design, implementation and impact of teacher evaluation systems. Two examples are Chile and Mexico. In Chile, various types of research have attempted to answer three questions: i) how the different instruments work and how is it possible to improve them, ii) what perception do teachers have of the evaluation system and the instruments used, and iii) the relevance of evaluations and instruments. On the first topic, Sun (2017) mentions at least two recent improvements that have had an evident impact on evaluation results: i) since 2010 the principals have had to substantiate the outstanding rating granted to their teachers, which determined that, compared with 2008 and 2009 when about 30% of teachers were rated at this level by the principals, this proportion fell to almost half and now on average only 15% of teachers get this rating; ii) since 2014 peer evaluation has included observation of actual practices, which has meant issuing specific skills with holistic rubrics which have to be taken into account. Since this change, the ratings given by evaluating peers has changed significantly in the upper tail, from almost 30% of teachers rated as outstanding in 2012 and 2013 to 15% in 2014 and 2015. A second type of study has done research on teachers' perception of the evaluation system and the usefulness and relevance of the instruments used. For this, the evaluations included optional complementary questionnaires asking teachers about the similarity between the portfolio and their daily tasks or their specific opinion of that instrument. Research shows that 72% of teachers agree that portfolio tasks are similar to those carried out in their daily work as teachers. Similarly, 86% of teachers believe that preparing the portfolio is useful or very useful for enhancing reflection on their teaching practices (Docentemás, 2017). The third type of research relates to the importance of evaluations and instruments. As mentioned in chapter 2, studies such as Taut et al. (2016) have found correlations between the rating obtained by teachers in the portfolio and value-added measures of the students.

The second experience is in Mexico which has investigated the operation of its mandatory evaluation systems and the impacts of implementation of these systems on progress in the teaching career. For example, after the performance evaluation implemented in

---

<sup>19</sup> In Ecuador, the regulation is not clear and there is no evidence that these promotions are actually taking place. In Honduras, the criteria used for horizontal promotions are based on service time and education level. Finally, in the Dominican Republic as yet there is no possibility of horizontal promotion.

<sup>20</sup> As mentioned earlier, El Salvador's legislation establishes that evaluations must be annual, but does not detail any characteristics for compliance. Moreover, in practice mandatory evaluations are not implemented in the country.

2015, INEE—in collaboration with the UNESCO Regional Office of Education for Latin America and the Caribbean (OREALC/UNESCO)—conducted an evaluation of its implementation, which included a satisfaction survey of 10,000 teachers and focus groups of about 410 teachers. The results of this research led to a reform of the mandatory evaluation model from 2017. In addition, research has been done to determine the impact of specific aspects of some reforms, such as Estrada (2015) and Santibáñez et al. (2006) referenced in Chapter 2 which evaluated the career entry exams and introduction of the Teacher Career Program, respectively.

Although all these efforts are valuable, more research is needed in the region to analyze the validity and consistency of the results produced by these new evaluations, as well as the impacts on the teachers and students.

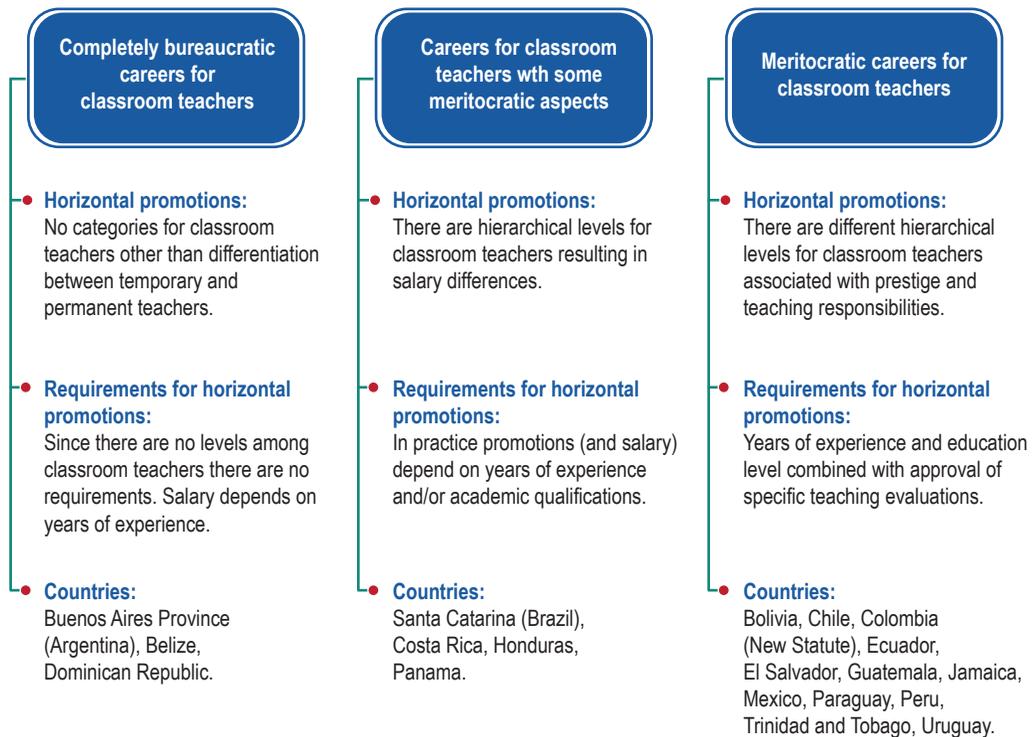
## EVALUATIONS FOR HORIZONTAL PROMOTIONS

The last type of evaluation analyzed in the study determines whether teachers have the skills needed to advance in their career. The existence of these evaluations depends, first, on the structure of the teaching career defined in each country's statutes or legislation. Therefore, the review of this last evaluation must begin with an analysis of the types of careers for which teachers in the region enter the profession.

The differences in the types of careers for teachers in the 19 education systems in the study are evident when analyzing their possibilities of promotion and the requirements for achieving it. To understand these options, it is useful to divide professional promotions into two classes: vertical and horizontal. Vertical promotions oblige teachers to leave the classrooms to access administrative work not necessarily directly related to teaching. For example, teachers may choose to advance to administrative positions inside or outside schools as assistant principals or principals, or to positions in the ministry of education at regional or national level. The statutes of the 19 education systems offer this possibility to their teachers who in general can reach these positions through special selection processes.

Horizontal promotions allow teachers to increase prestige, functions, responsibilities and salary level without the need to leave the classroom. Unlike vertical promotions, not all teachers in the region can be promoted horizontally in their career. According to the analysis of this aspect of teacher statutes, education systems can be divided into three groups, as shown in Figure 4.9, according to the possibilities of horizontal promotions and the requirements teachers must meet.

The first group consists of education systems whose teacher statutes do not establish differentiated categories of teachers, apart from distinguishing between permanently hired and temporary teachers. These are completely bureaucratic careers because their regulations do not give teachers the possibility of professional growth. Belize is the extreme case since there is no defined career, followed by the Dominican Republic, where the statute

**FIGURE 4.9 | Possibilities of Horizontal Promotions in Teaching Careers and Their Requirements, According to Teacher Statutes**

**Source:** Prepared by the authors based on data from Rodríguez and Rivera (2018).

only defines the teacher position, and Buenos Aires province (Argentina) where classroom teachers can be classified only as teachers or special teachers. The only thing that differentiates teachers in these three systems is years of teaching experience which defines their salary. It is worth mentioning that in 2016 the Dominican Republic designed a Guide for Certification and Development of the Teaching Career to classify teacher performance into five categories ranging from excellent to insufficient, and then link these categories to bonus pay and special recognitions. However, this certification is not yet being applied in the country, which is why the system was listed in this first group.

The second group contains education systems whose statutes explicitly define different levels or degrees of hierarchy for teachers who wish to remain in the classroom. However, these promotions are not meritocratic and depend mainly on requirements associated with years of experience in a specific category and/or the teacher's education level. This group includes Santa Catarina (Brazil), Costa Rica, Honduras and Panama. Although in Brazil there are examples of evaluations for promotion, such as the introduction of the

accelerated career in São Paulo, in many of the country's states and municipalities, in particular in Santa Catarina state analyzed in this study, the statute only mentions academic qualifications as a requirement.

The third group is made up of countries whose statutes offer their teachers meritocratic careers. As well as allowing them to rise in their profession with higher salary levels without leaving the classroom, these promotions are also generally associated with higher levels of prestige in the teaching profession because of their meritocratic nature. In some cases, the higher categories in the career give teachers the possibility of becoming teaching leaders or guides acting as tutors or mentors to their peers. In these countries the legislation requires teachers to pass some type of evaluation and obtain higher levels of education and experience to reach this level.

There is an additional difference between the countries listed in the third group related to the number of categories and, therefore, of horizontal promotions regulated in each statute (Table A11 of the appendix presents the categories defined by each law). While in most of the countries in this group the scales are numerical or designated by letters, in Chile and Jamaica horizontal promotions are associated with categories with a denomination clearly related to the teacher's experience and prestige. In Chile there are five categories, which the statute calls "sections": initial, early, advanced, expert I and expert II. Complying with the requirements to ascend in the first three sections is mandatory for all teachers in the country, while moving up to the last two sections is a voluntary process. In addition to completing the required years of experience, expert teachers need to obtain the best results in the two performance evaluation instruments: the portfolio, and the evaluation of content mastery and pedagogical knowledge. In Jamaica there are three categories for classroom teachers: beginner, effective and master. While ascending from beginner to effective is mandatory if teachers want to continue the career, moving up to the master category is optional.

The details of the career, chances of promotion and requirements are presented in Table A12 of the appendix for all systems. Evidently, it is only possible and useful to analyze the requirements for promotion for the countries listed in the third group of Figure 4.9. Table 4.7 summarizes the instruments that, according to the provisions of the legislation, must be used when evaluating horizontal promotions in these 12 countries, aside from the requirements related to time of service or experience in each system. It is noteworthy that about 70% of the countries include as requirement a component related to the regular mandatory evaluations described in the previous section, which makes it the most common input in the promotion evaluations in the region. Second place is taken by the specific tests designed for that purpose which evaluate content mastery and pedagogical knowledge. Except for these two instruments, requirements for teachers are very heterogeneous and there is no clear pattern in their use.

As with the previous evaluations, what is defined in the law is not necessarily what happens in practice. Many promotions do not depend on any type of evaluation based on

**TABLE 4.7 | Instruments for Horizontal Promotion Evaluations in Countries Where this Possibility Exists and the Legislation Includes Meritocratic Aspects**

Country	Average/ score or mandatory evaluation component	Self- evaluation	Classroom observation	Exam	Research	Student parent surveys	Global performance matrix	Interview
Bolivia								
Chile								
Colombia								
Ecuador								
El Salvador								
Guatemala								
Jamaica								
Mexico								
Paraguay								
Peru								
Trinidad and Tobago								
Uruguay								
<b>Percentage of countries</b>	<b>67%</b>	<b>8%</b>	<b>8%</b>	<b>58%</b>	<b>8%</b>	<b>8%</b>	<b>8%</b>	<b>8%</b>

Source: Prepared by the authors based on data from Rodríguez and Rivera (2018).

real evidence of teachers' skills and in the end are bureaucratic. In Bolivia, legislation specifies the requirements to be met to receive promotion, but there is no evidence to suggest that they are applied, which, as we have seen, also occurs with the other evaluations. In El Salvador, the review found evidence that teachers do not allow any evaluations to be implemented, therefore, again, promotions end up based on levels of education and experience. Both in Guatemala and in Uruguay, to determine promotions, the law requires use of regular mandatory evaluation scores, but as it became clear in the analysis of the previous section, they actually end up depending on a single instrument conducted by supervisors who do not provide detailed evidence of teachers performance. Likewise, in Trinidad and Tobago, no evidence was found that the interview required for promotion is adequately structured and needs to be complemented with evidence of teaching work, so it is not expected to provide accurate information on teaching practices in the country. In this context, only seven of the countries analyzed in this research have promotions that depend on specific evaluations based on evidence of pedagogical skills and content knowledge: Chile, Colombia, Ecuador, Jamaica, Mexico, Paraguay and Peru. Table A13 in the appendix details which instruments are included in the evaluations in this group of countries.

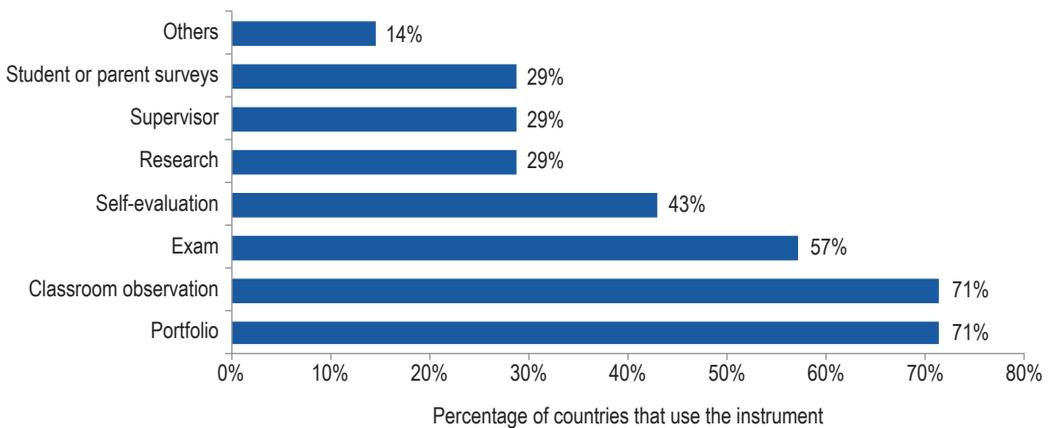
The analysis in Table 4.7 and Table A13 of the appendix reveals three interesting aspects. The first is that education systems that implement detailed regular mandatory evaluations use their specific results or components when considering the promotion possibilities of their teaching staff. It is expected that, after investing significant time and financial resources in these mandatory evaluations, countries use the information collected on teacher performance. In fact, within the group of education systems that implement mandatory evaluations with multiple instruments, only Honduras and the Dominican Republic do not use the results in their horizontal promotion processes. In Honduras, this is because the legislation only establishes experience and education as requirements for career progress. In the Dominican Republic the possibility of horizontal promotion is not defined in the statute.

The second relevant aspect has to do with the fact that in most of the seven countries: Chile, Colombia, Ecuador, Jamaica, Mexico, Paraguay and Peru, additional teacher promotion instruments are included. For Colombia and Jamaica, promotion evaluations include multiple additional instruments not included in the regular mandatory evaluations detailed in the next paragraph. For their part, Chile and Peru complement the mandatory evaluation information with additional standardized exams that teachers interested in promotion must take. In Mexico, teachers, as well as obtaining an “Outstanding” or “Excellent” result in the performance evaluation, had to obtain an “Outstanding” or “Relevant” result in an additional evaluation whose characteristics were determined by the INEE. Although in Ecuador the regulations establish that teachers must submit additional exams, the reforms in process of being implemented will include the rating obtained in the census-based examination applied to teachers as part of the mandatory evaluation in 2016. In Paraguay, the statute requires that teachers carry out an academic research project. However, there is no evidence that this requirement exists and, if it does, there is no information about the quality required for the research.

The third aspect, mentioned previously, is linked to the fact that in this group of countries Colombia and Jamaica have designed the largest number of additional specific processes for this evaluation, ensuring the use of multiple instruments. As in Colombia the regular evaluation is based on a single instrument that does not really provide detailed information on the teacher’s professional skills, it is necessary to include aspects such as classroom observation, self-evaluation and student surveys. In that respect, two points should be made: i) the new evaluation model for promotion, known as the Diagnostic and Formative Evaluation (ECDF) for teachers, emerged after a tough negotiation process with the teachers’ union, which advocated changing this test which was previously based on a standardized knowledge test finally eliminated (Figueroa et al., 2018) and ii) similarly for Chile, the classroom observation instrument is based on a video made by the teachers themselves complying with specific details of quality and content.<sup>21</sup> According to the

---

<sup>21</sup> For example, each video must focus on educational, learning, teaching and classroom practice in one of the classes taught by the teacher. The class must develop in three stages: 1) contextualization, ii) development of the class, 3) final reflection. Information retrieved from the Ser Maestro website on 2017.

**FIGURE 4.10 | Instruments Used in Horizontal Promotion Evaluations in Meritocratic Teaching Careers in Latin America and the Caribbean**

**Source:** Prepared by the authors based on data from Rodríguez and Rivera (2018).

instructions, this video is also evaluated by two peers: one from the teacher's regional context and one from the national context. In Jamaica, as detailed above, the legislation on regular evaluations is new and there is no prior information on it.

In that sense, Figure 4.10 summarizes the type of instruments that this group of seven countries uses for processes of promotion evaluation. They clearly attach importance to teaching practices, planning and content knowledge because these are the aspects most evaluated when defining horizontal movements in the career. Almost half the countries include as evaluation criteria a process of teacher self-evaluation and reflection and, to a lesser extent, some include the perception of supervisors, students and parents or the teacher's capacity to undertake research.

When there are multiple instruments, the question is raised about the weight assigned to each one by the education authorities. Unlike the case of regular mandatory evaluations, it is not possible to summarize the information to give each instrument a specific weight since, as mentioned before, many of the countries in this group use as input information previously collected in the mandatory evaluations. For example, in Chile, the level a teacher reaches on the career ladder depends on both the score obtained in the portfolio instrument (which in this report includes classroom observation and planning) and the knowledge test.<sup>22</sup> In Mexico, promotions depend exclusively on the final rating received by the teacher in the mandatory evaluation and on the additional examination that had to

<sup>22</sup> Figure A3 in the appendix shows the combinations of scores that teachers need to obtain in Chile to be promoted to each category.

be determined by the INEE, while in Ecuador they are based on the score achieved in the content knowledge exam. Only in Colombia and Paraguay is it possible to assign a specific weight to all instruments. In Colombia, classroom observation is weighted at 80%, scores obtained in the mandatory evaluation of the previous two years at 7.5% and student surveys and teacher self-evaluation take the remaining 12.5%. In Paraguay, weightings are distributed as follows: 30% for experience and professional certification, 50% for performance evaluation scores and 20% for academic research.<sup>23</sup>

Finally, it is interesting to note that promotions are not the only consequence linked to the results of this evaluation. In Chile, for example, if teachers fail the evaluation twice in a row and, therefore, do not move on from the initial or early section, they are removed from the position. Teachers who were already in the early section can return to the initial section two years later. However, if they do not move up to the advanced level, they are permanently dismissed. Also, as explained in the previous section, in-service training programs are generally linked to the results of regular evaluations. Colombia is the only country that links the results of the teacher evaluation for promotion to in-service training programs. All teachers who fail to pass the horizontal promotion evaluation must undergo a training program to overcome the weaknesses revealed by the evaluation. To do this, various faculties of education in national universities have created courses based on the criteria of the ECDF matrix in order to improve the classroom practices of teachers who failed the evaluations. However, there are no reports that provide information on their effectiveness and that confirm whether teachers' teaching practices improved.

As in the case of regular mandatory evaluation, research is needed to verify whether the instruments chosen to determine horizontal promotions in teaching careers in the region are effective and to identify their impacts on the quality of education. The evidence on these evaluations is still very scant and the review did not find any study that answered this question. This lack of analysis is probably due to the short time that has elapsed since countries such as Chile, Colombia, Ecuador, Mexico and Peru applied the reforms to the career and introduced the new horizontal promotions.<sup>24</sup>

---

<sup>23</sup> In the case of Jamaica, the legislation is not clear and does not assign a specific weight to the instruments used when selecting teachers.

<sup>24</sup> The only exception is the impact assessment of Santibáñez et al. (2006) related to the Teaching Career Program mentioned above; however, this program was eliminated and replaced in recent reforms by the Professional Teaching Service Law.





## CHAPTER 5

# FIVE KEYS TO A SUCCESSFUL TEACHER EVALUATION

Despite the progress made in recent years in terms of education system's coverage in Latin America and the Caribbean, the region urgently needs to implement policies to improve the quality of its education systems. One of the most important strategies for achieving this objective is to improve teacher quality, which can be achieved through systemic actions that promote: a high quality and relevant pre-service and in-service teacher training; a clear meritocratic regulatory framework that attract and select talented candidates to the teaching profession and offers the opportunities for career progress with competitive salary and benefits; and **an evaluation system that promotes the continuous improvement of the teaching force.**

The countries of Latin America and the Caribbean need to continue implementing and improving their teacher evaluation systems, with the aim of developing valid and reliable systems that use a multiplicity of instruments; with evaluators who understand the different aspects of the teaching processes; and that produce useful results to inform the policymaking process.

This study analyzes what we know about the structure and instruments that evaluations should have, their possible impacts, and the basic characteristics of evaluations in some high performing education systems. It also gives a detailed review of the current situation of teacher evaluations in 19 education systems in the Latin American and Caribbean region. The conclusion reached by this study is that there are five keys to a successful teacher evaluation system.

**First,** the design and implementation of any type of teacher policy, particularly in relation to evaluation, require clarity as to what is expected of a teacher of excellence, and a definition of what the teacher must *know* and *know how to do*. It is necessary to have

**teacher performance standards** that set out in detail the characteristics and skills that a teacher must have in terms of pedagogical skills and content knowledge. These standards are also useful for policymakers who can use them to regulate other aspects of the profession—including pre-service and in-service training—and for teachers themselves, so they know what the expectations are with respect to their knowledge and skills. The analysis finds that only five of the 19 education systems analyzed in Latin America have developed this type of teacher standards in detail through consultation processes with different actors of the education sector. It is important for the other countries of the region to move in this direction and define the ideal teacher required for their own systems and contexts.

**Second**, it is essential to be clear about **what is the objective to be achieved with the implementation of teacher evaluations**. Teacher evaluations can have two general purposes: i) improve the teaching practices, content knowledge and pedagogical skills of teachers, and ii) improve the composition and motivation of the teaching force through positive or negative incentives. Achieving these two objectives with a single tool or type of evaluation is difficult, since teachers will be willing to reveal their true weaknesses or skills depending on what the objective of the evaluation is. An evaluation that has a training objective can allow teachers to act more naturally and honestly during the evaluation, because they know that this can result in support for improving their teaching practice. On the other hand, an evaluation that has labor consequences can lead teachers to focus their teaching practice exclusively on the aspects to be evaluated in order to receive the benefits or avoid the negative consequences.

Therefore, it is not surprising that certain cases—both in successful education systems and in some of the region—implement different types of evaluations, depending on the teacher career stage and the objective they aim to achieve. The analysis suggests that this division of objectives needs to be strengthened in countries of the region where it already exists and introduced in those where evaluations that clearly differentiate objectives have not yet been implemented. Moreover, it would be desirable for regular mandatory teacher evaluations to be designed to detect the main pedagogical and content knowledge weaknesses of teachers, and to overcome them through professional development programs. In addition, separate evaluations should be carried out when results are linked to promotions, bonus pay or dismissals. These could be done with less frequency and with the use of complementary instruments that evaluate additional aspects of teacher performance.

**Third**, irrespective of the type of evaluation that policymakers decide to implement—entry, probationary period, regular mandatory or for promotion or bonus pay—their **design and implementation must produce reliable and valid results**. That is, they must consistently reflect aspects exclusively related to teaching practices and not subject to other factors (for example, they should not depend on who the evaluator is or on which day the evaluation takes place), and they must be highly correlated with student learning and performance.

These characteristics are essential for ensuring that teacher evaluation provides useful and relevant information for teachers and the education system, and for becoming a measure accepted by teachers as a good indicator of their performance. The complexity of the work that teachers do implies that a valid and reliable evaluation needs to use multiple instruments to capture some of the many tasks that teachers perform in the classroom. These tasks include, among others, the development of students' cognitive and socioemotional skills (see, for example, Jackson, 2018 and Kraft, 2019).

Evidence from one of the most comprehensive projects in the area (MET, 2013) suggests that a combination of three instruments—standardized classroom observations, value-added measures based on standardized student tests, and students' surveys—can be effective in detecting teachers who are outstanding in promoting students' cognitive learning in the United States. Use of these instruments, along with their weights in the evaluation, will depend, among other things, on the main objective pursued, on the teacher standards, on the state of teacher policy, on the quality of pre-service and in-service training, and on the budget available to implement the evaluations. In addition, as some studies have shown (Sartain and Steinberg, 2015; Stecher et al., 2018), the quality of the implementation of the instruments is also fundamental and will determine the possibility of achieving the expected objectives of the evaluation systems.

To ensure that teacher evaluation provides valid and reliable results, the selection of instruments used to measure its effectiveness is crucial. In this respect, it is worth highlighting some lessons related to the use and design of four of the most commonly considered instruments for measuring teachers' effectiveness: teacher value-added measures; classroom observation rubrics; teacher portfolios; and content knowledge exams.

With respect to **teacher value-added measures**—based on the results of multiple standardized student evaluations, usually administered at the beginning and end of a school year—the evidence suggests that, although they are adequate for identifying who the more (or less) effective teachers are, they are difficult and costly to implement. For instance, tests must be designed and administered to all students at least once at the end of each school year, and possibly also at the beginning of the year. The level of investment required for their implementation is so high that even in countries with more resources, such as the United States, only between 20%-25% of teachers have enough information to construct this measure. Moreover, these measures do not necessarily offer the information required to improve teacher effectiveness, because they do not detail many aspects of what teachers are doing in the classroom. Although these are instruments that identify teachers' effectiveness in promoting learning among their students, they do not necessarily provide information on what teachers are doing well or badly, or insufficiently in the classroom, which limits the possibility of identifying teachers' strengths and weaknesses and offering actions to improve their performance.

There are other additional instruments which can be very useful for identifying the strengths and weaknesses of teachers, and which can complement value-added measures.

These include **classroom observation tools**. The analysis suggests that strengthening efforts to implement and improve the use of this instrument would be beneficial for the region. The deficiencies in teaching practices in Latin America and the Caribbean are evidenced in the results of teacher evaluations in countries such as Chile, Ecuador and the Dominican Republic, as well as in regional studies that have observed teaching practices in several countries (Bruns and Luque, 2015).

The rubrics for classroom observation evaluate different aspects, from the use of classroom time by the teacher (e.g., Stallings) to the quality of teachers' interactions with their students (e.g. CLASS). But they can also measure different aspects of what the teacher knows and knows how to do. Although classroom observation is the most common instrument in the systems analyzed in this study, its implementation is complex. As mentioned earlier, the difficulty in using these types of instruments lies in generating classroom observations that are sustainable and scalable, due both to the cost of implementation and to the capacity needed to carry them out. Also, several elements of their design can influence their effectiveness: it is necessary to define how many observations are needed, if they must be applied in person or through videos, who will be responsible for carrying them out and coding/rating them, if they should be implemented with advance notice or not, among others.

None of these decisions have a single answer, and the design and implementation of the instrument will depend on its context and the resources available for it. For example, if it is decided to observe classes in-situ, given the low installed capacity in the region for applying this tool and the shortcomings mentioned in the teaching practice, implementation would have to start using simple, scalable and sustainable rubrics (Bruns, De Gregorio and Taut, 2016). These will identify the most basic weaknesses in teacher performance and those that urgently need to be improved and will make possible the use of more complete instruments over time. If, on the contrary, it is decided to do the observation of teaching practices through videos, it would be relatively easier to implement more complete rubrics, which measure additional aspects of the teachers' practices. However, it is essential to define how and by whom the videos would be made, and whether they would be the responsibility of the teachers themselves or of people external to the school.

The use of **teacher portfolios** to encourage teachers to carry out self-reflection processes and allow collection of direct evidence on how they plan, apply and evaluate knowledge has proven valuable both in the Latin American context and in developed countries. This tool is already being used in Chile and Honduras, countries whose experience can be used to learn important lessons in terms of its structure and implementation. It is also an alternative tool for identifying some aspects of teacher performance in systems with a large number of teachers, with difficult geographical characteristics, or with budgetary restrictions, which make it difficult to carry out a census-based evaluation using an observation instrument in the classroom. Like all other evaluation instruments, the way it is implemented is fundamental. For example, rating of portfolios must remain in the hands

of trained personnel, and to avoid subjectivity in the rating, it must follow a clear and systematic criteria. Similarly, there must be a guarantee that there is no possibility of creating “markets” for portfolios preparation by someone other than the teachers themselves, and that they are truly the result of each teacher’s own work and reflection.

Teacher evaluations may also include **content knowledge tests** to determine what level of knowledge teachers have in their specific area. By focusing on teaching practices and pedagogical skills, some of the instruments listed above do not allow for identification of shortcomings in this fundamental aspect of teaching. Evaluating teachers’ content knowledge is relatively less complicated and expensive than evaluating pedagogical skills. But, like with any teacher policy, its design and implementation matter, and it is necessary to develop content knowledge tests that match the learning objectives of each area and grade level.

These content knowledge exams should be included as part of the different types of teacher evaluations. First, as part of the teaching career entry exams to ensure that only individuals with a certain level of knowledge can enter and practice the profession. The loss of prestige of the profession, documented by Elacqua et al. (2018), is explained in part by the decline in the quality of pre-service training, which in turn has implications for teachers’ content knowledge. Therefore, it is also necessary to improve the quality of pre-service training programs and continue to attract candidates with higher performance levels to enroll in teacher training programs and enter the teaching career. A second point in which these tests could be implemented is in promotion exams, as in the case of Chile or Peru, to ensure that professionals at higher levels on the career ladder have the content knowledge that these levels require. This would ensure that teachers with strong content knowledge and, of course, with excellent teaching practices, can act as tutors or mentors to teachers who require help.

Another instrument that, in some cases, has been useful in identifying teacher effectiveness (for example, in the MET project) has been **student surveys** that evaluate teacher practices. However, application of this instrument is very sensitive to the type of the questions, since these can be difficult to understand particularly for younger students. On some occasions, student surveys have been accompanied by parent surveys, however, there is no evidence on their effectiveness in identifying teacher quality.

**The fourth key** to a successful teacher evaluation is that **its results should be used for making decisions about teachers’ professional development and their careers**. It is essential to make effective and appropriate use of the information produced by the evaluations. It will only be possible to improve the teaching quality through changes in teachers’ professional development and in the composition of the teaching force. If the main objective of the evaluation system is to improve the pedagogical skills and content knowledge of teachers, the evaluation must be accompanied by in-service training programs that directly aim to address the weaknesses detected in the evaluations. In the region, only Chile and Peru link

the results of regular mandatory evaluations to these types of programs. Although some outstanding initiatives have been implemented, there is a long way to go in this aspect in the region (Cortés et al., 2011; Schmelkes, 2017). Moreover, the planning and implementation of this type of training programs are complex, and require large investments in time and money, factors that could explain the limited progress the region has made on this front. However, it is only through these types of training programs that it will be possible to improve the skills of the teaching force in Latin America and the Caribbean countries.

Likewise, the use of information to create career incentives is not simple and a system of bonus pay, promotions or dismissals must be carefully designed so as not to lead to perverse incentives. For example, value-added measures can be used to generate changes in the composition of the teaching force by dismissing teachers who are consistently ineffective or are not promoting student learning. The same can apply to the results obtained from classroom observation tools, portfolios or content knowledge exams that suggest that a teacher does not have the pedagogical skills and content knowledge required for teaching. The results can also be linked to promotions, bonus pay or other incentives. However, the evidence suggests that these incentives need to be very carefully designed to ensure that they promote improvements in student learning, and that they do not simply become teaching strategies that prepare students specifically for the exams (*“teach to the test”*) or encourage copying or cheating by teachers to increase their performance scores in order to obtain some benefits.

When defining the use of the evaluation results, the political economy of implementing these measures need to be considered. For example, the use of the results defined by the authorities could be rejected by the teaching force if teachers perceive the evaluation more as a threat than an opportunity for continuous improvement. Acceptance or rejection of the final design of the evaluation and its consequences by the different actors in the education system could either strengthen or undermine the efforts of policymakers (Elacqua et al., 2018; Bruns et al., 2019). Although it is an issue beyond the scope of this study, it is worth noting that one of the richest experiences in terms of political handling of the teacher evaluation comes from Chile. There, the employment consequences of the evaluation were added gradually after long and extensive negotiations with the union, after teachers had understood and accepted the need for an evaluation, and an important part of the evaluation system was already operating (Mizala and Schneider, 2014).

The **fifth key** to implement a successful evaluation system is doing **research and continuously reviewing the instruments, processes, outcomes and impacts** of the evaluation. The decision on the instruments included in an evaluation, as well as their implementation, is not trivial. Implementation of successful evaluations also requires a great effort in time and money by all the actors in the education system. Consequently, research is important to ensure that there is a valid process for developing the instruments (i.e. that these are aligned with performance standards), that evaluations provide valid and reliable information, and

that the system favors improving educational quality. Although the lessons with respect to the different instruments are valuable, they mostly come from education systems very different from the ones in the region, so their use and adaptability to our contexts is very important. Also, as detailed in Chapter 4, research on the instruments, processes and requirements of evaluation systems in countries such as Chile has changed and improved them over time. Therefore, research efforts could have a positive influence on the correct functioning of the evaluation systems and the acceptance by all the stakeholders and could also contribute to complement the scarce literature of evaluation systems in the region.

Latin American and Caribbean countries still have a long way to go to implement systemic teacher evaluations that are reliable, valid, and whose results are used effectively to improve teacher quality. Several countries have taken important steps, with evaluations that include classroom observation of the teaching practices, content knowledge tests, or the use of portfolios. In each case different instruments have been used and have been implemented in a varied ways and with diverse results. **Even so, everyone agrees on something: only a successful teacher evaluation will improve teaching and learning systemically and open the path to improve the quality of education for all the children and young people in our region.**



## REFERENCES

- Aaronson, D., L. Barrow and W. Sander. 2007. Teachers and Students Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1): 95-135.
- Adnot, M., T. Dee, V. Katz and J. Wyckoff. 2017. Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1): 54-76.
- Alvarado, M., F. Duarte, C. Neilson and Mineduc (Ministerio de Educación de Chile). 2012. Efectos preliminares de la Beca Vocación de Profesor. Santiago de Chile: Centro de Estudios, Ministerio de Educación de Chile (Mineduc).
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. Standards for Educational and Psychological Testing. Washington, D.C.: American Psychological Association (APA).
- Aparecido Cordão, F. 2012. Carreira docente: análise jurídica e institucional. Document prepared for the Inter-American Development Bank. Unpublished.
- Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo and N. Schady. 2016. Teacher Quality and Learning Outcomes in Kindergarten. *The Quarterly Journal of Economics*, 131 (3): 1415-53.
- Avalos, B. 2016. "Learning from Research on Beginning Teachers". In: J. Loughran and M. L. Hamilton (eds.), *International Handbook of Teacher Education*. Nueva York: Springer.
- Banerjee, Abhijit Vinayak, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics* 122 (3): 1235-64.
- Banuri, S., P. Keefer and D. de Walque. 2018. Love the Job... or the Patient? Task vs. Mission-Based Motivations in Health Care. Working paper No. 8338. Washington, D.C.: World Bank.
- Barrera-Osorio, F. and D. Raju. 2017. Teacher performance pay: Experimental evidence from Pakistan. *Journal of Public Economics*, Elsevier, Vol. 148(C): 75-91.

- Benton, S. L. and W. E. Cashin. 2012. Student ratings of teaching: A summary of research and literature. IDEA paper no. 50. IDEA.
- Bertoni, E., G. Elacqua, A. Jaimovich, J. Rodríguez and H. Santos. 2018. Teacher Policies, Incentives, and Labor Markets in Chile, Colombia, and Peru: Implications for Equality. Working paper no. 9124. Washington, D.C.: Inter-American Development Bank (IDB).
- Bruns, B. and J. Luque. 2015. Great Teachers: How to Raise Student Learning in Latin America and the Caribbean. Washington, D.C.: World Bank.
- Bruns, B., S. De Gregorio and S. Taut. 2016. Measures of Effective Teaching in Developing Countries. RISE-WP-16/009.
- Bruns, B., I. H. Macdonald and B.R. Schneider. 2019. The politics of quality reforms and the challenges for SDGs in education. *World Development*, Elsevier, Vol. 118(C): 27-38.
- Brutti, Z. and F. Sánchez. 2017. Does Better Teacher Selection Lead to Better Students? Evidence from a Large-Scale Reform in Colombia. Documento CEDE no. 015350. Bogotá: Universidad de los Andes.
- Busso, M., J. Cristiá, D. Hincapié, J. Messina and L. Ripani. 2017. Learning Better: Public Policy for Skills Development. Development in the Americas series. Washington, D.C.: Inter-American Development Bank.
- Campbell, D. 1979. Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1): 67-90.
- Caraballo, D., A. García, M. E. Báez and F. Vargas. 2014. Estado de Situación de las Políticas Docentes en República Dominicana. Informe de seguimiento PREAL.
- Cevallos-Estarellas, P. 2016. Avances en las políticas docentes en Ecuador entre 2007 y 2016. Document prepared for the Inter-American Development Bank. Unpublished.
- Chetty, R., J. N. Friedman and J. E. Rockoff. 2014a. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9), 2593-2632. Available at: <https://doi.org/10.1257/aer.104.9.2593>.
- . 2014b. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633-79. Available at: <https://doi.org/10.1257/aer.104.9.2633>.
- Chingos, M. M. and P. E. Peterson. 2011. It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness. *Economics of Education Review*, 30 (3): 449-65.

- Cohen, J. and D. Goldhaber. 2016. Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher*, 45(6): 378–387.
- Cortés, F., S. Taut, M. V Santelices and M. J. Lagos 2011. Formación continua en profesores and la experiencia de los Planes de Superación Profesional (PSP) en Chile: Fortalezas and debilidades a la luz de la evidencia internacional. Segunda Reunión Anual de la Sociedad Chilena de Políticas Públicas, Santiago, Chile.
- CUREE (Centre for the Use of Research and Evidence in Education). 2018. International Review of Teacher Evaluation Systems: Technical Report and Comparative Analysis. Background study for the Inter-American Development Bank project “Measuring teacher effectiveness in Latin America”, 2017-18. Manuscript. Washington, D.C.: Inter-American Development Bank.
- Dee, T. S. and J. Wyckoff. 2015. Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, John Wiley & Sons, Ltd., Vol. 34(2), pp. 267-297.
- Docentemás. 2017. Portafolio. Uso Formativo Santiago de Chile: Ministerio de Educación. Available at: <http://www.docentemas.cl/pages/portafolio/uso-formativo.php>.
- Donaldson, M. L. and J. P. Papay. 2015. “Teacher evaluation for accountability and development.” In: H. F. Ladd and M. E. Goertz (eds.), *Handbook of Research in Education Finance and Policy*, pp. 174-193. Nueva York: Routledge.
- Elacqua, G., D. Hincapié, E. Vegas and M. Alfonso. 2018. Profesión: Profesor en América Latina. ¿Por qué se perdió el prestigio docente and cómo recuperarlo? Washington, D.C.: Inter-American Development Bank.
- Estrada, 2019. Rules versus Discretion in Public Service: Teacher Hiring in Mexico. *Journal of Labor Economics*, 37(2), 545–579.
- Evans, D. K. and A. Popova. 2016. What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. Oxford University Press. Washington, D.C.: World Bank. Available at: <https://openknowledge.worldbank.org/handle/10986/29308>.
- FEREMA (2015). Honduras el Estado de las Políticas Públicas Docentes. Tegucigalpa: Informes de Seguimiento PREAL.
- Figueroa, M., S. García, D. Maldonado, C. Rodríguez, A.M. Saavedra, G. Vargas. 2018. La profesión docente en Colombia: normatividad, formación, selección and evaluación. Working paper no. 54. Bogota: Escuela de Gobierno, Universidad de los Andes.

- Fryer, R. 2013. Teacher Incentives and Students Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*, 31(2): 373–427.
- Ganimian, A. J. and R. J. Murnane. 2016. Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, 86(3), 719–755. Available at: <http://dx.doi.org/10.3102/0034654315627499>.
- García, S., D. Maldonado, G. Perry, C. Rodríguez and J. E. Saavedra. 2014. *Tras la excelencia docente: Como mejorar la calidad educativa de todos los colombianos*. Buenos Aires: Fundación Compartir.
- Gertler, P. J., S. Martínez, P. Premand, L. B. Rawlings and C. M. J. Vermeersch. 2016. *Impact Evaluation in Practice (segunda edición)*. Washington, D.C.: Inter-American Development Bank and World Bank. Available at: <https://openknowledge.worldbank.org/handle/10986/25030>.
- Goldhaber, D. and E. Anthony. 2007. Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1): 134–150.
- Hanushek, E. A., J. F. Kain, D. M. O'Brien and S. G. Rivkin. 2005. *The Market for Teacher Quality*. Working paper del NBER No. 11154. Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E. A. and S. G. Rivkin. 2010. Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review*, 100 (2): 267–71.
- Harris, D. N. 2011. *Value-Added Measures in Education*. Cambridge, MA: Harvard University Press.
- INEE (Instituto Nacional para la Evaluación de la Educación). 2018. *La educación obligatoria en México, Informe 2018*. Mexico City: INEE.
- INEEd (Instituto Nacional de Evaluación Educativa de Uruguay). 2017. *Informe sobre el estado de la educación en Uruguay 2015–2016*. Montevideo: INEE.
- Ingersoll, R. M. and M. Strong. 2011. The Impact of Induction and Mentoring Programs for Beginning Teachers: A Critical Review of the Research. *Review of Educational Research*, 81 (2): 201–233.
- Inter-American Development Bank (IDB) – CIMA (Centro de Información para la Mejora de los Aprendizajes). 2019. *Attendance rate by age group*. Available at: <https://www.iadb.org/es/cima>.

- Jackson, C. K. 2018. What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*, 126, no. 5: 2072-2107.
- Jensen, B., A. Sandoval-Hernández, S. Knoll and E. González. 2012. The Experience of New Teachers: Results from TALIS 2008. Paris: OECD.
- Kane, T., K. Kerr and R. Pianta. 2014. Designing teacher evaluation systems: New guidance from the measures of effective teaching project. Hoboken, NJ: John Wiley & Sons.
- Kane, T. J., D. F. McCaffrey, T. Miller, T. and D. O. Staiger. 2013. Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Research paper. Seattle: MET Project, Bill & Melinda Gates Foundation. Available at: <https://eric.ed.gov/?id=ED540959>.
- Koretz, D. 2015. Adapting Educational Measurement to the Demands of Test-Based Accountability. *Measurement: Interdisciplinary Research and Perspectives*, 13(1): 1-25.
- Kraft, M. A. 2019. Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, 54(1): 1-36.
- Kraft, M. A. and A. F. Gilmour. 2017. Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness. *Educational Researcher*, 46(5): 234-249.
- Kremer, M., C. Brannen and R. Glennerster. 2013. The challenge of education and learning in the developing world. *Science*, 340(6130): 297-300.
- Lachlan-Haché, L., L. Matlach, A. Guiden and M. Castro. 2015. What We Know About SLOs: An Annotated Bibliography of Research on and Evaluations of Student Learning Objectives. Washington, D.C.: American Institutes for Research.
- Lai, F., L. Zhang, Q. Qinghe, X. Hu, Y. Shi, M. Boswell and S. Rozelle. 2012. Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China. Working paper REAP no. 237, Rural Education Action Program, Freeman Spogli Institute. Stanford, CA: Stanford University.
- Lockwood, J., D. F. McCaffrey, L. S. Hamilton, B. Stecher, V. N. Le and J. F. Martínez. 2007. The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44(1): 47-67.
- Luschei, T. F., A. Chudgar and W. J. Rew. 2013. Exploring Differences in the Distribution of Teacher Qualifications across Mexico and South Korea: Evidence from the Teaching and Learning International Survey. *Teachers College Record*, v115 n5 2013.

- Marcelo, C. and D. Vaillant. 2017. Políticas and Programas de Inducción en la Docencia en Latinoamérica. *Cuadernos de Pesquisa*, Vol. 74, pp. 1224-29.
- McEwan, P. J. 2015. Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, 85(3): 353-394. Available at: <https://doi.org/10.3102/0034654314553127>.
- MEC (Ministerio de Educación and Cultura de Paraguay). 2009. Reglamento del Concurso Público de Oposición para la Selección de Educadores. Approved by Resolution N° 9135 on October 12<sup>th</sup>, 2009. Asunción: MEC.
- \_\_\_\_\_. 2016. Resolution N° 26704 from December 1st, 2016. Asunción: MEC.
- MET. 2013. Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study. Policy and Practice Brief. Seattle: MET Project, Bill & Melinda Gates Foundation.
- Mizala, A. and B. Ross Schneider. 2014. Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990-2010). *Governance: An International Journal of Policy, Administration, and Institutions*, Vol. 27(1): 87-109.
- \_\_\_\_\_. 2019. Promoting quality education in Chile: the politics of reforming teacher careers. *Journal of Education Policy*, DOI: 10.1080/02680939.2019.1585577.
- Molina, E. et al. 2018. Measuring Teaching Practices at Scale: Results from the Development and Validation of the Teach Classroom Observation Tool. Policy research working paper no. 8653. Washington, D.C.: World Bank.
- Muralidharan, K., A. Singh and A. Ganimian. 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4): 1426-1460. Available at <https://doi.org/10.1257/aer.20171112>.
- Murnane, R. J. and A. J. Ganimian. 2014. Improving educational outcomes in developing countries: Lessons from rigorous evaluations. NBER Working paper no. 20284. Cambridge, MA: National Bureau of Economic Research.
- Näslund, E., H. Alonzo and D. Martin. 2013. Challenges and Opportunities in the Belize Education Sector. Technical note no. 538. Washington, D.C.: Inter-American Development Bank.
- Nogueira, D. X. P., G. Ribeiro de Jesus and S. Pereira da Silva Cruz. 2013. Avaliação de desempenho docente no Brasil: desvelando concepções e tendências. *Linhas Críticas*, Brasília, DF, v. 19, n. 38, pp. 13-32.

- OECD (Organisation for Economic Co-operation and Development). 2009. *Teacher Evaluation: A Conceptual Framework and Examples of Country Practices*. Paris: OECD.
- . 2013. *Teachers for the 21st Century: Using Evaluation to Improve Teaching*. Paris: OEDC.
- . 2015. *Who Wants to Become a Teacher? PISA in Focus 58*. Paris: OECD.
- . 2016. *Reviews of School Resources: Uruguay 2016*. OECD Reviews of School Resources. Paris: OECD.
- . 2017. *Education in Costa Rica, Reviews of National Policies for Education*. Paris: OECD.
- Ome, A. 2012. *The Effects of Meritocracy for Teachers in Colombia*. Informe de investigación no. 010260. Bogota: Fedesarrollo.
- Papay, J. P. 2011. Different Tests, Different Answers the Stability of Teacher Value-Added Estimates across Outcome Measures. *American Educational Research Journal*, 48: 163-193.
- Papay, J., Taylor, E. S., Tyler, J. and M. Laski. 2016. *Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data*. NBER Working paper no. 21986. Cambridge, MA: National Bureau of Economic Research.
- Polikoff, M. S. 2015. The Stability of Observational and Student Survey Measures of Teaching Effectiveness. *American Journal of Education*, 121(2): 183-212.
- Rivkin, S. G., E. A. Hanushek and J. F. Kain. 2005. Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2): 417-458. Available at: <https://doi.org/10.1111/j.1468-0262.2005.00584.x>.
- Rockoff, J. E. 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, 94(2): 247-252.
- Rockoff, J. E. and C. Speroni. 2010. Subjective and Objective Evaluations of Teacher Effectiveness. *The American Economic Review: papers and proceedings*, Vol. 100, No. 2.
- Rockoff, J. E., D. O. Staiger, T. J. Kane and E. S. Taylor. 2012. Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools. *American Economic Review*, 102(7): 3184-3213.
- Rodríguez, C. 2015. *El manejo docente en la República Dominicana: análisis y recomendaciones de política a partir de un estudio comparativo con países exitosos y algunos de América Latina*. Document prepared for the Inter-American Development Bank. Unpublished.

- Rodríguez, C. and M. C. Rivera. 2018. Sistemas de evaluación docente en diecinueve países de América Latina and el Caribe. Background study for the Inter-American Development Bank project “Measuring teacher effectiveness in Latin America”, 2017-18. Manuscript. Washington, D.C.: Inter-American Development Bank.
- Santibáñez, L., J. F. Martínez, A. Datar, P. J. McEwan, C. Messan-Setodji and R. Basurto-Dávila. 2006. Haciendo camino: Análisis del sistema de evaluación and del impacto del programa de estímulos docentes Carrera Magisterial en México. Mexico City: Secretaría de Educación Pública.
- Santibáñez, L., D. Rubio and M. Vázquez. 2017. Formación Continua de Docentes: Política Actual en México y Buenas Prácticas Nacionales e Internacionales. Mexico: INEE/IDB.
- Sartain, L. and M. P. Steinberg. 2016. Teachers’ Labor Market Responses to Performance Evaluation Reform: Experimental Evidence from Chicago Public Schools. *Journal of Human Resources*, Vol. 51(3): 615-655.
- Schleicher, A. 2016. Teaching excellence through professional learning and policy reform. Lessons from Around the World, International Summit on the Teaching Profession.
- Schmelkes, S. 2017. “La evaluación docente en México”. Presentation at Instituto Nacional Para la Evaluación de la Educación, Mexico.
- Smith, W. C. and K. Kubacka. 2017. The emphasis of student test scores in teacher appraisal systems. *Education Policy Analysis Archives*, 25.
- Springer, M. G. (ed.). 2009. Performance Incentives. Washington, D.C.: Brookings Institution Press.
- Stallings, J. and S. Knight. 2003. Using the Stallings Observation System to Investigate Time on Task in Four Countries. Background study for the World Bank project International Time on Task (ITOT). Washington, D.C.: World Bank.
- Stecher, B. 2002. Consequences of large-scale, high-stakes testing on school and classroom practice. In: L. S. Hamilton, B. M. Stecher and S. P. Klein (eds.), Making sense of test-based accountability in education, pp. 79-100. Santa Mónica, CA: RAND.
- Stecher, B. M., D. J. Holtzman, M. S. Garet, L. S. Hamilton, J. Engberg, E. D. Steiner, A. Robyn, M. D. Baird, I. A. Gutiérrez, E. D. Peet, I. Brodziak de los Reyes, K. Fronberg, G. Weinberger, G. P. Hunter and J. Chambers. 2018. Improving Teaching Effectiveness: Final Report: The Intensive Partnerships for Effective Teaching Through 2015-2016. Santa Monica, CA: RAND Corporation.

- Steele, J., M. Pepper, M. Springer and J. R. Lockwood. 2015. The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, 48(2015): 86–101.
- Steinberg, M. P. and M. L. Donaldson. 2016. The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era. *Education Finance and Policy*, 11(3): 340–359. Available at: [https://doi.org/10.1162/EDFP\\_a\\_00186](https://doi.org/10.1162/EDFP_a_00186).
- Steinberg, M. P. and L. Sartain. 2015. Does Teacher Evaluation Improve School Performance? Experimental Evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, MIT Press, Vol. 10(4), pp. 535–572.
- Sun, Yulan F. 2017. “La evaluación Docente en Chile: Trayectorias and Aprendizajes”. Presentation at Centro de Medición MIDE.
- Taut, S., E. Valencia, D. Palacios, M. V. Santelices, D. Jiménez and J. Manzi. 2016. Teacher performance and student learning: linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*, 23: 53–74.
- Taylor, E. S. and J. H. Tyler. 2012. The Effect of Evaluation on Teacher Performance. *American Economic Review*, 102 (7): 3628–51.
- UNESCO (United Nations Educational, Scientific and Cultural Organization). 2015. Las carreras docentes en América Latina. La acción meritocrática para el desarrollo profesional. Santiago de Chile: UNESCO.
- Weisberg, D., S. Sexton, J. Mulhern, D. Keeling, J. Schunck, A. Palcisco and K. Morgan. 2009. The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness. Nueva York: The New Teacher Project (TNTP).
- Wyckoff, J. and V. Katz. 2018. Review of Teacher Evaluation Systems in the United States: Technical Report and Comparative Analysis. Background study for the Inter-American Development Bank project “Measuring teacher effectiveness in Latin America”, 2017–18. Manuscript. Washington, D.C.: Inter-American Development Bank.
- World Bank. 2017. World Development Indicators. Washington, D.C.: World Bank.
- . 2018. World Development Report 2018: Learning to Realize Education's Promise. Washington, D.C.: World Bank.





Teacher evaluation is a key instrument to support the professional development of teachers and improve teacher effectiveness. This study analyzes teacher evaluation systems in Latin America and the Caribbean, and several high-performing education systems in other regions and presents practical guidance on how to develop a successful teacher evaluation system.

There are five keys to a successful teacher evaluation system. First, clarify what is expected of a teacher of excellence. Second, define what specific objective the teacher evaluation is pursuing. Third, use multiple instruments that provide reliable and valid information. Fourth, use the evaluation results efficiently and accordingly to the objectives. Fifth, conduct research into the evaluation processes and results to improve its design and implementation.

Latin America and the Caribbean countries still have a long way to go to implement teacher evaluation systems whose results can be used to improve teacher effectiveness. This is a necessary path to ensure that all children and youth in the region have access to effective teachers and a high-quality education.

**The Inter-American Development Bank (IDB)** is an international institution created in 1959 to foster economic and social development in Latin America and the Caribbean.

