



RELIEVE. Revista Electrónica de Investigación  
y Evaluación Educativa

E-ISSN: 1134-4032

relieve@uv.es

Universitat de València  
España

Martínez-Rizo, Felipe

PROCEDIMIENTOS PARA EL ESTUDIO DE LAS PRÁCTICAS DOCENTES. REVISIÓN DE LA  
LITERATURA

RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa, vol. 18, núm. 1, 2012, pp. 1-  
22

Universitat de València  
Valencia, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=91624440001>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

## PROCEDIMIENTOS PARA EL ESTUDIO DE LAS PRÁCTICAS DOCENTES. REVISIÓN DE LA LITERATURA

*[Procedures for study of teaching practices. Literature review]*

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.

### Keywords

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

### 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.

### 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.

La expresión *prácticas docentes* denota el conjunto de actividades que llevan a cabo los maestros, como parte de su trabajo en el aula o en relación directa con él, con el propósito de que los estudiantes alcancen los propósitos de aprendizaje establecidos en planes y programas de estudio. Dichas prácticas son un objeto de estudio complejo, en especial si,

además de las conductas en que se manifiestan, se quiere analizar las ideas y concepciones subyacentes que las explican, los factores que inciden en ellas o, incluso, los efectos que producen

En muchos sistemas educativos hay interés por el tema, en parte sin duda por la preocu-

pación por contar con información sobre los maestros que permita evaluar su desempeño con mayor precisión y objetividad que la que se suele alcanzar con los acercamientos tradicionales.

La educación consume cantidades importantes y crecientes de recursos, y los salarios de los maestros son el principal renglón del gasto. El carácter de *servicio intensivo en mano de obra calificada* que tiene la educación explica lo anterior y hace previsible que la tendencia no cambiará pronto. Por otra parte, los criterios tradicionalmente usados para definir el salario de los maestros no resultan satisfactorios —antigüedad, escolaridad— ya que no consideran la calidad misma del trabajo de los maestros, que debería ser el criterio fundamental.

La alternativa de partir de los puntajes obtenidos por los alumnos en pruebas de rendimiento para inferir la calidad del trabajo de sus maestros, mediante los llamados Modelos de Valor Agregado, parece atractiva a primera vista, ya que el sentido mismo del trabajo docente es, precisamente, que los alumnos desarrollen ciertos conocimientos y competencias.

En la práctica, sin embargo, incluso los modelos estadísticos más complejos dan resultados demasiado imprecisos para ser útiles para sustentar por sí mismos decisiones delicadas sobre maestros y escuelas individuales, por varias razones, como la diversidad de factores que influyen en el aprendizaje, el limitado campo del currículo que cubren las pruebas, y la dificultad de contar con bases de datos completas de los alumnos y sus maestros.

No es fácil encontrar alternativas adecuadas para captar adecuadamente un objeto de estudio tan complejo; los tres tipos de acercamientos que se presentarán en este trabajo tienen ventajas y limitaciones que el investigador interesado en el tema deberá tener claras, para seleccionar la combinación más

adecuada de herramientas, dados los propósitos y las circunstancias de cada estudio.

## **Acercamientos basados en información dada por algunos sujetos**

### **Cuestionarios y escalas**

Un cuestionario busca información a través de las respuestas que se den a las preguntas que los integran. En estudios de las prácticas docentes los interrogados podrán ser los maestros mismos, los directores o supervisores, los alumnos o los padres de familia. Cada sujeto podrá informar sobre distintos aspectos.

El uso de cuestionarios supone que los interrogados conocen la información, que están dispuestos a darla y que son capaces de entender de manera uniforme las preguntas que se les hagan, lo que depende a su vez de la manera en que se formulen y presenten las preguntas y, en su caso, las opciones de respuesta.

El supuesto sobre el conocimiento de la información por parte de los interrogados tiene que ver, entre otras cosas, con la complejidad de la información pedida y si se refiere a hechos actuales, recientes o de un pasado lejano. Muy importante es la distinción entre preguntas relativas a hechos o bien a conocimientos, opiniones subjetivas o actitudes. El supuesto relativo a la disposición de las personas a dar cierta información a quien la solicite tiene que ver con la naturaleza de dicha información (pública o privada, más o menos íntima, relativa a conductas lícitas o ilícitas, socialmente aceptables o no), así como a la probabilidad de que se difunda o la seguridad de que tal cosa no ocurrirá (anónimo, confidencialidad).

Estos prerrequisitos no dependen del investigador, sino de los sujetos que cubrirá el estudio, en dos sentidos: uno referido a los respondientes en lo individual y otro considerándolos en conjunto: unas personas son más capaces de informar sobre ciertos fenómenos; unas están más dispuestas a informar

sobre temas delicados o personales, pero además hay contextos culturales en los que ciertos temas son tabú, mientras que en otros se puede hablar libremente de ellos.

El tercer prerrequisito, sobre la manera en que se formulen las preguntas, sí depende de los investigadores y afecta a los otros dos: la formulación de una pregunta y la forma en que se asegure el anonimato pueden hacer que una pregunta sea más o menos comprensible y más o menos amenazante (Cfr. Converse y Prese, 1986; Fowler, 1995; Wolf, 1991; Sudman y Bradburn, 1987).

Cada requisito de una encuesta por cuestionario implica aspectos particulares que hay que atender, en función de las personas a quienes se dirija. Un cuestionario para alumnos deberá cuidar especialmente la claridad de las preguntas y la extensión, ya que instrumentos largos o complejos producirán información de baja calidad, sobre todo con niños de corta edad. En el caso de maestros y directores será importante cuidar la tendencia a responder lo deseable en lugar de lo real, así como los temas delicados o amenazantes.

En todos los casos se deberá garantizar que los encuestados conozcan lo que se les pregunta, lo cual implica tener en cuenta la diferencia entre la ignorancia propiamente dicha y el poco interés por información irrelevante. Deberá atenderse la dificultad que para muchas personas representa manejar cifras grandes, tasas, porcentajes y tendencias; datos comparativos o en perspectiva; la inexactitud de la memoria y su fragilidad con datos antiguos; la distinción entre hechos y opiniones.

Asegurar que quien responde esté dispuesto a dar la información solicitada implica considerar la posibilidad de que influya lo socialmente deseable; el carácter delicado o amenazante del tema; la confianza que se tenga en que se mantendrá el anonimato; el desinterés que puede producir un cuestionario largo o la inadecuada secuencia de las preguntas; o fenómenos que se pueden pre-

sentar en forma inconsciente, como la tendencia a la media, a afirmar o el efecto de halo.

Para estar seguros de que los informantes entienden bien las preguntas se debe considerar la equivocidad de todo vocabulario, dados los universos semánticos que implica la diversidad cultural; la complejidad de la sintaxis que puede oscurecer el sentido de las preguntas, por ejemplo por el uso de frases subordinadas y negaciones dobles; los problemas de la extensión y el fraseo; los contenidos dobles, entre otros puntos.

En el caso de preguntas cerradas, se debe considerar la calidad y exhaustividad de las opciones, y en el de preguntas abiertas el riesgo de interrogantes vagos como los por qué, o los puntos de referencia cuando se pregunta sobre apreciaciones en términos de mucho-poco, etc.

No es fácil asegurar que una pregunta será entendida exactamente igual por todos los sujetos y, si eso no se logra las respuestas no serán comparables, ya que finalmente los sujetos *no estarán respondiendo la misma pregunta*, sino la que cada uno entendió, y sus respuestas no se podrán agregar ni analizar como muestras de ciertos rasgos de los respondientes.

Un tipo de cuestionario, la *escala*, trata de explorar posturas subjetivas como opiniones o actitudes y no el conocimiento de hechos objetivos. Para saber algo de esos aspectos *ocultos* de la realidad, esos *constructos latentes*, es necesario hacer inferencias a partir de aspectos que se puedan observar, como la expresión verbal de opiniones o actitudes, o la manifestación de conductas que las reflejen.

Ni siquiera la persona que tiene cierta actitud o sentimiento puede necesariamente expresarlo bien verbalmente. Cuando alguien experimenta una emoción fuerte suele decir que *no tiene palabras* para expresarla. Por ello buscar información sobre esos aspectos

con una sola pregunta es poco fiable, ya que es difícil encontrar una formulación que diga exactamente lo mismo a todos los sujetos. La idea que subyace la construcción de escalas tiene que ver con eso: es más confiable la información derivada de un conjunto de preguntas que de una sola, con tal que se cumpla una condición básica: que todas se refieran realmente al mismo aspecto de la realidad, que la escala sea *unidimensional*.

Si todos los ítems de una escala pertenecen a la misma dimensión, si se refieren a manifestaciones del mismo *constructo latente*, cada ítem captará un matiz distinto del constructo y será más probable que el conjunto de las respuestas represente correctamente lo que pretende. El análisis factorial, los modelos de respuesta al ítem y otras técnicas estadísticas permiten verificar que una escala mida una sola dimensión y permita *observar o medir* lo que no es posible captar a simple vista (Cfr. Morgenstern y Keeves, 1997; de Vellis, 1991).

Las escalas tampoco son inmunes a sesgos sistemáticos; en estudios de la OCDE y la IEA, por ejemplo, los alumnos de países con resultados muy bajos en pruebas de competencias cognitivas dicen que tienen actitudes más favorables hacia las mismas competencias que los estudiantes de países en que los resultados cognitivos son mucho mejores.

Lo anterior hace dudar de la confiabilidad de las escalas utilizadas, dado que las actitudes podrían ser más susceptibles de influencias culturales que los conocimientos. En concreto, se puede hipotizar que rasgos de las culturas de algunos países podrían llevar a muchos alumnos a responder positivamente las preguntas que indagan sobre sus actitudes, en un grado mayor que el que ocurre con los chicos de otros medios culturales.

Un trabajo sobre los resultados de las pruebas PISA señala que se ha identificado desde hace tiempo la presencia de ciertos patrones característicos que suelen mostrar

las respuestas a preguntas de escalas tipo Likert:

*Desafortunadamente, hay un creciente cuerpo de evidencias que sugieren que muchas diferencias que se han observado en estudios internacionales o interculturales están realmente contaminadas por ciertas características de los instrumentos de medición utilizados (artifacts of measurement)... Buena parte de estas investigaciones se refiere particularmente a las diferencias interculturales asociadas al uso de escalas Likert, o de ítems categóricos individuales tomados de ese tipo de escalas (Buckley, 2009: 4).*

Las encuestas con cuestionarios y escalas constituyen una estrategia económica para el estudio de las prácticas docentes, pero muchas veces producen información de mala calidad, en general debido a deficiencias que podrían evitarse.

Trabajos recientes indican que la validez y la confiabilidad de la información derivada de encuestas pueden ser mejores de lo que suele pensarse. En un estudio se aplicó un cuestionario a maestros de álgebra; con sus respuestas se construyó una medida del grado en que sus prácticas eran congruentes con los estándares del *National Council of Teachers of Mathematics*.

Los escasos estudios anteriores valoraban la confiabilidad de los instrumentos ítem por ítem, y la validez comparando dos instrumentos basados en reportes del maestro (un cuestionario y un diario). La novedad de este trabajo consistió en que la confiabilidad se analizó para la medida compuesta basada en las respuestas a 13 ítems tipo Likert, con seis opciones de respuesta cada uno; el análisis de la validez se hizo comparando las respuestas dadas por los docentes al cuestionario con los resultados de observaciones de su trabajo en el aula.

La consistencia de la medida compuesta basada en los 13 ítems tipo Likert fue ade-



cuada ( $\alpha = 0.85$ ). El coeficiente de confiabilidad de la medida compuesta que se obtuvo comparando los resultados de la primera aplicación con una segunda hecha cuatro meses más tarde fue de 0.69; y la correlación entre esa medida y otra basada en observación en aula de los mismos maestros, fue de 0.85.

Las cifras anteriores son indicativas de que la calidad de la información obtenida con el instrumento respondido por los maestros fue bastante aceptable, aunque otras limitaciones del instrumento siguen presentes, en particular el que solamente permite saber si ciertas prácticas se utilizan con más o menos frecuencia, pero no informa sobre qué tanto tiempo se usa cada una, y menos aún sobre la calidad con que se hace (Mayer, 1999).

Los maestros no son los únicos informantes calificados sobre las prácticas docentes. Los directores y los supervisores son fuentes importantes en relación con aspectos como la planeación de las clases, ya que normalmente su función incluye el apoyar el trabajo de los maestros al respecto y revisarlo. En algunos sistemas es usual que estos actores también observen periódicamente el trabajo en aula, en cuyo caso también pueden informar al respecto.

Los alumnos, por su parte, no son adecuados para opinar sobre la planeación de las clases o el conocimiento que tengan los maestros sobre las materias que enseñan, pero sí lo son, incluso al menos desde los últimos grados de la primaria, para informar sobre los temas cubiertos, las estrategias docentes utilizadas o la retroalimentación que brindan los maestros, entre otros temas.

En la medida en que el papel del maestro incluye cierta relación con los padres de familia para informarles sobre el avance de sus hijos, saber de los problemas que se enfrenten en el hogar o pedir colaboración, los padres pueden ser también informantes valiosos sobre esos asuntos.

## Auto-reportes

Son una variante de los cuestionarios y, como el término mismo indica, consisten también en informes proporcionados por los maestros mismos. También pueden ser no estructurados (en los que los maestros describen con sus propias palabras lo que han hecho durante cierto lapso) o estructurados (con formatos predefinidos para informar sobre la frecuencia con que llevaron a cabo ciertas prácticas).

La limitación que se suele atribuir a los *auto-reportes* tiene que ver con el riesgo de que los maestros no informen sobre lo que realmente hicieron, sobre sus prácticas efectivas, sino sobre lo que consideran que debieron haber hecho, o sea sobre las prácticas consideradas deseables en la profesión.

Un estudio relativamente antiguo exploró el punto, analizando la calidad de la información reportada por nueve profesores de inglés, comparándola con la que dieron sus alumnos y con la que se obtuvo mediante observación de sus clases. El instrumento utilizado para los *auto-reportes* de los maestros estaba altamente estructurado, con 77 ítems que describían otras tantas prácticas, de las cuales 37 se referían a la enseñanza de vocabulario y deletreo, y 40 a gramática y sintaxis.

Los maestros hicieron *auto-reportes* semanales en seis ocasiones, y uno de resumen al final del período. Un instrumento similar al de resumen de los maestros se aplicó a los alumnos al final del período del estudio; además se hicieron observaciones semanales no estructuradas de una clase impartida por los maestros estudiados, que luego se codificaron.

La información sobre las prácticas de los maestros que se obtuvo gracias a los *auto-reportes* basados en el instrumento descrito mostró razonable consistencia interna y también fue congruente con la información que dieron los alumnos y la que se obtuvo me-

diantе observaciones en aula. (Koziol y Burns, 1986)

### Bitácoras y diarios

Variante de auto-reportes: textos en que el maestro describe sus actividades durante cierto tiempo. La diferencia es la periodicidad con que se pide información para aumentar la confiabilidad, ya que es menos probable que se distorsione la realidad si las actividades se reportan varias veces a lo largo del tiempo. Se puede dejar que cada informante reporte en forma libre sus actividades o usar una guía estructurada para que lo haga. Los pros y contras de estas alternativas son los mismos que en cualquier otro caso, pero la carga de trabajo que supone para un maestro llevar un diario abierto durante muchos días es pesada, mientras que llenar un formato muy estructurado durante el mismo lapso es mucho más sencillo.

Se ha utilizado esta herramienta para estudiar el *currículo implementado* y la *oportunidad de aprendizaje*: el grado en que la enseñanza cubre o no los temas que el *currículo planeado* establece, lo que incidirá en el *currículo alcanzado o logrado* (Rowan, Camburn y Correnti, 2004; Rowan y Correnti, 2009).

El *currículo implementado* se suele explorar con cuestionarios aplicado al final del curso en que los maestros informan retrospectivamente sobre los temas cubiertos, o mediante la observación de algunas clases. En ambos casos la información puede ser de baja calidad:

*Considérese, por ejemplo, que a lo largo de un año académico de nueve meses, el maestro de primaria típico dará 140 o más días de clase... a los 20 o 30 alumnos de su grupo, a veces con actividades diferenciadas para alumnos individuales o subgrupos. Aún más: durante un día cualquiera, las actividades de enseñanza de un maestro típicamente se desarrollarán según varias dimensiones; un maestro normalmente cubrirá varios objetivos con*

*distintos niveles de demanda cognitiva en un solo día, trabajando con distintos arreglos conductuales y utilizando una variedad de técnicas de enseñanza propias de cada tema... algunos rasgos se repetirán a lo largo del año, pero otros no, por lo que las prácticas son multidimensionales, pero además altamente variables a lo largo del año* (Según Rogosa, Floden y Willet, 1984, citados por Rowan y Correnti, 2009: 121).

La variabilidad de prácticas de los maestros en varios momentos y la que distingue unos maestros y otros harían necesario, en general, un número elevado de observaciones (15-30) para asegurar una razonable consistencia de la información; el costo de este tipo de estudios se eleva en consecuencia, lo que hace atractiva la opción de los diarios (Rowan, Camburn y Correnti, 2004: 14-17).

Al pedir a los maestros que reporten de inmediato lo que hicieron en el aula cierto día, los diarios ganan en confiabilidad respecto a encuestas aplicadas una vez al año, pues los problemas de recordar el pasado y basarse más en interpretaciones de lo deseable se reducen sustancialmente; y al pedir que los reportes se hagan durante muchos días la muestra de prácticas reportadas es más representativa del universo que la información que dan pocas observaciones al año. Pero si el número de días a reportar aumenta, la carga para los maestros se hace mayor.

Para estimular las respuestas se puede ofrecer algún pago a los maestros, como hicieron Rowan y colaboradores, quienes también pusieron un número de teléfono gratuito a disposición de los maestros para resolver sus dudas. Como resultado se obtuvieron tasas de respuesta del orden de 90% y, según reportan los investigadores, datos de calidad ligeramente inferior a la de los derivados de observaciones en el aula (Rowan y Correnti, 2009: 122).

Se construyeron escalas que combinaron las respuestas a varios ítems, y el análisis mostró que un 72% de la variación del tiempo de enseñanza se dio entre unos días y otros; 23% entre maestros de una misma escuela; y sólo 5% entre diferentes escuelas. La  $\sigma$  de la distribución del tiempo de enseñanza *entre días* fue de 45 minutos, o sea que en 15 de cada 100 días el tiempo dedicado a la lectura fue realmente de *cero minutos*, aunque en el sistema americano suele suponerse se dedican diariamente 90 minutos a la lectura y la lengua, y aunque en promedio los maestros estudiados dedicaban 80 minutos diarios a esa área.

Estos datos implican que, para tener información suficiente sobre las prácticas de un maestro son necesarios reportes de unos 20 días al año (Rowan y Correnti, 2009: 123); un número similar de observaciones, mucho más costosas, sería igualmente necesario para tener una muestra suficiente de lo que pasa en el aula.

### Diarios en red (blogs)

El manejo de diarios en línea para recabar información sobre prácticas docentes puede reducir de manera importante el peso del trabajo que cada maestro, así como el de los investigadores. Si se cuenta con una plataforma adecuada, esta forma de recabar la información permite enviarla rápidamente y hace innecesario que los investigadores deban capturarla posteriormente para su análisis, ya que se almacena en el sistema al mismo tiempo que es registrada por los informantes.

Los principios metodológicos utilizados para diseñar un instrumento de tipo diario para recabar información en línea sobre prácticas docentes no son distintos de los que se deben manejar para diarios de tipo tradicional, con lápiz y papel. Siempre será necesario basarse en un marco conceptual que precise los aspectos a estudiar, identifique sus dimensiones y defina la manera en que se medirán.

Se tuvo acceso a un solo trabajo que analiza este tipo de diarios de maestros, que se llevó a cabo en relación con el mismo proyecto del que se derivó la información del apartado anterior. El trabajo se llevó a cabo hacia 1990, cuando el desarrollo de herramientas digitales estaba menos avanzado; su uso con un gran número de maestros, en escuelas con diferentes condiciones de acceso a las tecnologías llevó a decidir no manejar la herramienta en línea, pero pese a ello los autores consideraron que los resultados fueron alentadores. (Ball et al., 1999)

### Viñetas

Las preguntas de un cuestionario pueden ser entendidas en forma distinta. Es frecuente, además, que se utilicen términos propios de ciertas teorías, con las que los investigadores están familiarizados, pero no necesariamente los maestros, alumnos u otros sujetos; con ello aumenta la probabilidad de que la comprensión de los respondientes no coincida con la de los investigadores. Abundan ejemplos en el campo educativo. Si se pregunta a unos maestros, por ejemplo, si utilizan *trabajo colaborativo* o *evaluación formativa*, y responden afirmativamente, es difícil saber qué entendían exactamente por esos términos, y es probable que haya una diversidad de comprensiones al respecto, lo que invalida seriamente las conclusiones que se saquen con base en tales respuestas.

Por ello se ha desarrollado las *viñetas* que, en lugar de pedir información sobre prácticas en términos teóricos y abstractos lo hacen con descripciones precisas de conductas concretas, contextualizadas, pidiendo a los respondientes que indiquen si sus propias formas de trabajar se aproximan más o menos a las descritas en la viñeta. Este tipo de preguntas se utiliza desde hace tiempo en estudios de otros temas, como sobre actitudes discriminatorias o laborales (Cfr. Martin et al., 1991; Martin, 2006), pero en educación su uso es más reciente y no hay mucha investigación sobre la calidad de la información que permiten obtener.



Un trabajo de Stecher et al. (2006) estudió las prácticas de enseñanza de algunos maestros, indagando si eran consistentes con orientaciones innovadoras definidas en estándares para la enseñanza de matemáticas y ciencias (*reform oriented*), en contraste con la enseñanza *tradicional*. Para validar la información obtenida se usó una combinación de cuestionarios simples, cuestionarios basados en viñetas, diarios de los maestros y observaciones en el aula. Se hicieron también entrevistas cognitivas a una submuestra de maestros, en relación con las viñetas.

La preparación de viñetas comenzó con la operacionalización de los conceptos de currículo y práctica de enseñanza innovadora (*reform oriented*), en una taxonomía de 23 *elementos agrupados en tres categorías*: naturaleza de las matemáticas, pensamiento matemático de los alumnos y enseñanza de matemáticas.

Se identificaron elementos susceptibles de medirse con viñetas, se escogieron *dos temas* de matemáticas (área-perímetro y multiplicación con números de dos dígitos) y *cuatro situaciones* en que se pueden manifestar prácticas innovadoras (introducción de una lección, reacción ante errores de los alumnos, conciliación de enfoques distintos y selección de objetivos de aprendizaje). Los elementos se integraron en *escenarios* más amplios, que ofrecieron el contexto para el conjunto.

La sección de viñetas comenzaba describiendo los escenarios y las instrucciones para responder. Cada escenario incluía una contextualización y presentaba las situaciones, seguidas por posibles formas de actuar en forma innovadora o tradicional, pidiendo al maestro indicar qué tan probable sería que actuara en la forma descrita en cada opción.

Las posibles acciones que eran las opciones de respuesta en las viñetas fueron 51; 27 de ellas, a juicio de expertos y maestros, describían prácticas innovadoras (*reform oriented*), en tanto que otras 24 se referían a prác-

ticas tradicionales. A partir de lo anterior se definieron dos métodos para asignar una calificación global a cada respondiente, situándolo en un continuo innovador-tradicional. Se hicieron varios análisis para estimar tanto la confiabilidad de la medición obtenida como su validez, contrastándola con la información obtenida mediante los cuestionarios simples, los diarios y, en especial, las observaciones del trabajo en aula.

Los resultados apoyan la idea de que la información obtenida mediante viñetas es de buena calidad en algunos aspectos, pero no en todos; las entrevistas cognitivas con maestros participantes mostraron que no siempre su interpretación de las descripciones de las viñetas coincidía con la de los investigadores. Se pidió a los maestros que leyeran cada posible acción descrita en una viñeta y expresaran su forma de entenderla y sus razones para escoger una opción, *pensando en voz alta* al hacerlo. Se vio que en algunos casos los maestros no entendían de inmediato lo que quería decir la posible acción, y que antes de escoger una respuesta debían reformular la idea en sus propias palabras. Los maestros necesitaban saber el nivel de rendimiento de los alumnos con los que se suponía trabajaban, ya que las acciones a emprender dependerían en buena parte de ello.

Una conclusión es que preparar buenas viñetas implica un trabajo mucho mayor al que supone hacer un cuestionario simple (Stecher et al., 2006: 120).

### **Observación de situaciones pedagógicas (*classroom observation*)**

El segundo grupo de acercamientos al estudio de la práctica docente no se basa en lo que digan los maestros mismos u otros informantes, sino en la observación por terceras personas de dichas prácticas.

Hace medio siglo Medley y Mitzel decían ya que *la investigación sobre las conductas que tienen lugar en el aula no es un pasatiempo para aficionados, sino una ocupación*

de tiempo completo para profesionales técnicamente competentes (1963: 253) y, en forma congruente, definían la observación como una técnica rigurosa en los términos siguientes:

*Una técnica observacional... es una en la que un observador registra aspectos relevantes de una conducta que tiene lugar en el aula a medida que ocurre o muy poco después (or within a negligible time limit after) con un mínimo de cuantificación entre la observación de la conducta y su registro... por consiguiente no se incluyen en esta definición esquemas en los que se pide al observador que califique al maestro, a los alumnos o a la clase en una o más dimensiones, aún si las calificaciones se basan en la observación directa de conductas específicas.* (Medley y Mitzel, 1963: 253)

Con una concepción más amplia la definición de observación puede incluir procedimientos con escalas de calificación, técnicas de registro y postcodificación (por ejemplo con grabaciones en video) y también técnicas cualitativas. Medley y Mitzel (1963) distinguían los sistemas de observación basados en tiempos (de categorías) y los basados en eventos (de signos). Una década más tarde Rosenshine y Furst (1973) distinguían técnicas de observación según el procedimiento de registro: sistemas de conteo (categorías o

signos) y de calificación (*rating*); según especificidad de los ítems: sobre conductas específicas (de baja inferencia) o más generales (de alta inferencia); y según el tipo de codificación, de una o más dimensiones.

Los instrumentos desarrollados recientemente suelen basarse en teorías actuales sobre los aspectos a observar y cuidan las propiedades psicométricas. La descripción siguiente se reduce a algunos de los instrumentos más estudiados, seleccionados para el *Proyecto sobre Medición de la Efectividad de la Enseñanza* (MET Project, 2010a y 2010c).

### Classroom Assessment Scoring System (CLASS)

Posiblemente la herramienta sobre la que hay más estudios. Es un instrumento para estudiar las actividades del docente y las interacciones que tiene con sus alumnos. Fruto de un trabajo de más de una década, a cargo de Robert Pianta y sus colegas, una versión previa se difundió con el nombre de *Classroom Observation System*, COS (Pianta y Hamre, 2009; MET Project, 2010d).

La conceptualización de las actividades e interacciones distingue tres dominios: uno relativo a la organización del aula; otro al apoyo instruccional que se ofrece a los alumnos; y uno más al apoyo emocional. La Tabla N° 1 las sintetiza.

Tabla N° 1. Dimensiones de la enseñanza efectiva evaluadas por el CLASS

|  | Apoyo Emocional   | Organización del aula  | Apoyo Instruccional   |
|--|---|--|---|
| <b>Preescolar y primeros grados de primaria</b>      | Clima positivo<br>Clima negativo<br>Sensibilidad maestro<br>Atención a punto de vista del niño    | Gestión de conducta<br>Productividad instruccional<br>Formatos aprendizaje | Desarrollo conceptual<br>Calidad retroalimentación<br>Modelamiento lenguaje                                 |
| <b>Últimos grados de primaria y nivel secundaria</b> | Clima positivo<br>Clima Negativo<br>Sensibilidad maestro<br>Atención a punto de vista adolescente | Gestión de conducta<br>Productividad instruccional<br>Formatos aprendizaje | Comprensión contenidos<br>Análisis-solución problemas<br>Calidad retroalimentación<br>Diálogo instruccional |

Fuente: MET Project, 2010d: 3

La observación se lleva a cabo en períodos de media hora; en cada período se dedican 20

minutos a observar y tomar notas, y el tiempo restante a la calificación. Según los auto-

res bastan cuatro ciclos de observación para contar con una muestra representativa de lo que ocurre en un salón de clases.

El sistema se desarrolló para observar grupos de preescolar o de los primeros grados de primaria; están en desarrollo versiones para grados superiores. El organismo que tiene los derechos del sistema capacita observadores. La calidad de la información derivada del uso del sistema CLASS ha sido explorada empíricamente en forma extensa para la versión enfocada a preescolar y los primeros grados de primaria, con resultados satisfactorios. El trabajo respecto a las otras versiones está menos avanzado (Goe, Bell y Little, 2008: 24).

El gran número de escuelas y distritos escolares que usan la herramienta refleja su atractivo para usos formativos, pero al mismo tiempo hace previsible que aumente la heterogeneidad de los cuidados que se tomen para asegurar la competencia de los observadores y la confiabilidad de su trabajo.

### **Marco de Referencia para la Enseñanza (*Framework for Teaching*).**

A partir de su trabajo en el ETS para la Versión III del Praxis, Charlotte Danielson y sus colaboradores, autores de esta herramienta, la describen como

*...estructurada según componentes de la instrucción derivados de la investigación, alineada a los estándares del consorcio interestatal (INTASC), y basada en una perspectiva constructivista del aprendizaje y la enseñanza...* (MET Project, 2010d; Goe, Bell y Little, 2008: 21-22)

El Marco tiene cuatro dominios: Planeación y preparación de la clase; ambiente del aula; instrucción; y responsabilidades profesionales. Estos dominios se dividen en 22 componentes y 76 indicadores. A guisa de ejemplo, los componentes y los indicadores del dominio de ambiente del aula son los siguientes:

- Creación de un ambiente de respeto y buena relación
  - Interacciones del maestro con los alumnos
  - Interacciones de los alumnos entre sí
- Establecimiento de una cultura de aprendizaje
  - Importancia de los contenidos
  - Expectativas de aprendizaje y logro
  - Orgullo de los alumnos por su trabajo
- Procedimientos de gestión del aula
  - Manejo de grupos para la instrucción
  - Manejo de las transiciones entre etapas de los procesos
  - Manejo de materiales
- Gestión de la conducta de los alumnos
  - Expectativas
  - Monitoreo de la conducta de los alumnos
  - Respuesta ante las conductas inapropiadas (MET Project, 2010b)

El FFT incluye rúbricas detalladas con base en las cuales se puede evaluar a un maestro en cuanto a cada uno de los 76 elementos en cuatro niveles de desempeño, que se definen como *insatisfactorio*, *básico*, *avanzado* (*proficient*) y *sobresaliente* (*distinguished*) (MET Project, 2010d).

### **Sistema MQI (*Mathematical Quality of Instruction*)**

Como indica su nombre, la herramienta se orienta precisamente a la observación de prácticas de enseñanza en el campo de las matemáticas, a partir de lo que la investigación al respecto ha encontrado (Hill et al. 2010a y 2010b). Se identificaron así cinco puntos, como los más relevantes al respecto:

- Uso preciso y rico del lenguaje matemático;
- Ausencia de errores e imprecisiones;
- Presencia de explicaciones matemáticas correctas y participación de los alumnos

en la construcción de significado y en el razonamiento matemático;

- Conexión del trabajo del aula con ideas importantes en el campo; y
- Esfuerzos para garantizar que todos los alumnos del grupo, y no sólo algunos, tengan acceso al conocimiento matemático (MET Project, 2010e).

El MQI explora tres tipos de relaciones: de los maestros con los contenidos; de los alumnos con los contenidos; y de los maestros con los alumnos. Al valorar esas dimensiones, el MQI sobresale entre los instrumentos del área de matemáticas, porque ofrece una visión completa y equilibrada de los elementos que, en conjunto, conforman la enseñanza de calidad en el área (MET Project, 2010e).

Durante el desarrollo del sistema MQI se encontró que, además de un buen manejo general de las matemáticas, el maestro deben dominar un tipo particular de conocimiento matemático, precisamente *el que hace falta para la enseñanza de esta área*, que tiene que ver con una comprensión precisa de los obstáculos que dificultan a los alumnos el aprendizaje de este campo. Por ello en paralelo se ha desarrollado otro tipo de herramienta para medir ese tipo de conocimiento: el MKT (*Mathematical Knowledge for Teaching*) (MET Project, 2010e; Hill et al., 2008).

### Sistema PLATO

Otro ejemplo de herramienta especializada en un área curricular particular es la denominada “Protocolo para la Observación de la Enseñanza de las Artes del Lenguaje” (*Protocol for Language Arts Teaching Observation*, PLATO).

Con base en investigaciones previas, el sistema PLATO se estructura a partir de cuatro factores que se considera subyacen la enseñanza: la demanda que el área plantea a las prácticas del aula y el discurso relativo; el andamiaje para apoyar la enseñanza de los

contenidos de lengua; las representaciones de los contenidos y el uso que se hace de ellos; y el ambiente del aula (MET Project, 2010f).

El sistema identifica 13 elementos que constituyen dimensiones independientes, y permite evaluarlos con una rúbrica, en una escala de uno a cuatro:

|  |                                  |
|--|----------------------------------|
| Propósito                              | Práctica guiada                  |
| Desafío intelectual                    | Discurso que se tiene en el aula |
| Representación del contenido           | Instrucción basada en textos     |
| Conexiones con el conocimiento previo  | Lenguaje académico               |
| Conexiones con la experiencia personal | Manejo de las conductas          |
| Modelamiento                           | Manejo del tiempo                |
| Estrategias explícitas de enseñanza    |                                  |

El sistema prevé observaciones independientes de 15 minutos cada una, durante una clase: dos en clases de 45 minutos y tres en las de 90. La investigación sobre la herramienta es más reducida que en los casos anteriores, y se ha desarrollado principalmente en escuelas de Nueva York. (MET Project, 2010d y 2010f)

### Sistema QST (*Quality of Science Teaching*)

Se trata de un nuevo protocolo de observación que está en desarrollo, en el marco del *Proyecto sobre Medición de la Efectividad de la Enseñanza* (MET Project).

El punto de partida para el desarrollo del QST es un sistema (*The Teaching Event*) para evaluar aspirantes a ocupar plazas de maestro en el sistema educativo de California, que comprende 13 áreas de enseñanza, mediante un portafolios con evidencias de la práctica docente que incluyen planes de clase, un video, muestras del trabajo de los alumnos y una reflexión del aspirante mismo. El sistema maneja cinco dimensiones – planeación, enseñanza, evaluación, reflexión y uso de lenguaje académico— a partir de las cuales se espera derivar constructos para



preparar protocolos y guías de codificación para observar sesiones de clase en aula (MET Project, 2010d).

### **Observaciones con grabación en video**

Un tipo especial de observación es el que se hace no directamente en el aula, sino en forma diferida, sobre registros de las mismas que tengan la fidelidad suficiente para permitirlo, por ejemplo con grabaciones de video.

La ventaja de este tipo de trabajos —en particular en gran escala, involucrando a decenas o centenas de escuelas en regiones o incluso en países diversos— es que es más factible contar en esos lugares con técnicos que operen una cámara de video de manera eficiente, que con observadores con la calificación necesaria para obtener información válida y confiable. Es el caso de los trabajos que usan grabaciones en video de clases conducidas por maestros, y un ejemplo destacado de ellos es el “estudio de clases grabadas en video” que formó parte del *Tercer Estudio Internacional sobre Matemáticas y Ciencias* (TIMSS) en 1995, y de su repetición en 1999 (TIMSS-R) (Stigler, Gallimore y Hiebert, 2000).

Otras ventajas son que se puede detener la reproducción de la grabación o repetirla mientras se califica, y que se puede volver a calificar con otros jueces y utilizando otros protocolos o escalas.

En el caso del TIMSS y el TIMSS-R sólo se grabó una clase por profesor, lo que obviamente no basta para tener una muestra suficiente de su forma de trabajar a lo largo de un curso. Por otra parte, las grabaciones en video, como la observación directa en vivo, no pueden captar lo que están pensando el maestro o los alumnos, sin lo cual no puede entenderse cabalmente el sentido de sus acciones. Esos aspectos “subjetivos” deberán atenderse con otros acercamientos.

Una desventaja de la grabación en video se deriva de que el ángulo que puede captar una cámara convencional es reducido, por lo que

muchas cosas que pasan en el aula no quedan registradas, fuera de las que ocurren en ese reducido campo de visión. La forma de subsanar esta limitación es usar simultáneamente varias cámaras. Además de aumentar los costos, esta solución hace también que la interferencia producida por la observación misma sea muy grande, lo que pone en riesgo la validez.

Recientemente se explora el uso de video-grabaciones mediante un sistema digital avanzado que, a diferencia de las cámaras convencionales, permite registrar lo que ocurre en el aula con una visión panorámica de 360°, y luego al reproducir la grabación hacer los acercamientos que se quiera. Sin minimizar las ventajas que ofrecen las poderosas tecnologías digitales a las que hoy se tiene acceso, debe reiterarse también que su uso no evita la necesidad de contar con observadores calificados, así como con esquemas de categorías que permitan estructurar lo observado según las dimensiones pertinentes.

Para valorar adecuadamente el potencial de los acercamientos a la práctica docente basados en distintos tipos de grabación de video sería necesario analizar en detalle los pros y contras, el costo y los beneficios de las distintas opciones, desde los videos con cámaras y micrófonos múltiples, y uno o incluso dos camarógrafos, hasta la grabación en 360°, cuya efectividad debe aún demostrarse, para ver si puede dar efectivamente información de mejor calidad que las otras opciones a menor costo o no.

### **Acercamientos basados en el análisis de productos de las prácticas**

Los acercamientos al estudio de las prácticas docentes que se revisan en este apartado se basan en el análisis de los productos de las prácticas, como planes de clase, cuadernos de alumnos, exámenes aplicados o tareas encomendadas. Se trata, pues, de una tercera categoría, distinta tanto de los trabajos basados en lo que los docentes mismos reportan, como de los que utilizan observación.



## **Análisis de tareas**

Un primer acercamiento de este grupo consiste en el análisis de las tareas que los alumnos llevan a cabo por indicación del docente. Es claro que la revisión de esos trabajos podrá revelar muchas cosas sobre la práctica del maestro, como la forma en que entiende su papel, el manejo que tiene de los contenidos, su concepción de la evaluación, el grado en que ofrece retroalimentación más o menos rica a los alumnos, entre otras. Por ello se entiende que un ejemplo temprano de este tipo de acercamiento lleve como subtítulo “Abriendo una ventana sobre las prácticas del aula” (Matsumura y Pascal, 2003).

Los trabajos en que se basa este aparatado definen la calidad de las tareas según tres dimensiones: el nivel de demanda cognitiva que representaban; la claridad de los objetivos de aprendizaje a los que una tarea debería contribuir; y la claridad con que se especificaban los criterios de calificación. Las puntuaciones asignadas en una escala de cuatro puntos en cada una de estas dimensiones se combinan para formar una puntuación de calidad global (Matsumura y Pascal, 2003).

De cada maestro participante (181 de 35 escuelas, de los grados 4°, 7° y 10°) se recogieron materiales de tres tareas (dos sobre comprensión lectora y una sobre expresión escrita), incluyendo información sobre la planeación, las instrucciones dadas a los alumnos, las tareas entregadas por cuatro alumnos (dos de alto y dos de medio rendimiento) y se observó a los maestros en clase dos veces. Se analizaron las tareas recolectadas y los resultados se sintetizaron para formar indicadores de la práctica docente. Se encontró un nivel aceptable de consistencia entre los analistas, siendo necesarias tareas de tres o cuatro alumnos de cada maestro para tener estimaciones estables, y sólo si las tareas habían sido diseñadas por el maestro, y no tomadas de otra fuente.

La calidad de las tareas se asoció con la de la enseñanza, según las observaciones de la

misma, y también con la calidad del trabajo de los alumnos; aquellos estudiantes cuyo maestro dejaba tareas cognitivamente más exigentes y manejaban criterios de calificación más claros, mostraron también mayor avance en evaluaciones externas. Sin embargo, en general la calidad de las tareas asignadas por los maestros no fue muy alta (Matsumura y Pascal, 2003).

## **El portafolio y sus variantes**

El portafolio es una herramienta para evaluar alumnos, maestros, escuelas, etc. En el caso de los maestros los portafolios consisten en conjuntos de evidencias como planes de clase, cuadernos de alumnos, registros de clases, entre otros.

Los portafolios contienen materiales seleccionados por el sujeto que va a ser evaluado, de quien se espera una *autoevaluación*, a partir de los materiales incluidos en el portafolios, que luego serán valorados por otras personas, que podrán ser autoridades, pares u otros evaluadores externos.

Al usar portafolios con el propósito de observar/medir/evaluar la práctica docente se aplica, como en los demás instrumentos, la necesidad de precisar el foco de atención al que se refieren, en el caso que nos ocupa el aspecto(s) de la práctica que se quiere observar. Deben también explicitarse las categorías que se usarán para sistematizar lo observado y la forma de pasar a la medición y al juicio.

Hay ejemplos destacados de sistemas de evaluación de maestros basados en la preparación de portafolios y en el análisis de su contenido. El sistema *Praxis* del *Educational Testing Service*, además de pruebas de lectura, escritura y matemáticas, así como de conocimientos pedagógicos y de las áreas de contenido en que trabaja un maestro, incluye un portafolios.

El grado en que los maestros cumplen los estándares de práctica profesional para los maestros desarrollador por el *Interstate New*

*Teacher Assessment and Support Consortium* (INTASC) se evalúa con un portafolios, y lo mismo ocurre con los estándares del *National Board for Professional Teaching Standards* (NBPTS) (Porter, Youngs y Odden, 2001: 263-265).

En Latinoamérica destaca el sistema de evaluación de docentes desarrollado por el Ministerio de Educación de Chile, que forma parte de un conjunto mayor de políticas en pro de la mejora de la calidad educativa, dentro de las cuales tienen un peso específico las que se refieren a la superación profesional de los maestros.

El *Sistema de Evaluación del Desempeño Profesional Docente* (SEDPD) comenzó a desarrollarse en 2003, con base en un conjunto preciso de estándares que constituyen el *Marco para la Buena Enseñanza*, preparado durante los dos años anteriores. El trabajo técnico se encomendó a una agencia independiente muy reconocida, el Centro de Medición de la Universidad Católica de Chile (MIDE), que trabajó en estrecha relación con el Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (CPEIP) del Ministerio de Educación de ese país.

Los documentos del SEDPD definen con precisión sus elementos centrales: para qué se evalúa; qué se evalúa; qué consecuencias tiene la evaluación para los docentes; quiénes deben evaluarse; y que instrumentos y fuentes de información deben utilizarse. En cuanto a esto último, los instrumentos de la evaluación son cuatro: una autoevaluación; una entrevista por un evaluador par; un informe de referencia de terceros (el Director y el Jefe de la Unidad Técnico Profesional de la escuela); y un portafolio de desempeño pedagógico. Para cada uno existen pautas perfectamente definidas y son aplicadas por personal entrenado para ello.

El cuidadoso trabajo de diseño y validación de los instrumentos de dicho Sistema hace de él un caso ejemplar en la región, y que se

compara favorablemente con los sistemas análogos de los sistemas educativos más desarrollados. Una obra colectiva reciente presenta amplia información sobre el desarrollo del sistema y sus resultados, tras varios años de aplicación. (Manzi, González y Sun, 2011)

Otros dos ejemplos de herramienta tipo portafolio son los siguientes.

*El sistema IQA (Instructional Quality Assessment)*

Es un desarrollo del trabajo de Matsumura sobre el análisis de las tareas hechas por los alumnos como forma de aproximarse al trabajo del maestro. Este nuevo sistema está en proceso, y centra la atención en comprensión lectora y matemáticas, en el nivel medio del sistema educativo.

El sistema comprende protocolos para evaluar la enseñanza con observaciones en aula, así como la calidad de las tareas asignadas por el maestro. Se reportan trabajos sobre la confiabilidad y potencial validez de la información obtenida.

Esos trabajos coinciden con otros en encontrar una amplia variación de la calidad de la enseñanza en las áreas estudiadas, con un nivel promedio no muy alto. La calidad de la información fue mejor en matemáticas que en lectura. Los resultados de la aplicación del IQA predicen los obtenidos por los alumnos en algunas evaluaciones externas, después de controlar otros factores posiblemente asociados. La solidez de los resultados es limitada porque se derivan de un número reducido de casos. (Matsumura et al., 2006)

*Paquetes de materiales (Artifacts Packages)*

Este tipo de instrumentos es una variante de portafolios y no se limita a la revisión de tareas, sino que incluye cualquier otro tipo de material (*artifact*) que informe sobre las prácticas que tienen lugar en el aula: planes de clase; notas que el maestro entregue a los alumnos; fotografías de la forma en que se

organiza a los alumnos para trabajar o de lo que escribe en el pizarrón; trabajos de los alumnos, sean o no tareas; grabaciones de sesiones, etcétera.

Los promotores del acercamiento usan la expresión *scoop notebook* para designar la libreta o carpeta en que se integran los materiales recogidos para su análisis posterior, a la manera de los biólogos que primero dedican cierto tiempo al trabajo de campo, durante el cual se dedican a recopilar el mayor número posible de especímenes con una red (*scoop*), para luego analizarlos con todo cuidado en el laboratorio. Así los investigadores de la práctica docente pueden primero dedicar tiempo a recolectar todas las evidencias que puedan (materiales, *artifacts*) del trabajo de docentes y alumnos, para luego estudiarlas en detalle.

El propósito del proyecto *Scoop* fue desarrollar un acercamiento alternativo al estudio de la práctica docente, empleando artefactos y materiales para alcanzar a hacer una representación de dicha práctica suficiente para que una persona, sin observar directamente al maestro ni al aula pudiera hacer juicios válidos sobre algunos aspectos particulares, únicamente con base en esos materiales.

Además de incluir instrucciones para los maestros sobre cómo reunir los materiales, un componente clave es el que constituyen las guías de calificación que los jueces utilizan para analizarlos. El trabajo desarrollado se refiere en particular a las áreas de matemáticas y ciencias (Borko, Stecher y Kuffner, 2007).

Con base en estudios previos, se definieron diez dimensiones de las prácticas de enseñanza en las dos áreas en las que se pudiera apreciar su consistencia con las orientaciones para una enseñanza innovadora, según los estándares nacionales de matemáticas y ciencias.

En el caso de matemáticas las dimensiones son:

- Tipos de agrupamiento utilizados para el trabajo de los alumnos.
- Estructura de la clase.
- Uso de formas diversas de representar las ideas y conceptos.
- Uso de herramientas matemáticas.
- Demanda cognitiva que se maneja.
- Presencia de un discurso matemático común.
- Manejo de explicaciones y justificaciones.
- Manejo de solución de problemas.
- Formas de evaluación utilizadas.
- Uso de conexiones con la vida real o aplicación de conocimientos.

Las dimensiones del área de ciencias son similares y, para cada una, se ofrece una rúbrica que define con precisión lo que se entenderá por una práctica de calidad alta, media o baja. Las guías elaboradas a partir de esas dimensiones se pueden usar tanto para hacer observaciones en aula, como para analizar los materiales recogidos en la “carpetared”, que los maestros reúnen mediante un trabajo intensivo durante una semana, siguiendo las directrices que se les dan.

No es una elección libre de cada maestro, sino una que busca representatividad, en una materia y una semana. Sí se pide al maestro que incluya su punto de vista y consideraciones sobre los materiales, el propósito de manejarlos, uso que les da, etc. Los materiales incluyen los que se usan antes de la clase (planes, notas, rúbricas para evaluar), en la clase (fotos del pizarrón, materiales proyectados, trabajos de alumnos) y después de ella (tareas, exámenes, ítems de portafolios).

Aunque se trata de un trabajo en desarrollo, los resultados muestran que los juicios sobre la práctica docente basados en el sistema tienen niveles aceptables de confiabilidad y

validez (Borko, Stecher y Kuffner, 2007; Borko et al. 2005).

### Para cerrar el apartado

Las herramientas revisadas en este punto, como las de los anteriores, tiene puntos fuertes y débiles, por lo que conviene verlas como complementarias, más que como alternativas.

La utilización de productos de las prácticas como fuentes de información comparte con la observación la ventaja de que no se ve afectada por el riesgo de captar lo deseable más que lo real; como la observación de grabaciones de observaciones, los productos de las prácticas pueden ser revisados una y otra vez, por distintos analistas y utilizando protocolos diferentes. Estos acercamientos también comparten con la observación su incapacidad de captar lo que pasaba por la mente de los maestros al realizar las actividades de las que se derivaron los productos, lo que sólo se podrá explorar a través de la versión que den los actores estudiados, con acercamientos del primer grupo.

Además de las herramientas revisadas, hay información sobre otras en la “Guía de productos para la evaluación de maestros” (NCCTQ, 2011, [www.tqsource.org](http://www.tqsource.org)) ya mencionada, que incluye instrumentos de observación y otras herramientas:

|                                    |                              |
|------------------------------------|------------------------------|
| De observación en aula: 41         | Artefactos de enseñanza: 8   |
| Portafolios: 6                     | Autorreportes de maestros: 4 |
| Encuestas de alumnos: 3            | Modelos de valor agregado: 6 |
| Medidas de desempeño de alumnos: 4 | Modelos combinado: 21        |

### Conclusión

La insatisfacción con los sistemas de evaluación de maestros –tanto los basados en indicadores como antigüedad y escolaridad, como los recientes que utilizan medidas del valor que se supone *agregan* el maestro y la escuela al aprendizaje de los alumnos— ex-

plica el interés por buscar mejores maneras de *observar y/o medir la instrucción efectiva*, sobre todo con propósitos de *evaluación sumativa* y rendición de cuentas, en concreto sustentando sistemas de remuneración.

Es importante destacar la *dimensión formativa* que puede tener la evaluación de maestros, al identificar las deficiencias particulares de las prácticas para poder diseñar actividades de actualización y otras formas de intervención, que lleven a mejorar las de los maestros menos experimentados y/o menos eficaces. Para contar con evaluaciones sumativas válidas y confiables de la eficacia de los maestros, o para tener buenas evaluaciones formativas que permitan ayudarles a mejorar, se necesitan *observaciones/mediciones* de buena calidad.

La aplicación de cuestionarios y herramientas similares puede ser barata, pero la información que ofrecen puede ser superficial y de baja confiabilidad, al depender de los informantes, pero sólo con esos acercamientos es posible captar lo que los actores estudiados saben, creen o piensan. La observación puede dar información rica, pero implica mucho tiempo de observadores calificados y no puede captar lo que piensan los sujetos. Los análisis de productos de las prácticas comparten algunos de los pros y contras de los otros dos tipos de acercamiento.

Ningún acercamiento es en sí mismo superior o inferior a otros; la calidad de la información que dará cada uno depende del objeto y el propósito del estudio, de la calidad del instrumento y la forma en que se aplique y se analicen sus resultados.

### La oposición cuanti-cuali

Retomemos las nociones de observación y medición y la supuesta oposición que habría entre ambos conceptos, análoga a la que distinguiría los acercamientos cuantitativos y los cualitativos. En estas páginas se comparte la idea de que no hay una oposición tajante entre ambos enfoques, sino una de grado.



La distinción de técnicas cuantitativas y cualitativas se entiende muchas veces como si fueran dos tipos totalmente distintos de acercamiento a la realidad, el primero que implicaría codificación y el segundo que la descartaría.

En realidad *toda técnica implica codificar o estructurar información*; lo que cambia es el momento en que se hace: en las técnicas llamadas *cualitativas* (observación participante, etc.) se codifica *después* de recoger la información; en cuestionarios con preguntas simples se hace *desde que se comienza a elaborar el instrumento*.

Siempre se codifica; los humanos no podemos aprehender la realidad de manera holística; conocemos *analíticamente*, con categorías. Al aplicar técnicas de *pensar en voz alta* para detectar la forma en que procede un observador que pretenda ser totalmente cualitativo se encontrará que de hecho utiliza categorías que, al no explicitarse, tienen el riesgo de inconsistencia incluso si las aplica un mismo observador, además de la imposibilidad de comparar lo registrado por varios.

Es imposible observar algo sin categorías. La diferencia real es si las categorías utilizadas lo son explícita o implícitamente; si la explicitación se hace antes de observar o después; si las categorías están bien o mal definidas; si son excluyentes o no; si pretenden conformar una escala coherente y consistente, etc.

La ventaja de trabajar con categorías previamente definidas es clara, pero siempre está presente la desventaja correlativa: lo que no encaje en las categorías pasará desapercibido. Por ello es crucial la forma en que se construyan. No definir las categorías previamente tienen la ventaja de no excluir *a priori* lo que no se ajuste a lo predefinido (lo que puede ser adecuado en una etapa exploratoria), pero el costo que se paga es alto: como cada observador aplica sus categorías implícitas, la información resultante es heterogénea en grados desconocidos, lo que hace im-

posible cualquier análisis riguroso posterior. Codificar *a posteriori* registros previos, incluso muy completos, enfrenta exactamente los mismos problemas.

La calidad de la información que se obtiene con un acercamiento estructurado depende de la conceptualización subyacente en los instrumentos; y la que se deriva de un acercamiento no estructurado, de la calidad de los observadores. Pero la calidad de éstos, finalmente, al hacer juicios de alta inferencia sobre la marcha, o al codificar *a posteriori* con base en los registros, depende a su vez del manejo de categorías comunes para asegurar su aplicación uniforme.

### **Las rúbricas como herramientas de observación-medición-evaluación**

Unas herramientas muy conocidas son las *rúbricas*, que se analizan en este punto porque los principios que implica su construcción son generalizables a cualquier instrumento de *observación-medición-evaluación*.

El término *rúbrica* es una traducción del inglés que resulta poco clara, ya que en español denota la *firma* que alguien estampa al final de un escrito para. En el sentido que nos ocupa, en cambio, una *rúbrica* es un esquema que los maestros usan, por ejemplo, para calificar trabajos de tipo ensayo o tareas que implican la realización o *ejecución* de algunas actividades. Para calificar un ensayo una *rúbrica* podrá indicar que se tome en cuenta el esquema que organiza el trabajo, la corrección gramatical, la ortografía, la claridad de las ideas expuestas, etc.

Más adecuada es la traducción del término como *matriz de valoración*, expresión que explicita una característica básica de tales herramientas, la que se refleja en su estructura *matricial*, de *tabla de doble entrada*, con renglones y columnas:

- Un eje contiene los aspectos o *dimensiones* del objeto de estudio. Por lo general estos aspectos se incluyen en los renglones de la tabla, cada uno de los cuales presenta



una *dimensión* de la realidad a evaluar, por ejemplo la corrección gramatical de una redacción, su ortografía, etc.

- El otro eje tiene los *niveles de desempeño o ejecución* de cada dimensión, con cierto *orden* de mejor a peor. Las columnas presentan una gradación de desempeños de distinta calidad, que se identifican con etiquetas como insatisfactorio, mínimo, suficiente, avanzado y sobresaliente, o similares.

Una matriz de valoración o rúbrica busca ser, pues, una herramienta de medición, lo que es prerequisite para que sea también herramienta de evaluación. Como se ha dicho, los principios de su construcción son aplicables a cualquier instrumento.

La realidad es siempre multidimensional, pero una buena medición debe referirse a una dimensión bien definida de la realidad: sólo así cumplirá el requisito básico de *unidimensionalidad*. Por ello preparar una rúbrica o cualquier instrumento supone un proceso de clarificación conceptual, de identificación y distinción de las dimensiones de la realidad a medir y evaluar (*operacionalización*). Y como para evaluar necesitamos medir al menor ordinalmente, para preparar una rúbrica o cualquier otro instrumento de evaluación es necesario definir niveles graduados de desempeño o ejecución, *mediciones ordinales* basadas en la observación de rasgos que se asocian a cada nivel.

### Consideraciones finales

Para tener información de buena calidad con procesos de observación/medición deberán cuidarse tres elementos:

- El trabajo conceptual necesario para establecer las categorías en que se clasificarán las realidades captadas, así como los niveles de desempeño que se considerarán para su medición.
- El trabajo técnico —con las herramientas psicométricas apropiadas— que se requie-

re para cuidar la validez y confiabilidad de las observaciones.

- El trabajo empírico que, aplicando los marcos conceptuales, y utilizando las herramientas psicométricas, permitirá calibrar los instrumentos y valorar la calidad de la información obtenida mediante su uso.

No parece razonable pedir a todos los investigadores que dominen las técnicas psicométricas más avanzadas, pero de todos se deberá esperar un trabajo conceptual y empírico cuidadoso. Los pros y contras de las técnicas que suponen codificación previa y las que codifican *a posteriori*, por otra parte, aconsejan utilizar preferentemente una combinación de ambas; es la idea que subyace el atractivo de los llamados métodos mixtos. Esta recomendación tiene sentido, pero también encierra un riesgo ya mencionado, que no se debe perder de vista.

Una conclusión de un trabajo sobre viñetas es aplicable de manera más amplia a los instrumentos de observación y medición en general: una combinación de acercamientos suele dar información más rica que cualquiera por separado, pero al mismo tiempo hace aumentar el riesgo de que los resultados no sean consistentes. En palabras de Stecher:

*Si bien se anima frecuentemente a los educadores a utilizar mediciones múltiples para obtener información más confiable, los investigadores deben tener conciencia de que el manejar formatos múltiples y diversos respondientes puede disminuir la consistencia de las respuestas, al incluir fuentes adicionales de variación* (Stecher et al., 2006: 120).

### Referencias

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.

## ARTICLE RECORD / FICHA DEL ARTÍCULO

|  |   |
|--|---|
| <b>Reference / Referencia</b>                  | Martínez-Rizo, Felipe (2012). Procedimientos para el estudio sobre las prácticas docentes. Revisión de la literatura. <i>RELIEVE</i> , v. 18, n. 1, art. 1. <a href="http://www.uv.es/RELIEVE/v18n1/RELIEVEv18n1_1.htm">http://www.uv.es/RELIEVE/v18n1/RELIEVEv18n1_1.htm</a>   |
| <b>Title / Título</b>                          | Procedimientos para el estudio sobre las prácticas docentes. . Revisión de la literatura. [ <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>                      | <p><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></p> <p>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.</p> |
| <b>Keywords / Descriptores</b>                 | <p><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></p> <p>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..</p>   |
| <b>Institution / Institución</b>               | Universidad Autónoma de Aguascalientes (México).  |
| <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)   |

## 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](#) 

## RELIEVE

**Revista E**lectrónica de **I**nvestigación y **E**valuación 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).