Chilean Evaluated Teachers Give Their Opinions about the National Teacher Evaluation System

Dante Cisterna-Alburquerque[1]

Michigan State University

Abstract

A study was conducted in a Chilean commune (district) in order to describe opinions of evaluated teachers and school principals about the national teacher performance evaluation system and the ways they are using the reported information about teachers' performance. Evaluated teachers and principals positively value the clear procedures and adequate organization of the system. By contrast, the overwhelming tasks required for responding to the instruments and the quality of the reports are scarcely evaluated. They also make little use of the information reported for teachers' improvement. The results of this study suggest that the national and large-scale assessment, whose official purpose is formative, has focused on the quality of its operative procedures and its technical aspects, at the expense of strengthening the use, details, and pertinence of the reported information for their intended users.

---

[1] 250-C Erickson Hall, Department of Teacher Education, East Lansing, MI 48424 USA
E-mail cisterna@msu.edu (517) 4887007

I. Introduction

The development of systems for evaluating teacher performance has become an important educational policy in the Latin-American context, especially since the massive reforms made at the end of the 20th century (Robalino, 2007). Chile was one of the pioneer countries to implement a complex, national, standards-based evaluation system. This was born in 2003, and even though it was politically negotiated by its stakeholders (Chilean government, municipalities, and teachers union), it was resisted by many schoolteachers, the intended users of the evaluation. This evaluation design has been recognized as well-thought-out and exhaustive (OECD, 2010), and after several years of operation, it is operatively consolidated and every year thousands of teachers in Chilean municipal (public) schools participate in the process.

This paper presents a study conducted in the schools of a Chilean commune (district). The study's purpose was to describe the knowledge and opinions of evaluated teachers and school leaders about the national performance evaluation system in which they participated. The study also described the use that teachers and school leaders make of the results provided by the evaluation system, given that the evaluation was officially designed to enhance teaching through continuous professional development (Docentemás, 2009).

Much of the information about the operation and implementation of the evaluation system comes from reports provided by the same system; by contrast, evidence gathered from intended users of the evaluation is scarce. Collecting data from different sources and actors is important to frame the picture of this teacher performance evaluation, especially when the Chilean law which regulates the process defines this process as formative, "which aims to improve teachers' pedagogical work and to promote their continuous professional development" (CPEIP, 2006).

The paper examines first some aspects of the teacher evaluation systems, with a special focus on the Chilean context. Research on teachers' perceptions of

performance evaluation and feedback is also reviewed. Second, the methodology for doing the study is described as well as its main results. Finally, the findings of this study are analyzed and discussed according to previous research and the context of the Chilean evaluation system.

II. Background

1. Teacher Performance Evaluation Systems. Teacher performance evaluation has been consolidated as a relevant process in Latin-America, and in countries such as Chile. Robalino (2007) considered that this topic raises notoriety since the educational reforms made in the decades of the 1980s and 1990s, which implied a shift in the traditional vision of teacher evaluation and mainly referred to the purposes of selecting, certificating, or contracting (Milman et al, 1997). In this new perspective, the variable "teacher professional performance" has been identified as relevant and influential to achieve a qualitative increase in school management, educational outcomes, and student learning achievements (Valdés, 1990). Consequently, educational policymakers have designed and implemented a set of initiatives aimed to enhance the professionalization of teachers, for example, through professional development opportunities, incentives for good teacher performance, performance standards, and teacher performance evaluation (CPEIP, 2006). Similarly, the establishment of teacher evaluation has emerged as an educational policy suggested by different institutions, because it is conceived as desirable and is also able to be applied in different educative contexts (World Bank, 2002).

According to a policy perspective, teacher performance evaluation contributes to identifying those qualities which characterize a good teacher, in order to establish a baseline for generating strategies for the improvement of the whole educational system (Valdés, 1990). Moreover, these processes help the system in assuring standards for teachers and in getting useful information for decision-making in various

aspects. An example of these decisions is gathering evidence about the performance of teachers who work in the public system, and then linking these data with student learning achievement, teacher education accountability, collaborative networks of teachers, and curriculum decision-making (Dwyer, 1998).

For the evaluated teacher, it is assumed that the benefit of teacher performance evaluation is related to the enhancement of one's own performance, professionalism, and public perception. This also can be related to the establishment of strategies of support and intervention to improve performance. Evaluating teachers is seen as key to enhance teacher professional development, through which teachers with a minimal level of competence reach higher levels of professional suitability and expand their understanding of themselves, their roles, context, and professional career (Riegle, 1987).

The complexities of teacher evaluation processes increase when considering the wide variety of intentions and interests of different stakeholders regarding to evaluation implementation and evaluation use. Such complexities also have a direct relationship with the demands of intended users and the clients of the process. For example, Murillo (2007) included as teacher evaluation those processes with in-service teachers whose goals are improving teacher performance; maintaining their motivation; and recognizing socially and monetarily their work. This conceptualization combines different evaluative purposes which have to be carried out in concrete and contextualized processes. Probably, that is the reason because the design and implementation of teacher evaluation occasions important levels of controversy and poses problematic issues for evidence of technical, methodological, administrative, normative-legal, cultural, and ethical aspects.

Generally speaking, there is not a common model of what is considered a "quality teacher" according to the international perspective. Every country or educational system creates its own teacher performance evaluation model, by

4

integrating different theoretical approaches and diversity of perspectives and instruments (Murillo, 2007). Therefore, it is essential that a teacher evaluation system makes explicit the theoretical model that supports it, and considers the need of creating a constructive image of the process, for their users as well as the public opinion, by spreading the idea that evaluation is in favor of teachers and of their professionalism.

In the Latin American context, teacher evaluation is officially conceptualized as a process aimed to improve teacher quality, and also to offer information for the design of policies, strategies, and programs to be implemented nationally, locally, and institutionally in order to contribute to professional development. However, evaluations and intended outcomes are not always consistent with the approach, methodology, instruments and use of the information (Robalino, 2007). In addition, Murillo (2007) said that the most frequent teacher evaluation systems impact on promoting the professional scale and salary, as a criterion for vertical promotion, or for setting negative consequences such as withdrawing tenure or permission to teach.

2. The Chilean System of Teacher Performance Evaluation In Chile, the system for evaluating teacher performance has been implemented since 2003. It is focused on the professional work of schoolteachers in municipal schools (in Chile public schools have a double dependency -administrative issues are mainly of municipal competence while curricular issues are mainly the competence of the Ministry of Education.) A political agreement (in Spanish: "Acuerdo Tripartito") signed in June of 2003 between the Chilean Ministry of Education, the Chilean Association of Municipalities, and the Chilean teachers union ("Colegio de Profesores de Chile") framed the bases for the evaluation system. This was declared as formative, i.e., oriented to improve teachers' pedagogical work and promote their continuous professional development (CPEIP, 2006). The system expects to enhance the teaching profession, by promoting teacher strengths and overcoming weaknesses in order to

5

achieve better student achievement (Docentemás, 2009). In addition, the system was declared as a formative process and not as punitive, because it did not pretend to sanction or expel teachers, even in extreme cases. The evaluation also intended connecting with initiatives of continuous teacher professional development, with a special emphasis on those teachers who evidence lower levels of performance (CPEIP, 2006).

The Chilean teacher performance evaluation is standards-based and nationally centralized. Its design, implementation, and administration are responsibility of the Chilean Ministry of Education, which contracts a private company for technical and operative support. The local component of the evaluation is administered by the Chilean municipalities. In each one a local commission of evaluation ("Comisión Comunal de Evaluación") is established whose responsibility is to carry out the procedures for the evaluation locally.

The reference for setting criteria to construct the assessment instruments is the "Framework for Good Teaching" (In Spanish: "Marco para la Buena Enseñanza"), a document published in September of 2003 which presented descriptions to characterize teacher performance, by including in a measurable or at least observable way the aspects that teachers must know and be able to do in their profession (MINEDUC, 2003). This framework was inspired by the report "A Framework for Teaching" (Danielson, 1996) which was proposed in the United States as a reference for school districts to design locally their own teacher performance standards. It is important to remark that, in the Latin American context, the "Framework for Good Teaching" is relevant and innovative because it is an explicit and theoretical reference to develop teacher performance evaluations. This framework is also considered as eclectic because it collects and integrates different models of teacher assessment, such as aspects related to teachers, classroom behaviors, task development, and professionalism development (Murillo, 2007).

6

Teachers who participate in this system are evaluated by four instruments which aim to gather different aspects of teacher performance, according to the constructs defined in the "Framework for Good Teaching." The instruments and their prompts are the same for all teachers in the country, regardless their level of teaching, subject, geographic location, and other contextual variables. The instruments are 1) a teacher portfolio, which contributes to gathering and ordering evidence about classroom practice and schoolwork, and it includes planning, teaching, assessing, and reflecting on a standardized teaching unit; 2) self-assessment, which invites one to reflect and provide the teacher's own perception of performance; 3) peer-interview, conducted by another teacher from the same commune and subject which questions the contextual-specific characteristics of teaching; 4) supervisor assessment, completed by the school leaders, i.e., the school principal and the leader of the "technical-pedagogical units", who supports classroom teachers in their pedagogical work.

Teachers are evaluated every four years. In the year of evaluation, teachers have to complete four instruments during the school year. In the case of teacher portfolios, they are later revised and scored by other school teachers who are contracted by "correction centers" which are administered by different Chilean universities (Docentemás, 2009). The ratings of the instruments are weighted in an integrated score which is transformed into general performance categories. They are named as Outstanding, Competent, Basic, and Unsatisfactory.

The evaluation results are then shared through reports. The system generates different types of report to the evaluated teachers, the school principal, and the educational authorities of the Chilean municipalities. Reports are locally revised by the local commission of evaluation in order to analyze the results according to the contextual variables which may affect the result of the evaluation. If pertinent, the

local commission may modify the general performance category assigned to the teacher by the centralized correction process.

The report designed for the teacher presents the scores obtained by the different instruments, a description of the main strengths and weaknesses in each category and indicator, and the teacher's general performance category. The school principal and other educational authorities receive reports which only include general results per school and teacher (Docentemás, 2009).

Evaluated teachers who obtain the two lower categories - Basic and Unsatisfactory- are eligible for professional development for overcoming their deficits. These plans are mandatory and are funded by the Chilean government. Teachers who are in the two higher performance categories –Competent and Outstanding- may apply for a variable monetary raise, after they pass a test of content and pedagogical knowledge. A special situation occurs for teachers who are in the category Unsatisfactory because they are obliged to be evaluated the next year, and the teacher who continues in that situation may have to teaching (Ley Nº 19,933).

Since the beginning of the implementation, this evaluation system has faced resistance from an important number of teachers who refused to participate in the process. They have criticized various aspects of the evaluation, refer to, in general, the evaluation purpose and methodology and teachers' participation in the evaluation design (Cárdenas, 2007). In the two first years of implementation the evaluation was voluntary and 5,400 teachers participated, but also an important number refused. Since the enactment of a set of legal regulations in January of 2005, in which the participation in this system was considered mandatory, the number of participating teachers has increased to more than 52,000, and now the system consolidated as a current practice for Chilean teachers who work in public schools.

This evaluation model has been described as well-known and exhaustive, with rich qualitative information, and as an excellent starting point to identify teachers'

individual strengths and weaknesses and to guide them towards tailor-made professional development programs (OECD, 2010). It has been described as objective, technically transparent, with a design which incorporates a good formative effort. The combination of evaluation, accreditation, professional development, and educative quality is one of its main strengths (Pizarro, 2007).

3. Meaningful and Useful Evaluations for Teachers Undoubtedly, the implementation of this complex evaluation system produces different types of consequences for teachers. First, they have to study the characteristics of the evaluation, a new process for them, and then to respond the assessment instruments. Second, even though it is argued that the instruments are similar to their current teaching practice, they have to adjust their practice to the standardized prompts required (an additional and external demand.) Third, after teachers receive their results report and are assigned into a specific performance category, and depending on their results, they have to participate in professional development initiatives or may apply for merit pay. Concerning the different stages of the evaluation, there are some preliminary findings about how evaluated teachers appreciate their evaluative experience and the usefulness of the information aimed for their professional improvement. Manzi (2007), from the perspective of the evaluation designers, reported the results of a questionnaire responded to by teachers when they gave back their portfolios. For those scales regarding the similarities of the assessment instruments to the current teaching practice, teachers' scores were close to the scales' central values. Teachers also said that the evaluation contributed highly to "recover practices", while other aspects such as systematizing procedures, improving practices, motivating students, and detecting one's own strengths and weaknesses had lower scores.

Cárdenas (2007), in a study with teachers, administrators, and educational authorities in a Chilean province, concluded that low participation and involvement in

the decisions are those aspects which caused reluctance to the evaluation. Also, the different visions of teachers, educational authorities, and policymakers about the process affected the evaluation operation, by evidencing some perceptions of ambiguity. Even though teachers considered teacher performance evaluation as necessary, it has been perceived by them as unclear in its purpose, methodology, and objectives. Moreover, teachers have expressed concerns about raters' suitability and ignorance about the procedures for assessment instruments' correction. Teachers have also argued the weights of each instrument, the existence of fears of the evaluation, and criticisms of the mass media by the way they covered teachers' reluctance to be evaluated. Teachers considered that this evaluation, due to its standardized nature, did not capture the particularities of each teacher's work. Concerning the mandatory professional development initiatives aimed for teachers with lower performance levels, Cortés (2007) said that the main factor that affected the operation of the system is lack of information about the procedures to work with the individual teacher reports. Many times, professional developers were sufficiently clear about neither objectives nor conceptual framework. An important proportion of the teachers have had a negative perception to participating in the plans. More than 40% of the participant teachers said the courses lacked depth, and almost a 30% criticized their length.

Findings in other contexts provide similar evidence. Ovando (2001) in a study of a local teacher evaluation system in the US, said that evaluated teachers positively appreciated opportunities for professional development that offered feedback, and self-assessment. On the other hand, teachers questioned descriptions of the performance categories used and certain degree of subjectivity of that evaluation. The Word Bank (2002) reported difficulties in the effectiveness of teacher evaluation, because the assessment instruments have been, in some cases, decontextualized from teacher practices. In general, few teacher evaluation systems distinguish between the improvement of professional performance and the administrative or monetary

assignment factors. These combinations cause that evaluation, as an opportunity for teacher development, to be interrupted because one purpose is prevailing over the others. Similarly, Dwyer (1988) emphasized that the results reported in the evaluations, if they are discussed remote from the values of the schools and communities, rarely are perceived as useful by the intended users.

Generally speaking, teachers consider positive the information reported by the evaluation. However, when the process has drawbacks concerning opportunity, usefulness, and credibility the results do not help teachers to change their teaching practices. Milanowski (2005) said that for accomplishing better results in performance evaluations, it is essential to focus on the quality of the communication among participants rather than on evaluation design and operation. For Magarrey (2002), the low effectiveness of teacher evaluations is related to the excessive emphasis of evaluators on organizational aspects of teaching, which also scarcely considers learning contexts and teacher pedagogical skills. Kimball (2002) considered that successful feedback requires information linked to concrete evidence and should be presented according to specific criteria. Brandshaw et al (2003) said that lack of time to gather information and to help teachers, when giving feedback, was a factor that influenced the scarce satisfaction that these evaluations caused in teachers. Also, they mentioned the importance of assuring quality communication because this contributes to create an environment of trust to help stakeholders discuss evaluative criteria and understand what good teaching is. Cabatoff (2000) emphasized that the language currently used in results reports uses an important number of terms with different meanings for participants, and many times reactions to these discrepancies may be interpreted as teacher resistance to incorporate the suggestions mentioned in the results reports.

Thus, these different findings provide evidence of the complexities associated with teacher evaluation. For the case of the Chilean teacher evaluation system, which explicitly declares a formative purpose, the impact of this process on teacher

11

improvement is an issue to be focused and enhanced. Moreover, since this evaluation system emerged from a political agreement in which the Chilean teachers union participated as a valid actor, the main intended users of the evaluation should be the school teachers; and therefore, the evaluation design and implementation also ought to respond their particular needs.

Despite the fact that the Chilean teacher performance evaluation system is recognized as a well-though-out process (OECD, 2010), there is no direct evidence of how teachers and school leaders are using the results to report on their own improvement, or how they are valuing the evaluation according to acceptance and usefulness. Furthermore, the information gathered in this study suggests that this national, standardized, and large-scale assessment has prioritized its operative procedures and technical aspects, at the expense of strengthening the use and pertinence of the reported information for their intended users. For evaluated teachers, participating in this complex process has implied simultaneously responding to different purposes, and as a result, the general performance category obtained by the teacher is more important than the intended formative purposes.

III. Research questions and methods

Focusing on the experiences of evaluated teachers and school leaders this study posed the following research questions:

1) What is the knowledge that evaluated teachers have about the Chilean teacher performance evaluation system?

2) What are the opinions that evaluated teachers and school leaders have about the Chilean teacher performance evaluation system?

3) What are the opinions that evaluated teachers and school leaders have about the results report?

4) Which uses do evaluated teachers and school leaders give to the information reported by the teacher performance evaluation system?

This quantitative study is defined as descriptive and has a non-experimental design. Five dimensions were defined which account for the evaluation different aspects and stages, and also are relevant to set the respective sub-dimensions and indicators aimed to guide and organize the procedures for gathering data. The five dimensions were 1) knowledge of the evaluation characteristics, 2) opinions about the evaluation implementation, 3) opinions about the evaluation instruments, 4) opinions about the results report, and 5) uses that teachers give to the reported information. The sample was 54 elementary school teachers who taught in classes between grades 1 to 8. They belonged to ten municipal (public) schools of an urban commune in Santiago de Chile. Previously, these teachers had been evaluated at least once in this district and had completed the process; i.e., they were assigned into one of the four general performance categories. Twelve teachers were "Basic", 38 "Competent", and 4 "Outstanding". Even though there were in the commune teachers categorized as "Unsatisfactory", they were not working in the schools when the data was gathered. In addition, three school leaders (one principal and two members of the technical-pedagogical unit) were included in the sample. Moreover, it is important to note that one evaluated teacher refused to participate in the study and thereby was not included in the sample.

Two procedures for gathering data were used. The first was a questionnaire which included 71 scales and 2 open-ended questions. The scales were organized in a fashion which helped the evaluated teacher review the different stages of the process. The questionnaire was constructed according to the different dimensions, sub-dimensions, and indicators set in the study, which are presented in Table 1. This assessment instrument was also revised and validated by three specialists in educational measurement and teacher evaluation.

13

Table 1 Dimensions, sub-dimensions, and indicators of the study

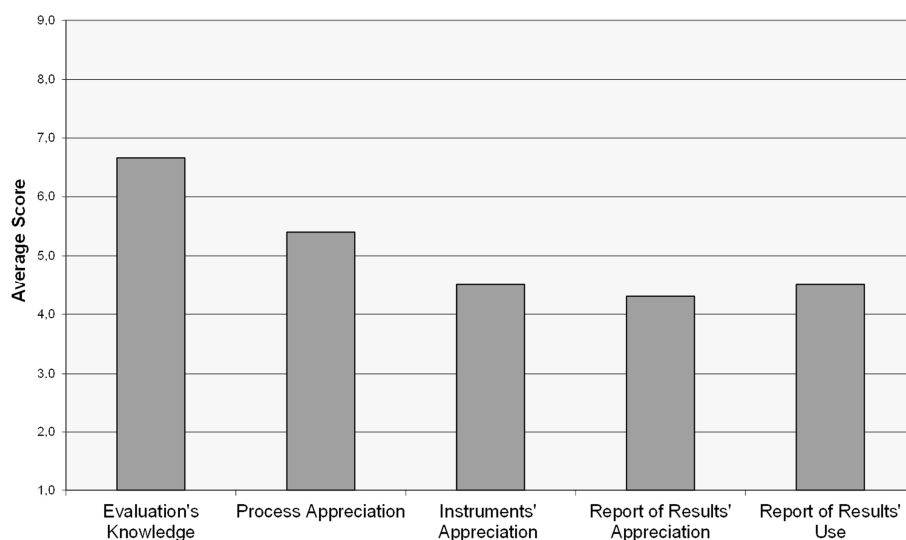| Dimension	Sub-dimension	Indicator
1. Knowledge of the evaluation characteristics	1.1 Knowledge of the regulatory framework	1.1.1. Knowledge of regulatory norms and laws
		1.1.2 Knowledge of the "Framework for Good Teaching"
	1.2 Knowledge of the evaluation procedures	1.2.1. Knowledge of the assessment Instruments
		1.2.2. Knowledge of the evaluation stages
		1.2.3 Knowledge of the local commission of evaluation
2. Opinions about the evaluation implementation	2.1 Process organization	2.1.1. Evaluation calendar
		2.1.2 Clearness of the evaluation Procedures
	2.2 Evaluation implementation	2.2.1 Respect to the evaluation stages
		2.2.2 Operation of the local commission of evaluation
		2.2.3 Procedures for supporting teachers
	2.3 Assessment instruments' administration	3.2.1 Time required to respond to the assessment instruments
		3.2.2 Procedures for administrating assessment instruments
		3.2.3 Respect of procedures by instruments' administrators
	2.4 Results report operational procedures	4.1.1 Procedures for handing over the results report
3. Opinions about the evaluation instruments		3.1.1 Teacher portfolio directions and characteristics
		3.1.2 Assessment instruments characteristics
		3.1.3 Agreement with the assessment instruments
4. Opinions about the results report		4.1.2 Type and quality of the information reported
5. Uses that teachers give to the reported information	5.1 Results report usefulness for teaching practices	5.1.1 Report usefulness for teaching practices
		5.1.2 Uses given to the report information
	5.2 Teaching practices after the evaluation	5.2.1 Professional contribution of the evaluation to teaching practices
		5.2.3 Participation in programs of improvement after the evaluation

The second procedure for gathering data was an interview protocol for teachers and school leaders. The purpose of the interviews was to deepen understanding of the opinions, ideas, and knowledge of evaluated teachers and school leaders, complementing the information collected by the questionnaire. The interview protocol was semi-structured and had eight topics, which covered the different dimensions of the study. In order to interview participants with different profiles, ten teachers were selected according to their scores on the questionnaire and their general performance categories. The criterion for selecting the three school leaders to be interviewed was the average score in the teacher questionnaire for the respective school.

The data gathered in the questionnaire was analyzed according to descriptive statistics (means; standard deviations; proportions for every dimension, sub-dimension, and indicator) and inferential procedures (comparison of means among groups of teachers.) The interviews were coded and analyzed according to the technique of content analysis (categorical semantic analysis.)
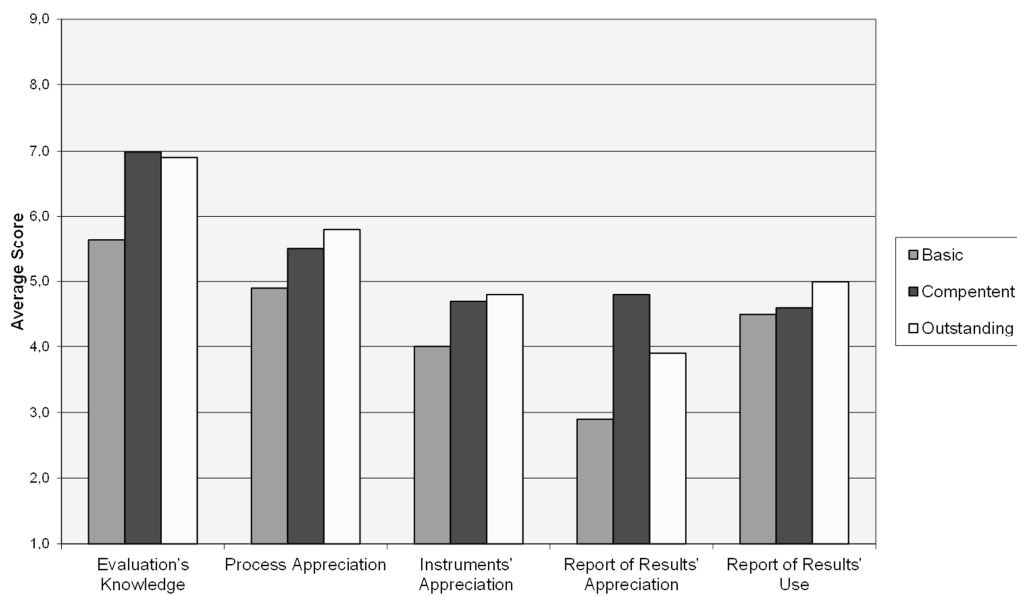
IV. Results

The results are presented according to the dimensions, sub-dimensions, and indicators of the study. The questionnaire's scales ranged between 1 and 9 (1 is the lowest value). Graph 1 shows the average scores according to the five dimensions of the study. The higher average scores were on the dimensions 1 and 2, which related to knowledge of the evaluation characteristics and opinions about the evaluation implementation, respectively. On the contrary, the lower scores were on the three dimensions regarding opinions about the evaluation instruments, opinions about the results report, and uses that teachers give to the reported information.

Graph 1 Questionnaire's scores according to study dimensions

For each dimension, scores were compared according to the teachers' general

performance levels. Graph 2 shows that teachers categorized as "Basic" tended to

have lower scores than "Competent" and "Outstanding" teachers. Furthermore,

comparisons of means between "Basic" teachers in one group and "Competent" and

"Outstanding" in the other group resulted in significant differences for dimensions 1 to

4 ($p < 0.05$), and non-significant for dimension 5 (uses that teachers give to the

reported information).

Graph 2 Questionnaire's score comparisons, according to study dimensions and teacher
performance levels

Concerning results per sub-dimension, Graph 3 shows that the highest average score

was for "knowledge of the regulatory framework", which referred to the laws, norms,

and the "Framework for Good Teaching." Also, knowledge about evaluation

procedures, opinions about assessment instruments administration, and opinions about

procedures for reporting results were highly appreciated. By contrast, the sub-

dimensions regarding process organization, opinions of the assessment instruments,

opinions of the results report, result report usefulness for teaching practices, and changes in teaching practices after the evaluation had lower scores.

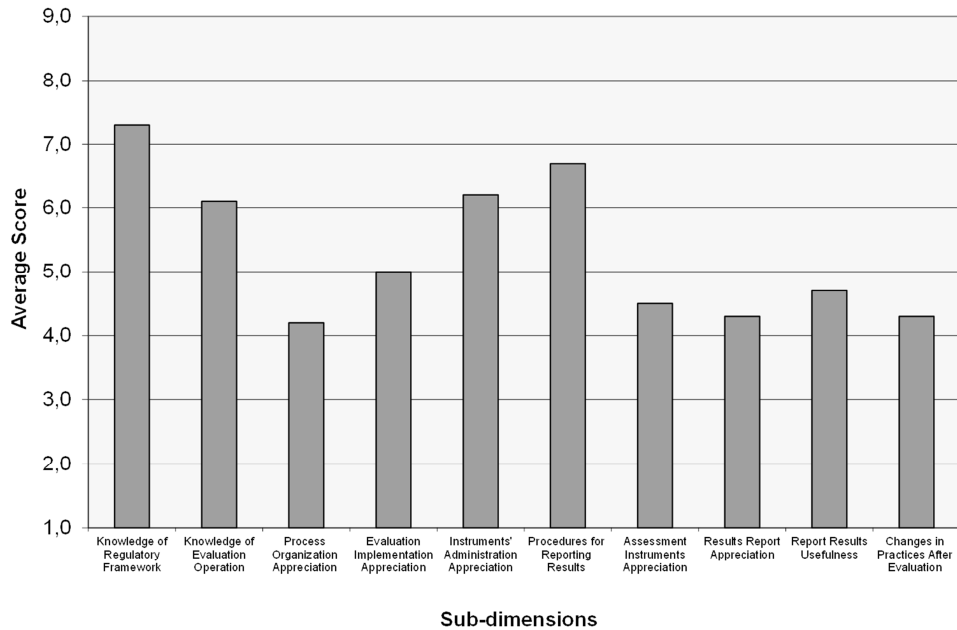Graph 3 Questionnaire's scores according to study sub-dimensions

Table 2 presents the average scores of the questionnaire, according to dimension, sub-dimension, and indicator. As a summary, the more appreciated indicators referred to Knowledge of the evaluation reference framework, knowledge of the regulatory framework and operative procedures, procedures for reporting results, and respect of procedures by instruments' administrators. Time required to respond to the assessment instruments, agreement with the assessment instruments, support of teachers during the process, and professional contribution of the reported information were less appreciated.

Table 2 Questionnaire's average scores, according to study dimensions, sub-dimensions, and indicators

Dimension	Score	Sub-dimension	Score	Indicator	Score
1. Knowledge of the evaluation characteristics	6,7	1.1 Knowledge of the regulatory framework	7,3	1.1.1. Knowledge of regulatory norms and laws	7,1
				1.1.2 Knowledge of the "Framework for Good Teaching"	7,5
		1.2 Knowledge of the evaluation procedures	6,1	1.2.1. Knowledge of the assessment Instruments	7,2
				1.2.2. Knowledge of the evaluation stages	7,2
				1.2.3 Knowledge of the local commission of evaluation	4,7
2. Opinions about the evaluation implementation	5,4	2.1 Process organization	4,2	2.1.1. Evaluation calendar	3,6
				2.1.2 Clearness of the evaluation Procedures	5,0
		2.2 Evaluation implementation	5,0	2.2.1 Respect to the evaluation stages	6,3
				2.2.2 Operation of the local commission of evaluation	5,4
				2.2.3 Procedures for supporting teachers	4,1
		2.3 Assessment instruments' administration	6,2	3.2.1 Time required to respond to the assessment instruments	3,5
				3.2.2 Procedures for administrating assessment instruments	6,6
				3.2.3 Respect of procedures by instruments' administrators	6,9
		2.4 Results report operational procedures	6,7	4.1.1 Procedures for handing over the results report	6,7
3. Opinions about the evaluation instruments	4,5			3.1.1 Teacher portfolio directions and characteristics	5,6
				3.1.2 Assessment instruments characteristics	4,4
				3.1.3 Agreement with the assessment instruments	3,9
4. Opinions about the results report	4,3			4.1.2 Type and quality of the information reported	4,3
5. Uses that teachers give to the reported information	4,5	5.1 Results report usefulness for teaching practices	4,7	5.1.1 Report usefulness for teaching practices	4,3
				5.1.2 Uses given to the report information	5,2
		5.2 Teaching practices after the evaluation	4,3	5.2.1 Professional contribution of the evaluation to teaching practices	4,1
				5.2.3 Participation in programs of improvement after the evaluation	4,4

V. Discussion

The findings of the study are discussed according to its research questions. The information gathered by the teacher questionnaire and the interviews of teachers and school leaders contributed effectively to describing knowledge, opinions, and uses of the Chilean teacher performance evaluation system, for evaluating teachers.

- Research question 1: What is the knowledge that evaluated teachers have about the Chilean teacher performance evaluation system?

The evaluated teachers had to get information and learn about this mandatory process. The majority of teachers declared having relevant knowledge of the evaluation's regulations and the "Framework for Good Teaching." They also declared having requested information about the characteristics of the different evaluation stages. For them, knowledge about the "Framework for the Good Teacher" was the characteristic with a higher score. This document, which is the reference for the assessment instruments, was essential to teachers to understand the context of the evaluation, to complete the tasks required, and to respond to the instruments. Moreover some teachers participated in informative meetings organized by the local teachers union and the educational authorities. According to the study of Cárdenas (2007), teachers also recognized the close relationship between this framework and the evaluation process.

It is not surprising that those teachers who were categorized as Basic had a significantly lower score ($p<0.05$) on the questionnaire than teachers categorized as Competent or Outstanding. Possibly, the lower declared knowledge of these teachers was related to their need to update their pedagogical knowledge, once they were aware of their deficits. This situation poses the question of how much of the evaluation results were determined by having specific knowledge regarding operative procedures

and formalities, which are essential to participate successfully in the evaluation but far from reflective of teachers' actual performance.

- Research question 2: What are the opinions that evaluated teachers and school leaders have about the Chilean teacher performance evaluation system?

Among the aspects of the evaluation which were positively appreciated by the teachers were mainly those referring to the operative and administrative components of the evaluation system. Examples of these aspects were administration of the assessment instruments, procedures for reporting results, and respect and fulfillment of the evaluation calendar and its stages. This shows that the evaluation was respecting its regulatory framework and procedures, with accounts of an operative and well implemented system whose participants (Chilean communes, Ministry of Education, and Technical Assistance) clearly knew their responsibilities. In general, the organization and implementation of the local (and more visible) components of the evaluation system were also well-valued by the teachers.

The administration of the assessment instruments is another aspect that was well-valued, and it also demonstrates fulfilling protocols and procedures. Since this evaluation is national and standardized, it must assure that every teacher be assessed under the same criteria and conditions. Therefore, this fact accounts for the technical quality of the evaluation system.

In general, these favorable opinions demonstrate an adequate operation and implementation of the evaluative system. This aspect is important because, according to Murillo (2007), one of the most difficult issues to achieve is technical quality-processes, especially because of the high cost that it implies. Therefore, an evaluation system which respects and emphasizes its procedures and rules would be better appreciated by its users. Interviews of teachers and school leaders suggested these

same findings, even though some teachers criticized particular aspects of the process, such as overdue provision of materials in the first years of implementation.

However, teachers' opinions tended to be less favorable as they talked about themselves as subjects being evaluated, such as their appreciation of the instances designed by the system to support their needs and the clearness of the directions to respond the assessment instruments, especially in those teachers whose performance categories were lower. The latter may be associated with two facts. First, the support provided was mainly in technical aspects rather than in pedagogical, and probably it was further from teacher needs. Second, teachers with low performance categories also declared having less knowledge about the evaluation process and being less updated, and probably they needed special types of support. Indeed, some teachers commented that local educational authorities should be more supportive and flexible in evaluating teachers.

The most struggling aspect of the evaluation was the time required to respond to the assessment instruments. Teachers declared that responding to the instruments, especially the teacher portfolio, was excessively time-demanding and overwhelming. A big proportion of teachers considered that these demands negatively affected their current teaching practice. For others, this situation was stressful because they should study and understand the directions of the prompts, a critical issue for those who declared not being update about the evaluation reference framework; and therefore, it implied participating in the process under conditions of anxiety and stress. Moreover, teachers cannot mainly respond to the instruments in their current workday, and this affected time destined to other activities. Some teachers declared that completing the portfolio negatively affected the quality of their teaching work, because they used the time currently assigned to their teaching responsibilities -such as planning lessons and revising student homework- in the evaluation. Moreover, it is important to remark that

not all teachers declared having these problems despite the fact they should complete the instruments after their workday.

Some teachers considered the prompts to be very far from their current teaching practices; and hence, they spent much time in studying the characteristics and requirements of the evaluation, and then they started responding to the assessment instruments. Thus, those teachers who had higher shortages in their pedagogical knowledge, reading skills, and terminology use had more problems to face the evaluation, aspects also mentioned by Conley et al (2006). Similarly, it is possible that the results of teachers who fell into the lower performance categories were influenced by lack of fit with the evaluation characteristics and assessment prompts, an issue exacerbated by the scarce opportunities of pedagogical support contemplated in the evaluation.

For most of the evaluated teachers, responding to the assessment instruments was the most relevant experience of the process. Some teachers considered positive those aspects which elicited professional learning and reflection. By contrast, the main criticisms were of the difficulties for the teacher to adjust and fulfill the strict requirements of the instruments, especially the portfolio, which contemplated standardized prompts for planning and implementing a teaching unit. This was mainly illustrated in one portfolio requirement which demanded videotaping a 40 minute lesson by an external cameraman contracted by the evaluation system. The directions for this lesson were very rigid and standardized and, according to the opinion of many evaluated teachers, it merely became an exhibition of teaching behaviors or even in a performance, rather than a current lesson. No less important, this lack of fit may be affecting the tendency to disagree with the instruments used. Teachers commented that the assessment instruments did not help them to evidence their own teaching practices or the particular context of their work. Some teachers even suggested that the instruments assessed the ability of responding to the prompts rather than the real

performance of the teacher. For Kimball (2002), this aspect was important, because the degree of trust and justice by which the evaluator is perceived affects teachers' perceptions of the evaluation.

Concerning performance categories, some teachers questioned the weights for each instrument, similar to what was reported by Cárdenas (2007). Additionally, there is some criticism of the procedures for correcting the instruments. Centralized correction, even though it enhances objectivity in the ratings, also limits knowledge and participation of evaluated teachers with regard to assessment instruments' construction and correction.

School leaders (school principal and members of the pedagogical technic unit) considered that the evaluation process occurred adequately, and its calendar and stages were respected. In some cases, they recognized they provided support to their teachers, especially in emotional fashion. Indeed, they declared that every teacher had some problems during the evaluation. Regarding timing for completing the instruments, they mentioned that it was an issue of criticism, even though they thought that time was sufficient and adequate. Moreover, school leaders did not express agreement with the evaluation instruments, but they assumed that these were part of the regulatory framework that affected teachers and thereby they had to accept it.

- Research question 3: What are the opinions that evaluated teachers and school leaders have about the results report?

According to the declared formative purpose of this performance evaluation, the stage of reporting results is crucial because the teacher receives information aimed to evaluate professional strengths and weakness, and guiding their own practice to better levels of professionalism. However, the results of this study indicated that this evaluation stage had the lowest appreciation of the entire process.

The most questioned aspect was the type and quality of the information contained in the results report. An open-ended question of the questionnaire, 22% of the teachers mentioned that the report needs to be more detailed. Even though the report provided some information about the topics they more and less accomplished, it lacks clear and precise specifications to help the teacher to self-evaluate his or her particular characteristics and to develop an improvement strategy. Teachers described the results report with general, unspecific, and standardized descriptions of the performance categories for each indicator; and, as expected in a centralized and large-scale evaluation system, feedback was not linked to the direct evidence provided by the teacher. Although the report helped the teacher to distinguish, in some way, strengths and weaknesses, its main problem was the absence of meaningful descriptions to justify the reasons why a teacher was scored in a specific performance category and not on other. Moreover, since the report did not mention the specific mistake that the teacher made, it is very difficult to determine the performance aspects that need to be improved. As Kimball (2002) said, successful feedback to teachers implies that the information reported be linked to concrete evidence. The latter aspect is especially relevant in a centralized evaluation system whose instruments are externally rated, and where the only opportunity of feedback corresponds to a written report. The report thereby needs a common language between teachers and evaluators, so that it can enable understanding assessment criteria and professional development (Rallis y Grossman, 2000; Magarrey, 2002; Milanowski, 2005).

According to the opinions of school leaders, the results report they received was important because it helped them to compare the results of their school with others. Since school leaders were focused on teachers' general performance rather than the specific information oriented to teacher improvement, they disregarded the fact that the report only had general information about their school teachers' abilities.

24

These aspects were more relevant, considering that this evaluation was designed for overcoming professional weaknesses and enabling professional development, in order to achieve better student learning (Docentemás, 2009). Accordingly, the degree of detail of the results report is a factor that may be affecting the success of evaluation regarding the accomplishment of its intended purpose. In addition, given the complexities of teacher evaluation, in which stakeholders interact with different values and purposes, the formative component of this evaluation may be weakened and, oppositely, the processes oriented to control teacher work enhanced. Probably, the ambiguity of an evaluation system which combines different purposes in its design is a factor that negatively affected teachers' opinions and appreciations (Cárdenas, 2007; Dwyer, 1988). Interviewed teachers reinforced that the most relevant aspect of the evaluation for them was the final performance category they were assigned and not the information concerning their professional strengths and weaknesses. Furthermore, it is not surprising that several teachers (14%) declared that the report of results did not contain any relevant aspect in order to improve their performance.

- Research question 4: Which uses do evaluated teachers and school leaders give to the information reported by the teacher performance evaluation system?

Evaluated teachers mentioned that their best perceived aspect was the consideration they gave to the reported information. They indicated that after the evaluative experience they used report feedback for accomplishing issues such as more confidence for teaching, better ability to reflect on the own professional practice and opportunities to self-evaluate their professional teaching. These elements are important because they provide some evidence of the formative purpose which is the official aim of this evaluation system. Teachers also declared having more knowledge of the evaluative process and the "Framework to Good Teaching," and that

participating in this evaluation provided them a reference to better structure their classes and update their pedagogical knowledge.

However, an important proportion of teachers declared scarce use of the reported information for changing their current practices, and some teachers even explicitly declared not having changed their practices. The latter can be attributed to different reasons such as prevalence of the general performance category obtained in the process instead of feedback content, lack of detail of the results report, or attitudes of rejection to the process. Some interviewed teachers continued perceiving this process as imposed and external to their current work.

In addition, teachers from the Chilean commune in which the study was conducted and whose general performance category was "Basic" were required to participate in the professional development plans. Teachers commented that the plans were scarcely useful for improving their performance. The activities were perceived as inadequately integrated to the particular contexts of participant teachers because they were not targeted to overcoming the specific weaknesses of the group, which is similar to that reported by Cortés (2007). Teachers also criticized the excessive length and workload of these plans, which affected teachers' motivation.

However, as these plans were consolidating, they partially improved with the incorporation of mentor teachers who also participated in the evaluation and had higher results. Participant teachers perceived this approach as more useful for their learning, paradoxically, how to increase their confidence and preparation for the next evaluation. Probably, that means participant teachers' perceptions of usefulness is related to their expectations of passing satisfactorily another evaluation rather than an instance for improving teacher performance. Murillo (2007) says that one of the main problems of teacher evaluation systems is determining teachers' behaviors, and therefore, one important risk of this type of initiatives is that they are only appreciated

to the extent they help teachers to learn the characteristics of the assessment instruments and raise their general performance category.

Similarly, school leaders said the information reported has been generally useful for teachers, for accountability purposes, and for the opportunity to assist professional development. In an incipient fashion, they recognized changes in some teachers in their schools, and they also planned to include evaluation results in their institutional administration processes.

VI. Concluding remarks

The evaluative experience of this group of school teachers and school leaders illustrates the complexity of implementing an evaluative system for enhancing teacher professionalism, but it also operated further from the intender users' context. First, it would be simplistic and erroneous to assume that teachers are against being evaluated, or their reluctance is the main factor that explains the problems of the evaluation system. The findings of this study coincide with the conclusions of the Cárdenas (2007) study, in which teachers did agree with the evaluation although they mainly expressed concerns about the purposes, methodology, and teachers' participation in the design. This study also shows that for many evaluated teachers, the process was mainly perceived as external to their context and imposed. Similarly, the evaluation was frequently seen as an additional task to accomplish, in a bureaucratic way; by spending a lot of time responding to prompts and performing in order to be professionally accountable. Despite these problems, more than 70% of teachers completed the evaluation process and, within this group, more than 65% have been in the two higher performance categories (Docentemás, 2009).

Second, to what extent is a centralized and standardized evaluation system able to effectively respond to their intended users' needs? Since this evaluation was politically negotiated, it was established as national and centralized. From the

27

educational policy perspective, it is highly desirable to assure national performance standards. However, from the perspective of the evaluated teachers the process is meaningless. The requirements of having standardized and generic prompts, regardless of the particular contexts of teachers and schools, limited using this evaluation to improve continuous professional development and to enhance better levels of professionalism. Not surprisingly, this centralized approach has prioritized developing the technical quality of the evaluation, obviously an essential baseline which contributes to the legitimacy and credibility to the system (Murillo, 2007), but is clearly not sufficient to assure a meaningful and useful process for its intended users.

This same tension between standardized systems and users' needs is evident in the evaluation's results report, whose lack of detail and specificity was severally criticized by those teachers who were looking for relevant information about their performance on the assessment prompts to which they responded. Concerning an evaluation system targeted for enhancing teacher professionalism and continuous improvement, a main challenge is to report quality feedback through a process whose evidence is rated nationally and massively. However, it is important to pose the question of whether this is really possible. Diverse studies have showed the complexities of providing quality feedback, especially when a written report is used (Cabatoff, 2000; Kimball, 2002; Magarrey, 2002; Brandshaw et al, 2003; Milanowski, 2005). Moreover, understanding a written report without using concrete evidence associated with the performance and without assuring a common meaning among stakeholders imply results reports becoming a mere bureaucratic procedure, far from the expectations of the intended users of the evaluation.

Similarly, the official statement of the purpose of an evaluation system was not a warrant of a design. This study provides evidence that the formative component of this Chilean teacher performance evaluation is undermined, at the expense of assuring technical quality and the procedural aspects of the process. Consequently, it also poses

the question whether school teachers are actually the main intended users of the evaluation results, especially considering its design, its procedural approach, and its scarce quality feedback as input for teacher professional development. From another perspective, the question if this evaluation system is another customized and sophisticated version of the old-style teacher evaluations aimed to select, promote, and certify needs to be addressed through future research.

This study has several limitations which restrict the generalization of its findings to larger populations. First, the study was conducted in only one Chilean commune and, possibly, there are several factors regarding the schools which may not be applicable in other settings. Second, the sample of evaluated teachers was not random, but intentional. Third, participant teachers belonged only to elementary levels (grades K1-8) because the system started evaluating this group of teachers first, but now it has expanded to other groups such as high school, special education, and early childhood education teachers, whose particularities may be very different from those who integrated the study sample. Finally, other limitations of the study were the impossibility of gathering data from the lowest performance category ("Unsatisfactory") because they were not working in the commune at the moment of data collection; the data were not analyzed according to the year of evaluation, because the national evaluation system has made changes and innovations in its procedures and assessment instruments; and the incorporation of school teachers and school leaders needs to be complemented in future research by including a broader number of stakeholders.

Bibliography

Boyd, R., (1989) Improving Teacher Evaluations. Practical Assessment, Research & Evaluation, 1 (7)

Bonifaz, R., (2007) Evaluación Docente y formación continua. In: "Evaluación Docente en Chile: Fundamentos, Experiencias y Resultados" Mide UC. Retrieved on Jan 17, 2008 from http://www.mideuc.cl/seminario/

Brandshaw, L., R. Joyner, and S. Colby, (2003) Can Teacher Evaluation Improve Teaching and Learning? Retrieved on May, 3 2008 from http://www.wmich.edu/evalctr/create/2003/CREATE03Lynn.ppt

Cabatoff, K., (2000) Translating Evaluation Findings into "Policy Language" New Directions for Evaluation, 86

Cárdenas, N. y C. Marín, (2007) Análisis crítico del Sistema de Evaluación del Desempeño Profesional Docente chileno: el caso de la comuna de Osorno. Paper presented at the XIX Chilean and V International meeting of Researchers in Education. ENIN 2007, Santiago, Chile.

Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (2006) Compromiso con el Desarrollo profesional Docente. CPEIP 2000 – 2005. Santiago, Chile.

Conley, S., D. Muncey, and S. You, (2006) Standards-based Evaluation and Teacher Career satisfaction: A Structural Equation Modeling Analysis. Journal of Personnel Evaluation in Education, 18.

Cortés, F., (2007) Planes de Superación Profesional: una alternativa de desarrollo profesional ligada a la evaluación docente. In: "Evaluación Docente en Chile: Fundamentos, Experiencias y Resultados" Mide UC. Retrieved on Jan 17, 2008 from http://www.mideuc.cl/seminario/

Delgado, J. and J. Gutiérrez, (1999) Métodos y técnicas cualitativas de investigación en ciencias sociales. Madrid, Spain. Síntesis

Danielson, C., (1996) Enhancing Professional Practice, A Framework for Teaching. Virginia Association for Supervision and Curriculum Development. Alexandria. VA.

Dwyer, C., (1998) The Evaluation of Teachers. Evaluation & Educational Reform. Retrieved on Jun 12, 2006 from http://www.reduc.cl/raes.nsf/0/49b262c2e43958010042568390073d7ab/$FILE/rae08.244.pdf

Kimball, S., (2002) Analysis of feedback, Enabling Conditions and Fairness Perceptions of Teachers in three School Districts with new Standards–based Evaluation Systems, Journal of Personnel Evaluation in Education, 16 (4).

Ley Nº 19.933. Diario Oficial de la República de Chile, Santiago, Chile, February 12, 2004

Magarrey, M., (2002) Changing Perceptions of the Nature of Teacher Evaluation in Ontario Secondary Schools. Retrieved on May 3, 2008 from http://www.oise.utoronto.ca/field-centres/TVC/RossReports/vol7no3.htm

Manzi, J., (2007) Evaluación Docente: Antecedentes, resultados y proyecciones. In: "Evaluación Docente en Chile: Fundamentos, Experiencias y Resultados". Retrieved on Jan, 17 2008 from http://www.mideuc.cl/seminario/

Millman, J. and L. Darling-Hammond (Eds.), (1997) Manual para la evaluación del profesorado. Madrid, Spain. La Muralla

Milanowski, A. and H. Heneman, (2001) Assessment of Teacher Reactions to a Standard-based Evaluation System: a pilot study. Journal of Personnel Evaluation in Education, 15 (3)

Milanowski, A., (2005) Split Roles in Performance Evaluation–A field Study Involving New Teachers, Journal of Personnel Evaluation in Education, 18

MINEDUC, (2003) Marco para la Buena enseñanza (MBE). Santiago, Chile. Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas

Murillo, J., M. Robalino, and A. Korner, (2007) Evaluación del Desempeño y Carrera Profesional Docente. Una panorámica entre América y Europa. OREALC – UNESCO. Second Edition. Santiago, Chile

Ontario Ministry of Education, (2004) Teacher Excellence–Unlocking Student Potential through Continuing Professional Development. Retrieved Jun 12, 2006 from http:www.edu.gov.on.ca/eng/general/elemsec/partnership/potential.html

OECD (2010), OECD Economic Surveys: Chile 2010, OECD Publishing. doi: 10.1787/eco_surveys-chl-2010-en

Ovando, M., (2001) Teachers' Perceptions of a Learner–centered Teacher Evaluation System, Journal of Personnel Evaluation in Education, 15 (3)

Pizarro, R., (2007) Análisis Crítico de la Evaluación de Desempeño Docente en Chile. 2nd International Conference of Teacher Function Evaluation. Temuco, Chile.

Rallis, S. and G. Rossman, (2000) Dialogue for Learning: Evaluator as critical friend. New Directions for Evaluation, 86

Riegle, R., (1997) Conceptions of Faculty Development. Educational Theory 37

Robalino, M., (2007) Tensiones y Oportunidades de la Evaluación Docente. 2nd International Conference of Teacher Function Evaluation. Temuco, Chile

Stiggins, R. and L. Duke, (1997) Más allá de la Competencia Mínima. Evaluación para el Desarrollo Profesional. In: Millman, J. and Darling-Hammond, L (comps.) Manual para la Evaluación del profesorado. Ed. La Muralla. Madrid. Spain

Sistema de Evaluación del Desempeño Profesional Docente, Docentemás (2009) Secciones "Sistema de Evaluación" and "Instrumentos de Evaluación". Retrieved on Mar 12, 2009 from http://www.docentemas.cl

World Bank, (2002) Teacher Evaluation as Part of Quality Assurance. Retrieved on Jun 12, 2006 from http://www1.worldbank.org/education/globaleducationreform/default.htm