

## PROCEDURES FOR STUDY OF TEACHING PRACTICES. LITERATURE REVIEW

*[Procedimientos para el estudio de las prácticas docentes. Revisión de la literatura]*

by/por

[Article record](#)

[About authors](#)

[HTML format](#)

**Martínez-Rizo, Felipe** ([felipemartinez.rizo@gmail.com](mailto:felipemartinez.rizo@gmail.com))

[Ficha del artículo](#)

[Sobre los autores](#)

[Formato HTML](#)

### Abstract

Interest in studying teaching practices has increased, because of the need to evaluate teachers and dissatisfaction with the usual ways of doing it. Recent approaches with Value-Added Models, based on students' results on achievement tests do not seem satisfactory. The article is based on a review of literature and classifies the approaches to practices in three groups: instruments based on information given by teachers; observation protocols; and approaches based on the analysis of products of the practices. Specific tools are described and advantages and disadvantages of the three approaches are discussed.

### Resumen

El interés por las prácticas docentes ha aumentado, en parte por la necesidad de evaluar a los maestros y la insatisfacción con las formas usuales de hacerlo. Los enfoques basados en Modelos de Valor Agregado según resultados de los alumnos en pruebas estandarizadas tampoco satisfacen. Con base en la literatura analizada se clasifican los acercamientos a las prácticas en tres grupos: instrumentos basados en información dada por los maestros mismos, protocolos de observación y acercamientos basados en análisis de productos de las prácticas. Se describen herramientas particulares y se discuten ventajas y desventajas de los tres tipos de acercamiento.

### Keywords

Teacher Effectiveness; Instructional Effectiveness; Teacher Evaluation; Teacher Surveys; Vignettes; Classroom Observation Techniques; Portfolios (Background Materials); Classroom Research; Data Collection; Research Tools; Alternative Assessment.

### Descriptores

Efectividad docente; efectividad instruccional; evaluación de maestros; encuestas de maestros; viñetas; técnicas de observación en aula; portafolios; investigación en aula; recolección de datos; instrumentos de investigación; evaluación alternativa.

“*Teaching practices*” denote the set of activities undertaken by teachers, as part of their work in the classroom or in direct connection with it, so that students achieve learning purposes set out in curriculum. Such practices are a complex object of study, especially if, in addition to the behaviors they manifest, the researcher wants to analyze the underlying ideas and concepts, the factors

which influence the practices, or even the effects they produce.

In education systems there is interest in the subject, because of the need to have information to assess teacher performance with greater precision and objectivity that is often achieved with traditional approaches.

There are at least two reasons behind this interest: a) education systems consume increasing amounts of resources, teacher salaries are the main item in their budget and the trend is not expected to change soon, as schools are labor intensive services which employ many skilled workers; and b) the criteria traditionally used to define teachers' salaries (age, schooling) are not satisfactory and do not consider the very quality of teachers' work, which should be the focus.

The alternative based on students' scores in achievement tests to infer the quality of their teachers, through so-called Value-Added Models, seems attractive at first sight, since the final purpose of teaching is precisely that students learn something.

In practice, however, even the more complex statistical models give results too inaccurate to be useful to reliably support sensitive decisions about individual teachers and schools for various reasons, such as the number of factors that influence learning, limited tests' curriculum coverage, and the difficulty of having data from all the students and their teachers.

It is not easy to find alternative ways to reliably capture such a complex object as teachers' effectiveness. This paper presents three types of approaches that have pros and cons that the researcher should understand in order to select the best combination of tools suited to the purpose and circumstances of each study.

## **Approaches based on information given by the subjects**

### **Questionnaires and scales**

A questionnaire seeks information through the answers given to some questions by respondents. In projects about teaching practices the logical informers are the teachers, but principals, supervisors, students and parents can also provide useful information about specific aspects.

The rationale of using a questionnaire is that respondents know the information; that they are willing to provide it; and that they are able to understand questions consistently; the last condition in turn depends on the way to formulate and present the questions and, where appropriate, the response options.

The assumption about the knowledge by respondents of some information has to do, among other things, with the complexity of the information requested and whether it relates to current or to past events, whether recent or distant. Very important is the distinction between questions relating to facts, in contrast to questions about knowledge, subjective opinions or attitudes.

The assumption on the willingness of people to give certain information to anyone who requests it has to do with the nature of such information (public or private, more or less intimate, relating to legal or illegal behaviors, socially acceptable or not), and with the likelihood of confidentiality and/or anonymity.

These prerequisites are not dependent on the researcher, but on respondents in two ways: one referring to the respondents individually and the other taking them as a whole: some people are better able to report certain events, some are more willing to reporting on personal or sensitive issues, but there also are cultural contexts in which certain topics are taboo, while others can speak freely of them.

The third prerequisite, on how questions are framed, does depend on the researchers and affects the other two: the phrasing of a question and measures to ensure anonymity can make a question more or less understandable and more or less threatening. (See Converse and Prese, 1986, Fowler 1995; Wolf, 1991, Sudman and Bradburn, 1987)

Each requirement of a questionnaire involves particular issues that must be addressed, depending on the respondents. A

questionnaire for students should take special care of the clarity of the questions and extension, as long or complex instruments produce poor quality information, especially with young children. In the case of teachers and principals it is important to avoid socially desirable answers and to be aware of sensitive or threatening issues.

Researchers must check that respondents know what they are asked, which means taking into account the difference between real ignorance and the lack of interest for irrelevant information; to recognize that for many people it is difficult to understand large numbers, percentages, ratios, trends or prospective data; to consider the memory inaccuracy and frailty with old data; and the distinction between facts and opinions.

Willingness to give the requested information involves considering the influence of what is socially desirable; the threatening or sensitive nature of the subject; the trust about maintaining anonymity; the lack of interest that can stem from a long questionnaire or from inappropriate sequencing of questions; or phenomena that may occur unconsciously, as the tendency to the mean or the halo effect.

To be sure that respondents understand the questions researcher must consider the unavoidable ambiguity of all terms—stemming from cultural diversity and its different semantic universes—the complexity of the syntax that can obscure the meaning of the questions, for example by the use of subordinate clauses and double negatives, as well as the problems of extension and phrasing, etc.

With multiple choice questions the quality and completeness of options is crucial, with open response items the risk of vague questions is always present, as is the ambiguity of points of reference in quantity or intensity judgments in terms of many or few, often or seldom, etc.

It is not easy to be certain that a question is understood exactly in the same way by all respondents. If this is not achieved *they will not answer the same question*, and it will not be possible to use the information as a valid indicator of certain characteristics or behaviors of the respondents.

One type of questionnaire, a scale, intends to explore subjective attitudes and opinions rather than knowledge of objective facts. To know something about those hidden aspects of reality or latent constructs, one has to make inferences based on something that can be observed, such as verbal expression of opinions or attitudes, or the manifestation of behaviors that reflect this.

Even the person who has a certain attitude or feeling may be unable to express it verbally. When someone experiences a strong emotion it is usual to say that he or she *have not words* to express it. To look for information on these aspects with only a single question is unreliable because it is difficult to find a formulation that says exactly the same to all respondents. The idea behind the construction of scales has to do with this idea; the information derived from a set of questions is more reliable than the answer to a single one, provided the set of questions meets a basic condition: that all questions really relate to the same aspect of reality, the one *dimension* that the scale is intended to measure.

If all items of a scale belong to the same dimension, if they refer to manifestations of the same latent construct, each item will pick a different nuance of the construct and it will be more likely that the set of answers would correctly represent it. Factor analysis or item response models can be used to check if a scale is measuring a single dimension, and so measures which is not possible to capture in a direct way. (Cf. Morgenstern and Keeves, 1997; de Vellis, 1991).

Scales are not immune to systematic bias. A paradoxical finding of international evalu-

ation studies by OECD and IEA is that students from some countries with very low scores in reading, math or science literacy express attitudes toward reading, math or science supposedly much more positive than attitudes of students in countries with much better levels of cognitive competencies.

This casts doubt on the reliability of the scales used, as attitudes may be more susceptible to cultural influences than knowledge. In particular, one can hypothesize that specific traits of the cultures of some countries may lead many students to respond positively to questions that probe their attitudes, to a greater degree than occurs with the students from other cultures.

A paper on the PISA test results indicates that it has long identified the presence of certain characteristic patterns that tend to show responses to Likert questions:

*Unfortunately, there is an increasingly large body of evidence that suggests that many observed cross-national or cross-cultural differences are, in fact, contaminated by artifacts of measurement... Much of this research focuses particularly on cross-cultural differences in the usage of Likert scales or individual categorical items drawn from such scales. (Buckley, 2009: 5)*

Questionnaires and scales are a low-cost strategy for the study of teaching practices, but often produce poor quality information, usually due to deficiencies that could be avoided. However, recent work indicates that the quality of the information derived from surveys may be better than commonly thought. In one case a questionnaire has been used with a sample of algebra teachers, and a measure of the degree to which their practices were consistent with the standards of the National Council of Teachers of Mathematics was constructed.

A few previous studies assessed the reliability of the instruments item by item, and

validity by comparing two instruments based on teacher reports (a questionnaire and a log). In this study reliability was analyzed for the composite measure based on responses to 13 Likert items with six response options each, and the validity analysis by comparing teachers' responses to the questionnaire with the results of observations of their work in the classroom.

The consistency of the composite measure based on the 13 Likert items was adequate ( $\alpha = 0.85$ ). The coefficient of reliability of the composite measure obtained by comparing the results of the first application with a second made four months later was 0.69, and the correlation between this measure and another based on classroom observation of teachers themselves, was 0.85.

The above figures show that the quality of information obtained with the instrument answered by the teachers was quite acceptable, although other limitations of the instrument are still present, in particular that it only shows whether certain practices are used more or less often, but provides no information on how long each practice is used, and even less about the quality of the practice. (Mayer, 1999)

Teachers are not the only informants on teaching practices. School principals and supervisors are important sources in relation to aspects such as lesson planning, because normally their role includes supporting and monitoring teachers' work. In some countries it is also usual that these actors periodically observe teachers' classroom work and to assess it.

Students are not able to comment on teachers' lesson planning, or knowledge about the subjects they teach but, at least since the late elementary school grades, they are reliable to report on the topics covered, teaching strategies used and the feedback provided by teachers, among other topics. And to the extent that the teacher's role includes some relationship with parents to inform them about

the progress of their children, to know the problems they face at home or ask for cooperation, parents can also be valuable informants.

### Self-reports

A variant of the questionnaires, self-reports are provided by teachers themselves, as the term indicates. They can also be unstructured (in which teachers describe in their own words what they have done for a certain period) or structured (with predefined formats to report how often they carried out certain practices).

A limitation generally attributed to self-reports has to do with the risk that teachers will not report about what they really did, but on what they believe they should have done, the practices considered desirable in the profession.

A relatively old study explored the point, analyzing the quality of information reported by nine teachers of English, compared with reports by their students and with information obtained by observing their classes. The instrument used for self-reports of the teachers was highly structured, with 77 items describing as many practices, of which 37 related to the teaching of vocabulary and spelling, and 40 to grammar and syntax.

The teachers submitted six weekly self-reports and a summary at the end of the period. An instrument similar to the teachers' summary was applied to students at the end of the study period, and weekly non structured observations of a class taught by the teachers studied were done, which were encoded later.

Information on teachers' practices obtained through self-reports based on the described instrument showed reasonable internal consistency and was also consistent with the information given by the students and that obtained through classroom observations. (Koziol and Burns, 1986)

### Logs

Logs are a variant of self-reports: texts in which teachers describe their activities for some time. The difference is the frequency with which information is sought, to increase reliability, since it is less likely that respondents will distort reality if they report their activities several times and not only once. You can let each informant report on his/her activities freely or using a structured guide to do it. The pros and cons of these alternatives are the same as in any other case, but the workload involved for a teacher to keep a log open for many days is heavy, while filling a very structured format for the same period is easier.

This tool is used to study enacted curriculum and learning opportunity: the degree to which teaching covers the topics planned – the intended curriculum— which will in turn affect the curriculum achieved or accomplished. (Rowan, Camburn and Correnti, 2004; Correnti and Rowan, 2009)

The enacted curriculum is often explored with questionnaires applied at the end of the course in which teachers report retrospectively on the topics covered, or by observing a few classes. In both cases, however, the information may be of poor quality in comparison to a very complex behavioral universe:

*Over a nine-month academic year, the typical elementary teacher will usually work 140 or more days of class, with 20 or 30 students, sometimes with distinct activities for individual students or subgroups; in any day teaching activities will typically be conducted according to several dimensions, as a teacher usually cover several objectives with different levels of cognitive demand in a single day, working with different behavioral arrangements, using a variety of teaching techniques for each subject, some features repeated throughout the year, but not others, so that practices are multidimensional and highly var-*

*iable throughout the year* (According to Rogosa, Floden, and Willett, 1984, cited by Correnti and Rowan, 2009: 121).

Because of the variance of teaching practices, inter teachers and within teachers, inter days and inter subjects, a large sample of observations (15-30 for each teacher) would be needed in order to ensure a reasonable consistency of information, and the cost of such studies would rise accordingly, making attractive the option of logs (Rowan, Camburn and Correnti, 2004: 14-17).

By asking teachers to immediately report what they did at classroom on a single day, logs gain in reliability in comparison to surveys conducted once a year, as the problems of remembering the past are substantially reduced, as well as tendency to produce socially desirable reports. Logs for a number of days produce a sample more representative of the universe of reported practices in comparison to a few observations.

As previously stated, if the number of days to report increases the burden on teachers is greater. One way to encourage respondents is to offer a payment, as did Rowan and colleagues, who also made a phone number available free of charge for teachers participating in the project. As a result, response rates were 90% and data quality slightly lower than information derived from observations in the classroom, according to the researchers. (Rowan and Correnti, 2009: 122)

Scales were constructed to combine answers to several items. Analysis showed that 72% of the variation of teaching time devoted to reading was between days, 23% between teachers within schools, and only 5% between schools. Standard deviation of the distribution of time teaching reading between days was 45 minutes, so that in 15 out of 100 days the time teaching reading was actually zero minutes, although then daily mean was 80 minutes, not far from the intended time of 90. These data imply that to

have sufficient information on teaching practices with logs, this kind of reports are needed about 20 days a year (Correnti and Rowan, 2009: 123). A similar number of observations, much more expensive, would also be needed to have a sufficient sample of what happens in the classroom.

### **Blogs**

Blogs to gather information about teaching practices can significantly reduce the workload on teachers, as well as on researchers. Information can be recorded by teachers themselves, eliminating the need for researchers to subsequently do it, since it is stored in the system at the same time that is recorded. The methodology to design a log for collecting online information about teaching practices is no different from that needed for traditional pencil and paper instruments. There will always be the need of a conceptual framework to specify which practices will be included, identify their dimensions and define how they will be measured. The project from which these ideas about online teacher logs are derived was carried out around 1990, a pre-historical time for digital tools. The risks of its use with a large number of teachers in schools with different conditions of access to ICT prompted researchers to decide not to handle the online log at that time, but the authors considered it a promising tool for a future. (Ball et al., 1999)

### **Vignettes**

Items in a questionnaire can be understood differently. If terms specific to certain theories are used, these are familiar for researchers, but not always for teachers, students or other subjects, and this increases the risk that the understanding of the respondents does not match that of researchers. Examples abound in educational research. If you ask some teachers, for example, if they use collaborative work or formative evaluation, and they answer affirmatively, it is quite possible that at least some of them did not understand exactly the same as the researcher, which

seriously invalidates the conclusions to be drawn based on those answers.

This is the reason behind the development of *vignettes*, a variant of questions that, instead of asking for information on practices in abstract, theoretical wording, does it with accurate descriptions of specific behaviors, in context, asking respondents to indicate if their own work is similar to the one described in the vignette. This type of question has been used in studies about discriminatory attitudes or conceptions about work (cf. Martin et al., 1991; Martin, 2006), but its use in education is recent and there is little research on the quality of information that is gained.

A study by Stecher et al. (2006) approached the classroom practices of a group of teachers, looking at their consistency with a set of standards for innovative (reform oriented) teaching in mathematics and science, in contrast to traditional teaching in both areas. To validate the information obtained, researchers used a combination of traditional questionnaires, vignettes based questionnaires, teachers logs and classroom observations. Cognitive interviews were also conducted with a subsample of teachers in relation to the vignettes.

The preparation of vignettes began with operational definition of curriculum content and innovative teaching practices (reform-oriented), which produced a taxonomy of 23 items grouped into three categories: nature of mathematics, mathematical thinking of students and teaching of mathematics.

Elements that could be measured with vignettes were then identified; mathematical topics were selected (area, perimeter, and multiplication with two digit numbers); and four situations defined in which innovative practices can be present were identified: introduction of a lesson, response to student errors, reconciling different approaches and selection of learning objectives. All these elements were integrated into broader sce-

narios, which provided the context for the whole.

The vignettes section of the questionnaire started with a description of a scenario and instructions to respond. Each scenario included a context and situations, followed by options to act with innovative or traditional teaching practices, asking each respondent to indicate how likely was that he/she would act in the way described in each option.

The response options or possible actions were expressed in 51 vignettes: 27 of them, according to experts and teachers, described innovative practices (reform-oriented), while another 24 were related to traditional practices. Two methods were defined to assign an overall score for each respondent, placing it in a continuum innovative-traditional. Several tests were made to estimate reliability and validity of the measures obtained, contrasting it with the information obtained through traditional questionnaires, logs and, in particular, observations of classroom work.

The results support the idea that the information obtained with vignettes is of good quality in some dimensions, but not all. Cognitive interviews with participating teachers show that their interpretation of the descriptions of the vignettes did not always match that of researchers. Teachers were asked to read each possible action described in a vignette and express their understanding and their reasons for choosing one option, thinking aloud while doing so. It was found that in some cases teachers did not understand immediately what possible action meant, and that before choosing an answer they need to rephrase the idea in their own words.

Moreover teachers said that, in order to choose an option, it was necessary for them to know the performance level of the students with whom they were supposed to work, as the actions to be taken depend in good part of it.

One important conclusion is that preparing good vignettes involves a much greater amount of work in comparison to what is needed to make a traditional questionnaire (Stecher et al., 2006: 120).

### **Classroom observation**

The second group of approaches to the study of teaching practice is not based on what teachers inform, but on observation of such practices by third parties.

Half a century ago Medley and Mitzel said that *research on behaviors that occur in the classroom is not a pastime for amateurs, but a full-time occupation for technically competent professionals* (1963: 253). They defined the observation as a rigorous technique as follows:

*...an observational technique which can be used to measure classroom behavior is one in which an observer records relevant aspects of classroom behavior as (or within a negligible time limit after) they occur, with a minimum of quantification intervening between the observation of a behavior and the recording of it ... Schemes in which the classroom visitor is asked to rate the teacher, class or pupils on one or more "dimensions", even when the ratings are based on direct observation of specified behaviors, are not included in this definition. (Medley and Mitzel, 1963: 253)*

With a broader definition, observation procedures may include rating scales, and post-codificación recording techniques (e. g. video recordings) and qualitative techniques. Medley and Mitzel (1963) distinguished observing systems in time-based (categories) and event-based (signs).

A decade later Rosenshine and Furst (1973) classified observation techniques in three ways: according to the recording procedure: counting systems (categories or signs) and rating systems; according to specificity of the items: on very specific behaviors (low inference) or more general (high inference); and according to the coding system, of one or more dimensions.

Recently developed tools are based on current theories on the issues to observe and are more sophisticated in dealing with psychometric properties.

The following descriptions are limited to some of the most researched recently developed observation tools, as selected in the Project *Measures of Effective Teaching* (MET Project, 2010a and 2010c).

### **Classroom Assessment Scoring System (CLASS)**

Possibly the more researched observation instrument. It is intended to study the practices of teachers and the interactions they have with their students. Product of a work by more than a decade, led by Robert Pianta and his colleagues, a previous version was circulated under the name of Classroom Observation System, COS. (Pianta & Hamre, 2009; MET Project, 2010d)

The conceptualization of the activities and interactions distinguishes three domains: one for classroom organization, one for the instructional support offered to students, and one for emotional support, as synthesized in Table No. 1.



Table No. 1. Effective teaching dimensions assessed by the CLASS

	<b>Emotional Support</b>	<b>Classroom Organization</b>	<b>Instructional Support</b>
PreK-K3	Positive climate Negative climate Teacher sensitivity Regard for child's perspective	Behavior management Instructional productivity Learning formats	Concept development Quality of feedback Language modeling
K6-K9	Positive Climate Negative Climate Teacher Sensitivity Regard for teen's perspective	Behavior management Instructional productivity Learning formats	Understanding content Analysis & problem solving Quality of feedback Instructional dialogue

Source: MET Project, 2010d: 3

CLASS observation is carried out in half-hour periods. In each period 20 minutes are devoted to observe and take notes, and the remaining time to the rating of practices in the dimensions considered. Authors state that four observation cycles are enough to have a representative sample of what happens in a classroom.

CLASS was developed to observe practices at preschool or early grades of elementary school. A version for higher grades is being developed.

### Framework for Teaching (FFT)

Charlotte Danielson and her colleagues describe this observation tool, a development of ETS Praxis III, as

*...a research based protocol... aligned with the Interstate New Teachers Assessment and Support Consortium (INTASC) standards, which represents the professional consensus about what a beginning teacher should know... (MET project, 2010d; Goe, Bell and Little, 2008: 21 -22)*

The Framework has four domains: planning and preparing a class; classroom environment; instruction; and professional responsibilities. These domains are divided into 22 components and 76 indicators. By way of example, components and indicators of mastery for classroom environment domain are:

- Creating an environment of respect and good interaction:

- Teacher with students;
- Student with each other;
- Establishing a learning culture:
  - Importance of content;
  - Learning & achievement expectations;
  - Pride of the students for their work;
- Classroom management procedures:
  - Management of instructional groups;
  - Management of transitions between stages of the process;
  - Handling of support material;
- Management of student behavior:
  - Expectations;
  - Monitoring of student behavior;
  - Response or inappropriate behavior (MET Project, 2010b)

FFT includes detailed rubrics for observers to assess teachers in each of the 76 elements with four performance levels, defined as unsatisfactory, basic, proficient and distinguished. (MET Project, 2010d)

### Mathematical Quality of Instruction (MQI)

A research based observational instrument for the study of teaching practices in mathematics, MQI was developed by Heather Hill with colleagues at the University of Michigan and Harvard University. (Hill et al. 2010th and 2010b)

Five aspects were identified as most relevant:

- Richness of the mathematics, use of mathematical language;
- No errors or inaccuracies;
- Presence of correct mathematical explanations and student participation in the construction of meaning and mathematical reasoning;
- Connecting the classroom work with important ideas in the field, and
- Efforts to ensure that all students in the group, and not just some, have access to mathematical knowledge. (MET Project, 2010e)

MQI explores three types of relationships: of teachers with content, of students with content, and of teachers with students. In assessing these dimensions, MQI stands out among the instruments in the area of mathematics because it offers a complete and balanced view of the elements that, together, make quality education in the area (MET Project, 2010e).

During the development of MQI was found that, in addition to a mastery of general mathematics, teachers must master a particular type of mathematical knowledge that is needed for the teaching of this area, which has to do with an understanding of the barriers that make students learning in this field so difficult.

As a companion to MQI, Hill et al. (2008) developed a tool for measuring that kind of knowledge: *Mathematical Knowledge for Teaching* (MKT) (MET Project, 2010e, Hill et al., 2008).

## PLATO

Another specialized tool for a particular curriculum area is the *Protocol for Teaching Language Arts Observation*, PLATO.

Also research based, PLATO is structured around four factors underlying teaching: cognitive demand that the area poses to classroom practice and discourse; scaffolding to support the teaching of language; repre-

sentations of contents and use made of them; and classroom environment (MET Project, 2010f).

The system identifies 13 elements that are independent dimensions, and a rubric designed to assess them, on a scale of one to four:

Purpose	Guided practice
Intellectual challenge	Classroom discourse
Representation of content	Text-based instruction
Connections to prior knowledge	Academic language
Connections with personal experience	Behavior management
Modeling	Time management
Explicit teaching strategies	

PLATO collected data are based on independent observations of 15 minutes each, during a class: two in classes of 45 minutes and three in classes of 90 minutes. The research to validate PLATO is not as large as in the previous cases, and has been done mainly in schools of New York (MET Project, 2010d and 2010f).

## Quality of Science Teaching (QST)

An observation protocol that is being developed as part of the project *Measures of Effective Teaching* (MET).

The starting point for the development of QST is *The Teaching Event*, a system to assess candidates to fill teaching positions in California.

QST includes information about 13 areas of teaching practices, through a portfolio with evidences that include lesson plans, a video, samples of student work, and a reflection made by the candidate himself.

The dimensions considered are five: planning, teaching, assessment, reflection and academic language. QST developers expect to derive constructs for developing protocols and coding guidelines to observe classroom practices. (MET Project, 2010d)

### **Videotaped observations**

A special type of observation is one that is not done directly in classroom, but on a deferred basis, based on records of practices that have sufficient detail to allow coding of recorded behaviors, for instance with video recordings.

The advantage of this type of work – particularly in large scale, involving tens or hundreds of schools in different regions or countries— is that in many cases it is more feasible to find technicians who operate a video camera in an efficient way, than observers with the necessary qualifications to obtain valid and reliable information. One special case is the use of video recordings of classes conducted by teachers. An outstanding example is the study of videotaped classes that was part of the *Third International Mathematics and Science Study* (TIMSS) in 1995, and its repetition in 1999 (TIMSS-R). (Stigler, Gallimore, & Hiebert, 2000)

A clear advantage of video recording is that it is possible to stop or playback the tape as it is coded, and also to recode with different raters and using different protocols or scales.

In the case of TIMSS and TIMSS-R only one class per teacher was recorded, which obviously is not enough to have a good sample of his/her way of teaching throughout a school year. Moreover, video recordings, such as direct in vivo observation, can't grasp teachers' or students' thinking, which is needed in order to fully understand the meaning of their actions. These "subjective" aspects should be addressed with other approaches.

A disadvantage of video recording is derived from the limited angle that a conventional camera can capture, with many things going on in the classroom not being recorded, occurring out of that narrow field of vision. The way to overcome this limitation is to simultaneously use multiple cameras. In

addition to increasing costs, this solution also increases the interference caused by the observation, which puts at risk the validity.

Recently the use of a digital video recording is explored with advanced cameras that can record with a panoramic view of 360°, which allows as many playbacks as desired, with a focus on different sections of the classroom. Recognizing the advantages of today's digital technologies, we must recall that its use does not obviate the need for qualified observers and for coding schemes in order to organize the information recorded according to relevant dimensions.

To properly assess the potential of approaches to teaching practice based on different types of video recording it would be necessary to scrutinize in detail the pros and cons, costs and benefits of different options, from multiple video cameras, microphones and cameramen, to 360° digital recording whose effectiveness has yet to be demonstrated, to see if it can effectively produce better information at a lower cost or not.

### **Approaches based on the analysis of products of the practices**

The approaches to the study of teaching practices reviewed in this section are based on the analysis of products of the practices, such as lesson plans, student workbooks, tests results or home work assignments. It is, therefore, a third category of techniques, different from those based on reports by teachers themselves, and those based on observation.

### **Teachers' Assignments and Student Work**

A first approach of this group is the analysis of the work that students perform after being assigned by the teacher. The scrutiny of this kind of work can reveal many things about the practice of the teacher, like his/her own understanding of the teaching role; his/her mastery of some content; his/her concept of assessment; the kind of feedback

he/she provides to students, among others. Therefore, it seems logic that an early example of this type of approach is subtitled "Opening a Window on Classroom Practice." (Matsumura and Pascal, 2003)

The examples reviewed in this section define the quality of teachers' assignments and student work according to three dimensions: the level of cognitive demand they represent; the clarity of learning objectives to which a particular assignment is supposed to contribute; and how the rating criteria are specified. Each assignment receives one to four points on a scale for each dimension, and the scores are combined to form an overall quality score. (Matsumura and Pascal, 2003)

181 teachers of 35 schools (grades 4°, 7° and 10 °) participated in the study; from each one samples from three assignments were collected (two on reading comprehension and one on writing), with information about planning, instructions given to students, work done by four students (two high and two low performers). Teachers' performance was observed twice. Materials collected were analyzed and the results synthesized to form indicators of teaching practice. An acceptable level of consistency among analysts was found. To have consistent estimates for each teacher it was necessary to analyze the work from three to four students and only if work has been designed by the teacher and not taken from another source.

The quality of the assignment was associated with that of teaching practice, according to classroom observations, and also with the quality of student work. Those students whose teacher assigned cognitively demanding work, and presented clearer scoring criteria, showed also more progress in external assessments. However, the overall quality of the work assigned by teachers was not very high. (Matsumura and Pascal, 2003)

### **Portfolio and its variants**

Portfolios are tools to evaluate students, teachers, schools etc. Teachers' portfolios

consist of sets of evidence such as lesson plans, student workbooks and school records. The portfolios contain materials selected by the subject being assessed, who is expected to engage in self-assessment; the materials included will then be assessed by authorities, peers or other external reviewers.

As occurs with any data collection instrument, when portfolios are used to observe/measure/evaluate teaching practices it is necessary to specify the focus of interest, in this case the relevant aspects of the practice to observe. The categories to be used to systemize data collected should also be made explicit, as well as the way to proceed from measure to value judgment.

There are outstanding examples of teachers' evaluation systems based on portfolios. ETS Praxis system includes a portfolio, as well as tests in reading, writing and math, pedagogical skills and content areas.

The extent to which teachers meet the standards of professional practice developed by the *Interstate New Teacher Assessment and Support Consortium* (INTASC) is valued with a portfolio, as is the case for the standards of the *National Board for Professional Teaching Standards* (NBPTS). (Porter, Youngs & Odden, 2001: 263-265)

In Latin America an outstanding teacher evaluation system has been developed by Chile's Ministry of Education, within a larger set of policies for improving the quality of education that recognize the importance of teachers' professional development.

The development of the *System for Teachers' Performance Appraisal* (*Sistema de Evaluación del Desempeño Profesional Docente*, SEDPD) began in 2003, starting with a set of standards that provide a *framework for good teaching*, prepared during the previous two years. The technical work was entrusted to a prestigious independent agency, the Center for Measurement of the Catholic University of Chile (MIDE), working

with the *Center for Training, Experimentation and Pedagogical Research* of the Ministry of Education.

SEDPD defines its core elements: the purpose and content of the evaluation, what implications are for teachers, who should be assessed, and what tools and information sources should be used. Evaluation instruments are four: a self-evaluation; an interview by a peer evaluator; a benchmarking report from third parties (the principal and the person in charge of the technical unit at the school); and a portfolio of educational performance. For each one there are well-defined guidelines that are applied by trained personnel.

Because of the careful design and validation of the instruments, Chile's system is a model in Latin America, and compares favorably with similar evaluation models in many education systems of more developed countries. A recent book presents extensive information on Chile's teachers' evaluation system development and its results, after several years of implementation. (Manzi, Gonzalez and Sun, 2011)

Two other examples of portfolio are as follows.

#### *Instructional Quality Assessment (IQA)*

Developed from previous research by Matsumura on teachers' assignments and student work, as a strategy to explore teachers' practices, this new system is under development and focuses on reading and math, at high school level.

The system includes protocols for assessing teaching practices with classroom observations, as well as with analysis of the quality of student work assigned by teachers. There are reports on reliability and validity of the information obtained, with results consistent with previous research, with findings about a wide variation in the quality of teaching in the areas studied, and an average level not too high. The quality of information was

better in math than in reading. The results of the IQA predict those obtained by students in external assessments, after controlling for other factors possibly associated. The robustness of the results is limited because they are derived from a small number of cases. (Matsumura et al., 2006)

#### *Artifacts Packages*

This type of instrument is a variant of portfolios and is not limited to the review of teachers' assignments, but includes any other material (artifact) that can provide information on practices that take place in the classroom: lesson plans, teacher handouts given to students, photographs of how are organized students to work in the classroom, or of blackboard notes, student work, whether or not assigned by the teacher, video recording of classroom sessions, and so on.

To designate a notebook or folder in which materials collected for further analysis are integrated, proponents of this approach use the expression "*scoop notebook*" as a reminiscence of a biologist who first spends some time doing field work, collecting as many specimens as possible with a net (scoop), then carefully analyzing them in the laboratory. In a similar way, researchers interested on teaching practices may first devote time to collect all kind of evidences (materials, artifacts) of teachers' and students' work, and then study them in detail.

The purpose of the Scoop Project was to develop an alternative approach to the study of teaching practice, using artifacts and materials in order to represent the practice with details enough to make valid judgments about it, based solely on these materials without directly observing the teacher and the classroom.

In addition to precise instructions for participant teachers on how to gather materials, other key components of the system are the scoring guides that raters use to analyze them. The work reviewed refers in particular

to the areas of math and science. (Borko, Stecher and Kuffner, 2007)

Based on previous research, ten dimensions of teaching practices in the two areas were identified; in each dimension it is possible to appreciate if there is consistency with guidelines for innovative teaching, according to national standards for math and science. In the case of mathematics, the ten dimensions are:

- Types of grouping used for student work.
- Structure of the class.
- Use of different ways to represent ideas and concepts.
- Use of mathematical tools.
- Cognitive demand.
- Presence of a common mathematical discourse.
- Explanations and justifications.
- Problem solving.
- Forms of assessment.
- Connections with real life or application of knowledge.

The dimensions for science are similar and a rubric is provided for each one, with a precise definition of practices of high, medium or low quality.

The guidelines developed can be used either for classroom observation or to analyze materials collected by teachers in the "scoop-notebook" during a week, seeking representativeness in a subject. Teachers are asked to include their point of view and reflections about the materials, the purpose and use of each one, and so on. The materials include those used before class (plans, notes, rubrics to assess), during class (photos of the chalkboard, slides or power points, student work) and after (homework, tests, portfolio items).

Although this is a work in progress, the results show that judgments about the teaching practice based on the system have good lev-

els of reliability and validity. ((Borko, Stecher and Kuffner, 2007; Borko et al. 2005)

### Remarks

As those listed in previous sections, the tools reviewed in this one have both, strengths and weaknesses, so it is better to see them as complementary rather than as alternatives.

The use of products of teaching practices as sources of information has in common with observations the advantage that they are not affected by the risk of picking up more the socially desirable than the real. As occurs with recorded observations, products of practices can be reviewed again and again, by different analysts and using different protocols. These tools also share with observations the inability to grasp what was going on in the minds of teachers that performed the activities from which products reviewed were derived; those thoughts can only be explored via the version given by actors themselves with approaches of the first of our three groups.

In addition to the tools reviewed, information about many others can be found in the "*Product Guide for Teacher Evaluation*" (NCCTQ, 2011, [www.tqsource.org](http://www.tqsource.org)), which includes measuring instruments, and other tools, as follows:

Classroom observation: 41	Artifacts of Teaching: 8
Portfolios: 6	Teachers Self-Reports: 4
Students' Surveys: 3	Value-Added Models: 6
Measures of Student Performance: 4	Combined Models: 21

### Conclusion

Dissatisfaction with teacher evaluation systems—either traditional, based on indicators such as schooling and seniority, or more recent that intend to measure the value that is supposedly added to student learning by teachers and schools—explains the interest in finding better ways to observe and/or measure effective instruction, especially for purposes of summative assessment and ac-

countability, particularly as a part of compensation systems.

It is important to underline the possible formative dimension of teacher evaluation, if particular deficiencies of teaching practice are identified in order to better inform the design of professional development activities and other interventions which will improve the practice of less experienced and/or less effective teachers.

In order to develop valid and reliable summative evaluations of the effectiveness of teachers, or to have good formative evaluations that will help them to improve their practice, observations/measurements of good quality are needed.

The use of questionnaires and similar tools can be inexpensive, but the information provided may be superficial and of low reliability, as it relies on informants; however, these approaches are the only able to capture what the actors know, believe or think. Observations can provide rich information, but involve time and skilled observers and can't grasp what the subjects think. The analysis of products of the practices shares some of the pros and cons of the other two approaches.

No approach in itself is better or worst; the quality of information that each one provides depends on the purpose of the study, the quality of the instrument and how it is applied and the results are analyzed.

### **Qualitative/quantitative opposition**

The distinction between quantitative and qualitative techniques is often understood as two totally different types of approach to reality, the first supposing categories and coding, and the second ruling them out.

Actually any technique involves encoding information, what changes is the moment to do it: in approaches usually considered qualitatives (participant observation and similar) coding is done and categories are constructed *after* collecting information; when using

questionnaires categories to encode information are defined since the development of the instrument.

Coding is always present. Humans can't grasp the reality in holistic ways. We always know analytically, with categories. By applying techniques of thinking aloud to detect the way of doing of a qualitative observers, it is possible to find that they actually use categories that, not being explicit, increase the risk of inconsistency even if the same observer works at different times, in addition to the impossibility of comparing information recorded by different observers.

It is impossible to observe something without categories. The real difference is whether the categories used are explicit or implicit; whether they are made explicit before or after observing; whether categories are well or ill-defined; whether they are exclusive or not; if they are to form a coherent and consistent scale, etc.

The advantage of working with pre-defined categories is clear, but there is always the corresponding disadvantage: things that do not fit into the categories go unnoticed. Therefore it is crucial how they are built. Not to define categories at the beginning has the advantage of not excluding a priori things that do not conform to them, which may be appropriate in an exploratory stage, but the cost to pay is high: as each observer applies his/her implicit categories, the resulting information is heterogeneous in unknown degree, making impossible any rigorous analysis. Post-encoding previous records faces exactly the same problems.

The quality of the information obtained with a structured approach depends on the underlying conceptualization of the instrument; the quality of information derived from an unstructured approach depends on the capacity of the observers. But finally the ability of these to make high-inference judgments on the fly, or to postcode based on the records, depends in turn on the mastery

of common categories to ensure uniform application.

### **Rubrics as tools for observation, measurement and/or evaluation**

Another well-known tool is a rubric, which is discussed here because the principles involved in its construction can be generalized to any instrument for measurement, monitoring or evaluation.

In Spanish, the term rubric denotes someone stamped signature to a written piece, but in the sense that concerns us, a rubric is a framework that teachers use for grading papers or essay-type assignments. To grade an extended response test, for instance, a rubric may take into account the scheme that organizes the work, grammar, spelling, clarity of ideas, and so on.

More appropriate is the translation in Spanish of rubric as “evaluation framework”, expression that underlines a basic feature of such tools, which is reflected in its matrix structure, double entry table, with rows and columns:

.. One axis contains aspects or dimensions of the object of study. Usually these issues are included in the rows of the table, each of which has a dimension of the reality to assess, for example wording, or spelling, etc.

.. The other axis contains the performance levels in each dimension, with an order from best to worst: the columns present a different gradation of quality performances, which are identified with labels such as unsatisfactory, sufficient, advanced, outstanding, and the like.

A rubric is intended to be a measurement tool, which is a prerequisite to be also an evaluation tool. Its principles of construction are applicable to any instrument.

Reality is always multidimensional, but a good measurement should refer to a well-defined dimension of reality in order to meet

the basic requirement of one-dimensionality. Developing a rubric or any measurement instrument involves a process of conceptual clarification, identification and distinction of the dimensions of reality to measure and evaluate (operationalization).

To evaluate something we need to measure it at least at an ordinal level. Accordingly, to develop a rubric or other assessment tool it is necessary to define performance levels, ordinal measures based on observation of traits associated with each level.

### **Final Thoughts**

To have good quality information by means of observation or measurement, three steps of the process have to be done:

- Conceptual work to define categories to classify the captured realities and performance levels to be considered for measurement.
- Technical work with appropriate psychometric tools for assuring validity and reliability.
- Empirical work –with conceptual frameworks and psychometric tools— to calibrate the instruments and assess the quality of the information obtained through their use.

It is not feasible for all researchers to master the most advanced psychometric techniques, but all should do a careful conceptual and empirical work.

Besides that, the pros and cons of techniques involving a priori or a posteriori encoding suggest that it is better to use a combination of both approaches, indeed the idea behind the appeal of so-called mixed methods. This suggestion makes sense, but it also represents a risk: a combination of approaches often gives richer information than each one separately, but at the same time increases the risk of the results not being consistent. In the words of Stecher:



*While educators are often encouraged to use multiple measures to provide more reliable information, researchers should be aware that multiple formats and multiple respondents can reduce consistency of responses by adding additional sources of variation.* (Stecher et al., 2006: 120).

## References

- Ball, D. L., Capurn, E., Correnti, R., Phelps, G. & Wallace, R. (1999). *New Tools for Research on Instruction and Instructional Policy: A Web-based Teacher Log*. Seattle: Center for the Study of Teaching and Policy, University of Washington.
- Borko, Hilda, Stecher, B. y Kuffner, K. (2007). *Using Artifacts to Characterize Reform Oriented Instruction: The Scoop Notebook and Rating Guide*. (CSE Technical Report 707). Los Angeles: UCLA.
- Borko, H., Stecher, B. M., Alonzo, A., Moncure, S., & McClam, S. (2005). Artifact Packages for Characterizing Classroom Practice: A pilot Study. *Educational Assessment*, Vol. 10 (2): 73-104.
- Buckley, Jack (2009). *Cross-national Response Styles in International Educational Assessments: Evidence from PISA 2006*. Original no publicado.
- Converse, J. M. & S. Presse. (1986). *Survey questions. Handcrafting the standardized questionnaire*. Series Quantitative Applications in the Social Sciences N° 63. Beverly Hills. Sage.
- Fowler, Floyd J. (1995). *Improving survey questions. Design and evaluation*. Applied Social Research Methods Series N° 38. Newbury Park: Sage.
- Goe, Laura, C. Bell & O. Little (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. Washington: NCCTQ.
- Hill, H. (2010a). *Mathematical Quality of Instruction (MQI). Coding Tool*. University of Michigan, Learning Mathematics for Teaching.
- Hill, Heather C. Ball, D. L., Bass, H., Blunk, M., Brach, K. et al. (2010b). Measuring the mathematical quality of instruction. *Journal of Mathematics Teacher Education*. Springer.
- Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G., Sleep, L., et al. (2008). Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An Exploratory Study. *Cognition and Instruction*, Vol. 26 (4): 430-511.
- Hill, Heather C., Stephen G. Schilling y Deborah L. Ball (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, Vol. 105: 11-30.
- Koziol, S. M. & Burns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *Journal of Educational Research*, 79(4): 205-209.
- Manzi, Jorge, R. González & Sin Y. (2011). *La evaluación docente en Chile*. Santiago: MIDE-Universidad Católica.
- Martin, Elizabeth (2006). *Vignettes and Respondents Debriefing. for Questionnaire Design and Evaluation*. Washington. U. S. Bureau of Census. Research Report Series, Survey Methodology N° 2006/8.
- Martin, Elizabeth A., Campanelli, P. C. & Fay, R. E. (1991). An application of Rasch analysis to questionnaire design: using vignettes to study the meaning of "work" in the Current Population survey. *Journal of the Royal Statistical Society. Series D. Special Issue*. Vol. 40 (3): 265-276.
- Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M. (2006). *Measuring Reading Comprehension and Mathematics Instruction in Urban Middle Schools: A Pilot Study of the Instructional Quality Assessment* (CSE T. Report 681). Los Angeles: UCLA.
- Matsumura, Lindsay C. y J. Pascal (2003). *Teacher's Assignments and Student Work: Opening a Window on Classroom Practice*.

- (CSE Technical Report 602). Los Angeles: UCLA.
- Mayer, D. (1999). Measuring Instructional Practice: Can Policy Makers Trust Survey Data? *Educational Evaluation and Policy Analysis*, 21(1): 29-45.
- Medley, Donald M. & Harold Mitzel (1963). Measuring classroom behavior by systematic observation. Gage, N.L. Ed. *Handbook of Research on Teaching*. Chicago. Rand McNally, pp. 247-328.
- MET Project (2010a). *A Composite Measure of Teacher Effectiveness*. MET Project Research Paper. Bill & Melinda Gates Foundation.
- MET Project (2010b). *Danielson's Framework for Teaching for Classroom Observations*. Bill & Melinda Gates Foundation.
- MET Project (2010c). *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*. MET Project Research Paper. Bill & Melinda Gates Foundation.
- MET Project (2010d). *Overview: Teacher Observation Rubrics*. Measures of Effective Teaching. Teachscape.
- MET Project (2010e). *The MQI Protocol for Classroom Observations*. Bill & Melinda Gates Foundation.
- MET Project (2010f). *The PLATO Protocol for Classroom Observations*. Bill & Melinda Gates Foundation.
- MET Project (2010g). *Validation Engine for Observational Protocols*. Bill & Melinda Gates Foundation.
- Morgenstern, C. & J. P. Keeves (1997). Descriptive scales. En J. P. Keeves, *Educational research, methodology & measurement*, 900-908. Oxford: Elsevier
- NCCTQ (2011). *Guide to teacher evaluation products*. National Comprehensive Center for Teacher Quality. Recuperado el 18 de enero de 2011, de [www.tqsource.org](http://www.tqsource.org).
- Pianta, Robert C. & Hamre, Bridget K. (2009). Conceptualization, Measurement and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, Vol. 38 (2): 109-119
- Porter, Andrew C., Youngs, P. & Odden, A. (2001). Advances in teacher assessments and their uses. En Virginia Richardson (ed.), *Handbook of Research on Teaching*, pp. 259-297. Washington: AERA.
- Rogosa, D., R. Floden & J. B. Willet (1984). Assessing the stability of teacher behavior. *Journal of Educational Psychology*, 76, 1000-1027.
- Rosenshine, Barak & Norma Furst (1973). The use of direct observation to study teaching. En Travers, Robert M. W. Ed. *Second Handbook of Research on Teaching*, pp. 122-183. Chicago: Rand McNally College Publ. Co.
- Rowan, Brian & Correnti, R. (2009). Studying Reading Instruction With Teacher Logs: Lessons From the Study of Instructional Improvement. *Educational Researcher*, Vol. 38 (2): 120-131.
- Rowan, Brian, Camburn, E. & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: a study of literacy teaching in third-grade classrooms. *The Elementary School Journal*, 105, 75-102.
- Stecher, B., Le, V., Hamilton, L., Ryan, G., Robyn, A., & Lockwood, J.R. (2006). Using Structured Classroom Vignettes to Measure Instructional Practices in Mathematics. *Educational Evaluation and Policy Analysis*, Vol. 28 (2): 101-130.
- Stigler, James W., Gallimore, R. y Hiebert, J. (2000). Using Video Surveys to Compare Classrooms and Teaching Across Cultures: Examples and Lessons from the TIMSS Video Studies. *Educational Psychologist*, Vol. 35 (2): 87-100.
- Sudman, S. & N. M. Bradburn (1987). *Asking questions. A practical guide to questionnaire design*. San Francisco: Jossey-Bass.
- De Vellis, Robert F. (1991). *Scale development. Theory and applications*. Applied Social Research Methods Series, Vol. 26. Newbury Park: Sage.

Wolf, R. M. (1991). Cuestionarios. En T. Husén y T. N. Postlethwaithe (Eds.), *Enciclopedia Internacional de la Educación*,

Vol. 2: 1002-1006. Barcelona: Vicens Vives.

**ABOUT THE AUTHORS / SOBRE LOS AUTORES**

**Martínez-Rizo, Felipe** ([felipemartinez.rizo@gmail.com](mailto:felipemartinez.rizo@gmail.com)). Licenciado en Ciencias Políticas y Sociales por Universidad De Lovaina (Bélgica) y profesor en la Universidad Autónoma de Aguascalientes (México), de la que llegó a ser Rector. Ha sido el fundador (y director desde 2002 hasta 2008) del *Instituto Nacional para la Evaluación de la Educación* (INEE) de México. Miembro de la Academia Mexicana de Ciencias, del Consejo Mexicano de Investigación Educativa (México) y del *National Council for Educational Measurement* (USA). Recibió en 2011 un doctorado *honoris causa* por la Universidad de Valencia (España). Su dirección postal es Depto. de Educación, Edif. 13. Av. Universidad N° 940. C. P. 20100 Aguascalientes, Ags. (México). [Buscar otros artículos de este autor en Google Académico / Find other](#)

[articles by this author in Scholar Google](#) 

**ARTICLE RECORD / FICHA DEL ARTÍCULO**

<b>Reference / Referencia</b>	Martínez-Rizo, Felipe (2012). Procedures for study of teaching practices. Literature review. <i>RELIEVE</i> , v. 18, n. 1, art. 1. <a href="http://www.uv.es/RELIEVE/v18n1/RELIEVEv18n1_1eng.htm">http://www.uv.es/RELIEVE/v18n1/RELIEVEv18n1_1eng.htm</a>
<b>Title / Título</b>	Procedimientos para el estudio sobre las prácticas docentes. [ <i>Procedures for study of teaching practices. Literature review</i> ].
<b>Authors / Autores</b>	Martínez-Rizo, Felipe
<b>Review / Revista</b>	RELIEVE (Revista Electrónica de Investigación y Evaluación Educativa), v. 18, n. 1
<b>ISSN</b>	1134-4032
<b>Publication date / Fecha de publicación</b>	2012 ( <b>Reception Date</b> : 2011 November 11 ; <b>Approval Date</b> : 2012 May 23. <b>Publication Date</b> : 2012 May 23).
<b>Abstract / Resumen</b>	<i>Interest in studying teaching practices has increased, because of the need to evaluate teachers and dissatisfaction with the usual ways of doing it. Recent approaches with Value-Added Models, based on students' results on achievement tests do not seem satisfactory. The article is based on a review of literature and classifies the approaches to practices in three groups: instruments based on information given by teachers; observation protocols; and approaches based on the analysis of products of the practices. Specific tools are described and advantages and disadvantages of the three approaches are discussed.</i> El interés por las prácticas docentes ha aumentado, en parte por la necesidad de evaluar a los maestros y la insatisfacción con las formas usuales de hacerlo. Los enfoques basados en Modelos de Valor Agregado según resultados de los alumnos en pruebas estandarizadas tampoco satisfacen. Con base en la literatura analizada se clasifican los acercamientos a las prácticas en tres grupos: instrumentos basados en información dada por los maestros mismos, protocolos de observación y acercamientos basados en análisis de productos de las prácticas. Se describen herramientas particulares y se discuten ventajas y desventajas de los tres tipos de acercamiento.
<b>Keywords / Descriptores</b>	<i>Teacher Effectiveness; Instructional Effectiveness; Teacher Evaluation; Teacher Surveys; Vignettes; Classroom Observation Techniques; Portfolios (Background Materials); Classroom Research; Data Collection; Research Tools; Alternative Assessment.</i> Efectividad docente; efectividad instruccional; evaluación de maestros; encuestas de maestros; viñetas; técnicas de observación en aula; portafolios; investigación en aula; recolección de datos; instrumentos de investigación; evaluación alternativa..
<b>Institution / Institución</b>	Universidad Autónoma de Aguascalientes (Mexico).

<b>Publication site / Dirección</b>	<a href="http://www.uv.es/RELIEVE">http://www.uv.es/RELIEVE</a>
<b>Language / Idioma</b>	Español & English version (Title, abstract and keywords in English & Spanish)

## RELIEVE

**Revista ELectrónica de Investigación y EValuación Educativa**  
*E-Journal of Educational Research, Assessment and Evaluation*

[ISSN: 1134-4032]

© Copyright, RELIEVE. Reproduction and distribution of this articles it is authorized if the content is no modified and their origin is indicated (RELIEVE Journal, volume, number and electronic address of the document).

© Copyright, RELIEVE. Se autoriza la reproducción y distribución de este artículo siempre que no se modifique el contenido y se indique su origen (RELIEVE, volumen, número y dirección electrónica del documento).