

# EVALUACIÓN DE LOS MAESTROS

*Carol Anne Dwyer*

*La tercera parte del libro se inicia con una revisión crítica del tema de la evaluación de los maestros dentro de la perspectiva general del mejoramiento de la educación, que es el propósito final de la evaluación. Este capítulo presenta un análisis de los propósitos de la evaluación de los docentes, los estándares a los que debe responder y las metodologías disponibles. También identifica aspectos a considerar cuando se ejecutan evaluaciones de los maestros, tales como la selección del método, la inclusión de los resultados de la investigación, los énfasis entre la teoría y la práctica, las diversas perspectivas involucradas y el alcance de los criterios de evaluación.*

## **INTRODUCCION**

En este documento se analizarán los aspectos siguientes de la evaluación de los maestros:(1)

- Utilización de la evaluación de los maestros para el mejoramiento educativo en un sistema de evaluación integrado.
- Relaciones entre metas, estándares y pruebas en la evaluación de los maestros.
- Estándares para: validez, calidad técnica y utilización; currículo; conocimiento y aptitud docente; nivel de desempeño; y oportunidad para aprender.
- Fines de la evaluación de los maestros.
- Críticas de las alternativas de evaluación de los maestros.
- Análisis del campo de la enseñanza para fines de evaluación.
- Papel de un concepto rector de la enseñanza para la evaluación de los maestros.
- Selección de los métodos de evaluación.
- Determinadas cuestiones y debates en la evaluación de los maestros.

## **EVALUACIONES DE LOS MAESTROS Y MEJORAMIENTO EDUCATIVO**

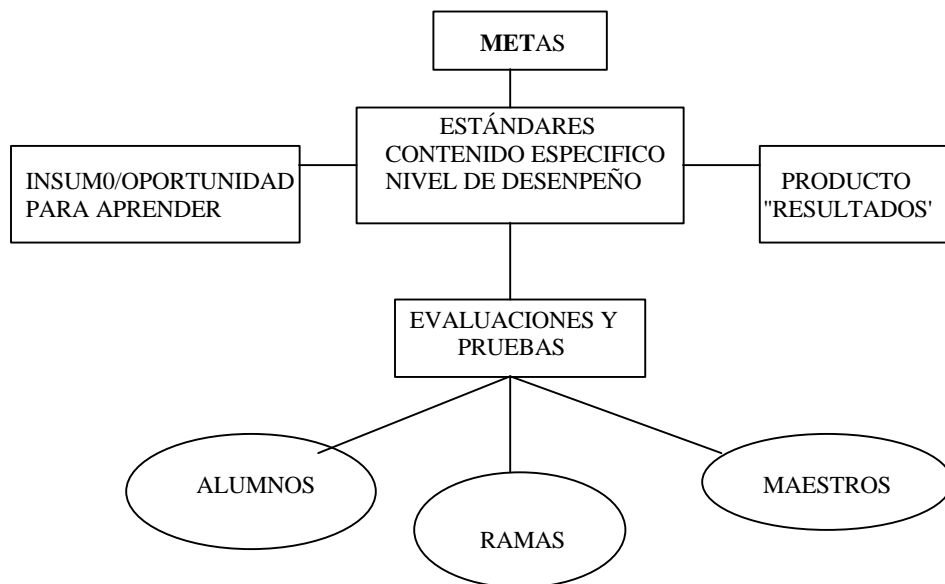
Para ser eficaz en el mejoramiento de la educación, la evaluación de los maestros ha de tratarse como elemento intrínseco del sistema educativo. Este es un principio general de la evaluación eficaz y un importante factor en determinar la validez de las evaluaciones utilizadas. Las evaluaciones de los maestros tienen vínculos importantes con los valores sociales del público en general, los de quienes gobiernan y los de los educadores profesionales. Las evaluaciones de los maestros también están vinculadas con las políticas y prácticas de la formación de docentes; con la adquisición de competencia por parte de los maestros que ejercen la profesión; y con las decisiones acerca de los currículos escolares. Es una condición absolutamente necesaria para las evaluaciones de los maestros y los alumnos, en términos de validez, justo equilibrio, carácter práctico y utilidad, hacer claros y coherentes estos vínculos.

Amén de las necesidades técnicas y éticas que contribuyen a satisfacer una visión unificada de la educación, existe una necesidad práctica que ha de satisfacerse también. Tomando como base muchos años de experiencia como diseñadora y redactora de evaluaciones educativas, he llegado a la conclusión de que los datos obtenidos con las evaluaciones que se tratan al margen de los valores de la sociedad y de la vida de las escuelas y las comunidades en las que funcionan, rara vez serán percibidos como útiles por quienes deberían beneficiarse de ellos. A fin de ser beneficiosas para el sistema educativo en su conjunto, estas evaluaciones han de estar cuidadosamente vinculadas con estas entidades mayores. Las características técnicas de las evaluaciones de los maestros y otros tipos de evaluaciones no nos dicen, de por sí, si valdrá la pena realizarlas. El valor de la información derivada de las evaluaciones será determinado, en última instancia, por el grado en el que se vea influenciado el sistema educativo, bien sea positiva o negativamente.

### **METAS, ESTÁNDARES Y EVALUACIONES**

Para comprender el papel que desempeñan las evaluaciones de los maestros en las mejoras educativas se requiere una clara comprensión de las relaciones y distinciones que existen en el sistema educativo en su conjunto. Forman parte integral de este proceso los tres elementos siguientes: metas, estándares y evaluaciones. Estos tres elementos contribuyen de forma importante al mejoramiento de la educación. Si bien las diferencias en su finalidad, aspecto y utilización pueden parecer evidentes, otras no lo son. Es vital establecer vínculos lógicos entre estos elementos y formular políticas y prácticas que concuerden con esta visión más amplia, pero a menudo son pasos que se ignoran al establecer mecanismos eficaces para planificar, impartir y evaluar la instrucción. La Figura 1 proporciona un cuadro simplificado de las interconexiones entre metas, estándares y evaluaciones en el mejoramiento educativo.

**Figura 1. Proceso de utilización de las evaluaciones para el mejoramiento educativo**



Las metas son inherentemente exposiciones de las aspiraciones. Como tales, tienen tres características que son importantes para el presente análisis: (1) en relación con los estándares y evaluaciones, el acuerdo sobre las metas es fácil de obtener; (2) las metas no siempre se alcanzan completamente y (3) siempre habrá metas educativas importantes, o aspectos de determinadas metas que, por distintas razones prácticas o conceptuales, no se medirán. Independientemente de lo cuidadosa que sea la especificación de las metas, no son, por naturaleza, estándares ni evaluaciones. El proceso de pasar de metas a actividades concretas de evaluación es característicamente complejo y abarca las numerosas cuestiones de los recursos y la ejecución. Así pues, en último análisis, las evaluaciones nunca representarán las metas deseadas con una cobertura completa o una fidelidad absoluta. No obstante, el proceso de establecimiento de metas es una empresa valiosa al igual que lo es un proceso bien concebido de evaluación.

Aun cuando resulta que muchas metas importantes no son directamente medibles,

la existencia de metas concretas sirve como función útil en la coordinación de las actividades y la planificación de la utilización de fuentes de datos indirectas y complementarias para seguir el proceso hacia el logro de las metas. Las metas explícitas también pueden ser útiles para resolver las diferencias acerca de políticas y practicas concretas. Por ejemplo, gran parte del esfuerzo que ahora se está dedicando a mejorar la enseñanza de las matemáticas y las ciencias en los Estados Unidos emana de las comparaciones internacionales del desempeño de los alumnos en estas áreas con resultados desfavorables para los Estados Unidos. Estos datos permitieron a los educadores en los Estados Unidos proponer una meta de mejoramiento en su conjunto que fue aceptable para el público en general y, en el curso del tiempo, obtener recursos considerables para el largo proceso de fijar los estándares, crear las evaluaciones y proporcionar los recursos que se necesitarán para alcanzar esta meta.

Una vez que se ha llegado a un acuerdo sobre las metas, han de establecerse estándares concretos para alcanzarlas. Volviendo a mi ejemplo precedente, el progreso hacia una meta nacional consistente en mejorar la comprensión de las matemáticas y las ciencias por parte de los alumnos requiere establecer estándares curriculares en ambas disciplinas que especifiquen qué deberían aprender los alumnos y qué deberían enseñar los maestros y el nivel en el que debería demostrarse el rendimiento. En los Estados Unidos en nuestros días, este proceso incluye al menos dos perspectivas diferentes: las de cada uno de los estados que aplicarán los estándares y las de los profesionales de la educación, tales como las asociaciones nacionales de materias cuyos miembros en el campo de la docencia, formación de docentes e investigación tienen experiencia curricular apropiada. Los estándares para los maestros también han de formularse y coordinarse con los estándares para resultados de los alumnos, ya que es razonable suponer que los maestros no pueden enseñar con eficacia lo que ellos mismos no conocen. Los estándares sobre resultados para los maestros pueden tener repercusiones sobre el contenido de sus programas de formación de docentes. Los estándares sobre el contenido para los maestros se ocupan de la materia y el conocimiento pedagógico que los maestros deberán poseer y de su capacidad para utilizar este conocimiento en el aula. El insumo, o la oportunidad para aprender los estándares, asegura que los maestros tienen acceso a experiencias educativas continuas, materiales y otros recursos que ellos y sus alumnos necesitan para producir los resultados educativos requeridos.

En los estándares que actualmente se establecen en los Estados Unidos, han surgido varios temas importantes que probablemente sean aplicables a otros lugares. El proceso de establecimiento de estándares no sólo aclara lo que se ha de enseñar y aprender, sino que también eleva el nivel de expectativas acerca del desempeño. Los estándares establecidos en un proceso consensual, que incluya a los profesionales de la educación, tienden a ser refinados, exigentes y progresivos. Si bien la creación de estándares altos constituye un resultado altamente valorado del proceso de establecimiento de estándares en términos de creación de un clima de altas aspiraciones para aprendizaje de los alumnos, también crea cierto número de

problemas prácticos que hay que resolver. Los maestros, las organizaciones profesionales de maestros y los educadores de maestros estarán justificablemente preocupados sobre cómo alcanzarán estos altos estándares. Como cuestión práctica, estarán preocupados acerca de cómo obtener los recursos que pueden requerirse para ello.

Sin embargo, estas preocupaciones no han reducido su interés en formular y aplicar los estándares. Concretamente, muchos maestros no están preparados en la actualidad para enseñar al nivel que implican los estándares, por lo que se necesitará una extensa educación en el desempeño de las funciones para los maestros (y para otros educadores en sus programas de formación de docentes) a fin de cumplir los estándares. Además, si se va a esperar de los maestros en una amplia gama de escuelas y alumnos el mismo nivel alto de desempeño, hay que prestar atención a las disparidades en los recursos (Banks, 1997). Esto significa que los estándares han de preocuparse no sólo del producto final de la enseñanza de los alumnos (producto educativo), sino también de los recursos a disposición de los alumnos y maestros para ayudar en la enseñanza y el aprendizaje (estándares de insumos educativos u oportunidad).

Esta atención a los insumos puede parecer evidente en teoría (si se va a enjuiciar a las personas por lo que han aprendido, deberán tener la misma oportunidad para aprenderlo), pero la ejecución es muy controvertida. Entre las áreas de desacuerdo que han surgido a este respecto figuran quién debería tener la responsabilidad de asegurar la calidad de los insumos y cómo deberían proporcionarse los fondos. Además, han proliferado debates metodológicos en torno al tema de la oportunidad para aprender. Algunos aspectos de los estándares sobre insumos son relativamente fáciles de medir, tales como la disponibilidad de libros y la calidad de las instalaciones escolares. Sin embargo, otros aspectos muy importantes son mucho más difíciles de medir, tales como la calidad de los maestros. En los Estados Unidos, el establecimiento y la aplicación de estándares sobre insumos han sido muy controvertidos incluso en las etapas de planificación y promete seguir siéndolo, dada la amplia variación en la forma en la que se están conceptualizando estos estándares.

Debido a que las evaluaciones emanan directamente de los estándares, en principio deberían crearse sólo después de haber concluido la especificación de las metas y los estándares. En la práctica, no obstante, el proceso tiende a ser iterativo. Por ejemplo, a menudo la primera vez que las aspiraciones de estándares educativos se ponen en operación es cuando se traducen a tareas de evaluación o preguntas de prueba. Aun cuando esto es lamentable en algunos aspectos, la situación puede tener aspectos positivos si existe cierta cantidad de flexibilidad en el sistema educativo. La experiencia con la redacción y el ensayo de evaluaciones puede proporcionar una idea de cómo deberían perfeccionarse los estándares, ya que las evaluaciones tratan de traducir los estándares a indicadores medibles del desempeño mientras que se aspira a preservar la intención original. Cuando los estándares se traducen a tareas y preguntas concretas para incluirlas en una prueba, entonces la

gente critica las tareas como demasiado difíciles, demasiado fáciles, irrelevantes, injustas o políticamente correctas o incorrectas. No obstante, esta crítica puede proporcionar una oportunidad para agudizar y concentrar el significado de metas, estándares, evaluaciones y sus interrelaciones.

Han de establecerse los vínculos entre todos los elementos que contribuyen al logro de metas educativas-alumnos, maestros y currículos-a fin de que las evaluaciones de los maestros tengan máxima eficacia. Es importante que estos vínculos se establezcan al nivel de metas y estándares, no sólo a nivel de evaluación. En esta etapa del establecimiento de estándares en los Estados Unidos, a pesar de la cuestión separada de las dificultades creadas por las numerosas diferencias en formato, idioma, nivel de especificidad e intención, se están identificando numerosos casos de discrepancias sustantivas más amplias que necesitarán resolverse: discrepancias entre los estándares de grupos profesionales, los de los estados y los nacionales; discrepancias entre lo que se está pidiendo a los alumnos que aprendan y a los maestros que preparen; y, lo que es aún más importante, discrepancias entre los requisitos de hoy y de mañana para maestros y alumnos. Estándares

Existen numerosos tipos de estándares que son aplicables a la evaluación de los maestros y hay mucha confusión generada por la falta de atención a los diferentes fines a los que sirven. A continuación presentamos ejemplos de los distintos tipos de estándares.

### *Estándares para validez, calidad técnica y utilización*

La calidad de las evaluaciones de los maestros puede enjuiciarse por los estándares técnicos que se aplican ampliamente a cualquier evaluación educativa o psicológica. La teoría moderna de la validez (por ejemplo, APA, 1985; Cole y Moss, 1989; Messick, 1989; Moss, 1992) subraya el amplio contexto en el que se formulan y utilizan las evaluaciones y proporciona una gula conceptual para considerar la validez de las evaluaciones de actividades complejas tales como la enseñanza. Este criterio de la calidad de la evaluación pone de relieve la importancia de demostrar los vínculos lógicos entre partes del sistema educativo y un examen sistemático de las consecuencias de realizar evaluaciones. Este criterio es ahora ampliamente aceptado por los especialistas en mediciones. La serie de estándares para evaluaciones a los que se hace referencia más ampliamente, *Standards for Educational and Psychological Testing* (Asociación Norteamericana de Investigaciones Educativas, Asociación Psicológica Norteamericana y Consejo Nacional de Medición de la Educación, 1985) refleja este punto de vista al igual que lo hace la revisión de estos estándares que ahora está en vías de realización (la fecha prevista para conclusión es 1998). No obstante, estos estándares son tan generales que dejan mucho a la discreción de quienes elaboran y utilizan las evaluaciones de los maestros. Entre las publicaciones adicionales sobre estándares y expectativas técnicas en esta esfera figura la publicación del Comité Conjunto sobre Estándares de Evaluación Educativa titulada *Personnel Evaluation Standards* (1988). Estos

estándares incluyen no sólo la atención a la calidad de las evaluaciones sino también a las características del contexto educativo y el empleo ético de las evaluaciones.

La junta Nacional de Estándares de Enseñanza Profesional (NBPTS) también ha ofrecido una perspectiva sobre los estándares técnicos y otros específicos para la evaluación de la enseñanza que han tenido influencia en los recientes debates de las evaluaciones de los maestros. La NBPTS ha aplicado a sus propias actividades el estándar de que: "Las evaluaciones han de ser profesionalmente creíbles, públicamente aceptables, legalmente defendibles, administrativamente fiables y económicamente asequibles" (NBPTS, 1991, página 53).

Hay una creciente disponibilidad de los estándares propuestos relacionados específicamente con la calidad técnica de las evaluaciones del desempeño aunque ninguno ha alcanzado aún el nivel de aceptación general de los estándares de la AERA/APA/NCME o el Comité Conjunto. Algunos estándares propuestos relacionados con los distintos aspectos de la evaluación del desempeño de los maestros han sido ofrecidos por Claxton, Murrell y Porter (1987); Linn, Baker y Dunbar (1991); Moss (1992); Miller y Legg (1993); y Quellmalz (1991). Dwyer (1994) brinda un análisis adicional de esta literatura y su aplicación a las evaluaciones de los maestros.

### *Estándares curriculares*

Los estándares curriculares se refieren a series de declaraciones de conocimiento y aptitudes que han de aprender los alumnos, presentados de ordinario en un marco que indica el alcance del contenido que se tratara y la secuencia en la que se ha de enseñar por lo general el material. Los estándares curriculares son pertinentes para las evaluaciones de los maestros en dos formas. Primero, series de currículos de los alumnos, tales como las elaboradas por el Consejo Nacional de Maestros de Matemáticas (1989) al definir lo que han de aprender los alumnos, tienen repercusiones directas para lo que han de conocer y ser capaces de enseñar los maestros. Por implicación, dichos estándares curriculares orientados al alumno también guardan relación con las actividades de desarrollo profesional de los maestros que se necesitan. Segundo, los estándares curriculares pueden formularse también para la formación de docentes. Dichos estándares se refieren directamente al contenido, al conocimiento y a las aptitudes pedagógicas específicos para el contenido que han de poseer los maestros. Para que dichos estándares sean eficaces deberían vincularlos con el conocimiento y las aptitudes que han desarrollarse en los alumnos.

### **Estándares para conocimiento y aptitudes de los maestros**

Tal como se indicó arriba, los estándares curriculares pueden especificar directa o indirectamente el conocimiento y las aptitudes que se requerirán de los maestros. Los estándares para el conocimiento y las aptitudes de los maestros están vinculados conceptualmente con estos estándares curriculares, pero abordan la cuestión de

forma que (1) guarda relación directamente con los maestros y (2) especifica conocimiento y aptitudes que trascienden del conocimiento relacionado con el currículo de determinadas disciplinas. Los estándares para el conocimiento y las aptitudes de los maestros pueden abordar el conocimiento del contenido de la materia por parte de los maestros; el conocimiento por los maestros de cómo enseñar ese contenido (pedagogía específica para el contenido); el conocimiento por los maestros de principios pedagógicos generales (por ejemplo, desarrollo del idioma, recompensas y castigo); y la capacidad de los maestros para aplicar este conocimiento y estas aptitudes en un entorno de aula (desempeño pedagógico).

Según Bridges (1986), en la práctica, las escuelas evalúan a los maestros ya empleados de acuerdo a cinco criterios amplios: (1) conocimiento de la materia; (2) capacidad de impartir el conocimiento; (3) capacidad de mantener la disciplina en el aula; (4) capacidad de mantener un clima apropiado en el aula y (5) capacidad de establecer una relación con los padres y alumnos. No obstante, como regla general, el personal escolar no tiene una comprensión concreta del significado de estos criterios y, por tanto, con frecuencia los maestros no están seguros de lo que se espera de ellos o cómo mejorar su actuación. En el marco escolar, la determinación de la competencia de los maestros se realiza a menudo a través de las observaciones de los rectores u otro personal supervisor, aunque eso no constituya el mejor método para obtener información sobre todas las cinco áreas identificadas por Bridges. Es raro que el aprobar una prueba estandarizada sea uno de los criterios para seguir en el empleo pero así ha ocurrido en varias jurisdicciones en los Estados Unidos.

Las evaluaciones de los maestros para fines distintos de la evaluación anual de los ya empleados presentan diferentes opciones de evaluación. Se ha creado un ejemplo de un enfoque diferente para uso en certificar que los maestros cumplen estándares profesionales muy altos y son profesionales excelentes. La NBPTS ha diseñado evaluaciones complejas e integrales, para los maestros experimentados, que están vinculadas a niveles de edades concretas de los alumnos y a áreas de materias concretas. Las evaluaciones de la NBPTS son, característicamente, una combinación de métodos de evaluación (fuertemente orientados hacia la evaluación del desempeño) que son específicos para la enseñanza en un determinado contexto, tal como las aptitudes del lenguaje de los primeros años de la adolescencia. Las evaluaciones son externas en el sentido de que son enjuiciadas por evaluadores especialmente capacitados que son educadores experimentados y no forman parte del entorno de empleo del maestro. Las evaluaciones de la NBPTS pueden obtenerse ahora en ciertas áreas de materias básicas y edades que incluyen a un gran número de maestros; otras están aún en proceso de desarrollo. Además del desarrollo de evaluaciones, la NBPTS también ha realizado una gran cantidad de investigaciones en cierto número de áreas relacionadas con el uso de simulaciones para la evaluación de los maestros (por ejemplo, desempeño grabado en videocinta; el uso de portafolios de trabajo; ejercicios en centros de evaluación; y el impacto de las evaluaciones de la NBPTS sobre los maestros que participan en ellas).

Otro modelo, elaborado originalmente para la evaluación de los maestros que comienzan, pero que se ha ampliado ahora para incluir también a los maestros experimentados, es el de la Serie de Práctica elaborada por el Servicio de Pruebas Educativas Dwyer, 1994; Dwyer y Ramsey, 1995; Dwyer y Villegas, 1993). La Serie de Práctica ofrece actividades de desarrollo profesional y evaluación que utilizan distintos métodos de evaluación para recopilar datos para fines de toma de decisiones personales e institucionales. Las áreas de evaluación son:

- Aptitudes básicas (lectura, escritura y matemáticas) requeridas de todos los posibles maestros para el éxito en la formación de docentes y buen desempeño posterior.
- Conocimiento de la materia y conocimiento pedagógico específico para el contenido. Estas evaluaciones utilizan métodos de recopilación de datos por escrito, basados en computadora, orales y de otra índole.
- Conocimiento de los principios Pedagógicos básicos (utilizando métodos de recopilación de datos tales como los descritos arriba).
- Aplicación de conocimientos y aptitudes en el aula. Estas evaluaciones son predominantemente evaluaciones del desempeño y dependen de la observación en la propia aula del maestro, las entrevistas con el maestro y los documentos de recopilación de datos.

### *Estándares de nivel de rendimiento*

Además de especificar el "qué" en los estándares de enseñanza y aprendizaje, también ha de abordarse la cuestión de "cuánto" Los estándares de conocimiento y aptitudes curriculares y del maestro abordan a menudo esta cuestión por implicación y con menos frecuencia la abordan directamente. Un ejemplo de abordar los estándares de nivel de conocimiento y aptitudes por implicación, incluido el cálculo como elemento de los estándares de matemáticas, implicarla de ordinario un nivel más elevado de resultados de desempeño que los estándares que especificaban sólo el contenido de matemáticas enseñado habitualmente antes del cálculo.

No obstante, la cuestión del nivel de estándares de desempeño se aborda de forma apropiada, independientemente de dichas consideraciones. Los estándares pueden establecerse en tres niveles (individualmente o en combinación). Por ejemplo, puede establecerse un estándar general para determinar, sobre una base de si o no, si un maestro reúne las condiciones para enseñar o para que se le renueve un contrato de empleo. A un nivel más detallado, los estándares pueden establecerse para determinar si un maestro posee conocimiento adecuado de un aspecto en particular de la literatura o la capacidad de mantener la disciplina en su aula.

Dwyer (1997) proporciona un análisis general de los principios fundamentales utilizados para establecer dichos estándares. La naturaleza de las puntuaciones límite no está en hallar la puntuación "correcta", sino en el enjuiciamiento de uno o más individuos habilitados o autorizados acerca de la cuestión "¿Cuánto es

suficiente?" La respuesta no está en la prueba o en el método utilizado para establecer la puntuación límite. Esta es una pregunta que sólo puede responderse con respecto a factores que trascienden de la propia prueba, con referencia al contexto en el que se utiliza la prueba y las percepciones de los jueces acerca de dicho contexto.

En las evaluaciones de los maestros, tales estándares pueden ser muy controvertidos. El mismo estándar puede ser considerado simultáneamente por personas diferentes como (1) tan alto que representa una barrera desmedida del empleo de personas calificadas que se convertirían en maestros aceptables; o (2) tan bajo que presta un mal servicio a los alumnos al permitir a personas no calificadas para enseñar. Dwyer advierte que esta situación no resulta de haber seleccionado la puntuación límite "incorrecta" o de deficiencias técnicas de la prueba en la que se utiliza la puntuación límite. Resulta de claros desacuerdos acerca de la importancia relativa de diferentes clases de error de clasificación y acerca de la importancia relativa de las políticas educativas tales como la disponibilidad de maestros, la elevación de los estándares educativos y las preocupaciones acerca de la equidad para maestros y alumnos.

Tres aspectos centrales y mutuamente relacionados acerca de tales estándares son pertinentes para la evaluación de los maestros:

1. Todos los métodos de establecimiento de puntuaciones límite dependen de juicios. Los juicios pueden efectuarse acerca de las personas o de las preguntas de la prueba (u otros aspectos de la propia prueba), pero el juicio con influencias independientes de la prueba es una característica intrínseca de cualquier puntuación límite.
2. El establecimiento de puntuaciones límite conducirá invariablemente a errores al determinar si una persona ha cumplido o no el estándar. Las puntuaciones límite imponen casi siempre diferenciaciones externas sobre una distribución continua. Muy pocas evaluaciones de cualquier clase que sean pueden distinguir de forma fiable entre personas con puntuaciones adyacentes, pero, al aplicar una puntuación límite se obliga en realidad a efectuar dicha distinción.
3. No existe una puntuación límite "verdadera" que pueda hallarse con la aplicación del método correcto o una muestra de jueces suficientemente grande. La determinación de una puntuación límite es similar a la estimación de un parámetro de población. El uso de muestras mayores de jueces, o jueces mejor capacitados, en estudios destinados a establecer puntuaciones límite mejorara ciertos aspectos técnicos concretos (Por ejemplo, reducir el error de muestreo (Lawshe, 1975), pero esto no resultará en una puntuación límite que sea superior a las puntuaciones límite que pudieran haberse establecido con otra selección de jueces calificados o con otra metodología igualmente justificable para la selección de la puntuación límite.

## *Estándares de oportunidad para aprender*

En un contexto que se ocupa principalmente de estándares para los alumnos, Banks (1997) advierte la importancia de la oportunidad al especificar los estándares. Indica que, al establecer los estándares educativos nacionales *Goals 2000* en los Estados Unidos, entre las clases de estándares considerados figuraron estándares de contenido, de desempeño, de evaluación y de oportunidad para aprender. Los estándares de oportunidad para aprender fueron los más controvertidos de las cuatro clases. Estos últimos abordan condiciones en las escuelas y las comunidades que limitan la capacidad de los alumnos y maestros para alcanzar los otros tipos de estándares. Entre los ejemplos de dichas variables en cuanto a insumos figuran la calidad de las instalaciones escolares, la disponibilidad de materiales docentes y el nivel de experiencia pedagógica disponible en las escuelas. La Figura 1, que antecede, muestra el papel de los estándares de oportunidad en el proceso de mejoramiento de la educación. A pesar de la lógica para tener dichos estándares, las críticas que implican de los mecanismos actuales de apoyo de las escuelas los hacen con frecuencia poco apetecibles por razones políticas. Los detractores del empleo de dichos estándares argumentan que reducen el énfasis en los resultados de desempeño y la responsabilidad al concentrarse de forma inapropiada en los recursos y los insumos.

Cabe advertir también que, en el caso especial del *licenciamiento* de los maestros que comienzan a ejercer la profesión, el concepto de la oportunidad para aprender no es relevante. En los contextos de licenciamiento, la decisión que ha de hacerse es la de si el maestro posee el conocimiento y las aptitudes necesarios para ser considerado apto en situaciones de enseñanza concretas. Debido a que estas situaciones son aquellas en las que se hace hincapié en la protección del público contra el daño que puede producir un posible maestro incompetente, se pide al posible maestro que demuestre la posesión del conocimiento y las aptitudes estipulados. El método mediante el cual se obtuvieron esas aptitudes o las circunstancias para obtenerlas o no obtenerlas no es estrictamente pertinente en este contexto limitado. Si se desean antecedentes adicionales sobre esta cuestión, véase Shimberg (1985).

## **FINES DE LA EVALUACION**

El proceso integrado que acabo de describir de pasar de metas a evaluaciones puede coadyuvar al logro de cierto número de objetivos en la evaluación de los maestros. Lamentablemente, por lo general ocurre que la finalidad de la evaluación no es universalmente clara para todas las partes interesadas cuando se adopta inicialmente la decisión de crear evaluaciones de los maestros. En la práctica, a menudo hay diferentes puntos de vista acerca de la finalidad de las evaluaciones, aun después de haberse comenzado a utilizarlas. Un paso esencial en el diseño de la evaluación consiste en adquirir una clara comprensión de las razones para la evaluación y los usos que se harán de los datos de la misma, pero, equivocadamente, este paso se da a menudo por sentado. Así pues, es importante ser claros acerca de los fines

perseguidos con las evaluaciones de los maestros.

### *Fines de la evaluación*

Diferentes puntos de vista sobre la finalidad de las metas de la mejora educativa entrañan a menudo diferentes puntos de vista de las evaluaciones de los maestros. Dichos puntos de vista sobre las evaluaciones de los maestros pueden ser bastante diferentes con respecto a la sustancia de las evaluaciones; los defensores y detractores de las evaluaciones; la razón fundamental para las evaluaciones; y los enfoques prácticos para las evaluaciones (Dwyer y Stufflebeam, 1996). Entre algunas de las metas principales de las evaluaciones de los maestros que con frecuencia se observan en la práctica y se citan en la literatura, figuran las siguientes:

- Meioramiento de la enseñanza en el aula. Los educadores profesionales son importantes defensores de este punto de vista. Entraña un desarrollo educativo continuo a lo largo del cual una persona puede mejorar, una preferencia por evaluaciones formativas en vez de sumativas, y un fuerte vínculo con las actividades de desarrollo profesional. Si se desea un análisis adicional de esta materia, véanse Hunter (1988); Duke y Stiggins (1990); y Shulman (1986).
- Responsabilidad y desarrollo profesionales. Los maestros y sus asociaciones profesionales son los defensores principales de este punto de vista. Entraña una fuerte visión de la enseñanza como profesión con sus propios estándares, ética e incentivos intrínsecos para la persona consagrada. Aun cuando las responsabilidades son un elemento clave en este punto de vista, la responsabilidad es a la profesión y a sus estándares de práctica y ética, más que a entidades externas tales como empleadores o el estado. Véase un análisis adicional de esta materia en la junta Nacional de Estándares Docentes Profesionales (1991).
- Control administrativo. Los administradores de escuelas son los defensores principales de este punto de vista. Entraña considerar la enseñanza como una situación de empleo que requiere supervisión y control del maestro por la unidad administrativa. En el ámbito de las escuelas públicas en los Estados Unidos, la base para este criterio se fundamenta en la protección del público contra práctica negligente. Para un análisis adicional de esta materia, véanse Andrews (1985) y Redfern (1963, 1980).
- Pago por mérito: En este punto de vista, que puede considerarse como subserie de la responsabilidad y el control administrativo que describimos arriba, se estima que los maestros necesitan el reconocimiento y la motivación que proporcionan los incrementos salariales. Los defensores principales de este punto de vista son el público en general y los funcionarios del gobierno que lo representan. Los defensores de este punto de vista desean a menudo utilizar el rendimiento de los alumnos como indicador del mérito para el que los aumentos

de sueldo sirven de recompensa. La aplicación de este criterio puede entrañar la creación de una escala profesional con una serie de pasos vinculados con el desempeño. Si se desea un análisis adicional de esta materia, véanse Webster, Mendro y Almaguer (1993).

### *Responsabilidad y mejora educativa*

El seguimiento/responsabilidad, el mejoramiento y la instrucción y la evaluación programática son los tres fines más frecuentemente citados para la evaluación de los maestros en los Estados Unidos en nuestros días. Aunque están estrechamente relacionados entre si conceptualmente, la evaluación para estos fines difiere bastante en diseño y en sus repercusiones en cuanto a la forma en que se utilizarán los datos obtenidos con estas evaluaciones. Por ejemplo, la evaluación para responsabilidad y evaluación programática puede utilizar una muestra de personas y preguntas o ejercicios de evaluación en vez de pedir a cada individuo que complete cada una de las partes de la evaluación. En contraste, las evaluaciones concebidas directamente para mejoramiento de la instrucción se administran de forma más apropiada a cada maestro o alumno y deberían estar estrechamente vinculadas con las áreas del currículo que son más importantes para estos alumnos y sus maestros.

La responsabilidad es, por tanto, no una meta en si sino, más bien, un medio hacia el fin del mejoramiento de la educación. A menos que se planifiquen con cuidado las consecuencias del buen y el mal desempeño de los maestros y, a menos que el sistema de evaluación sea justo de hecho y en aspecto, la evaluación de la responsabilidad puede, en realidad, subvertir su mira pretendida de mejoramiento de la instrucción. El uso generalizado de las evaluaciones para responsabilidad de los maestros en los Estados Unidos ha creado gran cantidad de controversia. La investigación ha demostrado que, sin planes cuidadosos para recopilación y utilización de los datos, pueden obtenerse resultados que chocan directamente con sus fines originales. Por ejemplo, los maestros y administradores que no valoran el contenido de las evaluaciones pueden recurrir a medios ilícitos para obtener puntuaciones aceptables.

Surgen problemas tanto éticos como prácticos cuando el contenido de las evaluaciones no se vincula estrechamente con aspectos importantes de la enseñanza. En tales circunstancias, el tiempo empleado en la preparación de las evaluaciones se considerará con razón como tiempo perdido. Cuando estos vínculos están ausentes entre las evaluaciones y las metas y estándares educativos más amplios, hay también razón para preocuparse de que «se obtiene lo que se evalúa»-es decir, el tiempo y la atención dedicados a los temas que aparecen en las evaluaciones, a expensas de otros temas más importantes que no aparecen. Como cuestión a este respecto se cita a veces el formato de las evaluaciones. Una dependencia excesiva en una forma única de pruebas puede hacer que los maestros (o los alumnos) adquieran habilidad en la clase de reflexión que requiere esa forma de pruebas a expensas de otras clases.

Para establecer fuertes vínculos entre la responsabilidad de los maestros y el

mejoramiento educativo se requiere prestar atención a los mecanismos y recursos de que se dispone para resolver los problemas cuando los revelan los datos de las evaluaciones. La correspondencia entre las metas y los estándares educativos, por un lado, y el formato y el contenido de las evaluaciones de los maestros, por el otro, ayuda a los responsables de la política educativa y al público a comprender claramente dónde existen problemas, lo que a su vez puede sugerir sus soluciones.

Diferentes fines para las evaluaciones de los maestros implican diferentes estándares, métodos de evaluación y énfasis. Un factor importante en la evaluación de los maestros para fines relacionados con las decisiones administrativas, tales como la responsabilidad y el pago por mérito, es la existencia de escaseces reales o potenciales de maestros. Rara vez se hallarán útiles las evaluaciones que exacerban las escaseces de maestros en áreas vitales. Históricamente en los Estados Unidos, dichas escaseces han ocurrido en materias técnicas tales como las matemáticas, en las que la educación compite, de ordinario sin éxito, con la industria por personas cualificadas. Para análisis de éstas y otras cuestiones afines sobre la oferta y demanda de maestros, véanse Murnane y Schwinden (1989) y Sedlak y Schlossman (1986).

#### *Consideraciones especiales para las evaluaciones de los maestros principiantes*

Las evaluaciones de los maestros principiantes y experimentados difieren típicamente en términos de las metas que se pretende alcanzar con las evaluaciones y en la forma que adoptan estas últimas. Esto no quiere decir que no estén relacionadas ambas; un distintivo de una evaluación cuidadosamente planificada, de cualquier clase que sea, es el cuidado con el que se ha alineado con etapas previas o posteriores en la carrera del maestro. Una visión común del maestro debería regir la evaluación y su contenido en todas las etapas.

Las evaluaciones de los maestros que comienzan su carrera tienen de ordinario el fin primordial de proteger los intereses de los alumnos y del público de los efectos nocivos de una docencia subestándar. Dichas evaluaciones están orientadas característicamente a asegurar que el posible maestro posee el conocimiento y las aptitudes que le cualificarían para empleo como maestro pero, de ordinario, no garantizarían a la persona un cargo como maestro (Madaus y Mehrens, 1990; Mehrens, 1987; Rebell, 1990). El conocimiento y las aptitudes abarcados se consideran indispensables pero no necesariamente suficientes para fines de determinar la idoneidad de un cargo docente en particular.

En los Estados Unidos, las licencias de los maestros se otorgan típicamente bajo la autoridad de un gobierno estatal o una entidad que ha sido habilitada para desempeñar esta función.

Las evaluaciones de la Práctica para el licenciamiento de los maestros que comienzan su carrera son un ejemplo de las pruebas impartidas para tal finalidad.

En contraste con las evaluaciones de maestros experimentados, las evaluaciones de los maestros principiantes abarcan aptitudes habilitadoras de lectura, escritura y matemáticas. El conocimiento de la materia y la pedagogía específica para el contenido en las evaluaciones de los maestros principiantes concuerda con las expectativas de los maestros experimentados. La suficiencia en la aplicación del conocimiento y las aptitudes en el aula es de naturaleza análoga a la que se espera de los maestros experimentados, pero a un nivel inferior.

Un aspecto crítico de la evaluación de los maestros que comienzan su carrera es el de que las evaluaciones se utilizan para adoptar una decisión dicotoma. Aunque sirven también para otros fines, las evaluaciones de los maestros principiantes resultan característicamente en una decisión afirmativa o negativa acerca de la capacidad del posible maestro para avanzar en la profesión docente. Una implicación de medición de esta característica de las evaluaciones de los maestros principiantes es la de que la evaluación ha de estar concebida para facilitar esta decisión dicotoma, en vez de describir graduaciones de desempeño igualmente bien a lo largo de todo el continuo del conocimiento y aptitudes docentes. En la práctica, esto significa que los diseñadores de la evaluación han de comprender las cualidades que diferencian al candidato aceptable del candidato no aceptable y proporcionar tareas de evaluación que permitan con exactitud hacer diferenciaciones en el desempeño a este nivel en particular, en vez de a través de una gama más amplia de desempeño.

### *Crítica de la evaluación de los maestros*

La evaluación de los maestros ha sido un área de las pruebas educativas especialmente controvertida casi desde sus comienzos a principios de la década de 1930. Las críticas de las evaluaciones de los maestros han diferido durante los años y con fines y métodos diferentes de evaluación.

En general, las críticas de la evaluación de los maestros principiantes para entrada en la profesión han subrayado a menudo la imparcialidad de los instrumentos y procedimientos, en especial con respecto al desempeño de los posibles maestros que son miembros de grupos minoritarios (por ejemplo, Darling-Hammond, 1986; Garcia, 1985; Haney, Madaus y Kreitzer, 1987; Hood y Parker, 1991), alegando que las prácticas de evaluación de los maestros se llevan a cabo a menudo de forma poco uniforme que infringen los derechos de quienes son evaluados. Las críticas basadas en el daño hecho a los alumnos por la falta de conocimientos docentes del maestro son bastante raras en esta esfera.

Las críticas de la evaluación de los maestros en el ejercicio de su profesión para fines de responsabilidad, pago por méritos, etc., también han suscitado preocupaciones acerca del impacto real o potencial de las prácticas de evaluación de los maestros sobre los derechos de los posibles maestros a licencias docentes o acceso a empleo; sobre los currículos de las instituciones de formación de docentes; y sobre el ejercicio de la propia profesión docente. La mayor parte de las críticas que aparecen

en la literatura profesional guarda relación con el uso de pruebas de opciones múltiples para el conocimiento de la materia. Desde fines de la década de 1980, con el mayor uso de las evaluaciones de desempeño tanto para maestros como para alumnos, las críticas en la literatura profesional han tendido cada vez más a concentrarse en los aspectos prácticos de las evaluaciones del desempeño. Dwyer y Stufflebeam (1996) han identificado ocho cuestiones principales en las evaluaciones de los maestros que guardan relación principalmente con el uso de las evaluaciones para la selección y el empleo de los maestros:

- El uso de sistemas de evaluación no validados.
- El uso insuficiente de estándares profesionales para la planificación y el mejoramiento de los sistemas educativos.
- Opciones ineficaces de criterios claros, válidos y aplicables para determinar el desempeño de los maestros.
- La necesidad de técnicas y materiales para realizar los pasos básicos de la evaluación de los maestros.
- La falta de capacitación en evaluación para los evaluadores.
- La falta de gula sobre la evaluación de maestros de poblaciones especiales, tales como alumnos con impedimentos o los que necesitan educación bilingüe.
- La no consideración del contexto del aula y la escuela en la realización de las evaluaciones.
- La ausencia de una base teórica para las evaluaciones.

A pesar del apoyo profesional generalizado para la validación de las evaluaciones de los maestros (por ejemplo, Linn y colaboradores, 1989; Madaus, 1990), la mayoría de las evaluaciones de los maestros utilizadas por las escuelas no han sido rigurosamente validadas (Burry, Chisholm y Shaw, 1990; Seriven, 1987; Streifer e Iwanicki, 1987).

En una crítica relacionada con las dos últimas cuestiones de esta lista, algunos detractores, reconociendo la complejidad de la enseñanza, están específicamente preocupados con la factibilidad de determinar el conocimiento pedagógico junto con la comprensión de las materias. Las preguntas de opciones múltiples que tratan de evaluar la aplicación del conocimiento pedagógico fracasan en la cuestión del contexto: lo que constituye buena docencia varía con la materia que se está enseñando y con los antecedentes y las características individuales de los alumnos a los que se enseña. En ausencia de información muy detallada acerca del contexto del aula, que es imposible en términos prácticos en el formato de opciones múltiples, no se puede llegar a una conclusión con certeza acerca de lo que constituye instrucción apropiada. Las evaluaciones de las actitudes docentes, en contraposición al conocimiento necesario de las materias, necesita por tanto depender de formas más complejas de recopilación de datos.

Las evaluaciones del desempeño de los maestros principiantes creadas por los estados, aunque son nuevas en relación con las pruebas de opciones múltiples, no

han escapado a su porción de críticas. Dichos sistemas se han basado en observaciones de baja inferencia del comportamiento de los maestros; es decir, la tarea de observación se limita a comportamientos claramente observables que requieren un bajo nivel de inferencia por parte de los evaluadores y, por tanto, puede preverse que presenten un alto nivel de acuerdo entre evaluadores. Aunque este alto nivel de acuerdo es una característica deseable de las calificaciones, no es tan deseable que se tenga que obtener en aras de la validez de las calificaciones. Las calificaciones de baja inferencia, por su propia naturaleza, no pueden hacer frente a las facetas importantes pero inobservables de la enseñanza, tales como la toma de decisiones del maestro. Muchos detractores de las evaluaciones de los maestros basadas en el desempeño han indicado la deficiencia de los sistemas de baja inferencia-tanto los basados en las simulaciones como en las interacciones directas en el aula-que no toman en cuenta el contexto del aula como parte de la evaluación.

Otras críticas de la evaluación de los maestros guardan relación con el impacto de la evaluación de los maestros sobre los currículos de formación de docentes (Milner, 1991) y sobre la propia profesión pedagógica (Darling-Hanimond, 1986). La preponderancia de esta crítica concierne al empleo de pruebas estandarizadas de opciones múltiples del conocimiento de la materia y pedagógico, más que a las evaluaciones del desempeño relacionadas con la capacidad de aplicar este conocimiento en el aula. Una cuestión importante para comprender las críticas de las evaluaciones de los maestros, y para formular políticas que eviten críticas legítimas, es hacer una distinción uniforme entre las dos clases de pruebas. Lamentablemente, las evaluaciones de las aptitudes básicas de habilitación (lectura, escritura y matemáticas) y el conocimiento de la materia, la pedagogía específica para el contenido y las aptitudes pedagógicas generales se tratan con demasiada frecuencia como si fuesen intercambiables con las evaluaciones del desempeño en el aula. Las inferencias hechas sobre la base del conocimiento de un maestro atañen a una condición necesaria *pero no suficiente* para una enseñanza eficaz en el aula. Si no se hace esta distinción simple pero vital se oscurecerán muchos debates acerca de las políticas y prácticas de evaluación de maestros.

Muchas de las deficiencias documentadas de las evaluaciones de los maestros pueden atribuirse a la falta de recursos, tales como procedimientos de capacitación adecuados para los evaluadores, instrumentos de evaluación e información procedente de estudios de validación. Sin embargo, otras críticas son el resultado del estado actual de la teoría y práctica de la evaluación de maestros. La ausencia de bases teóricas es un problema importante para muchos sistemas al evaluar tanto a los maestros principiantes como a los experimentados (Scriven, 1988a, 1988b, Wise, Darling-Hammond, McLaughlin y Bernstein, 1984). Lamentablemente, la literatura sobre investigación de la docencia ofrece poca orientación sobre cómo utilizar su teoría para formular y aplicar mejores evaluaciones de los maestros (Dawyer, 1994)\_

## DEFINICION DE LA ENSEÑANZA

Un reto importante en la creación de una serie bien formulada de evaluaciones de los maestros es la tarea de determinar qué constituye una buena enseñanza y qué estándares deberían utilizarse para determinar si se ha medido ésta de forma eficaz. La idoneidad de medir qué debería enseñarse no puede determinarse significativamente sin formular estándares explícitos, en un marco de referencia integral que abarque cuestiones tradicionales de preocupación para la educación, psicología y medición. Esta evaluación también ha de estar enmarcada en un contexto en relación con un determinado punto de vista de la enseñanza y el aprendizaje. Todos estos factores constituyen una parte apropiada de un concepto de la enseñanza que se utiliza para orientar decisiones posteriores acerca de los estándares y las evaluaciones. Las conclusiones acerca de la calidad de las evaluaciones, a su vez, han de guardar relación con el concepto de la enseñanza y los fines de las evaluaciones. Algunos acontecimientos recientes en el campo de las mediciones apoyan decididamente este punto de vista.

La teoría moderna de la validez, en su énfasis en un contexto amplio para el establecimiento de la validez de las evaluaciones, proporciona orientación conceptual al considerar la validez de las evaluaciones en desempeños complejos tales como la docencia y en los aspectos técnicos de la determinación del contenido. Dicho trabajo sobre la teoría de la validación pone de relieve las repercusiones para las mediciones de las interconexiones que existen en el seno de todo el sistema, del cual la evaluación es una parte, y el valor inherente en la medición directa del desempeño cuando esto es factible. Con el énfasis en el contenido y las consecuencias, las evaluaciones se examinan para determinar en términos amplios su beneficio para el sistema educativo del que forman parte. Por ejemplo, en la evaluación de los maestros, el contexto amplio de la evaluación de un maestro que comienza a enseñar incluirla factores tales como el distrito escolar, el estado y las políticas federales que afectan al ejercicio por el maestro de su profesión; el marco de la escuela y aula en el que ocurre la enseñanza; los recursos materiales a disposición del maestro; la naturaleza del currículo y la lección que se ha de enseñar; las características y el conocimiento previo de los alumnos; y la calidad de la preparación del maestro para la enseñanza.

Las consecuencias del empleo de las evaluaciones de los maestros se concentrarían análogamente en el impacto del empleo y resultados de la evaluación sobre la política educativa (¿Se crean nuevas políticas, basadas en los resultados de la evaluación? En caso afirmativo, ¿puede demostrarse que estas nuevas políticas son beneficiosas para los alumnos?); los materiales e instalaciones a disposición de los maestros y alumnos (¿Se identifican las necesidades materiales a través de la evaluación y, luego, se satisfacen dichas necesidades? ¿Desvía el proceso de evaluación la atención de las necesidades materiales de forma que se agravan dichas necesidades?); la calidad de la preparación de los maestros (¿Proporciona la evaluación información suficiente para los programas de formación de docentes a

fin de ayudar a los maestros a mejorar su programa?); la calidad del currículo (¿Conduce el proceso de evaluación y los resultados a la identificación de áreas del currículo que funcionan bien y otras que no funcionan bien?); y el aprendizaje de los alumnos (¿Conduce la evaluación de los maestros a un aprendizaje mejorado o interfiere con el desempeño en el aula de forma que los estudiantes aprenden menos?).

Si bien ahora hay una base teórica ampliamente aceptada para un criterio amplio y generalizado de la validez, incluyendo su énfasis en las consecuencias de la evaluación, y en realidad está codificado en la revisión más reciente de *Standards for Educational and Psychological Testing* (Asociación Norteamericana de Investigación de la Educación, Asociación Psicológica Norteamericana y Consejo Nacional sobre Medición en la Educación, 1985), hay aún un vacío considerable entre la literatura sobre validez y la serie diversa e interesante de cuestiones que se afrontan en la formulación de evaluaciones de los maestros, en particular cuando tienen lugar en ambientes en los que es mucho lo que está en juego y cuando se utilizan medios múltiples de recopilación de datos. Por ejemplo, los investigadores y elaboradores en esta esfera carecen de guía definitiva de la literatura sobre cuestiones tan importantes como las variables humanas y la materia en el aula en las evaluaciones de los maestros (cómo hacer frente a diferencias contextuales); y sobre las repercusiones para la validez de crear evaluaciones que toman como punto de partida el criterio de que el aprendizaje es un proceso activo consistente en construir significado a partir de experiencias previas (evaluaciones con una base constructivista). Estas cuestiones son vitales para determinar los criterios mediante los cuales se evaluará el desempeño docente pero se requiere llenar un vacío para pasar de la base teórica sólida en la literatura de investigación a la práctica real de investigación y desarrollo.

El cúmulo de publicaciones sobre validez, estándares, expectativas y aspiraciones si bien pone de relieve importantes cuestiones conceptuales y prácticas crea una pesada carga de interpretación y extrapolación para los responsables de formular las evaluaciones de los maestros, en particular quienes incluyen la atención al desempeño real en el aula. Es evidente que existe un acuerdo general en estas publicaciones en el sentido de que los responsables de elaborar las evaluaciones y los evaluadores deberían tomar en cuenta las consecuencias amplias de las evaluaciones; incorporar elementos del contexto en el proceso de evaluación; concentrarse explícita y exactamente en el conocimiento y aptitudes acerca de los cuales se desea sacar inferencias; e incluir en las evaluaciones toda la gama de contenido acerca de qué inferencias deben sacarse. A pesar de este alto nivel de acuerdo en principio, necesariamente se dejan a cada uno de los creadores de las evaluaciones y a quienes evalúan sus esfuerzos una gran cantidad de discreción y responsabilidad. La base para sus decisiones, si se considera cuidadosamente, puede ser un recurso importante para alinear las evaluaciones de los maestros con otros elementos del sistema educativo a fin de mejorar los resultados de la educación y alcanzar otras metas importantes.

Considerados como proceso, los elementos principales en la definición de la enseñanza para fines de evaluación son, entre otros, los siguientes:

- Determinar un concepto rector de la enseñanza.
- Formular un plan metodológico integral para definir la enseñanza.
- Vincular la definición de la enseñanza con las prácticas de evaluación.

### *Concepto rector de la enseñanza*

Es de vital importancia para la formulación de evaluaciones eficaces de los maestros formular un concepto que guíe la enseñanza y el aprendizaje el cual reconozca explícitamente la conexión entre enseñanza y aprendizaje. Es imposible debatir qué es fundamental acerca de la primera sin considerar el segundo.

Un concepto eficaz que oriente la enseñanza debería poseer las siguientes características:

- Ser explícito y claro acerca de dónde proceden los criterios, con qué valores están de acuerdo y qué posiciones de valor rechazan.
- No confundir temas que son inherentemente cuestiones de estilo docente con los que son cuestiones importantes de sustancia (Scriven, 1988a, 1988b).
- Conducir a inferencias acerca del contenido de las evaluaciones y los métodos utilizados para recopilar los datos.

Por estas razones, el concepto, que guíe la enseñanza deberá exponerse en una etapa inicial del proceso de desarrollo, antes de adoptar las decisiones finales de diseño.

Nuevamente, tomaré la Serie de Práctica como ejemplo para la formulación de un tal concepto rector. Este concepto ha sido formalizado por Dwyer y Villegas (1993). En el desarrollo de la Práctica, se hizo hincapié en la toma de decisiones por el maestro y en la importancia del contexto estudiantil, escolar y curricular al evaluar esa toma de decisiones. Este punto de vista implica decididamente el valor de la recopilación de datos en el marco real del aula en contraposición a simulaciones. Una segunda repercusión es la de que las evaluaciones deberían incluir oportunidades para que el evaluador y el evaluado interactúen en relación con el evento docente que se esté considerando a fin de que el evaluador comprenda plenamente las decisiones en las que se fundamentaron las acciones observables del maestro. Una tercera implicación es la de que, al no haber "una respuesta correcta única a la pregunta de qué constituye buena enseñanza (debido a que la enseñanza se considera como inherentemente sensible al contexto), la puntuación de las evaluaciones ha de permitir formas múltiples de respuestas «aceptables», en tanto explican claramente lo que constituye práctica profesional inaceptable.

Una cuarta implicación relacionada con el concepto de la enseñanza como actividad cognoscitiva completa es la de que el proceso de evaluación dependiendo,

como lo hace, de la consideración de una serie compleja de datos, requerirá un enjuiciamiento profesional sustancial para su ejecución. Por lo tanto, los evaluadores habrán de ser profesionales experimentados que han recibido capacitación para alcanzar una comprensión común de los criterios de evaluación y otras consideraciones para su aplicación.

Este concepto de la enseñanza y el aprendizaje, y los criterios que emanan de él, tienen fuertes vínculos con las teorías y prácticas psicológica, educativa y de medición. Especifica un desempeño objetivo cognoscitiva y comportamentalmente complejo y proporciona un marco para examinar el impacto de las evaluaciones sobre el sistema educativo del que forma parte (alumnos; maestros que se están evaluando; la profesión docente; la formación de maestros y la capacitación del personal).

En el caso de la Práctica, también se identificaron explícitamente los valores siguientes como parte de la formulación del concepto rector y se utilizaron como guía para la formulación de la evaluación propiamente dicha:

- Compromiso al tratamiento equitativo de los maestros.
- Estándares y técnicas de evaluación que aborden las acciones y toma de decisiones de los maestros.
- Especificación de ciertas prácticas docentes como inaceptables en cualquier contexto, en tanto se permiten muchas modalidades diferentes de prácticas aceptables, permitiendo la creación de estándares de conocimiento y prácticas docentes específicas y significativas.
- Creación de una experiencia de aprendizaje positiva tanto para el maestro que está siendo evaluado como para el evaluador.

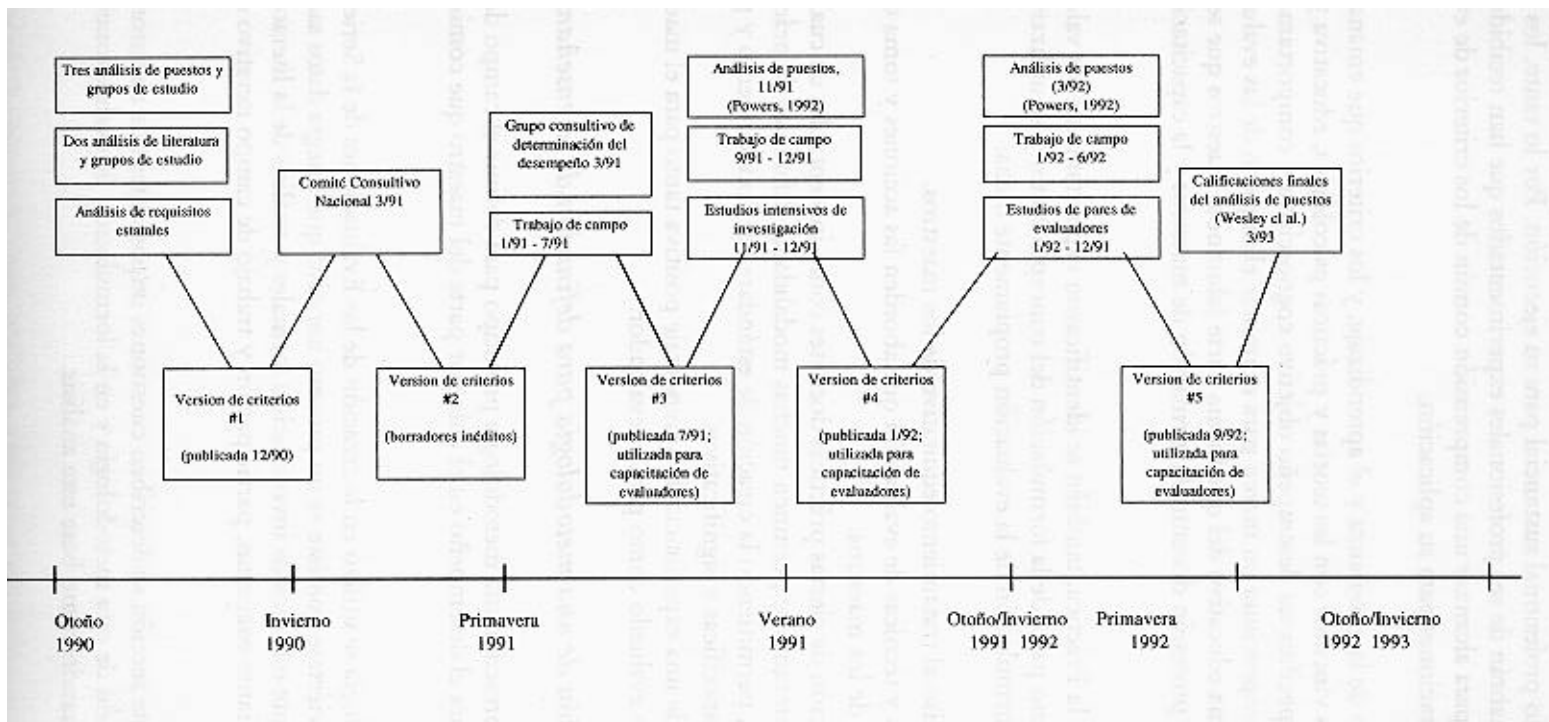
### *Especificación de una metodología para definición de la enseñanza*

La Figura 2 presenta una metodología prototipo para definir el campo de la enseñanza para el desempeño en el aula por parte del maestro que comienza su carrera.

Esta metodología se utilizó en la creación de las Evaluaciones de la Serie de Práctica. Adviértase que éste es un proceso iterativo que integra datos sacados de la teoría mediante estudios de investigación formales y análisis de la literatura; y práctica mediante encuestas, participación y trabajo de campo iterativo de los maestros.

En la siguiente sección se describen cuestiones importantes que entraron en juego en la aplicación de esta metodología y en la formulación de evaluaciones de los maestros tomando como base este análisis.

Figura 2. Práctica III: Evaluaciones del desempeño en el aula. Esquema de formulación de criterios



## **CUESTIONES RELACIONADAS CON LA FORMULACION DE EVALUACIONES DE LOS MAESTROS**

### *Selección de los métodos de evaluación*

La mejor estrategia para seleccionar los métodos de evaluación de los maestros puede resumirse muy simplemente: identificar primero qué es importante conocer acerca del maestro y, luego, determinar el procedimiento óptimo para obtener la información necesaria. Las posibilidades incluyen pruebas de opciones múltiples, otras formas de evaluaciones escritas y muchos tipos de determinación del desempeño. En la mayoría de los casos, una cobertura adecuada de los conocimientos y aptitudes necesarios requerirá una combinación de técnicas de evaluación. Los criterios para seleccionar el método óptimo para evaluar áreas específicas de contenido y aptitudes deberían ser integrales e incluir lo siguiente:

- Concordancia con el concepto rector de la enseñanza.
- -Relación lógica con el área objetiva de conocimiento o aptitud.
- Impacto educativo positivo.
- Factibilidad operativa y económica.
- Imparcialidad demostrable.

Tal como indicamos antes, las evaluaciones básicas y de alta calidad deberían resultar en un efecto educativo positivo tal como se expresa en el criterio de Messick (1989) sobre validez de construcción. De acuerdo con este criterio, una prueba válida conduce a buena práctica educativa y aprendizaje fructífero y una prueba inválida conduce a una adquisición subóptima de aptitudes.

La imparcialidad para con los maestros que están siendo evaluados es un criterio clave para enjuiciar cualquier posible metodología de evaluación. La imparcialidad puede considerarse sobre la base de características tales como el género, la raza o la etnicidad; o sobre la base de otros factores tales como los antecedentes educativos, experiencias pertinentes de la vida y familiarización con la materia y los alumnos a los que se va a enseñar.

Otras características relacionadas con la imparcialidad para el individuo son la fiabilidad de las evaluaciones, que han de enjuiciarse por medidas apropiadas para cada método empleado en la evaluación. Diferentes formas de evaluación requieren diferentes métodos para establecer la fiabilidad así como diferentes estándares para designar niveles apropiados de fiabilidad. Concretamente, los métodos y niveles empleados para enjuiciar pruebas de opciones múltiples no son necesariamente apropiados para aplicación a otras formas de evaluación.

Una consideración importante al seleccionar un método de evaluación es la de que el método ha de ser operativo y económicamente factible para uso en gran escala.

La economía y eficiencia a corto plazo de las pruebas de opciones múltiples son evidentes. Sin embargo, también es aconsejable, al estimar nuevos métodos de evaluación, comparar el impacto total de las distintas formas de evaluación. Dentro de límites cuidadosamente definidos, el valor agregado de los aumentos en la validez, el carácter directo de la medición, la motivación de los candidatos y otras medidas de calidad de la evaluación pueden contrarrestar los mayores costos y complejidad operativa de otras formas de evaluación. Estándares de calidad de pruebas centrales tales como la fiabilidad, la validez y la imparcialidad, son aún pertinentes para las distintas formas de evaluación que se seleccionen.

La economía de las evaluaciones del desempeño y las pruebas de respuestas construidas en general presentan retos formidables. En comparación con las pruebas de opciones múltiples, estos métodos de evaluación son inherentemente métodos de alta densidad de mano de obra y, por tanto, son más costosos de poner en práctica. Por esta razón, lo que se mide con los métodos de respuestas construidas y evaluaciones del desempeño ha de ser más significativo para justificar los gastos adicionales requeridos. Las pruebas de respuestas construidas, tomadas en su conjunto, sirven a menudo de experiencia educativa para quienes efectúan la evaluación y son los objetos de ella. Este criterio de la evaluación como actividad en el desempeño de las funciones es a menudo parte de la razón fundamental para los gastos adicionales de las evaluaciones del desempeño.

En el caso de las evaluaciones de la Práctica, se seleccionaron distintos métodos de evaluación. La Práctica I, Evaluaciones de las Aptitudes Básicas de Lectura, Escritura y Matemáticas, se utiliza para las personas que entran en la profesión docente evaluadas mediante una combinación de preguntas basadas en computadoras, de opciones múltiples, abiertas y ensayos. La Práctica II, Evaluaciones del Conocimiento de la Materia, difiere en cuanto a la metodología según la disciplina. Casi todas las pruebas de la Práctica II contienen algunas preguntas de opciones múltiples y abiertas, pero ciertos aspectos difieren mucho. Una evaluación de los maestros de idiomas, por ejemplo, utiliza modalidades de respuesta audiovisual para evaluar los aspectos receptivos y productivos de la competencia en comunicación. Una evaluación de los maestros de educación física utiliza estímulos de video para evaluar aspectos importantes de los movimientos corporales. La Práctica III, Evaluaciones del Desempeño en el Aula, utiliza una combinación de entrevistas, observaciones y documentación escrita para determinar la eficacia del maestro en un entorno de clase real.

### *¿Adelanto o atraso?*

Los elaboradores de sistemas de evaluación de maestros también hacen frente a lo que se denomina a menudo el dilema de "adelanto/atraso". En imparcialidad para con los maestros que están siendo evaluados, los criterios empleados para enjuiciarlos han de reflejar la práctica profesional actualmente aceptable. Esto asegura también que las evaluaciones concuerdan lógicamente con los fines de éstas. Sin embargo, un valor competidor es el que, en vista del largo tiempo requerido

para formular evaluaciones de alta calidad y la probabilidad de que continúen siendo utilizadas por cierto número de años, también es importante no crear evaluaciones que, en realidad, queden anticuadas antes de que se concluyan o que fomenten una continuación de prácticas docentes que incluso ahora son marginalmente aceptables para la profesión.

Es central para esta cuestión el hecho de que la Práctica profesional actualmente aceptable no es un concepto estático y que casi diariamente se crea nuevo conocimiento acerca de la docencia. Al evaluar, si puede considerarse que un determinado aspecto de la enseñanza es apoyado por la investigación, es por tanto necesario hacer cierto número de enjuiciamientos complejos acerca del estado de la investigación y tomar en cuenta el consenso profesional acerca de tendencias futuras en esa esfera. También es necesario considerar cuán definitiva es a la sazón un área de investigación. Por ejemplo, el área del comportamiento de los maestros y sus vínculos con el aprendizaje de los alumnos ha sido objeto de extensa investigación durante muchos años. En determinadas subáreas, los campos están bien trazados. Son numerosos los estudios bien concebidos y, en algunos casos, se ha llegado a conclusiones definitivas.

En contraste, el área cognoscitiva de los maestros y sus vínculos con el aprendizaje del alumno es relativamente nueva y se encuentra en estado de evolución. Aunque, por lo general, no se disputa la importancia de este campo de la investigación para la práctica docente, la naturaleza cognoscitiva del maestro si es aún algo incierta. Cierta número de principios importantes, aunque lógicamente bien expresados y convincentemente bien demostrados en los mejores estudios de investigación, no han sido reproducidos aún ampliamente. Para la elaboración de evaluaciones de los maestros es necesario efectuar enjuiciamientos basados en dichas observaciones acerca de qué incluir para los fines de evaluaciones de los maestros. Un factor que complica el problema es el de que la investigación sobre el comportamiento de los maestros y sobre el aspecto cognoscitivo de ellos tiende a utilizar diferentes metodologías, creando así otra dificultad en la evaluación de la investigación más 91 nueva conforme a estándares tradicionales.

En el caso de la elaboración de las evaluaciones de la Práctica, la mayor parte del contenido se sacó de la práctica actual, pero ciertos aspectos se enjuiciaron como parte de tendencias actuales. Por ejemplo, al determinar el desempeño en el aula para la Práctica III, se hizo mayor hincapié en la importancia de que los maestros trabajen entre si y con las familias de los estudiantes que el que las encuestas indican que es la práctica actual. La inclusión de esta faceta de la enseñanza se basó en el consenso profesional de que dichas interacciones son beneficiosas para el aprendizaje de los alumnos y en la investigación que apoya este hecho.

#### Énfasis prácticos y teóricos

Otro dilema que afrontan muchos elaboradores de cualquier sistema de evaluación de maestros es el mérito relativo del conocimiento teórico y práctico de la

enseñanza. Sternberg y Wagner (1993) han hecho una distinción útil entre los problemas académicos y los prácticos que puede ser interesante con respecto a la evaluación de los maestros. De acuerdo con Sternberg y Wagner, los problemas académicos tienden a (a) ser formulados por otras personas, (b) estar bien definidos, (c) ser completos con respecto a la información requerida para resolverlos, (d) tener sólo una respuesta correcta única, (e) tener sólo un método único para obtener la respuesta correcta, (f) no estar incorporados en la experiencia ordinaria y (g) tener un valor intrínseco escaso o nulo. En contraste, los problemas prácticos tienden a (a) requerir reconocimiento y formulación del problema, (b) estar mal definidos, (c) requerir la búsqueda de información, (d) tener soluciones aceptables múltiples, (e) permitir múltiples caminos que lleguen a la solución, (f) estar incorporados en la experiencia cotidiana y requerir dicha experiencia y (g) exigir motivación y participación personal (página 2). Es claro que, de acuerdo con Sternberg y Wagner, las evaluaciones integrales de maestros abordan los problemas prácticos que han de resolver estos últimos, más que los problemas académicos.

Además, tal como se indicó antes, es importante, por razones tanto prácticas como teóricas, que los criterios sobre los cuales se fundamentan las evaluaciones de los maestros y su marco de organización, guarden relación con la propia comprensión que tienen los maestros de su trabajo. Sin embargo, al propio tiempo, los criterios también deberán basarse en la teoría educativa y psicológica para establecer un marco lógico. También mostraron una generalización mayor a través de los contextos docentes y una mayor probabilidad de pasar la prueba del tiempo en el uso real en el aula. Nuevamente, la resolución de este dilema no es un proceso mecánico. El buscar los aportes de los maestros que ejercen su profesión y los teóricos de la educación para repasar y revisar los materiales de evaluación hasta que se perciba ampliamente que son aceptables desde ambos puntos de vista ha sido una estrategia eficaz en la formulación de las evaluaciones de la Práctica.

*¿Desde qué perspectiva debería considerarse la base de conocimientos?*

Tal como se indicó antes, las fuentes principales de datos para la base de conocimientos acerca de la enseñanza son, entre otras, los maestros en el ejercicio de su profesión y la perspectiva de investigación/teórica. Las fuentes adicionales pueden consistir en disposiciones que sustituyen a las escuelas rectoras y a los maestros. Cada una de estas perspectivas no proporciona simplemente un punto de vista diferente del mismo fenómeno; cada una formula preguntas diferentes, emplea métodos diferentes para llegar a conclusiones y tiene una serie diferente, aunque a menudo relacionada, de preocupaciones acerca del significado y la utilización del conocimiento sobre la enseñanza. Estos puntos de vista representan fundamentalmente paradigmas diferentes, en el sentido de que las hipótesis, las metodologías y los valores básicos difieren.

Por tanto, no es un proceso algorítmico o mecánico para llegar a criterios que incorporan datos procedentes de estas tres fuentes. Como ejemplo de una solución a este problema, los elaboradores de la Serie de Práctica realizaron un procedimiento

iterativo consistente en crear proyectos de criterios y presentarlos luego para análisis a representantes de esos tres puntos de vista principales. Se preguntó a los analizadores y miembros del grupo, en esencia, si los criterios representaban la base de conocimientos para la enseñanza como ellos la velan. Los criterios se sometieron a cierto número de importantes revisiones como resultado de este proceso. Con cada una de estas importantes revisiones, se emprendieron ciclos de trabajo de campo cada vez más largos para proporcionar datos adicionales. Los criterios de la Práctica III se consideraron definitivos cuando cada una de las partes interesadas a quienes se consultó, quedó satisfecha de que se hablan abordado sus preocupaciones principales.

### *Alcance de los criterios de evaluación*

Necesitan abordarse dos temas interrelacionados que pudieran caracterizarse como cuestiones relacionadas con el ámbito de los criterios en el proceso de traducir el conocimiento del campo de la enseñanza a criterios específicos de evaluación del maestro: determinar el «tamaño» o nivel apropiado de generalidad para los criterios; y determinar la gama de contextos de enseñanza a los que se aplican los criterios.

La enorme variabilidad entre los contextos del aula (Shulman, 1988a, 1988b; Stodolsky, 1988) presenta retos importantes para toda actividad de evaluación de los maestros. En el caso de la Práctica III, Evaluaciones de Desempeño en el Aula, los criterios de enseñanza se diseñaron como aspectos de esta; es decir, como principios que deberían aplicarse en una amplia gama de contextos docentes (incluida la variabilidad en la materia y el nivel de grados enseñado, el estilo docente y las características individuales y antecedentes de los alumnos) en vez de como «reglas» específicas que deberían seguirse o comportamientos que deberían demostrarse. Esto entraña una necesidad de atención a la uniformidad entre los criterios.

Una cuestión importante para poder diseñar evaluaciones prácticas que puedan aplicarse en una amplia gama de contextos del aula ha sido la determinación del concepto del tamaño adecuado para los criterios. Algunos investigadores (Kagan, 1990; Katz y Ratlfs, 1985) han denominado a este reto el «Principio de Goldilocks). En la evaluación de los maestros, este principio significa que, si los criterios son demasiado extensos, es decir, demasiado vagos y generales, es difícil formular y aplicar equilibradamente estándares significativos; los evaluadores no pueden aplicar una serie uniforme de enjuiciamientos al proceso de evaluación. Por el contrario, si los criterios son demasiado estrechos (demasiado específicos), pueden enjuiciarse con gran uniformidad, pero no captarán la esencia de la buena enseñanza y pueden promover un enfoque fragmentado y prescriptivo de la enseñanza. Así pues, los criterios de evaluación, como el cocido de oso que probó Goldilocks, han de ser «perfectos» si se desea alcanzar las metas de una evaluación justa y una mejor práctica docente.

Como cuestión práctica, para encontrar el nivel de especificidad «perfecto» se

requieren muchas interacciones de trabajo de campo y análisis, tales como las que se ilustran en la Figura 2.

En los informes del trabajo de campo para las evaluaciones de la Práctica, hay numerosos casos de estas experiencias que llevan a los elaboradores a la conclusión, por ejemplo, de que lo que ha sido un sólo criterio debería dividirse en dos criterios separados para ayudar a los evaluadores a comprender mejor cómo un determinado aspecto de la enseñanza se lleva a cabo realmente en el aula y ayudarles a reconocer la evidencia relacionada con este aspecto de la enseñanza cuando la observan.

Dicho trabajo de campo también permite a los evaluadores y elaboradores de las pruebas ver la forma en que los criterios guardan relación unos con otros en la práctica, y utilizar esta información como base para efectuar cambios en la forma en que los criterios se organizan, ordenan y describen. La organización y redacción de los criterios de forma que sean claros y lógicos, desde el punto de vista de quienes los van a utilizar, debería recibir alta prioridad en la labor de formulación.

### *Evaluadores y enjuiciamiento*

Tal como se ha indicado antes, el enjuiciamiento desempeña una función importante en la elaboración de cualquier instrumento para la evaluación de los maestros. En las pruebas de opciones múltiples, el enjuiciamiento se ejerce al nivel de decidir cuestiones tales como qué someter a prueba, qué preguntas deberían presentarse, qué constituye una respuesta correcta y el número de preguntas que necesitan contestarse correctamente para que se considere que un candidato ha aprobado la prueba. Así pues, incluso en esta forma familiar de pruebas de «objetivos», el enjuiciamiento desempeña un papel importante, aunque no en la propia puntuación. Las evaluaciones del desempeño, por el contrario, incorporan enjuiciamiento en muchas formas que son similares, pero también en algunas que son diferentes. Al utilizar las evaluaciones del desempeño para evaluar a los maestros, es vital comprender el papel desempeñado por el enjuiciamiento profesional de quienes los evalúan.

La especificación de los criterios de evaluación es una parte importante de la formulación de cualquier prueba de desempeño, pero el éxito de esta actividad en su conjunto sólo puede evaluarse a la luz de la capacidad de los evaluadores de utilizar estos criterios a fin de llegar a conclusiones técnica y profesionalmente defensibles.

Como en las pruebas tradicionales de opciones múltiples, en las que la gran preponderancia del enjuiciamiento profesional entra en juego durante la etapa previa a la administración de la formulación de la prueba, el enjuiciamiento profesional con las evaluaciones del desempeño se requiere tanto en las etapas de formulación como de utilización de la prueba. La calidad del enjuiciamiento profesional surte efecto sobre muchos aspectos importantes de la validez de la prueba, incluyendo, aunque sin limitarse a ello, la imparcialidad, complejidad

cognoscitiva y representación de los elementos. También guarda relación con las preocupaciones de generalización, aun cuando estas no sean las mismas para una evaluación del desempeño natural en el aula que para la evaluación basada en la simulación.

En la evaluación del desempeño en el aula, la preocupación con la generalización gira en torno a otros eventos docentes, no a otros aspectos de la enseñanza. En este sentido, los eventos docentes son análogos a los ejercicios o tareas en otras clases de evaluaciones del desempeño. Un ejemplo de esto puede encontrarse en la Práctica III, Evaluaciones del Desempeño en el Aula (véase el Cuadro 1 donde se da una descripción de los criterios para evaluación en la Serie de Práctica).

En el modelo de la Práctica, la "puntuación" de las "tareas" se mantiene constante a través de los criterios y sus estándares de puntuación conexos. Al evaluar el desempeño docente en vivo, la variabilidad entre las "tareas" es un fenómeno natural y aceptable y, por tanto, se espera que las inferencias basadas en una serie dada de eventos docentes sean generales para un universo intrínsecamente variable de eventos docentes que define el concepto. Por tanto, la generalización entre las tareas no es problemática en el mismo sentido que cuando las tareas se consideran como elementos parciales o indirectos del concepto.

Tal como se indicó anteriormente, los criterios de la Práctica 111, Evaluaciones del Desempeño en el Aula, tienen por fin ser interpretados como aspectos interrelacionados de un desempeño complejo, no como entidades funcionalmente independientes. Como tales, no trataríamos de generalizar de un aspecto de la enseñanza a otro como prueba de la validez sino, más bien, investigar los patrones de calificaciones dadas a través de distintas ocasiones y dentro de una ocasión dada por dos o más evaluadores (suponiendo que se espere que las ocasiones sean muy variables, en relación con la variabilidad del asesor o entre los asesores).

Los criterios de evaluación no son independientes, ya que constituyen aspectos de la enseñanza y no comportamientos particulares. Han de ser interpretados a la luz del contexto real en el aula, que incluye tanto a los alumnos como a la materia que se está enseñando. Los criterios sirven de guía para estructurar los enjuiciamientos de los evaluadores, asegurando que un marco común de referencia en vez de la preferencia personal constituya la base para las conclusiones de los evaluadores y las calificaciones. Por tanto, los enjuiciamientos profesionales de los evaluadores son la piedra angular del defendimiento de las puntuaciones del maestro que comienza a ejercer su profesión.

Utilizando los métodos que acabamos de describir, los evaluadores recopilan y organizan datos que inciden en cada uno de los criterios; efectúan enjuiciamientos críticos acerca de la importancia de la evidencia y su pertinencia para determinados criterios; luego llegan a una conclusión acerca del nivel de desempeño del maestro principiante en cada criterio basándose en esta evidencia y su interpretación de la misma. Los evaluadores documentan estos enjuiciamientos citando evidencia

concreta y vinculándola con una escala de calificaciones que describe niveles cada vez más suficientes de desempeño con respecto a cada uno de los criterios. La legitimidad del proceso de evaluación se basa por tanto en la calidad de este argumento (enjuiciamiento estructurado, documentado, profesional) más que en una ausencia pretendida de toma de decisiones humana (objetividad). Mediante estudios especiales (tales como las comparaciones por dos evaluadores), el trabajo de campo en distintos marcos docentes y el uso operativo, puede hallarse que los métodos de recopilación de datos distintos de los que ahora utilizamos producen mejor mención - es decir, enjuiciamientos exactos o detallados de los criterios, mejor documentación y efectos más positivos sobre el sistema de que forma parte la evaluación. Al determinar el valor y la validez de las evaluaciones, no obstante, los propios métodos de recopilación de datos están claramente subordinados a la calidad de los criterios y los enjuiciamientos y a la aplicación sistemática por los evaluadores de los procedimientos de evaluación.

## **CONCLUSION**

La evaluación de los maestros es un elemento clave de todo esfuerzo por mejorar la educación. Los sistemas bien concebidos y ejecutados para evaluación de los maestros son una parte integral del sistema educativo y coadyuvan al logro de sus metas principales. Los sistemas mal concebidos y ejecutados pueden tener un efecto deletéreo sobre el sistema mismo al que pretenden mejorar. Así pues: (1) una atención cuidadosa al diseño de las evaluaciones de los maestros, (2) seguimiento del cumplimiento con los planes durante la etapa de ejecución y (3) evaluación sistemática de todo el sistema de evaluación y sus consecuencias a través de su operación, son tres factores esenciales para obtener los beneficios previstos de las evaluaciones de los maestros.

Los educadores han tenido históricamente que luchar con la tensión entre las preocupaciones del equilibrio, la excelencia y la eficiencia que compiten entre sí. En distintos momentos, por razones comprensibles, el equilibrio ha cambiado a la preocupación acerca de uno de estos elementos a expensas de los otros. A fin de atender óptimamente las necesidades de la educación y de todos aquellos a quienes esta afecta, no obstante, necesitamos esforzarnos por ir más allá de nuestras propias preocupaciones diarias e individuales, y mantener nuestra mirada concentrada en el cuadro más grande a fin de obtener el mejor equilibrio posible entre estas tres preocupaciones urgentísimas.

### **Cuadro 1: Criterios de evaluación para la Serie de Práctica: Evaluaciones del aula para maestros principiantes**

Campo A: Organizar el conocimiento del contenido para el aprendizaje de los alumnos. El conocimiento del contenido que se ha de impartir sustenta todos los aspectos de la buena instrucción. El campo A se concentra en cómo los maestros utilizan su comprensión de los alumnos y las asignaturas para decidir acerca de las metas de

aprendizaje; diseñar o seleccionar actividades y material de instrucción apropiados; ordenar la instrucción de forma que ayude a los alumnos a alcanzar metas curriculares a corto y largo plazos; y concebir o seleccionar estrategias de evaluación informativa. Todos estos procesos, comenzando con las metas de aprendizaje, deben alinearse entre sí y, debido a las distintas necesidades representadas en toda aula, cada uno de los procesos mencionados debe llevarse a cabo de forma que responda a la serie de conocimientos y experiencias que los estudiantes traen al aula. Por consiguiente, el conocimiento de información pertinente acerca de los propios alumnos es parte integrante de este campo.

El campo A se refiere a cómo piensa el maestro acerca del contenido que va a enseñar. Este pensamiento es evidente en la forma en que el maestro organiza la instrucción en beneficio de sus alumnos. Las principales fuentes de evidencia en este campo son el Perfil del Aula, el Perfil de la Instrucción y la Entrevista Previa a la Observación. La observación del aula puede contribuir también a evaluar el desempeño en este aspecto.

*Criterios de evaluación para el campo A:*

A 1:Familiarizarse con los aspectos relevantes de las experiencias y los conocimientos básicos de los alumnos.

A2:Formular metas claras de aprendizaje de las lecciones que sean apropiadas para los alumnos.

A3:Demostrar una comprensión de los vínculos entre el contenido que se aprendió con anterioridad, el contenido actual y el que queda por aprender en el futuro.

A4:Crear o seleccionar métodos de aprendizaje, actividades de aprendizaje y material u otras fuentes de instrucción que sean apropiados para los alumnos y que armonicen con las metas de la lección.

A5:Crear o seleccionar estrategias de evaluación que sean apropiadas para los alumnos y que armonicen con las metas de la lección.

Campo B: Crear un entorno para el aprendizaje del alumno. El campo B se relaciona con los componentes sociales y emocionales del aprendizaje como requisitos y contexto del rendimiento académico. Así pues, la mayor parte de los criterios de este campo se concentran en las interacciones humanas en el aula, en los vínculos entre maestros y alumnos y entre los alumnos. El campo B aborda las cuestiones relacionadas con la imparcialidad y la concordia, con ayudar a los alumnos a convencerse de que pueden aprender y pueden responder a los retos de establecer y mantener normas constructivas de comportamiento en el aula. También incluye el "entorno" de aprendizaje en el sentido más literal: el marco físico en el que tienen lugar la enseñanza y el aprendizaje.

Un entorno de aprendizaje que proporciona seguridad tanto emotiva como física a los alumnos puede propiciar una amplia gama de experiencias de enseñanza y aprendizaje. Los maestros deben estar en condiciones de utilizar el conocimiento que tienen de sus alumnos para interpretar con exactitud el comportamiento de ellos y responder de manera apropiada a y ofreciendo apoyo. Al hacerlo, sus interacciones con los alumnos promueven continuamente el sentido de autoestima en ellos. Además, los esfuerzos de los alumnos por crear un sentido de comunidad en el aula con estándares claros no deberían ser nunca arbitrarios; todos los estándares de comportamiento e interacciones entre el maestro y los alumnos deberían fundamentarse en un sentido de respeto hacia todos los miembros de la comunidad en el aula.

Las pruebas de los criterios del campo B provienen principalmente de la observación del aula; las pruebas de apoyo pueden provenir de entrevistas tanto anteriores como posteriores a la observación. El Perfil del Aula proporciona información de contexto pertinente para los criterios que encierra este campo.

*Criterios de evaluación del campo B.-*

B1: Crear un clima que promueva la imparcialidad. B2: Establecer y mantener una relación con los alumnos. B3: Comunicar expectativas de aprendizaje interesante para cada alumno. B4: Establecer y mantener estándares uniformes de comportamiento en el aula. B5: Hacer el entorno físico lo más seguro y conducente que sea posible al aprendizaje.

Campo C: Enseñanza para el aprendizaje de los alumnos. Este campo se concentra en el acto de enseñar y en su meta general: ayudar a los alumnos a conectarse con el contenido. Tal como se utiliza aquí, el "contenido" se refiere a la materia de una disciplina y puede incluir conocimientos, aptitudes, percepciones y valores en cualquier campo: cognoscitivo, social, artístico, físico y demás. Los maestros dirigen a los alumnos en el proceso de establecer conexiones individuales con el contenido, haciendo así "encajar" bien el contenido en el marco de los conocimientos, intereses, aptitudes, antecedentes culturales y antecedentes personales de los alumnos. Al mismo tiempo, los maestros deberían ayudar a los alumnos a sobrepasar los límites de sus conocimientos o comprensión actuales. Los maestros supervisan el aprendizaje, asegurándose de que los alumnos asimilen con exactitud la información y que comprendan y puedan poner en práctica lo que han aprendido. Los maestros deben estar seguros también de que los alumnos comprendan lo que se espera de ellos en cuanto a procedimientos durante la lección y que el tiempo de clase se utilice provechosamente.

Gran parte de esta evidencia para el desempeño del maestro con respecto a estos criterios provendrá de la observación del aula. Ésta puede aumentarse o iluminarse con la evidencia proveniente de entrevistas anteriores o posteriores a la observación, del Perfil de la Instrucción y del Perfil de la Clase.

*Criterios de evaluación del campo C.*

C1: Aclarar a los alumnos las metas de aprendizaje y los procedimientos de instrucción. C2: Hacer el contenido comprensible para los alumnos. C3: Alentar a los alumnos a ampliar su capacidad de reflexión. C4: Vigilar la comprensión del contenido por parte de los alumnos mediante una gama de medios, proporcionando intercambio de información a los alumnos con el fin de apoyar el aprendizaje y adaptando las actividades de aprendizaje según lo exija la situación. C5: Utilizar *eficazmente el tiempo* dedicado a la instrucción.

Campo D: Profesionalismo de los maestros. Los maestros han de poder evaluar su propia eficacia en la instrucción a fin de planificar lecciones futuras específicas para determinadas clases y mejorar su enseñanza a través del tiempo. Deberían poder hablar del grado en que los distintos aspectos de una lección tuvieron éxito en términos de métodos de instrucción, respuestas de los alumnos y resultados del aprendizaje. Los maestros deberían poder explicar cómo procederán para trabajar en pro del aprendizaje de todos los alumnos. Las responsabilidades profesionales de todos los maestros, incluidos los maestros

principiantes, también entrañan compartir información adecuada con otros profesionales y con las familias de forma que apoyen el aprendizaje de diversas poblaciones de alumnos.

#### *Criterios de evaluación del campo D.*

D1: Reflexionar sobre el grado en que se alcanzaron las metas de aprendizaje.

D2: Demostrar un sentido de eficacia.

D3: Crear relaciones profesionales con colegas para compartir conocimientos acerca de la enseñanza y coordinar las actividades de aprendizaje de los alumnos.

D4: Comunicarse con los padres de familia o tutores acerca del aprendizaje del alumno.

### NOTAS

1. En el original inglés se utiliza el término *evaluation* a lo largo de todo este documento para indicar una gama de actividades emprendidas con el objeto de recopilar información para fines de toma de decisiones. El término se utiliza sin referencia a distinciones entre dichos fines e incluye, pero sin limitarse a ello, debates de técnicas específicas de evaluación. El término *assessment* se utiliza para referirse a estas técnicas específicas de recopilación de datos. En la actualidad, el uso de estos dos términos en los Estados Unidos es algo controvertido, refiriéndose el primero principalmente a actividades que realizan con maestros practicantes quienes los emplean, con el fin de tomar decisiones de personal aisladas.

2. El término *puntuación límite* se utiliza en esta sección para referirse a todo estándar que indique un nivel de rendimiento. El uso de este término se aplica a cualquier clase de evaluación en la que las personas evaluadas van a ser divididas en grupos para fines de toma de decisiones.

### REFERENCIAS

- American Educational Association, American Psychological Association, y National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association (APA).
- Andrews, H. A. (1985). *Evaluating for Excellence. Addressing the Need for Responsive and Effective Faculty Evaluation*. Stillwater, OK: New Forums Press.
- Banks, C. A. M. (1997). The challenges of national standards in a multicultural society. *Education Horizons*, 75, 3: 126-132.
- Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education*, 1: 345-363.
- Bridges, E. (1986). *The Incompetent Teacher: The challenge and the response*. Filadelfia: Falmer Press.
- Burry, J. A., Chissom, B. S., y Shaw, D. G. (1990, March). *Validity and Reliability of Classroom Observations: A Paradox*. Paper presented at a meeting of the National

- Council on Measurement in Education, Boston.
- Claxton, C., Murrell, P. H., y Porter, M. (1987). Outcomes Assessment. *AGB Reports*, 29, 5: 32-35.
- Cole, N. S., y Moss, P. A. (1989). Bias in Test Use. En: R. L. Linn (Ed.), *Educational Measurement* (3rd ed., págs. 201-219). Nueva York: Macmillan.
- Darling-Hammond, L. (1986). A proposal for evaluating in the teaching profession. *Elementary School Journal*, 86: 531-569.
- Duke, D., y Stiggins, R. J. (1990). Beyond Minimum Competence: Evaluation for Professional Development. En: J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation: Assessment Elementary and Secondary School Teachers* (págs. 116-132). Newbury Park, CA: Sage.
- Dwyer, C. A. (1993b). Innovation and Reform: Examples from teacher assessment." En: R. E. Bennett y W. C. Ward (Eds.), *Construction Versus Choice in Cognitive Measurement* (págs. 265-289). Hillsdale, NJ: Lawrence Erlbaum
- Dwyer, C. A. (1993c). Teaching and diversity: Meeting the challenges for innovative teacher assessments. *Journal of Teacher Education*, 44(2), 119-129.
- Dwyer, C. A. (1994). Criteria for Performance-Based Teacher Assessments: Validity, Standards, and Issues. *Journal of Personnel Evaluation in Education*, 8, 135-150.
- Dwyer, C. A. (1996). Cut-scores and Testing: Statistics, judgment, Truth, and Error. *Psychological Assessment*, 8 (4).
- Dwyer, C. A., y Ramsey, P. A. (1995). "Equity Issues in Teacher Assessment." En: M. T. Nettles and A. L. Nettles (Eds.), *Equity and Excellence in Educational Testing and Assessment*. Boston: Kluwer Academic.
- Dwyer, C. A. y Stufflebeam, D. (1996). Teacher Evaluation. En: D. Berliner y R. Calfee (Eds.), *Handbook of Educational Psychology* (págs. 765-786). Nueva York: Macmillan.
- Dwyer, C. A. y Villegas, A. M. (1993). *Guiding Conceptions and Assessments Principles for The Praxis Series.\* Professional Assessments for Beginning Teachers* (RR 93-17). Princeton, NJ: Educational Testing Service.
- Evertson, C., y Green, J. (1986). Observation as Method and Inquiry. En: M. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., págs. 162-213). Nueva York: Macmillan.
- Hunter, M. (1988). Effecting a Reconciliation Between Supervision and Evaluation. *Journal of Personnel Evaluation in Education*, 1: 275-279.
- Garcia, P. A. (1985). *A Study of Teacher Competency Testing and Tests Validity with Implications for Minorities: Final Report*. (NIE Grant No. NIE-G-85-0004). Edinburg, TX: Pan American University.
- Glass, G. V. (1990). Using Student Test Scores to Evaluate Teachers. En: J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation* (págs. 191-215). Newbury Park, CA: Sage.
- Haney, W., Madaus, G., y Kreitzer, A. (1987). Charms Talismanic: Testing Teachers for the Improvement of American Education. En: E. Z. Rothkopf (Ed.), *Review of Research in Education* (Vol. 14, págs. 169-238). Washington, DC: American Educational Research Association.
- Hood, S., y Parker, L. (1991). Minorities, Teacher Testing, and Recent U.S. Su

- preme Court Holdings: A Regressive Step. *Teachers College Record*, 92: 603-618.
- Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards*. Newbury park, CA: Sage.
- Kagan, D. M. (1990). Ways of Evaluating Teacher Cognition: Inferences Concerning the Goldilocks Principle. *Review of Educational Research*, 60, 419-469.
- Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology* 28: 563-575.
- Linn, R. L., Baker, E. L., y Dunbar, S. B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*, 20 (8): 15-21.
- Madaus, G. E (1990). Legal and Professional Issues in Teacher Certification Testing: A Psychometric Snark Hunt. En: J. V. Mitchell, S. L. Wise, and B. S. Plake (Eds.), *Assessment of Teaching.- Purposes, Practices, and Implications for the Profession* (págs. 209-259). Hillsdale, NJ: Lawrence Erlbaum.
- Madaus, G. y Melirens, W A. (1990). Conventional Tests for Licensure. En: J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation on: Assessing Elementary and Secondary School Teachers 's. 257-277*). Newbury Park, CA: Sage.
- Madaus, G. E, y Pullin, D. (1987, September). Teacher Certification Tests: Do They Really Measure What We Need to Know? *Phi Delta Kappan*, 69(1), 31-38.
- Mehrens, W A. (1987). Validity Issues in Teacher Licensure Tests. *Journal of Personnel Evaluation in Education*, 1: 195-229.
- Messick, S. (1989). Validity. En: R. L. Linn (Ed.), *Educational Measurement* (3rd ed., págs. 13-103). Nueva York: Macmillan.
- Miller, M. D., y Legg, S. M. (1993). Alternative Assessment in a High-Stakes Environment. *Educational Measurement.- Issues and Practice*, 12 (2): 9-15.
- Millman, J. (1981a). *Handbook of TeacherEvaluation*. Beverly Hills, CA: Sage.
- Millman, J. (1981b). Student Achievement as a Measure of Teacher Competence. En: J. Millman (Ed.) *Handbook of Teacher Evaluation* (págs. 146-166). Beverly Hills, CA: Sage.
- Milner, J. O. (1991). Suppositional Style and Teacher Evaluation. *Phi Delta Kappan*, 72: 464-467.
- Moss, P. A. (1992). Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. *Review of Educational Research*, 62: 229-258.
- Murnane, R. J., y Schwinden, M. (1989). Race, Gender, and Opportunity: Supply and Demand for New Teachers in North Carolina, 1975-1985. *Educational Evaluation and Policy Analysis*, 11: 93-108.
- National Board for Professional Teaching Standards (1991). *Toward High and Rigorous Standards for the Teaching Profession* (3rd ed.). Detroit, MI: Autor.
- National Council of Teachers of Mathematics. (1989, March). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Autor.
- Quellmalz, E. S. (1991). Developing Criteria for Performance Assessments: The Missing Link. *Applied Measurement in Education*, 4: 319-331.
- Rebell, M. A. (1990). Legal Issues Concerning Teacher Evaluation. En: J. Millman and L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation*:

- Assessing Elementary and Secondary School Teachers* (págs. 337-355). Newbury Park, CA: Sage.
- Redfern, G. B. (1960). *How to Appraise Teaching Performance*. Columbus, OH: School Management Institute.
- Redfern, G. B. (1980). *Evaluating Teachers and Administrators: A Performance Objectives Approach*. Boulder, CO: Westview Press.
- Scriven, M. (1987). Validity in Personnel Evaluation. *Journal of Personnel Evaluation in Education*, 1: 9-23.
- Scriven, M. (1988a). Duties-Based Teacher Evaluation. *Journal of Personnel Evaluation in Education*, 1: 319-334.
- Scriven, M. (1988b). Evaluating Teachers as Professionals: The Duties-Based Approach. En: S. J. Stanley y W J. Popham (Eds.), *Teacher Evaluation: Six Prescriptions for Success* (págs. 110-142). Alexandria, VA: Association for Supervision and Curriculum Development.
- Sedlak, M., y Schlossman, S. (1986). ¿Who Will Teach? *Historical Perspectives on the Changing Appeal of Teaching as a Profession*. (Report No. 3472-CSTP). Santa Mónica, CA: RAND Center for the Study of the Teaching Profession.
- Shalock, D., y Shalock, M. (1990, September). Extending Teacher Assessment Beyond Knowledge and Skills: An Emerging Focus on Teacher Accomplishments. *Issues and Practices in Performance Assessment*, 2: 81-126.
- Shimberg, B. (1985). Testing for Licensure and Certification. *American Psychologist*, 36: 1138-1146.
- Shulman, L. S. (1986) Paradigms and Research Programs in the Study of Teaching: A Contemporary Perspective. En: M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., págs. 3-36). Nueva York: Macmillan.
- Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, 57. 1-22.
- Shulman, L. S. (1988b). A Union of Insufficiencies: Strategies for Teacher Assessment in a Period of Educational Reform. *Educational Leadership*, 46: 36-41.
- Shulman, L. S. (1988a). A Union of Insufficiencies: Strategies for Teacher Assessment in a Period of Educational Reform. *Educational Leadership*, 46: 36-41.
- Shulman, L. S. (1988b). The Paradox of Teacher Assessment. En: *New Directions for Teacher Assessment. Proceedings of the 1988 ETS Invitational Conference*. Princeton, NJ: Educational Testing Service.
- Shulman, L. S. (1989). The Paradox of Teacher Assessment. En: *New Directions for Teacher Assessment.- Proceedings of the 1988 ETS Invitational Conference* (págs. 13-27). Princeton, NJ: Educational Testing Service.
- Stodolsky, S. S. (1988). *The Subject Matters*. Chicago: University of Chicago Press.
- Streifer, P., y Iwanicki, E. (1987). The Validation of Beginning Teacher Competencies in Connecticut. *Journal of Personnel Evaluation in Education*, 1: 33-55.
- Webster, W J., Mendro, R. L., y Alinaguer, T. O. (1993). *Effectiveness Indices.- The Major Component of an Equitable Accountability System*. Estudio presentado en la reunión anual de la American Educational Research Association, Atlanta, GA.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W, y Bernstein, H. T. (1984a). *Case Studies for Teacher Evaluation: A Study of Effective Practices*. Santa Mónica, CA: Rand Corporation.
- Wise, A. E., Darling-Hammond, L., McLaughlin, M. W, y Bernstein, H. T. (1984). *Teacher Evaluation: A Study of Effective Practices*. Santa Mónica, CA: Rand Corporation.