Specific Policy Recommendations on the Development of a Comprehensive In-Service Teacher Evaluation Framework

Carlos Mancera and Sylvia Schmelkes

June 2010

TABLE OF CONTENTS

I. INTRODUCTION: WHY TEACHER EVALUATION?

Mexico is a country that, despite phenomenal progress in the last 60 years in increasing education enrolment from three million to over 30 million students, still has only 60 percent of young people enrolled in upper-secondary education. But if Mexico aspires to develop a 21st century economy, it will need not only to boost its secondary school attainment rate but to ensure that all young people leave secondary schools with the knowledge and skills to become lifelong learners.

Although there are many factors that contribute to a country's education performance, we now know from a growing international body of research that building a highly skilled professional educator workforce is central to a country's ability to improve the outcomes of schooling for its young people (Sclafani and Manzi, 2010; OECD, 2005). We also know that having robust and accurate data and measurement systems that allow not only to track the progress of students and schools but to intervene in a timely way with appropriate support is a key for creating a culture of continuous improvement and of accountability for performance (McKinsey and Company, 2007).

It is for this reason that we propose the development of a comprehensive in-service teacher evaluation system for Mexico. This recommendation needs to be seen in the context of the broader set of recommendations offered by the OECD on teacher professional development and school leadership given that focusing on strengthening teacher evaluation alone will have limited results in raising student performance. The converse, however, is equally true: Mexico can improve its teacher preparation programs, its processes for assigning teachers to schools, its approach to professional development, its leadership in schools, but if these changes are not accompanied by a thoughtful and carefully implemented teacher evaluation system, these other changes are unlikely to have much impact on student performance.

We understand that according to international experience and the nature of the task itself, the road towards the development of an in-service teacher evaluation is very challenging. Even though it is a difficult task, this endeavour is critical to enhance education quality in Mexico. Early commitment by major stakeholders will be needed to ensure that the teacher evaluation system is built with their support. Federal and state educational authorities, the national teachers union (*SNTE*), academics, individual prestigious teachers, and leaders of civil society should all help carry out this important project that has the ultimate goal of improving the quality of education.

In this paper, taking into account the international experience and evidence on teacher evaluation, we will outline the major options, and recommend a broad path and implementation strategy for Mexico to follow.

Among the larger issues that countries debate as they develop and revise their teacher evaluation policies, which we will discuss below, are the following:

- What should be the different components of a fair teacher evaluation system;

- How to balance the formative and summative purposes of evaluation;

- How to engage teachers in the design and implementation of teacher evaluation systems;

- How to develop, implement and evaluate reliable standards of teaching practice that can form the basis for such evaluations;

- How to use student assessment results in evaluating teachers;

- What kind of stakes or consequences to attach to the results of teacher evaluations.

II. MAIN COMPONENTS OF A TEACHER EVALUATION SYSTEM

1. Why to evaluate

The purpose of all education systems, beyond assuring access for all students to basic education, should be to improve student learning. The quality of student learning depends on various factors, but most importantly on the quality of teaching. Although seeming obvious, it is important to keep in mind that what does not happen in the classroom does not actually happen in education. As the McKinsey report points out (2007: 12), "the available evidence suggests that the main driver of the variation in student learning at school is the quality of the teachers". This report even states that "the quality of an education system cannot exceed the quality of its teachers" (McKinsey and Company, 2007:16).

The results of educational research are also conclusive when locating the centrality of schools in the processes for the improvement of the educational system. The school, and not just the individual classroom, must be the focus of improvement in order to achieve systemic reform. The school is the unit of accountability; its quality is essential for the quality of education. Schools in which the principal focuses intensively on the improvement of teaching, based upon careful diagnosis and the development of a school-wide improvement plan, are much more likely to improve teaching practice and student learning that those without such leadership. Studies on effective schools (Sammons, Hillman and Mortimore, 1995; Posner, 2004), summarized in the work of OECD Steering Group on School Management and Teacher Policy in Mexico (2010), present strong arguments to support the importance of the school. They also point out the common factors that appear in all good schools. These studies complement other studies carried out in Latin America (Fernández, 2005; UNESCO-LLECE, 1998 and 2008) on school factors that influence school performance.

The main purpose of teacher evaluation is to improve teaching and thus the educational student results. Therefore, teacher evaluation should always have formative purposes, that is, should supply proper feedback to teachers on their needs for improvement. In addition, it should identify support and teacher training mechanisms that are needed to ensure improvement.

Teacher evaluation can also be summative in nature, that is, lead to positive and/or negative consequences for the teacher, or both. It is important for good teachers, and particularly for teachers that show improvement in their practices and results, to be recognized by receiving monetary and/or non-monetary incentives as a consequence. It is also important for an educational system to have solid bases for identifying teachers that show no interest or little capacity of improving their teaching practices and results in spite of support and training received as a consequence of formative evaluation, and to make the necessary decisions.

2. What to evaluate

It is especially important to evaluate teaching practices. Good educational results for all students are the ultimate goal of teaching, and certainly student learning results have to be taken into account. But simply knowing the results, or even their changes, will not by itself transform educational practices. So in addition to identifying student learning results and their change over time, mechanisms to foster improvement in teaching practices also need to be in place. If teacher practices are not transformed, it will

be difficult to achieve substantial improvements in student learning outcomes. In order to improve teaching practices, these have to be evaluated and a system must be put in place to support teachers for the improvement of student outcomes. To do this effectively, it is necessary to first address what to evaluate, i.e., to describe the attributes of good teaching performance and to define an evaluation framework.

3. Standards of teaching practice

Based on international experiences and pertinent literature, a common starting point for teacher evaluation is to accurately describe the elements of good teaching performance. These qualities or attributes, expressed as expectations of how teachers are expected to perform and teach, are commonly described in standards of teaching practices. The standards define the expected teacher performance needed for students to learn. Standards have two functions: they serve as a framework to guide teachers towards better performance, and they enable the assessment and evaluation of individual teacher performance.

4. Teaching standards should at least have the following characteristics:

Be understandable and aligned with the tasks involved in teaching, the functioning of schools, and student learning standards. Standards should reflect what teachers identify as good teacher performance.

1. **Cover all of the teaching domains defined**. Every one of these domains must be subdivided into components and indicators.

2. **Establish different levels of competency for each specific aspect that defines the domains of teacher and school work**. Standards should be used as a framework to guide teachers in the right path. The association of a teacher with a certain standard indicates the level of achievement regarding that standard. When standards are too high they become unattainable and lose their purpose for teacher guidance. Conversely, if standards are set too low, they no longer present a challenge for teachers, and therefore lose their value. Ideally, all standards should have different levels for every domain of teacher performance: the highest level would represent what expert teachers do; the lowest level would reflect the work of a beginner.

3. **Reflect a nuclear group of performances that should be observable in all teachers and all schools, irrespective of their conditions and circumstances**. If every group of teachers or every kind of school had a specific group of standards, then this would be contrary to the very nature of standards and the existence of attributes of teaching that pertain to this profession. This, however, does not mean that standards have the intention or the effect of homogenizing teaching. What it does imply is that standards must refer to the more abstract nature of the domain expected of all teachers. The standard itself not only accepts but invites its diversified application to different contexts and situations.

4. **Define and operationalize intended goals and outcomes of good teaching, while not prescribing specific practices for teachers to attain these goals**. Indeed, standards should allow for creativity and individuality in the classroom. For example: there is no unique way to motivate students to participate in the classroom and learn. This depends on the characteristics of students, on the subject that is being taught, on the time of day when the lesson takes place, on the closeness to vacation periods, and on many other factors. This is why teachers must know how to use their knowledge and skills on a day to day basis. The object, however, is constant: the motivation and participation of students. This is what a standard must state and scale. Also, standards have to be worded in such a way that they include and take into account very diverse contexts. In Mexico teachers have strongly differentiated functions due to important diversity among regions and schools. For example, a teacher in a one-teacher school does not carry out the

same tasks as a teacher in an urban school, nor does a teacher in an indigenous community do exactly the same things as does a teacher in a rural non-indigenous school. Standards relative to teacher performance for innovating and adapting to specific contextual and cultural situations in order to better engage students in learning activities and thus produce higher learning outcomes should be included. Standards relative to teacher performance for attending to special needs of diverse students should also be part of the set of standards that are being defined.

5. **Be dynamic**. Periodic revisions with the participation of teachers should be carried out in order to ensure that standards are properly scaled, and that all aspects of good teaching are being considered, as the teacher evaluation system is fine-tuned and broader educational changes take place.

Teacher participation in the process of defining the framework and the standards of good teaching is critical to ensure buy-in to the evaluation system. They also should be trained on the framework and on how it translates into teaching practices in the classroom. It is essential that teachers receive sufficient feedback based on rubrics connected to the standards-based framework. Also, the support systems for teachers that are capable of reaching individual schools and teachers have to be in place. Teachers know that only judging their strengths and weaknesses is not enough to change behaviours; feedback and support are also needed.

III. EVALUATION FRAMEWORK

The question is what distinguishes a good teacher and good teaching. There are many frameworks that can be used internationally to evaluate teacher practice (see, for example, Perrenoud, 2004; Rewards and Incentives Group, 2009; Ontario Ministry of Education, 2009; Santiago and Benavides, 2009; Khim Ong. 2008; Singapore Ministry of Education, 2006). Typically, they propose that at least these four main areas be covered in professional performance, as outlined in Danielson (2000)[i]:

- **Planning and preparation.** This domain has several components that describe how a teacher organizes the content that students are to learn, that is, how a teacher designs instruction. Components include a deep understanding of content and pedagogy and an understanding and appreciation of the students and what they bring with them. The content must be transformed through instructional design into sequences of activities and exercises accessible to the students. This area also covers assessment plans. Assessment techniques must reflect instructional outcomes and document student progress. Assessment must be used for formative purposes and provide diagnostic opportunities for students to demonstrate their level of understanding. This domain is based on the principle that a teacher's role is not so much to teach as it is to arrange for learning. The plans and the students' assignments may be included in a teacher's professional portfolio. The plan's effects must be observed through action in the classroom and are reflected in student learning outcomes.

- **The classroom environment.** These are aspects that set the stage for all learning. Components have to do with the way a teacher establishes a comfortable and respectful classroom environment that cultivates a culture for learning and creates a safe place for risk taking. This must lead to student behaviour that is cooperative and non-disruptive, and to a physical environment that is supportive to these purposes. Caring, high expectations, commitment of teachers to students, are included in the components. Master teachers in this domain are able to create an atmosphere of excitement about the importance of learning. The skills are demonstrated through classroom interaction and through interviews with or surveys of students.

- **Instruction**. The components of this domain are at the heart of teaching and they refer to the actual engagement of students in content. What matters is to get children to develop a complex understanding and to participate in a community of learners. It refers to the implementation of the plans designed in the first domain. Good teachers in this domain have finely honed instructional skills. Their work in the classroom is fluid and flexible. Their questions probe student thinking and serve to extend understanding. They are attentive to different students in the class. Skills in this domain are demonstrated through classroom interaction, observed either in person or on videotape, as well as through student learning outcomes.

- **Professional responsibilities.** These include roles assumed outside and in addition to those in the classroom. They involve professional engagement with the school, families and the communities, as well as their work for the school as a whole. They also include professional growth. Teachers that excel in this domain contribute to the general well-being of their institutions. The skills in this domain are demonstrated through teacher interaction with colleagues, families, other professionals and the larger communities.

The quality of teacher practice should always be evaluated with reference to equity, that is, to the ability of the teacher to plan, create an environment, differentiate instruction based on individual student needs, and strive for positive learning outcomes for all students. All classes are heterogeneous: students have different talents, intelligences, and rates of learning. In Mexico many classes have students from different cultures with different native languages. Teachers should strive to achieve desired learning results with all of them, and this requires routinely giving special attention to each student or subgroup of students. Each of the domains defined should address the issue of equity, and provide a definition of good teaching practice regarding equity, with progressive levels of achievement for each domain.

In the case of Mexico, where evidence suggests that the time allotted for teaching is not used optimally by teachers, criteria for judging teacher performance should include attendance and punctuality, as well as time-on-task. These criteria, like others, can be modified as the system progresses and average levels of teacher performance improve substantially.

1. How to evaluate

The main strategy for evaluating teachers is to visit them periodically in their work places. These visits should be carried out by carefully selected and well-trained evaluators, at least two per school. During these visits, all teachers that work in the school are evaluated. The evaluation draws and triangulates information from the following instruments:

- Classroom observation. The evaluator might be present in the classroom, or might review a class that has been videotaped. In Delaware, for example, two visits are made in the case of novice teachers: one visit is planned, the other unannounced. Classroom observations are carried out in both visits (Delaware Department of Education, 2008). This allows evaluators the opportunity to assess different aspects of teachers' performance.

- Teacher portfolios. These may include evidence on how classes are prepared, on critical reflection on the work carried out, on student deliverables, on classroom discussion. Teacher portfolios, by their very definition, can be easily adapted to different contexts.

- Teacher self-evaluation. This includes a critical reflection on the part of the teacher of his/her practice in relation to each of the standards in the framework. When a previous evaluation has been carried out, this self-evaluation should be based on a rubric agreed upon between evaluators and the teacher for the improvement of his/her teaching practice.

- Evidence of student learning. Such as samples of student work and other evidence such as student performance data. Both formative and summative evaluations carried out by the teacher should be included for revision. Information on student results on standardized tests, such as ENLACE, can be used as an input for both the evaluator and the teacher. Unsatisfactory results in ENLACE should perhaps imply more frequent evaluation visits. They could also be used to question favourable results in the four domains suggested in the evaluation framework, and lead to delving more deeply into the other instruments. In the case of the teacher, ENLACE results might be used as a baseline for goal setting (as in Delaware Department of Education, 2008), and as evidence of improvement when the information is available (in the case of teachers that remain for more than one year with the same group).

- Objective setting and/or teacher interviews. A face-to-face dialogue makes it easier for teachers to exchange information and explain why they chose specific strategies which may not be apparent through observation. Interviews may also serve the purpose of giving useful feedback to teachers.

- Interviews with the principal and the supervisor, and reports from the two and other educational authorities.

- Teacher knowledge tests, when available.

- Student and parent information. These can be obtained from surveys, focus groups or other mechanisms.

The deployment and use of these instruments has difficulties that are common to processes based on qualitative data, and therefore they must be carefully designed, well calibrated, and tested for reliability among evaluators. The fairness of the system rests on the fact that evaluation results do not depend on a single evaluation instrument, but on a wide array of tools that allow for the cross-referencing of information and incorporate different sources of evidence about teacher practice.

2. Who will evaluate

The availability of sufficient numbers of trained and competent evaluators is a key requirement for the successful evaluation of teachers based on teaching standards. Unless a considerable investment is made on the training of evaluators, a formal teacher evaluation may fail as the legitimacy of the evaluators is not recognised by the teachers. Evaluators should have, at minimum:

- **Knowledge of the work teachers carry out.** It is not uncommon that well-known and experienced teachers become evaluators beyond the boundaries of their own districts or states.

- **Training to make the expected observations** in accordance with established teacher standards and procedures for evaluations.

- **Autonomy in relation to the evaluated teacher.** This is necessary in order to avoid conflict of interest. When an evaluator has a personal relationship with the teacher he/she is evaluating, it is possible to foresee a conflict of interest that could damage the objectivity the evaluation needs.

3. Framework, standards, teacher evaluation, and equity

A scheme such as the one proposed has the great advantage of allowing for improvement in educational equity – as measured both by inputs and by results – for several reasons: a) it sets common performance measures while at the same time recognizing a heterogeneous reality (different standard levels); b) it identifies differences in quality that are not explained by contextual diversity, but by differences in capacities or training, and c) it addresses some of the problems that lead to these differences by investing in teacher training, which literature shows is the most important internal factor explaining student achievement.

Education in Mexico is extremely unequal. According to *the Instituto Nacional para la Evaluación de la Educación (INEE)* (2008:191), for example, the difference in student achievement between indigenous and private schools is almost two standard deviations. *INEE* (2008) also documents severe difference in the number and quality of inputs: schools in poorer areas receive fewer inputs from the system than those in more disadvantaged regions. Inequality is undoubtedly one of the most important problems of the Mexican educational system. The fact that a teacher evaluation system can help reduce this inequality should not be disregarded.

4. The centrality of schools

Teaching is almost never a solitary activity. It takes place within a school, with a formal or acting principal, with a group of teachers who work collectively. The school environment should be favourable to good teaching and to teacher development. The school, through its annual planning, should establish clear goals related to quality teaching and learning that involve all teachers. Frequent meetings for reviewing progress towards those goals and for defining areas of needed support to those teachers having difficulties meeting them are basic for fostering critical reflection on both individual and collective professional performance. Relationships with parents and community depend on the school as a whole, and research also tells us that good relationships make for better quality in both teaching and learning.

The centrality of schools has been extensively dealt with in the OECD Steering Group on Teacher Policy and School Management in Mexico report (2010), Schools at the Centre, Teachers at the Heart. While this document on teacher evaluation does not deal with schools as such, it is important to recognize the importance of their evaluation. Good schools build good teachers.

We are proposing the school as the unit of the visit as the main strategy for teacher evaluation. All teachers in a certain school are evaluated in the same visit. Therefore, it seems adequate to include the evaluation of the principal in the context of the school-based evaluation framework, and to train evaluators in both teacher and school evaluation[ii].

The OECD document we have referred to (2010) proposes the establishment of a set of standards for principals. These standards are the basis for the development of a set of instruments for carrying out this evaluation, which should also involve observation, portfolios, in-depth interview with the principal, parent and student surveys, interviews with the supervisor and other education authorities, interviews and/or focus groups with teachers, and a self evaluation of the school itself. As in the case of teachers, results of the school in standardized tests should be considered as a relevant input and might perhaps indicate which schools should be first in the evaluation cycle. Also as with the teachers, failing schools should be evaluated more frequently than those that are working satisfactorily. The evaluation cycle can have the same length as in the case of teachers.

School and principal evaluation should also be formative during at least the first cycle. After that, summative evaluations with consequences for principals and schools as a whole can form part of the evaluation system.

IV. ELEMENTS OF A TEACHER EVALUATION SYSTEM FOR MEXICO

Based on international evidence and on the analysis of the current education context in Mexico, this section proposes the main specific features of an in-service teacher evaluation system for Mexico.

Goal and objectives

There are few available teacher evaluation instruments in Mexico. Evaluation has usually been the responsibility of school principals, and to a lesser extent, of supervisors, or other educational authorities. Other efforts, like the development of standards, are incipient.

The construction of a teacher evaluation framework, as proposed in the previous section, demands a long and systematic effort to be undertaken together with the teachers, with the principals and other school management staff, the support of educational researchers, and the participation of society. Isolated initiatives already underway should be studied and built upon. For example, the *Subsecretaría de Educación Básica* commissioned a large study to develop teaching and school standards, as well as another study on standards of students' outcomes. The findings of these studies should be reviewed and taken into account as Mexico moves forward to construct a teacher evaluation framework.

As suggested in the previous section, a good evaluation system must address four questions: **why evaluate** (goals and purposes); **what to evaluate** (standards of practice); **how to evaluate** (methods); and **who should evaluate.** In our view, the principal purpose of any evaluation system (the "why") must be to improve teaching practice to enhance student learning.

In this sense, the overriding goal is formative: to enable the identification of weaknesses in teaching practice so as to be able to intervene in a timely fashion to remedy them. Such systems must ensure — mainly through school-based professional development, mentoring and courses-- that these formative processes really result in the transformation of teacher practice, and ultimately in the improvement of student learning. Monitoring this transformation is part of what the system must ensure.

Evaluation is essential to assess the quality of teacher performance and can also be a powerful instrument to underpin teacher professional development. When there are clear references as to what to expect of a good teacher, this encourages teachers to reflect on their own practice, and to continue their own development as professionals. The TALIS report (OECD, 2009) shows that more than 80% of teachers in participating countries consider evaluation and feedback very important to improve their teaching and also to innovate in the field.

A formative evaluation system, designed and oriented mainly to improve the quality of instruction, can also be a vehicle for recognising good teachers. Recognition programs can make teaching a more attractive profession, provide it with a higher social status, and increase the likelihood of retaining good teachers. In the TALIS survey (2009), most teachers report that they did not receive any kind of feedback as a result of evaluation and consequently viewed evaluation more as an administrative requirement than a way to improve their performance. A good evaluation system should serve to identify teachers whose practice needs strengthening and to identify and reward teachers whose practice is outstanding. Equally distributing benefits to all teachers, without considering the quality of performance, is not fair to good teachers, and it is neither effective nor efficient in terms of improving student learning.

Summative evaluation procedures, however, tend to be resisted by teachers in general due to significant consequences on their careers or livelihoods. Therefore, we believe that the formative goals and procedures of the evaluation system should be well established and valued by the teaching profession before introducing higher stakes summative evaluation.

1. Standards of practice and evaluation framework

Teacher performance standards should be the product of an exercise that reflects what Mexican teachers consider to be good teaching practices. The standards will be accepted and considered as valid guidance only if they emerge from the vision and practice of Mexican teachers and schools. The international references or the academic contributions are useful only if they are analyzed, and, as needed, modified and accepted by Mexican teachers. Nevertheless, Mexico has the opportunity to take advantage of the best international work in this realm. Sclafani and Manzi (2010) recommend using Danielson's "Framework for Teaching" (2000) and Chile's "*Marco para la Buena Enseñanza*" (Manzi, 2009; Avalos and Assael, 2006) as good starting points of the process.

To develop teacher performance standards, the process should take into account the desired characteristics of those standards as described in the previous section.

In the absence of a framework for good teacher performance, little has been done in Mexico to train evaluators. Teachers in Mexico are typically evaluated by their principals, and seldom by a person not within the school. Schools are not evaluated at all.

Potential evaluators, to be selected and trained, in the case of Mexico are to be found among the *Asesores Técnico-Pedagógicos (ATPs)*, the personnel of the Teacher Centres (*Centros de Maestros),* and the professors in the normal schools and in the National Pedagogical University. The first two of these potential groups of evaluators are trained teachers; they are well distributed geographically across the country, and are not currently employed in teaching positions. Although no accurate date is available yet, it can be stated that there are approximately 50,000 *ATPs*, while the system we are proposing would require around 2,500 full-time evaluators if the school evaluation cycle is every four years; it would take approximately 10,000 evaluators to have every school visited once a year by a team of two evaluators. Selection can be rigorous, and training, if carried out at the state level, can have the needed quality. The advantage of resorting to *ATPs* and other figures proposed lies in the fact that they are already hired by the educational system and therefore have no cost implications. On the opposite, they would find a clear and favourable framework to better conduct the kind of work they already have to do. The labor status of these evaluators would have to change, especially in the case of the *ATPs* and personnel in charge of the *Centros de Maestros*, so it properly reflects the nature of their new duties. The role of supervisors would also become more academic and less administrative driven as they interact with the evaluators and discuss their recommendations together with schools and teachers.

For all cases, however, there has to be a strict selection process, guided exclusively by technical criteria, and a substantial training program. There should be clear mechanisms in place to address situations in which there are potential conflicts of interest between evaluators and teachers, or where other conditions needed for the evaluation process are not met.

For teachers to accept a new evaluation system, they must receive training on the framework and what it looks like in the classroom. This training must clarify for teachers what knowledge, skills, behaviours, attitudes, and results they are expected to demonstrate to be considered an effective teacher. Once the evaluation system is in place, teachers should receive adequate feedback on their content knowledge and content pedagogy based upon rubrics connected to the standard-based framework. This will enable teachers to determine a personal development plan and identify appropriate professional development

options. School-based training and professional development plans should then be designed and supported to ensure proper implementation. The point is that teachers at all levels of proficiency and at all stages of their careers should have access to support and professional development to improve their practice.

The school leader should also be included in the evaluation. In fact, it has been demonstrated that one of the core leadership responsibilities that has consistently been identified as associated with improved student results is that of supporting, evaluating and developing teacher quality (OECD, 2009). They play a key role in informal evaluation annually and therefore can be key players in the formal process. The evaluation system should be designed with a clear view of the role the principal is expected to play. Principals who have a clear understanding of their role as instructional leaders know that regular evaluation of their teachers provides a key opportunity to discuss and agree on professional goals and assess the progress that teachers are making toward these goals. Preparation and training programs for principals should be added to the instructional improvement agenda in the near future.

External certification of teacher knowledge and skills is very helpful to enhance teaching and produce internal changes in schools. With an external reference, teachers are in a better position to confront their own educational practices, have fruitful discussions, and analyse ways to improve. Teachers who have higher awareness of their own standing are likely to be more effective in developing higher performance competencies.

While the framework for defining good teacher performance is under construction, we recommend that it be tried out within different teacher groups on a small scale. These pilot programs would bring valuable information to enable timely adjustments without significant cost. Above all, this would prevent difficulties in correcting the model at a more consolidated stage.

2. School visit as a key strategy for teacher evaluation

As stated in the previous section, teaching takes place within a school, with a formal or acting principal, with a group of teachers that works collectively. We are proposing the school as the unit of the visit as the main strategy for teacher evaluation. All teachers in a certain school are evaluated in the same visit. Therefore, it seems adequate to include the evaluation of the principal in the context of a school-based framework, and to train evaluators in both teacher and principal evaluation.

The result of a visit is twofold: a) the evaluators write a formative evaluation report based on the framework and the standards. This report clearly states the strong points of the teacher, as well as his/her weaknesses and the areas in need of improvement. The report finalizes with the proposal of an improvement plan that must be discussed with the teacher. The improvement plan suggests measures that the teacher has to take – such as giving more importance to daily planning, incorporating more feedback to students in formative evaluation, and giving more attention to students lagging behind —. The improvement plan also sets clear goals that the teacher must achieve by the next evaluation. b) The evaluators also identify areas where external support is needed in the form of mentor teachers, frequent in-class supervision with feedback, or special courses, for example. These requirements are communicated to the regional supervisor for their adequate implementation. The improvement plan with both components is discussed with the teacher and the principal, and necessary adjustments are made. This improvement plan serves as the basis for the next evaluation cycle.

After two evaluation cycles have been carried out, the evaluators will also be asked to write a summative report to be handed in to the educational authorities for the definition of pertinent consequences. This report clearly states the strong points of the teacher, as well as his/her weaknesses and the areas in need of improvement.

Frequent evaluation cycles have the advantage of giving close follow-up to the improvement plans and the work done towards the achievement of goals stated in those plans. The cost of the teacher and school time needed for an evaluation is worth paying *vis-à-vis* the benefits it can provide. The ideal is to have yearly evaluations. However, this can be very costly. Principals in schools can assume the responsibility of supporting, evaluating and developing teacher quality, based on periodic but not yearly external evaluations. Nevertheless, it seems necessary for evaluations to be more frequent in schools with lower student performance or in schools with a large concentration of novice teachers.

In both cases (formative and summative reports), teachers should have the possibility of contesting them, providing evidence in support of the areas of disagreement.

3. Evidence of teacher performance and student learning

A very important question in every evaluation system, of course, is how to assess impact on student learning. Gathering multiple sources of evidence about teacher practice meets the need for accuracy and fairness in the evaluation process, taking into account the complexity of what a good teacher should know and be able to do (Santiago and Benavides: 2009). A standards-based approach with multiple sources of evidence offers guidance on how to replicate teacher effectiveness. Complementary student performance-based systems can identify teachers associated with learning improvements, but need to take into account other information sources to yield insights about the reasons for the effectiveness of the individual teachers. (Sclafani and Manzi, 2009).

For individual teacher evaluation, therefore, a range of instruments and information sources should be used. The obvious options here include classroom observations, teacher portfolios, teacher knowledge tests, evidence of student learning and growth that may include student performance data, and interviews with peers and supervisors.

The use of teacher portfolios would include examples of the assignments teachers have given to students, sample student papers or other responses to those assignments, and teacher commentary on the student work. Portfolios, if done well, can provide an evaluator rich information about the quality of teaching and learning in the classroom and should be an important element in a formative evaluation process. In the case of Mexico, two factors that could provide valuable information on teacher performance could be teacher attendance and punctuality. According to TALIS (OECD, 2009), 69 percent of principals report that teachers arriving late hindered the provision of instruction in their schools a lot or to some extent; sixty seven percent report the same effect with regard to teacher absenteeism. Both figures are well above the TALIS average of 15 percent in the case of teachers arriving late and 26 percent in the case of absenteeism.

Mexico has been using teacher knowledge tests in a large scale for two factors in the *Carrera Magisterial* program: professional training (*Preparación Profesional*) and certification in training courses (*Acreditación de Cursos de Actualización y Superación del Magisterio*). The test on *Preparación Profesional* is related to the general knowledge and abilities required by each type of teacher. The assessment instrument is elaborated by the *Secretaría de Educación Pública*, and its application requires the coordination of the *Secretaría* with the state educational authorities. Different exams are created depending on the curricular demands. The *Acreditación de Cursos de Actualización y Superación del Magisterio* is based on the results of teachers in courses designed to strengthen specific knowledge and competences. At the end of the courses teachers are examined with national tests. The design of these courses takes into account results on the assessment of teacher training and student performance in order to overcome the weaknesses of the system.

Another possibility is to use ENLACE as part of a collective assessment process, with the school as the unit of accountability. We support the recommendation of the OECD Steering Group on Evaluation and Incentives Policy for Mexico that ENLACE results be a central part of a **school** evaluation process and associated incentives at the school level. In order to attribute value-added learning to schools and teachers, it is essential that ENLACE and its underlying administration procedures and data systems be further developed along the lines recommended by the OECD Steering before ENLACE can be used to assess individual teacher performance. In the meantime, ENLACE results can be used to stimulate continuous improvement processes of teacher and schools, as well as an independent check on the results of evaluations based primarily on evidence of teacher practice. Large and consistent discrepancies between these two sources of information should trigger some form of administrative review of the teacher evaluation process.

In the Delaware case, teachers whose students do not show improvement in standardized tests during an academic year are not classified as "effective" teachers even if they had strong results in all other areas of teacher performance; inversely, teachers whose students achieve high results are not classified as "ineffective" teachers even if they had "unsatisfactory" results in all other areas of teacher performance In other words, student growth is the minimum requirement to be considered a good teacher (Delaware Department of Education, 2010). Delaware has been able to introduce student achievement as part of teacher evaluation because of its very high quality database that enables the link between the trajectories of students and teachers.

Currently, applications of ENLACE cover students who can be reasonably and adequately tested by using multiple-choice forms. This kind of test would be inappropriate for assessing learning in preschool, in the first and second grades of primary school, and for assessing learning performance in certain specialised areas of the curriculum. Therefore, it would be necessary to develop and implement alternative student assessment methodologies for some of the areas not yet covered by ENLACE. This will require, among other things, adequate training of those conducting the assessments and the availability of effective procedures for collecting, processing and analysing information (Zúñiga and Gaviria, 2010).

As the teacher performance evaluation matures and gains recognition among teachers and society, it will be possible to link more consequences to it. Confidence and soundness in the evaluation is fundamental for the rewards to have significance and meaning. Therefore, a strategic component of the implementation plan must be the development of technical expertise for teacher evaluation and for supporting the improvement of teaching practice. This development requires time, financial resources and international collaboration. Time is probably the most critical and hard to manage resource. No immediate final solutions can be expected. But the system will have to show partial results in the short run, such as those that could stem from the school-based incentives and improvement processes that use the results of ENLACE.

The evaluation could start with existing elements of teacher evaluation and its scope could be progressively enlarged as other methods and tools are developed. ENLACE is already being used in a number of states as a tool for teacher and school formative evaluation.[iii] The evolving experiences in the states could provide various models that could be honed, replicated, and used to incorporate all teachers participating in ENLACE and their school peers in analysing their strengths and weaknesses, at least once a year. New materials and enhanced methods could be developed at the regional, state, or national level to support improvements in teaching as a result of processes based on ENLACE results. In the short-term, for example, Mexico has the opportunity to monitor the collective assessment and incentives for teacher performance, with the school as the unit of accountability and using student performance data from ENLACE.

4. Making ENLACE evolve into a "value added" test

ENLACE is a major asset of the Mexican education system. In order to enhance its usefulness for school and teacher evaluation purposes, however, it will need further development. The OECD will present advice regarding the constraints and opportunities for the further development of the ENLACE system in a separate document in June 2010. Based on preliminary findings, three strategic approaches are proposed:

- **The need for complete coverage of curriculum content.** If ENLACE is to be considered a high-stakes test for teacher evaluation and school-performance assessment, one of the issues to be considered is the extension of coverage of the curriculum content in the applied learning tests.

- **Reference to the curriculum.** Mexico has a clearly defined national curriculum for basic education, and ENLACE is designed to assess the degree to which students have mastered that curriculum. Any modification of ENLACE should be explicitly linked to the ongoing processes of curriculum reform in the different levels of basic education. To the extent that the curriculum reform implies more focus on the development of generic and transferable skills and competencies, ENLACE must be adapted to reflect these changes.

- **Technical development of the ENLACE assessment**. A more powerful design of ENLACE would need to address technical challenges such as vertical comparability, analysis that can reliably isolate contributions to learning, and determination of effects on learning from different potential influences. A robust design of ENLACE relies on prerequisites in at least three categories: administrative, logistical, and technical. A detailed treatment of these categories for the further development of ENLACE will be provided to SEP by the OECD in a separate document.

Challenges involving the development of ENLACE require careful thought on where to spend energy. Rather than dispersing energy with extending assessment to other grade levels and disciplines, at this point efforts would be better directed at strengthening validity in the existing ENLACE measures and improving the administration of the tests. At present ENLACE is the most relevant instrument of education evaluation in Mexico. Extending it to other grades or subjects would most likely not be compatible with the need of reinforcing the backbone for standards of educational quality in Mexico to which other parts of the evaluation strategy and instruments can then be anchored.

5. Teacher evaluation and its links to Carrera Magisterial system

A reliable teacher evaluation system must lie at the foundation of teacher incentive programs as a condition for incentives to accurately reflect the quality of teaching. Measuring the quality of teaching should embrace the multiple dimensions of the recommendations that are set forth in the previous sections.

Making *Carrera Magisterial* rely on a solid teacher evaluation system would help the program revamp the factors upon which its evaluation lies today. The tests on "*Preparación Profesional*" and "*Acreditación de Cursos de Actualización y Superación del Magisterio*" measure some of the areas of teacher competence — disciplinary and pedagogical content — but cannot assess many other areas of the professional practice of teachers as described in the proposed framework. Training courses may lead to improved teacher performance but this is not necessarily the case. "*Desempeño escolar*" which relates to teacher performance as judged by principals or supervisors does not rely on external evaluation methods that require judgment by independent evaluators.

It should also be mentioned that in the current system, progression through the levels of *Carrera Magisterial* does not require evidence of improved teacher performance, as measured by evidence of

increased student learning. The instruments to enter the Program are the same as for promotions. An evaluation system that ties financial incentives to promotion should also demand more from those that reach the higher levels.

Carrera Magisterial is well liked by teachers — no surprise, since it has led to salary increases for a majority of those who have applied. The federal government is investing over 40 billion Mexican pesos in the program annually. The size of *Carrera Magisterial* and its impact on the teaching profession make it necessary to ensure that participating teachers are evaluated using the criteria and methods of a robust teacher evaluation system like the one proposed above. This is a major reason to accelerate the design and implementation of a solid teacher evaluation system. Revamping *Carrera Magisterial* could first be done in the states that have the strongest commitment to developing the evaluation system that underpins this Program.

V. IMPLEMENTATION ISSUES DERIVED FROM THE INTERNATIONAL EXPERIENCE

The international experience sheds light on key implementation issues that could be valuable for the Mexican case. Countries across the world are trying to establish teacher evaluation systems with the ultimate goal of improving student achievement in ways that prepare students for successful lives. Empirical research is starting to provide evidence about what is needed, but one must consider whether systems created in one culture are appropriate in others. Perhaps the clearest international lesson is that creating a new evaluation system requires a great deal of thought and must include participation from the stakeholder groups that will be affected by the results.

As stated in the OECD report on international practices for evaluating and rewarding the quality of teachers (OECD, 2009), the returns of a well-educated population to national economic competitiveness are only likely to increase. At the same time, the capabilities of countries to measure the effectiveness of their education systems, and the contributions of teachers, are also expected to increase. In short, the likelihood of further development of performance-based policies is high. Teachers and their unions, of course, have special concerns. They want to ensure that whatever reforms are established, it is fair to all teachers. With any major policy shift there are likely to be disruptions for at least some of the current participants in the system. But both teachers' unions and governments have a common interest in making the education system more productive and should join together to design and test the viability of different performance-based policies. What seems clear is that the politics of performance-based evaluation in education are changing, and so too should our conceptions of the teachers, their unions, and the officials involved in the debate.

We draw key implementation issues from the OECD report on international practices for evaluating and rewarding the quality of teachers that seem relevant for the Mexican case and adjust them to fit in the local context. Instead of simply reproducing issues from international experience, we attempt to select, adjust and provide implementation points that can be useful for the Mexican case. We organize them in five categories which generally should run sequentially, albeit with some overlapping phases: starting general steps, creation of the evaluation framework, preparing schools, piloting and full implementation.

1. Starting general steps

Making the case for a teacher evaluation system. Student attainment can only substantially rise with better teachers. Evaluating teachers is a most valuable way to help them improve their practice. Since the purpose of teacher evaluation is to improve practice in order to improve student learning, the main goal of the system is formative in nature. We recommend to stress that formative evaluation will be established and tested before introducing consequences of significance to individual teacher's conditions. Later, as the evaluation system is accepted it will be of use to distinguish between good and bad teachers, and particularly recognizing good teachers. It should be said that it is task of the key stakeholders to develop the teacher evaluation system, and that government will not design it on its own and later try to impose it. Teachers and society at large should be convinced that developing a transparent and fair teacher evaluation system, supported by opportunities for teachers to improve their practice, is a necessary and worthwhile endeavour.

In Latin America the education sector involves multiple stakeholders, each of which has significant political weight (Romaguera, 2008). They also have very strong ideological positions about issues of evaluation, performance pay, working conditions and the individual rights of teachers. One of the challenges in implementing an ambitious teacher evaluation system is to identify high-level political supporters who are able and willing to galvanize broad political support for the program.

Involving stakeholders. A teacher evaluation system can only be built with the participation of stakeholders. From the very beginning they need to be involved in all major decisions that lead to the establishment of the system. Negotiations will be almost inevitable and the creation of a committee will help process them in an orderly manner. Its outcomes will depend on the clarity of the case for a teacher evaluation system and the early commitment of all parties involved in making concessions in favour of the public good. The work of the committee, successfully done, will have forged a strong advocacy team that will work together to sell the program to their constituency groups.

A crucial responsibility of the stakeholders' committee is the development of a communications strategy. As soon as a compensation committee is appointed, people within the system will begin to offer opinions and concerns about an incentive program. To provide valid and up-to-date information on the purpose of the committee and its progress, the committee must establish a communications strategy for keeping constituencies informed. Regular communication throughout the development and implementation of the system will dissipate any doubts about anything that could be interpreted as hidden. When communication remains open and regular, stakeholders have greater trust in the system. A plural and broad representation on the committee will help balance strong ideological positions.

A teacher evaluation plan should be designed with educators at all levels. It is crucial that they understand the plan and its focus on improving student achievement. Teachers also need to believe that the system is fair and transparent.

Involving local authorities. Designing and implementing a teacher evaluation will not be an easy task. System design can take place at the national or state level, or with the local education authority. If done at the national level, the components of the plan should be designed with input of local education authorities so that its incentives match their needs. Involving state authorities makes it easier to introduce flexibility in the choice of models; this may facilitate public support for the evaluation system.

Identifying a "champion" for the plan. Throughout the period from initial marketing through full implementation, the system needs a champion. The chief educational leader needs to be on board and work closely with the teacher union. The champion and the stakeholders' committee need a technical team to present proposals and process issues that are brought up as the evaluation system is developed. It should include research and evaluation experts for collecting and analyzing data.

Ensuring funding. A teacher evaluation system is expensive. Funds are needed to pay for rubrics, a rating system, hiring and training evaluators and collecting, analyzing and reporting data. Additional costs may also result from better controlled standardized student assessments. If the evaluation systems are tied to incentives on top of ongoing compensation, additional resources will have to be budgeted. The history of merit-pay plans in the United States in the 1970s-1990s, when continued funding for incentive programs was often vulnerable, has made educators cynical about discussions of new systems.

2. Creation of the evaluation framework

- **Developing standards for teaching**. The development of standards for teaching is the cornerstone for teacher evaluation and teacher training. The creation of the framework should start with the standards for teaching.

- **Developing valid performance measures**. Along with the data system, a school system must have valid and reliable measures of teacher performance that teachers agree are fair and accurate. This may include assessments of teacher performance that are based on multiple observations by trained evaluators using a standards-based rubric that teachers believe reflects good instructional practices. Other measures of teacher performance may include contributions to the school-improvement efforts and student performance measures, as has been described in the previous sections.

- **Building a robust data management system**. As a starting minimum the information system requires teacher and student data; as the system matures it should be able to connect both kinds of data. The recently launched teacher census in Mexico (*plantillas docentes*) is a good starting point, although the challenges ahead are huge.

- **Training competent evaluators**. It is a key requirement for the successful evaluation of teachers based on teaching standards. Unless a considerable investment is made on the training of evaluators, a formal teacher evaluation may fail as the legitimacy of the evaluators is not recognised by the teachers.

3. Preparing the schools

- **Training of school leader and teachers on the evaluation system**. Prior to implementation of the pilot, it is essential to prepare teachers and principals on the consequences of evaluation, starting, as previously stated, with formative issues. This should make information available on the support services available to teachers and principals. Training should be provided on the analysis of data from student assessments in all schools, both summative standardised assessments and formative assessments created and used at the school and classroom levels. Teachers need to understand how they can use the data from assessments to improve their individual and collective practices. In addition, training is needed on what the system considers effective teacher practice, and it should be provided to teachers and principals in the pilot schools before the program is extended. The training should be provided to both teachers and principals on what the expectations of the standards are and how the evaluation will be implemented. Training should also be provided to prepare teachers to use the support systems that should be available for them.

- **Creating access to feedback and improvement**. It is important to ensure that teachers have access to feedback, support and professional development to improve their knowledge and skills. If teachers are doing the best they know how to do, the system must provide feedback on areas of strength and weakness as well as support to overcome weaknesses. This includes access to professional development on content knowledge and effective content strategies shown to improve student performance; analyses of student work; and student performance, interim assessments done in the classroom or school-wide during the school term, and summative assessments to evaluate progress on the year's curriculum; as well as mentors and coaches who can assist the teacher in improving his or her performance through modelling and observations of master teachers.

- **Implementing a communication plan for teachers and principals**. Teachers and principals will want to know the consequences of the evaluation system on them and the professional development opportunities they are provided, among other issues that should be conveyed to them. Parents and the public at large will also want to have information. An effective communication strategy would require meetings with teacher and principals, multiple means of

distributing information to educators and the public (e.g. brochures, pamphlets, newsletters, website), trained individuals at schools to answer questions, and a well-informed media.

4. Piloting

Designing of a pilot program to test design, instruments and evaluators. While the perfect plan may have designed, the plan should be tested in a pilot to identify any challenges with the plan itself or its implementation before rolling it out to the entire system. If the data-management system cannot actually access and analyze the data in a timely and effective manner, or if the training for the new evaluation system did not prepare teachers and principals for effective implementation, appropriate changes can be made prior to the full program launch. A pilot also enables plan designers to try variations of the plan in different settings to see which has the most beneficial effect. A pilot also allows the system to implement the specific learning measures and performance measures within a small number of representative schools within the states or regions and analyze the results to ensure they are what the committee anticipated. Careful monitoring of the responses of teachers, administrators, unions, and external stakeholders to the plan from announcement of the program throughout the pilot period will provide important information. If valid concerns are expressed, the system can work to accommodate appropriately. Additionally, teacher surveys and principals' evaluations can be used to evaluate the effectiveness of the professional development program on an ongoing basis.

5. Full implementation

Preparing transitions from pilot to full implementation. To ensure the success of the full-scale implementation, weaknesses identified in the pilot must be addressed before implementation moves forward. Once the system is confident that it has addressed, it is ready to begin full implementation. Ongoing evaluation of the program will provide critical feedback for system improvement.

It would be unwise to move into full implementation before the pilot is considered successful in the eyes of the teachers, administrators, union leaders, academics and civic leaders who have been involved in its design, implementation and support. Even with a successful pilot, full implementation may place new strains on the administrative systems.

Attaching consequences to the evaluation system. Once the evaluation system has been tested, adjusted and accepted higher stakes could be attached to the system. The nature of the consequences and the best possible ways to implement them will require additional and new dialogue among stakeholders.

Creation of an evaluation plan. Without good evaluation, teacher evaluation systems cannot improve over time or determine the most effective options. Effective professional development needs to become a part of the system's operations, and regularly assessed and improved, if the system is to sustain improved instructional practices. Should the program not be contributing to the improvement of student achievement, the system would have to be adjusted. Even if student achievement is improving, it will not be possible to know what caused the improvement without an evaluation that isolates the effects of the teacher evaluation from other influences and activities.

VI. RECOMMENDATIONS

Building a framework for teacher evaluation and making it operational is a lengthy and complex endeavour. We can recommend a broad path and implementation strategy for Mexico to follow, but the ultimate decisions on how to go about making reform happen will have to be taken by the Mexican education authorities and stakeholders, including lawmakers and parents. In what follows we make general policy recommendations that address the central issues involved in establishing a teacher evaluation system. The recommendations do not go into practical and operational details. These will have to be defined once the decision to carry out a reform process in teacher evaluation has been taken and the main stakeholders have been involved and consulted. It also requires the initial establishment of the technical bodies and due consideration to political implications.

In Chile the process of putting a teacher evaluation system in place took approximately ten years, if not more. Mexico should be able to improve on that timetable, but it is imperative to start quickly. It is important to ensure that a clear course is set to navigate amidst political and administrative changes. There is a serious demand to improve education in Mexico. The role of teachers is paramount as is the will to better reward them for better performance and increased student learning.

The Sclafani and Manzi report (2010) outlines six short-term steps and four longer-term actions needed to put in place a teacher evaluation system for Mexico (See report). We concur with these recommendations, and with the recommendations outlined in the report of the OECD Steering Group on School Management and Teacher Policy in Mexico. We develop below some short-term recommendations for the Mexican context and complement them with others to suggest ways to move the agenda forward. These recommendations take into account the implementation issues derived from the international experience set forth in the preceding section.

Recommendation 1. Establish a leadership structure and clear rules for the governance of the evaluation system.

This is a critical decision that involves delicate political and institutional arrangements. A sound normative body and set of governance mechanisms is essential to provide guidance and direction to the road ahead. In order to produce this result, all major stakeholders need to agree on the key rules. A dialogue is needed between the educational authorities, the national teachers' union (SNTE), educational researchers, and civil society leaders to move this agenda forward. Well defined roles, and rights and duties for those involved should be made explicit.

The basic rules would provide a proper framework within which to have all future discussions and make operational decisions. We believe that the recommendations made in this paper could provide much of the core of these rules. The cornerstones are governing body, framework of effective teaching, evidence of teacher practices, multiple evaluators of teachers, use of student performance results, training of evaluators, training of teachers to understand the system, training of school principals to support the system, external certifications of teacher knowledge and skills, and school-based and complementary support for improvement of teaching practice.

The basic rules must pay attention to the distribution of rights and duties among stakeholders. Of special importance is the division of federal and state government responsibilities. National responsibilities

should be within the federal government's sphere; local responsibilities, within the sphere of local governments. Fruitful cooperation will depend on respecting this division of responsibilities. Defining them clearly from the beginning will make the process easier for all parties involved.

A governing body or a steering committee needs to be established composed of ex-officio members from the central *Secretaría de Educación Pública* and state educational authorities, a member of the Teacher Union (SNTE) and perhaps the president of the Federation of Parent Associations. Representatives of leading civil society organizations should also be involved. Besides the ex-officio members, nominal members from the academia, and respected members of the teaching professions must also be included. This body would set the governance rules, with special attention placed on the division of federal and state government responsibilities. Special care should be taken to decentralize decision making and the operations of the system, but keeping the national lead and coherence of the endeavour.

Among the first tasks of the steering committee is to design a strategic and operational implementation plan. This steering committee could be assisted by subcommittees for certain specialized tasks, such as the training of evaluators, the use of students' standardised tests, the development of the school-based teacher professional development support mechanisms, and the quality assurance of the whole process. Lawyers would have to work together with the steering committee to draft legislation to ensure the continuation of the process beyond the current administration. Budget offices will have to calculate the costs of the processes involved, for the short, mid and long terms, to ensure resource availability during this administration and beyond.

Given the high stakes of the proposed teacher evaluation system, it would be highly desirable to have the basic rules for governance made law. This would provide much needed stability. An executive order or other types of non-legislative procedures could be used to start deploying the proposed teacher evaluation system, but legislation should ideally also be introduced in Congress. The law would state the objectives of teacher evaluation, the principles that would govern it, the key aspects for the functioning of the system, its connection to school evaluation and other provisions such as the existence of the governing body, its composition or the role of the states. The purpose of the law would not be to create a discussion process among stakeholders to create the rules of the system.

The process to create the rules that govern the system would by itself foster wide communication to the broader public. Teachers and society at large should be informed and aware of the purposes that underlie the creation of the teacher evaluation system.

We recommend that the rules for the governance of the evaluation system be set as a condition before initiating full implementation of the system. Mexico now has a one-time-opportunity to build a comprehensive teacher evaluation framework. It will be successful with the concurrence of all major stakeholders, but will be fragile if they are not all included in the process. This recommendation does not imply that no progress can be made while the governance rules are written. In the meantime ENLACE should continue to be used as a tool to support teacher and school-wide improvement and for collective teacher rewards. Local developments of teaching standards could also be carried out without the need of the fully deployed program. The rules for the governance of the system, however, are needed for a sustained and medium-term reform process that will yield substantial results.

The appointment of an official from the *Secretaría de Educación Pública* to act as the on-going "champion" is also crucial. This champion would draft proposals for the steering group, and plan tactical steps to make sure that the implementation plan is approved according to proposed deadlines.

In addition to the "champion", other departments of the Ministry need to be involved throughout the design process to ensure that the teacher evaluation system is budgeted for, meets legal requirements, and

is aligned to the human resources systems. All of these investments of time will require funding over and above the usual allocations. The champion would be in charge of involving these other departments of the *Secretaría* whose support will be needed throughout the process.

The first task of the steering committee should be the writing of a strategic and operational implementation plan. It should include detailed short terms tasks and envision mid and long-term activities. Making the plan public would help it gain support from stakeholders and society at large.

A major issue related to the feasibility of putting the evaluation system in place is the willingness and readiness of individual states to join-in. A point of flexibility should be the speed to make the system cover the states. One model that we recommend is not to move resources in all states at the same time. The commitment of the states to become part of the evaluation system will be a major variable for success. We recommend that the implementation starts in those states where conditions are more favourable. Gradually, as progress is made and lessons learnt, it will be easier for others to participate.

Having the basic rules of the system made law, as previously said, is highly desirable. However, in the event that legislative action is not possible in the near term, the system could be created by other means, such as executive orders or agreements among the main stakeholders. The proposed speed of implementation could vary depending on restraints that are put forward by stakeholders. But it should be kept in mind that in any resulting scheme the clear definition of rules should not be avoided.

Recommendation 2. Establish a technical unit that will be responsible for the implementation of the evaluation.

Once the governance rules are set, considerable resources will have to be devoted to the development of a national evaluation system. In fact these rules should include the creation of the technical unit responsible for the implementation. It could be set as a technical secretariat within the *Secretaría de Educación Pública* and serve as the technical body of the steering committee. Recruiting top personnel with technical profiles is a first priority. If hiring is subject to authorization of a technical advisory board appointed by the steering committee, technical personnel would be better protected from political changes.

The tasks assigned to the technical unit represent a considerable challenge that requires a clear, consistent and timely development of the technical expertise involved. Budgeting needs should be arranged as part of the establishment of the unit.

In more advanced phases of the program, funds will be necessary for the implementation of a teacher evaluation system. Once the framework and teacher evaluation system have been designed and the instrumentation developed, there will be on-going costs: training of teachers and evaluators; deployment of evaluators, possibly multiple evaluators, in all schools; time of evaluators to conduct formative and summative evaluations, and analyse the evidence and provide feedback, and support to teachers for practice improvement, amongst others.

We suggest that the technical work that is concurrently done in the states rely on local technical units. Signed agreements between the *Secretaría de Educación Pública* and the state governments to create such local units and specify the profile of its personnel, key functions and relationship to the national unit are highly advisable. At the initial stages, these units are especially critical in the states where the standards instruments are being built and tested.

Although, eventually, technical units will be needed in all states for the operation and development of the evaluation system, they do not have to be simultaneously created. It should be expected that state governments will have different reactions to the proposal of the evaluation system. It would be wise to start setting it in those states where conditions are more favourable and have it progressively extended.

Recommendation 3. Develop standards for teaching.

International experience shows that the development of standards for teaching represents a fundamental basis for teacher evaluation and teacher training. The first product of the technical unit must be the framework on which the evaluation system will be based. A good starting point might be to review Chile's "*Marco para la Buena Enseñanza,*" or Danielson's "Framework for Teaching," although other frameworks should be considered as well. The instrumentation, training, and professional development content for the evaluation system must await the framework decisions.

The first step would be for the educational system to clearly define the knowledge and skills that students are expected to acquire at different educational levels, as these must be related to the standards of teaching. The *Subsecretaría de Educación Básica* has already been working on defining the standards of student knowledge and skills ("*competencias*"). In addition, as part of the work of OECD with SEP, further to the recommendation to develop teacher standards valuable for teacher training, for the processes of allocation of teaching *plazas* and for in-service teacher evaluation, a short report will provide evidence-based analysis defining what is a *standard* and which *components* standards may have for different purposes[iv].

Standards need to cover all of the important teaching domains. They should reflect what teachers identify as good teacher performance. Standards establish different levels of competency for each of these domains, and care should be taken to ensure that they are neither too low nor too high. A nuclear core of standards should be applicable to all teachers and all schools. Standards are not aimed at "standardizing" teacher behaviour; rather, they should enhance the capacity of teachers to innovate and adapt to different contexts. This is especially important in a diverse and multicultural country like Mexico, and for an educational system as diverse and complex as ours.

The evaluation framework should define standards in at least the following domains: use of instructional time (attendance, punctuality, time on task), planning and preparation (the design of instructional activities and evaluation procedures for all students), classroom environment (making the classroom a safe place for risk-taking), instruction (adapted to different students, engaging and challenging), and professional responsibilities. Special care should be placed on the ability of the teacher to strive for equity, that is, to attend to the needs of different students in order to achieve learning outcomes for all.

The technical unit would have to organize the process to develop the standards of teaching and develop and pilot test evaluation instruments. It would need some technical capabilities of its own, and would have to draw expertise from national and foreign experts. The characteristics of good standards described in previous sections should always be kept in mind.

We recommend that the teaching standards be developed along two parallel paths that start from an expert proposal. The expert proposal would be initially drafted by the experts of the unit, experts from the *Subsecretaría de Educación Básica* and external advisors. The first path would follow a format of testing the proposal with guided discussion with teachers; the second route would allow for teachers to freely review the expert proposal. In both cases, teachers must be representative of regions, the three levels of basic education, and their different modalities.

Once both processes have been completed, results should be compared to produce further discussions. The technical unit would then produce a second draft that would have to be filtered by experts and later tested with teachers. This national trial would serve not only to make final adjustments, but also to give legitimacy to the standards

This dual process combines the indispensable acceptance of standards by teachers and the knowledge of experts. In both cases states should be considered as key stakeholders and get involved accordingly. The three levels of basic education and different modalities should also actively collaborate.

Recommendation 4. Design an in-service teacher evaluation model that gradually evolves from a purely formative system to one that combines formative and summative aspects.

As suggested in the Sclafani and Manzi Report (2010), teacher evaluation systems are more likely to overcome resistance and become accepted by the teaching profession when teachers know that their main goal is to promote professional development, in the context of enhancing student learning. Therefore, we consider it fundamental to establish a system that during its first years of operation follows a formative goal.

Since the main purpose of teacher evaluation is to improve practice in order to improve student learning, the overriding goal of the teacher evaluation system must be formative. Teacher evaluation should provide feedback to each teacher on how to improve his or her practice, and the system should support mainly school-based professional development mechanisms to ensure that improvement is in fact possible. Teachers should receive adequate feedback on their content knowledge and content pedagogy based upon rubrics connected to the standard-based framework. This will enable teachers to determine a personal development plan and identify appropriate professional development options. School-based training and professional development plans should then be designed and supported to ensure proper implementation. The point is that teachers at all levels of proficiency and at all stages of their careers should have access to support and professional development to improve their practice It is important to define and clarify what the feedback process for teachers will be. Ultimately, summative as well as formative objectives of the teacher evaluation system can be sought, but we strongly believe that in order to build understanding and support among teachers, it is critical that a formative evaluation system be established and accepted by teachers before introducing any consequences or stakes.

Recommendation 5. Define the instruments for the in-service teacher evaluation system.

Based on international experience, a system that includes several evaluation instruments can be most effective, albeit challenging. The obvious options here include classroom observations, teacher portfolios, evidence of student learning and growth that may include student performance data, interviews with peers and supervisors, and tests of teacher knowledge. Each of them provides specific information about the work of teachers and when combined should give a clear picture of teaching performance.

Testing the teaching standards will require instruments. It should therefore be expected that the technical unit prepares pilot evaluation instruments (and trains evaluators) at an early stage. This will require extensive technical support from the *Subsecretaría de Educación Básica* personnel as well as external experts. It is important that instruments are consistent with the teaching standards. Instruments that seem indispensable are handbooks and rubrics for external evaluators who visit the schools and view the teaching practices, questionnaires for interviews, teacher knowledge tests, evaluation manuals for supervisors and principals and others. These instruments should be adjusted after an initial trial period.

As previously mentioned, Mexico has few instruments for teacher evaluation. Most of the evaluation is done by the principal without any specific guidance or instruments. Principals need to be better prepared to support this task. ENLACE is being increasingly used by teachers and schools as an indicator of performance. In the short and medium term, this can be strengthened and improved. As said before, we support the recommendation of the Mexico/OECD Steering Group on Evaluation and Incentives Policy that ENLACE results be a central part of a *school* evaluation process and associated incentives at the school level. In order to attribute value-added learning to schools and teachers, it is essential that ENLACE

and its underlying administration procedures and data systems be further developed along the lines recommended by the Mexico/OECD Steering before ENLACE can be used to assess individual teacher performance. In the meantime, ENLACE results can be used to stimulate continuous improvement processes of teacher and schools, as well as an independent check on the results of evaluations based primarily on evidence of teacher practice.

Teacher knowledge tests used for Carrera Magisterial —*Preparación Profesional and Acreditación de Cursos de Actualización y Superación del Magisterio*— also offer a departing platform for teacher evaluation within the proposed framework. In the future, as the teacher evaluation system develops, these two tests should be revised to make sure they are aligned to the evaluation framework.

Recommendation 6. Develop a support system for school-based professional development that leads to the improvement of teacher practice, and a system that monitors this improvement.

This is the most important consequence of formative evaluation, because it is the way of ensuring a continuous improvement of teaching practice that leads to improving student learning. This represents a major reform in the Mexican Educational System, because it implies three things:

Selecting and training a body of mentor teachers and coaches, as recommended by the OECD Steering Group in School Management and Teacher Policy (2010), and funding them with resources for visiting schools and working with teachers;

Redefining the role of school principals to focus on supporting, evaluating and developing teacher quality, among one of their key roles, training them, and supporting them to become instructional leaders, as recommended by the OECD Steering Group in School Management and Teacher Policy (2010);

Re-training school supervisors so that they can play an active role in monitoring improvements in teaching practice and principal leadership, and authorizing them to intervene with support mechanisms when needed.

School-based professional development means that schools be evaluated along with teachers. Teaching takes place within a school, with a formal or acting principal, and collaboration among teachers. As part of its annual planning, the school should establish clear goals related to quality teaching and learning that involve all teachers. Meetings for reviewing and for defining areas of needed teacher foster critical reflection on both individual and collective professional performance. Broader school autonomy seems a necessary complement for effective formative evaluation to happen in schools.

Recommendation 7. Train Evaluators.

The implementation of the teacher evaluation system will have to rely on a robust cadre of evaluators — supervisors, technical pedagogical advisors (ATP), principals, master teachers, retired educators, and teachers in the "Normales" or the "Universidad Pedagógica Nacional" in different content areas and levels of schooling — that can be trained to conduct effective evaluations using the adopted framework of effective teaching. They system would require around 2,500 full-time evaluators if the school evaluation cycle is four years; it would take approximately 10,000 evaluators to have every school visited once a year by a team of two evaluators. The figures are challenging, but are within reach if it is taken into account that only in the ATP there are around 50,000 teachers already being paid and who would most likely perform more effectively if working within the proposed teacher evaluation framework. The conversion of *ATPs* into evaluators would require the definition of a new labor status that responds to the duties and status of an evaluator.

Evaluators should not only be trained but also be periodically assessed on their competencies to accurately and consistently measure performance against the standards of the framework. Training and monitoring are important for all evaluations, but especially important when evaluations are used for summative purposes.

Although the recruitment and training of a cadre of evaluators should come mostly after the framework decisions and the definition of instruments for evaluation, it is nonetheless an activity that should start as the standards and the instruments are tested. The steering committee would define the guidelines for choosing, training, and assessing evaluators. Following these guidelines, the technical unit and the local technical units should agree on the profile of possible evaluators, training schemes, and assessment of those who take the training before they go into service.

Selecting, recruiting and training of evaluators is also a considerable challenge in itself and all the more so due to the enormous diversity of the Mexican Educational System to which we have referred to above. It is essential for evaluators to be accepted by the teachers, equipped to be fair in their duties, sensitive to organizational, contextual and cultural differences, and able to fulfil their role objectively and responsibly.

Recommendation 8. Reduce administrative duties of supervisors and principals, and increase school autonomy.

Supervisors, principals and even teachers are overloaded with administrative work. According to TALIS, Mexico is the participating country in the study where teachers devote most time to administrative tasks (17 percent). There should be a shift from an administrative leadership to an instructional leadership in schools.

From the beginning, the teacher evaluation framework should not be seen as one more useless burden. Explaining the framework to teachers and training them as well as the rest of the participants is imperative but not enough if administrative work is not made lighter. In fact, the heavy share of administrative work in the educational structure reflects practices that are insufficiently student-learning-driven. Inertial continuity of such practices could well hamper the framework from operating effectively.

Even though overcoming such difficulties is beyond the responsibility of the steering committee and the technical units, we suggest that they also monitor the reduction of administrative burdens on supervisors, principals and schools. Credibility in the teacher evaluation framework will hinge on the system's ability to make it look different from other administrative duties that have little or no impact on learning.

Broader school autonomy seems a necessary complement for effective formative evaluation to happen in schools. It implies, among other things, identifying, selecting, and training mentor teachers and coaches; reducing the administrative load of the school principal, and training them to become instructional leaders involved in the development of the teachers in their schools. It also requires supervisor retraining so that they mentor and monitor actual classroom practice improvement. This is a major reform in itself, key to improving student learning in Mexico.

Recommendation 9. Prepare a program for ENLACE to be further developed into a "value-added" test.

As mentioned in the preceding section, the task of modifying ENLACE to make it suitable for evaluation requires a specific program. This is technically complex but feasible. Above all, the benefits to assess teaching and other aspects of the educational system could be significantly expanded.

Additionally, some of the existing shortcomings of ENLACE for teacher evaluation would be surmounted by identifying options to implement assessments of student learning and growth to preschool, the first two grades of primary school, for example. Where standardized written student testing is not advisable, expert observation of student performance and evidence of their learning are possible substitutes. Sampling students in the classroom would diminish costs and still provide evidence of the group's progress and insights about teaching performance. It is worthwhile to review some of the alternative methodologies that are already being developed and pilot-tested in Mexico and in other educational contexts.

The case of special education is worthwhile mentioning. As ENLACE evolves into a high-stakes test, schools and teachers will have an incentive not to take children with special education needs since they would tend to lower their results. The same can occur with children that speak a native or foreign language and do not speak Spanish Such practices should be avoided for the protection of the children's rights to education and without affecting the fairness of the teacher evaluation system. One option in the case of special education needs would be to identify these children in regular schools to have their ENLACE results computed separately from those of their group. Personnel in charge of the nation-wide existing *Unidades de Servicio de Apoyo a la Educación Regular (USAER)* could identify those students. The development of these children could be used as evidence of the results produced by the USAER personnel.

Making reform happen requires that there is a perception that things are moving in the direction of wider and more in-depth evaluation. Working on the evolution of ENLACE gives the right kind of signals on the seriousness of the teacher evaluation purpose.

Recommendation 10. Gain momentum towards establishing the teacher evaluation system.

As previously mentioned, the road to build the system will be complex and long. The nature of the challenge makes it not advisable to take shortcuts. At the same time, even with the right rules for the governance of the system, there is the risk that initial enthusiasm may fade.

Consequently, it is imperative that the system shows partial results in the short run. Two strategies may be particularly useful to this end. One is the adoption of the OECD Steering Group on Evaluation and Incentives Policy recommendation that ENLACE results be a central part of a *school-level* evaluation process and associated financial, material and other type of incentives at the school level (OECD, 2010) The second is that further attention be given to the use of ENLACE to stimulate discussion and analysis of teacher and school performance on a regular basis. A number of states have realized the potential of the latter strategy and have been supporting the discussion and analysis of ENLACE outcomes within schools and regions. Should this strategy be technically, and financially supported by the federal government, teachers and schools would further benefit from this approach. An immediate consequence would be a clearer focus on learning. Existing school advisors would gain experience in relating student performance to the work done by teachers in schools. This paves the way for the kind of practices that teacher evaluation requires in schools.

Recommendation 11. Pilot and evaluate. The design, instruments and evaluators should be tested in pilots, in different contexts, before rolling out the evaluation system to the entire system of large tracts of schools.

A pilot also allows the system to implement the specific learning measures and performance measures within a small number of representative schools within the states or regions and analyze the results to ensure they are what was anticipated. A pilot also enables plan designers to try variations of the plan in different settings to see which has the most beneficial effect. An evaluation plan is needed not only for the initial stages of the program, but also for the mid and long-term.

VII. CONCLUSION

The purpose of all education systems is to improve student learning. The quality and equity of student learning depends on many factors, but most importantly on the quality of teaching. Therefore, the development of a comprehensive in-service teacher evaluation system, in line with the larger framework of teacher professional development proposed by OECD (2010), is recommended for Mexico.

Defining and putting in operation a teacher evaluation system, such as the one described above, demands a long-term and complex process capable of navigating across government administrations. This has delicate political implications and implies technical as well as legislative and regulatory requirements. In our view, there are two clear short-term goals that should define the activities of the present administration: (a) to define the Teacher Evaluation Framework, with teaching standards, and reach an agreement with all major stakeholders, and particularly teachers; and (b) to identify the best possible design and implementation strategies for incorporating student learning and growth within teacher evaluation and continuous improvement processes including technical work to further develop ENLACE.

The goal should be that by the end of this Administration, a framework and its standards will have been defined and accepted by key stakeholders. Also, by the end of the administration, the instruments will have been designed and groups of evaluators selected and trained; the support process strategically planned with mentor teachers and coaches selected for school clusters. Decisions will have been taken as to whether and how and when to use student learning and growth, including but not limited to students' standardised test results, in teacher evaluation.

The willingness and readiness of state educational authorities should be taken into account as the teacher evaluation system is built. The implementation strategy is likely to be more successful with the cooperation of the individual states. Active and committed states should be the first to operate teacher evaluation system.

We believe it is feasible that by the end of the present administration in-service teacher formative evaluation with trained evaluators and appropriate instruments will have been applied in clusters of schools in at least 10 states. The formative evaluation could cover 25 states by 2015, and by 2020 all of the states. The support system for school-based professional development and a system that monitors this improvement could be put in place accordingly. Summative evaluation that relies on the comprehensive framework could well be in use in some states by 2015. In these more advanced and quality-committed states, the evaluation system that underpins *Carrera Magisterial* could be fully revamped.

Finally, it is our hope that by the end of the administration legislation will be enacted to extend the development process sufficiently to enable the establishment of a comprehensive in-service teacher evaluation system. If this were not possible, the agreement with all major stakeholders should be very clear and solid for the indispensable long-term construction process of the evaluation system to hold.

This proposal will not be inexpensive. There are costs involved in every step of the process that are new to the Ministry and to the states. But without a highly skilled and motivated teaching force capable of dramatically improving educational outcomes for Mexico's young people, Mexico's goals of increased economic opportunity and social mobility will be more difficult to achieve. Recent studies published by the OECD show that improving student learning can have dramatic effects on Mexico's GDP in the medium and long-term. Not to invest in developing a teacher evaluation system well-aligned with the more comprehensive reforms proposed by the OECD for teacher policy and school leadership would, in the long run, be much more costly to Mexico's future than the investments we are recommending now.

[i]Evaluation studies of the use of the Danielson model in the US have demonstrated that the system is both valid and reliable, with the group of teachers who scored well on the evaluation system also showing greater student growth on tests than teachers in the control groups (See Milanowski, 2004; Milanowsky and Kimbell, 2003, and Milanowski, Kimbell and Odden, 2005).
[ii]The school is the context in which the teacher works. As such, it has great influence on the quality of teaching practices. The quality of the school is an important factor in determining the quality of the teaching that can take place within it. Therefore, the school should also be evaluated periodically. It is perhaps not advisable to carry out the evaluation of teachers and schools at the same time. Nevertheless, Mexican education authorities should consider the importance of school evaluation as part of an integral system of the evaluation of the quality of education.

[iii] A review of state-level teacher evaluation and rewards practices has been commissioned by the OECD and findings from this study will be available in June 2010.

[iv] The first draft of this analysis will be ready at the end of June 2010. The organization of a workshop during the first half of September 2010 to discuss this analysis and related issues is under discussion within SEP.

REFERENCES

Avalos, B. And J. Assael (2006). "Moving from Resistance to Agreement: The Case of the Chilean Teacher Performance Evaluation". International Journal of Educational Research 45, pp. 254-266.

Danielson, C. And T. McGreal (2000) "Teacher Evaluation to Enhance Professional Practice". Association for Supervision and Curriculum Development (Ascd), Alexandria, Virginia.

Danielson, C. (2007). "Enhancing Professional Practice: A Framework for Teaching". Association for Supervision & Curriculum Development (Ascd); 2 edition. Alexandria, Virginia.

Danielson, C. (2008). "The Handbook for Enhancing Professional Practice: Using the Framework for Teaching in Your School". Association for Supervision & Curriculum Development (Ascd); 2 edition. Alexandria, Virginia.

Delaware Department of Education (2008). DPASII Delaware Performance Appraisal System. Dover, Delaware.

Delaware Department of Education (2010). Race to the Top. Application for Funding. CFDA Number 84395A. Narrative. The State of Delaware.

Khim Ong, K. Et al. (2008). "Teacher Appraisal and its Outcomes in Singapore Primary Schools". Journal of Educational Administration, pp. 39.54.

Instituto Nacional para la Evaluación de la Educación (2008). "*¿Avanza o Retrocede la Calidad Educativa? Tendencia y Perspectivas de la Educación Básica en México*". Informe Anual. INEE, México.

Manzi, J. (2009). "Individual Incentives and Teacher Evaluation: The Chilean Case" (Revised Draft February 2009). Universidad Católica de Chile. Santiago.

McKinsey & Company (2007). "*How the world's best-performing school systems come out on top*". Mckinsey & Company.

Milanowski, A. (2004). "The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence from Cincinnati". Peabody Journal of Education, Vol. 15, No. 3, pp. 193-212.

Milanowsky, A. and S. Kimball (2003). "The Framework-Based Teacher Performance Assessment Systems in Cincinnati and Washoe". CPRE Working Paper Series TC-03-07.

OECD. (2009). "Teaching and Learning International Survey (TALIS)". Paris.

OECD (2009). "*Mejorar el liderazgo escolar vol. 1: Políticas y prácticas*". Mexico.

OECD (2009). "*Evaluating and Rewarding the Quality of Teachers. International Practices*". Paris.

OECD (2010). "Students and Schools at the Centre, Teachers at the Heart: School Management and Teacher Policy for 21st Century Schools in Mexico". OECD Steering Group on School Management and Teacher Policy. Paris. OECD.

Ontario Ministry of Education (2009). "Overview of the Ontario Teacher Performance Appraisal (TPA) System".

Perrenoud, P. (2004). "*Diez Nuevas Competencias para Enseñar: Invitación al Viaje*". Madrid: Grao.

Romaguera, P. (2008). "Incentives and Stimuli for Teachers: The Case of Chile". Paper presented at the International OECD Mexico Joint Conference, The Quality of Education, Mexico City.

Rewards and Incentives Group (2009). "Teachers' and Head Teacher's Performance Management: Guidance". Available from www.teachernet.gov.uk/management/payandperformance/performancemanagement.

Santiago, P. et al. (2009) "Teacher Evaluation in Portugal". OECD Review.

Santiago, P. and F. Benavides (2009). "Teacher Evaluation. A Conceptual Framework and Examples of Country Practices". Paris: OECD Secretariat.

Sclafani, S. and J. Manzi. (2010). "Report on In-Service Teacher Evaluation and Development Practices in a Comparative Perspective." Paper prepared for the work of the OECD Steering Groups on Evaluation and Teacher Incentives and School Management and Teacher Policy.

Singapore Ministry of Education (2006). "Singapore Staff Appraisal (Education Service)." MOE, Singapore.

Zuñiga, L. and J.L. Gaviria (2010). "Could ENLACE be transformed into a main component of a new system of incentives for teachers, in Mexico?"