



ISBN: 978-607-02-9094-7

Universidad Nacional Autónoma de México

Instituto de Investigaciones
sobre la Universidad y la Educación

www.iisue.unam.mx/libros

José María García Garduño (2017)

“Los estándares de evaluación docente y la eficacia
escolar: desarrollo, tendencias”

en *Docencia y evaluación en la Reforma Educativa 2013*,

Ángel Díaz-Barriga (coord.),

IISUE-UNAM, México, pp. 283-326.

Esta obra se encuentra bajo una licencia Creative Commons
Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional
(CC BY-NC-ND 4.0)

Los estándares de evaluación docente y la eficacia escolar: desarrollo, tendencias

*José María García Garduño*¹

El propósito de este capítulo es realizar un análisis de la evolución de la investigación sobre eficacia docente, su relación con los estándares de evaluación docente y analizar de manera preliminar su situación en México.

La globalización económica que comenzó a ocurrir durante la década de los ochenta ha traído consigo competencia por los mercados entre los países del orbe. Debido a la creencia de que existe una relación entre la robustez de la economía y la calidad educativa, la competencia económica se ha trasladado a la escuela. A esta institución se le ha asignado la misión universal de preparar a las nuevas generaciones para el mundo del trabajo. Ante esa competencia y la pérdida de mercados y puestos de trabajo, en la década de los ochenta, los Estados Unidos se sintieron amenazados por el poderío y competitividad de Japón y otros países. Como señala Pinar (2004), cuando la economía va mal o hay más crimen no se culpa a las escuelas de economía ni a las iglesias. En cambio se hace responsable a la escuela –sin serlo– de la mayoría de los males que aquejaban la economía de ese país.

En esa misma década y la siguiente, y por influencia de la industria, la escuela comenzó a adoptar los métodos de calidad total –posteriormente se le denominó mejoramiento o aseguramiento de la calidad. La calidad total fracasó en la escuela (Mosadeghrad, 2014); las instituciones educativas, tanto de Estados Unidos como de otros países, se dieron cuenta de que la escuela no es una fábrica que puede

1 Académico de la Universidad Autónoma de la Ciudad de México.

seleccionar la materia prima (alumnos), controlar y determinar los tiempos y movimientos complejos de la vida escolar; costó trabajo entenderlo, el movimiento de calidad total no tuvo el éxito buscado en la mejora de la calidad educativa.

La globalización no sólo trajo comparaciones de la competitividad económica entre países, sino también de su desempeño educativo. A mediados de la década de los noventa comenzaron a surgir las comparaciones internacionales, una competencia educativa entre naciones –en términos llanos– por medio de los resultados de las pruebas internacionales. La prueba TIMSS (Trends in International Mathematics and Science Study) comenzó a aplicarse a mediados de esa década en diferentes países. A partir del año 2000, cada tres años, los países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y otros países invitados –hasta llegar a ser más de 60– evalúan el desempeño de sus escolares a través del examen Programa Internacional de Evaluación de los Alumnos (PISA, por sus siglas en inglés). Es tal su influencia que los países que participan en la evaluación toman como referente los resultados en el diseño de sus políticas educativas y programas curriculares tendientes a elevar la calidad de la educación. Por ejemplo, en México la evaluación de los estándares curriculares de la educación básica están diseñados tomando en consideración los resultados de PISA (SEP, 2011). Los países desarrollados, como Estados Unidos, también muestran preocupación por seguir ocupando lugares promedio en los resultados de esta prueba y tratan de mejorarlos (OECD, 2013a). En un estudio realizado con datos de 48 países, se afirma que hay una relación positiva entre crecimiento económico y ganancias en el desempeño escolar (medido a través de los test). Asimismo, que a pesar de las ganancias (proporcionales) que los Estados Unidos han logrado, éstas son menores que las de 24 países, entre ellos Colombia, Brasil y Chile (Hanushek, Peterson y Woessmann, 2012).

Si bien las políticas y procedimientos para evaluar el desempeño docente en la educación básica y secundaria han estado presentes desde antes de la competencia económica y educativa, a partir de los resultados de PISA y estudios concurrentes, la evaluación docente se está propagando entre los sistemas educativos de diferentes partes

del mundo. El diseño de políticas de evaluación docente descansa en la premisa generalizada que además de los estudiantes, el principal insumo para mejorar la calidad educativa son los docentes. Entonces se deben diseñar políticas de rendición de cuentas capaces de monitorear, retroalimentar y recompensar su desempeño profesional, a través de evaluaciones formativas y sumativas. Esta creencia es impulsada por los organismos y estudios internacionales, de ahí que la política de evaluar a los docentes por medio de estándares se esté generalizando en la mayoría de los países desarrollados y en otros emergentes, como Chile y México (OECD, 2013a).

La popularidad que está alcanzando la evaluación docente como estrategia para mejorar la calidad educativa ha despertado el interés por estudiarla. Se han hecho revisiones del estado de la evaluación docente en los países de la OCDE (OECD, 2013b; Figazzollo, 2013). El Programa de Promoción de la Reforma Educativa de América Latina y el Caribe (PREAL) (Isoré, 2010) publicó en español una revisión de las prácticas de evaluación en los países de esa organización. Si bien estas revisiones proveen de un panorama del estado de la evaluación docente en los países miembros, su análisis es muy general, sólo provee de un panorama del estado de la evaluación y no diferencia su nivel de desarrollo entre los países seleccionados. Por ejemplo, en el análisis de la OCDE (OECD, 2013b) México es catalogado por tener un sistema de evaluación docente; sin embargo, en el momento en que este organismo levantó la encuesta, la evaluación de los docentes sólo se efectuaba por medio del programa Carrera magisterial, el cual valora básicamente los conocimientos y resultados en la pruebas, además de otros factores. La evaluación docente que se ha realizado en México está apenas iniciándose, no se puede comparar con la que se efectúa regularmente en otros países, principalmente los anglosajones. Se puede concluir de los informes mencionados arriba que en los países anglosajones y otros desarrollados y emergentes como Chile, la evaluación docente se lleva a cabo por medio de estándares. Los estándares se pueden conceptualizar como los conocimientos y habilidades que un docente debe ser capaz de demostrar (Shakman *et al.*, 2012).

Otro factor que ha impulsado la evaluación docente basada en estándares es el cúmulo y avances de la investigación sobre eficacia

docente. La investigación sobre este tema comenzó a hacer progresos palpables en los Estados Unidos, y posteriormente en Inglaterra, a principios de los años setenta, y ha sido un pilar en el que se ha sustentado la creación de los estándares. Éstos descansan en los hallazgos de la investigación sobre eficacia docente. Darling-Hammond y Loewenberg-Ball (1998), con base en la investigación de Ronald Ferguson (citado por ellos), sostienen que el conocimiento y pericia del docente (*expertise*) explica el logro de los estudiantes más que cualquiera otra variable. Recientemente, un informe del Banco Mundial (Bruns y Luque, 2014: 5) sostiene que “una vez que los niños ingresan a la escuela, ningún otro factor es más importante que la calidad de los profesores”.

Sobre el tipo de estándares, procedimientos e instrumentos empleados en los sistemas de evaluación docente, no existen estudios que los aborden en conjunto con la eficacia escolar o docente. La eficacia docente es una de las piedras angulares para el desarrollo de los estándares de evaluación. Sólo fue posible identificar un estudio, el cual compara los estándares de los sistemas de evaluación docente de cinco estados de la Unión Americana (Shakman *et al.*, 2012). México y otros países apenas están comenzando a crear sistemas de evaluación docente. La investigación sobre el tema es escasa, por tal motivo se consideró relevante hacer un análisis preliminar de las tendencias de la evaluación docente en el medio internacional.

Las preguntas que guiaron este trabajo fueron: ¿Cuáles han sido la evolución, situación actual y avances recientes de la investigación sobre la eficacia docente? ¿Cómo se ha desarrollado el movimiento de evaluación con base en estándares? ¿Cuáles son los instrumentos y métodos que se emplean para valorar a los docentes a través de estándares? ¿Cuáles son las nuevas tendencias en el desarrollo de estándares de evaluación docente?

MÉTODO Y ORGANIZACIÓN DEL TRABAJO

Se realizó una búsqueda por Internet de documentos que estuvieran relacionados con los estándares de evaluación docente. Asimismo, se

investigó sobre otras publicaciones en revistas y otros medios electrónicos en relación con el tema. El trabajo se organizó de la siguiente manera. En la primera parte se hace un análisis de los antecedentes, situación y tendencias sobre la eficacia escolar; en la segunda parte se hace un análisis histórico de la evolución de los estándares educativos y de evaluación docente. La siguiente sección está dedicada al análisis, diseño y selección de instrumentos empleados en la evaluación, así como sus ventajas y desventajas. Posteriormente se hace un análisis de las nuevas tendencias en la evaluación docente por medio de estándares y su relación con la formación, certificación y evaluación de los docentes. Finalmente, el trabajo termina con las conclusiones y recomendaciones que se desprenden del análisis realizado.

LA INVESTIGACIÓN SOBRE EFICACIA DE LA ESCUELA Y DEL DOCENTE

Además de las políticas para elevar la calidad y equidad de la educación, las cuales se han generalizado a la mayoría de los sistemas educativos, los avances de la investigación docente han sido la piedra angular que ha guiado el diseño de los estándares de evaluación docente. En esta sección se realizará un breve análisis de los antecedentes, situación actual y nuevas tendencias. De acuerdo con Walberg (1986: 215) la investigación científica aplicada, como es el caso de la investigación docente, se lleva a cabo en etapas progresivas que van desde la identificación de resultados valiosos hasta su adopción y evaluación:

1. *Identificación de resultados valiosos.* Disciplinas como la medicina y la educación buscan resultados valiosos como: salud, productos y aprendizaje.
2. *Identificación de factores plausibles.* La investigación significativa busca las causas de los resultados valiosos.
3. *Observación de factores y efectos.* Los factores pueden ser manipulados experimentalmente u observados en su estado natural donde ocurren.

4. *Replicación de covariaciones.* Se replican las relaciones con diferentes sujetos y en una variedad de circunstancias.
5. *Síntesis de la investigación.* Estudios comparables se evalúan y resumen sistemáticamente.
6. *Adopción y evaluación.* Los profesionales pueden razonablemente aceptar las implicaciones de los hallazgos de la investigación que la síntesis probó como válidos.

Estas etapas corresponden a la misma evolución que ha tenido la investigación sobre eficacia docente, principalmente en Estados Unidos, país que ha sido el pionero y líder. En este respecto es posible identificar cinco etapas que ha seguido la investigación sobre eficacia docente, tanto en Estados Unidos como en el medio internacional: 1) los inicios de dicha investigación a finales del siglo XIX y hasta mediados de la década de los sesenta; 2) la segunda, el inicio de la investigación de James Coleman y colaboradores a mediados de los sesenta; 3) inicio del movimiento de eficacia escolar en la década de los setenta; 4) la fusión del movimiento de investigación sobre la eficacia escolar con el de mejora de la eficacia escolar, y 5) la consolidación de la investigación sobre eficacia docente.

La primera etapa comprende los inicios de la investigación sobre eficacia docente a finales del siglo XIX y hasta mediados de la década de los sesenta. La investigación sobre eficacia docente nació en las postrimerías del siglo XIX; de acuerdo con Medley (1972 y 1977) la investigación sobre los factores que hacen una buena enseñanza o docencia apareció en 1896. El autor (Medley, 1972) señala que una de las investigaciones más importantes fue la realizada por Hart, *Teachers and teaching: By ten thousand High School seniors*, publicada en 1934 en un libro con el mismo nombre. El investigador solicitó a los estudiantes de bachillerato que describieran al mejor docente que hubieran tenido y a otro que no les hubiera gustado. Los hallazgos de la investigación de Hart indicaron que un buen docente es aquel que posee habilidades para enseñar: claro, organizado (usa ejemplos), amigable, bien intencionado, interesado en los alumnos, imparcial (sin estudiantes favoritos) y justo al evaluar.

En este periodo comenzaron a realizarse las primeras síntesis de investigaciones. La más reconocida, y posiblemente la primera realizada, es la de Barr (1948). El autor hizo una síntesis de 150 estudios realizados entre 1905 y 1945 en Estados Unidos. En esa época aún no se contaba con las técnicas de metanálisis iniciadas por Gene Glass (1976) a mediados de la década de los setenta. Barr denominaba *eficiencia* a lo que ahora se conoce como eficacia o efectividad; el autor la definía como “los controles comunes sobre el comportamiento tales como: conocimiento específico, habilidades generales, intereses, actitudes e ideales” (Barr, 1948: 205). Asimismo, señalaba que los métodos para identificar estos aspectos eran inadecuados y su relación con la eficiencia no está bien establecida. Aún no aparecían en el escenario de la investigación docente los métodos cualitativos. El grueso de los instrumentos empleados en los estudios analizados fueron test y escalas de observación.

Barr llegó a la conclusión de que existía una gran variedad de vocabulario para caracterizar la eficiencia docente y sus prerrequisitos; que al menos había cinco maneras diferentes de hacerlo: en términos de cualidades de la persona; en términos de las competencias esperadas; en términos de los resultados deseados de los estudiantes; en términos de controles de comportamiento (conocimientos, habilidades generales, intereses, actitudes e ideales); y en términos de una colección miscelánea de datos sobre antecedentes personales. El autor también concluyó que ninguna de estas caracterizaciones parecía adecuada para todos los propósitos pero que no eran, en general, incompatibles y que hasta que se vislumbraran mejores sistemas todas éstas se debían emplear para caracterizar la eficiencia docente. Sin embargo, en esta revisión se aprecia que el grueso de las investigaciones analizadas se enfocaron a concebir la “eficiencia” docente en términos de las cualidades personales o rasgos de personalidad del docente.

En la década de los sesenta y principios de la siguiente se podría ubicar el desarrollo de la investigación docente tal y como la conocemos hoy. Una de las piedras de toque de esta nueva era fue la publicación en 1963 del *Handbook of Research on Teaching*, el cual sentó las bases de la investigación contemporánea de la investigación docente.

De acuerdo con Klette (2007), se pasó de una concepción conductista a una funcional de la docencia.

La segunda etapa se puede ubicar con la célebre investigación de James Coleman y colaboradores, *Equality and educational opportunity*, realizada en 1966 bajo las disposiciones señaladas en el Acta de los Derechos Civiles de 1964 (Coleman, 1990). Los hallazgos de esta investigación señalaron que el nivel socioeconómico y la raza son las variables que explican en gran medida el desempeño de los alumnos y que los efectos o la contribución de la escuela son menores en la explicación de ese desempeño. En esta etapa aparecen los primeros estudios cualitativos, como el bien conocido de Jackson *La vida en las aulas*, obra publicada en inglés en 1968 y en español en 1975. El estudio de Jackson tuvo una gran influencia no sólo en la investigación docente sino también en el currículum.

La tercera etapa inicia en la década de los setenta, con la aparición de la investigación sobre eficacia escolar. Algunos investigadores empezaron a cuestionar los resultados del Informe Coleman; intentaban mostrar la contribución de la escuela en el desempeño de los alumnos. Desde esta década, y hasta principios de los noventa, la investigación se enfocó a la identificación de los factores o efectos de la escuela en el desempeño escolar, más allá de los determinantes socioeconómicos y raciales del desempeño académico que el Informe Coleman señalaba.

Una de las primeras investigaciones que se realizó fue la de Weber (1971); el autor llevó a cabo un estudio sobre habilidades de lectura en tres escuelas de barrios de clase trabajadora en las ciudades de Los Ángeles, Nueva York y Kansas. El autor concluyó que no son los orígenes de clase sino los efectos de la escuela lo que determina el desempeño académico de los alumnos. El estudio de Weber identificó que las características comunes de este tipo de escuelas son: un fuerte liderazgo, fuerte énfasis en la práctica de la lectura y uso de la fonética, buena atmósfera (clima escolar) y una evaluación cuidadosa de los progresos del alumno, entre los factores principales.

Edmonds (1979) es considerado el padre del movimiento de eficacia docente; este autor realizó la primera síntesis de las investigaciones producidas en esa década (incluyendo algunas de su autoría) sobre los factores que promueven la eficacia docente en escuelas de

bajo nivel socioeconómico. Los hallazgos de Edmonds son similares a los que encontró Weber; agrega que otra de las características de las escuelas eficaces es que están dispuestas a evitar cosas que no funcionan y a comprometerse a implementar cosas que sí funcionan.

Además de estas investigaciones, el marco sistémico para la investigación docente creado por Dunkin y Biddle (1974) proporcionó elementos teórico-metodológicos que impulsaron la investigación. El modelo conocido como proceso-producto, que en realidad es un marco de investigación sistémico para la investigación sobre las variables presagio-contexto-proceso-producto, dio nuevos bríos a la investigación docente. Shulman (1986) señala que este marco metodológico se inspiró en Mitzel.

Los avances en esta línea de investigación en la década de los ochenta se ven reflejados en el tercer *Handbook of Research on Teaching* (Wittrock, 1986). El artículo de Shulman (1986) sobre los “Paradigmas de los programas de investigación en el estudio de la docencia: una perspectiva contemporánea” analizó los principales paradigmas para el estudio de la docencia, lo cual proporcionó, por primera vez, un conjunto sistemático de principios teóricos para su estudio e investigación. En ese mismo *Handbook* se sintetizaron los avances de la investigación sobre eficacia escolar y los efectos de la escuela; destacan los trabajos de Walberg (1986) y de Brophy y Good (1986).

La cuarta etapa es la fusión del movimiento de investigación sobre la eficacia escolar con el de mejora de dicha eficacia. El primer movimiento está centrado, como se ha visto, en identificar los factores que hacen eficaz a una escuela, el segundo en mejorar los resultados de la escuela a través de intervenciones locales realizadas por directivos y docentes (Murillo, 2004). La mejora de la eficacia escolar comenzó en la década de los ochenta y se consolidó en las dos siguientes (Hopkins *et al.*, 2011). El movimiento de mejora de la eficacia escolar –señalan los mismos autores– comenzó con la introducción del desarrollo organizacional y los movimientos de auto-gestión y mejora del liderazgo escolar en Estados Unidos, Australia e Inglaterra. La obra de Scheerens (2000) es representativa de este tipo investigación y tendencias de este movimiento. La estrategia de la

mejora de la eficacia escolar es resumida por Hopkins *et al.* (2011) y consta de cinco fases:

1. Comprensión de la cultura organizacional de la escuela.
2. Investigación-acción e iniciativas individuales.
3. Gestionar el cambio y énfasis en el liderazgo.
4. Construcción de la capacidad para el aprendizaje en el nivel local.
5. Rumbo al mejoramiento sistemático.

Este movimiento ha tenido una influencia notable en la construcción de los estándares, pues centra las acciones de mejora de la escuela en el nivel local, bajo el liderazgo del director y docentes que trabajan en equipo, crean comunidades de aprendizaje y diseñan estrategias de desarrollo profesional en equipo e individualmente, aspectos que se dejan notar en los estándares adoptados por varios países.

La última y quinta etapa es la consolidación de la investigación sobre eficacia docente. Durante la última década la investigación retornó, de alguna manera, a sus inicios; su objeto de estudio cambió de la eficacia de la escuela a la eficacia del docente. La investigación y las nuevas síntesis ahora se enfocaron a identificar y explicar los factores que hacen a un docente exitoso. La investigación sobre eficacia docente está alcanzando su madurez. Parece existir consenso sobre lo que constituye una docencia eficaz. Sin embargo, aún quedan preguntas por responder, como las que se plantean en la reciente investigación sobre eficacia escolar (Ko y Sammons, 2013): ¿Cómo vamos a definir la eficacia docente? ¿Debería estar restringida sólo a la docencia en el aula? ¿La mejor visión de la eficacia es en relación con la influencia en el aula o sobre los resultados de los estudiantes? ¿Cómo podemos medir sus efectos? ¿Qué instrumentos deberíamos emplear? Éstas y otras preguntas aún no se han respondido cabalmente.

Ko y Sammons (2013: 2) resumen los factores que la investigación ha determinado como los más importantes para caracterizar la eficacia del docente:

- Tiene claridad sobre los objetivos de la enseñanza.

- Tiene conocimiento sobre los contenidos del currículum y estrategias para enseñarlos.
- Comunica a sus estudiantes lo que espera de ellos, y por qué.
- Emplea de manera experta los materiales didácticos para dedicar más tiempo a las prácticas que clarifican y enriquecen el contenido.
- Conoce a sus estudiantes, adapta la enseñanza a sus necesidades y anticipa las falsas concepciones que pueden existir sobre el conocimiento existente.
- Enseña a los estudiantes estrategias metacognitivas y les da oportunidades para dominarlas.
- Aborda objetivos de alto como de bajo nivel cognitivo.
- “Monitorea” la comprensión de los estudiantes a través de una retroalimentación adecuada y frecuente.
- Integra su docencia con la de otras áreas y materias.
- Acepta responsabilidad por los resultados de los estudiantes.

Existe un amplio consenso en la investigación acerca de lo que es la eficacia docente. Darling-Hammond y Baratz-Snowden (2007: 112) resumen estos factores de la siguiente manera:

Los docentes eficaces involucran a los estudiantes en un aprendizaje activo –debatiendo, discutiendo, investigando, experimentando y construyendo modelos, trabajos escritos y productos, además de escuchar y leer información, observando demostraciones y practicando habilidades. Los docentes tienen expectativas muy claras sobre una alta calidad en el trabajo y proporcionan modelos a los estudiantes para que cumplan con esos estándares. Los docentes también proveen de una retroalimentación constante y continuamente revisan sus trabajos para que cumplan con esos estándares.

Como puede apreciarse en las citas de arriba, la eficacia está relacionada con tres grandes dimensiones de la actuación del docente: el *conocimiento del currículum*, *conocimiento, interacción y evaluación de los estudiantes*, y la *eficacia de las estrategias docentes empleadas en el aula*. A partir de estas dimensiones varios sistemas

educativos han construido los estándares contemporáneos de evaluación docente a los cuales se han agregado *el trabajo en equipo* y *desarrollo profesional*, es decir, la colaboración del docente con colegas y otros grupos o instancias de la comunidad escolar y una preocupación activa de los docentes por desarrollarse profesionalmente. La síntesis de investigación de Goe, Bell y Little (2008: 11) proporciona una definición comprensiva de la eficacia docente que incluye estas dimensiones:

1. Los docentes eficaces tienen altas expectativas de todos los estudiantes y los ayudan a aprender, midiendo el aprendizaje a través del valor agregado y otros test basados en el progreso de los estudiantes o por otros instrumentos alternativos.
2. Los docentes eficaces contribuyen a lograr en todos los estudiantes resultados académicos, actitudinales y sociales positivos tales como: asistencia regular, promoción en tiempo al siguiente grado, graduación en tiempo, sentido de autoeficacia y comportamiento cooperativo.
3. Los docentes eficaces emplean recursos diversos para planear y estructurar oportunidades para un aprendizaje comprometido; monitorean formativamente los progresos de los estudiantes, adaptan la enseñanza tanto como sea necesario; y evalúan el aprendizaje usando fuentes múltiples de evidencia.
4. Los docentes eficaces contribuyen al desarrollo de las aulas y las escuelas que valoran la diversidad y la actitud cívica.
5. Los docentes eficaces colaboran con otros docentes, administradores, padres y profesionales de la educación para asegurar el éxito de los estudiantes, particularmente de estudiantes con necesidades especiales y aquellos que están en riesgo de fracasar.

Vale la pena reiterar que la definición de eficacia docente que se adopte es un paso necesario para el diseño y definición de los estándares de evaluación. Sin una definición apropiada, un sistema de evaluación docente basado en estándares pondría en duda su validez de contenido.

LOS ESTÁNDARES DE EVALUACIÓN DOCENTE. ANTECEDENTES

Los estándares de evaluación docente se iniciaron en Estados Unidos; su inicio no tiene una fecha marcada. Se considera que desde mediados del siglo XIX ese país mostró preocupación por definir las normas y expectativas de la formación de docentes en las escuelas normales (Lee, 2007). No obstante, es a partir del siglo XX que se aprecia consistentemente la necesidad por definir las normas o estándares de la preparación docente (Edelfelt y Rath, 1998). Desde entonces, la preocupación que ha guiado la creación de estándares de formación de docentes es uniformar una gran diversidad de requisitos de admisión, currícula y requerimientos para ingresar a la profesión. Los esfuerzos más notables y sostenidos comenzaron a principios del siglo XX. Edelfelt y Rath (1998) informan que se realizaron varios estudios, entre ellos destacan los siguientes.

En 1929 la Carnegie Foundation patrocinó un estudio para determinar lo que los docentes realizaban bajo las condiciones “modernas de práctica”. Parece ser que la primera vez que apareció la concepción de la palabra *estándares* en un informe sobre este tema fue cuando la Association of Colleges for Teacher Education (actualmente la American Association for Colleges of Teacher Education) creó en 1945 el Comité sobre Estándares y Encuestas, cuya intención fue determinar la disposición del campo de la formación docente para mejorar los estándares e identificar los aspectos en los cuales tales disposiciones podrían adoptarse. En 1946 la Commission of Teacher Education de la American Council on Education publicó un reporte sobre el mejoramiento de la formación docente.

En 1961 la National Commission on Teacher Education and Professional Standards estableció una definición amplia de competencia docente.

Aunque aún existe debate sobre el origen del modelo de enseñanza de las competencias, es posible que el artículo de Elam (1971), publicado a nombre del Comité sobre Formación Docente Basada en el Desempeño (Committee on Performance-Based Teacher Education), integrado por la Asociación Americana de Colegios de Formación Docente (AACTE, por sus siglas en inglés), sea el primer trabajo que pos-

tula la enseñanza basada en competencias. Entonces, se puede colegir que este enfoque tuvo sus orígenes formales dentro del campo de la formación docente. Elam (1971), en representación de esa comisión, señala que la “formación docente basada en el desempeño”, como denominó a este enfoque, se caracterizaba por cinco elementos básicos:

1. Las competencias docentes a demostrar son derivadas del rol (docente) y especificadas en términos conductuales, y son públicas.
2. Los criterios de evaluación se basan en competencias, especificados en niveles de maestría y se hacen públicos.
3. Su evaluación requiere que se tome al desempeño como la principal evidencia y toma en cuenta el conocimiento de los estudiantes.
4. El ritmo de avance de los estudiantes depende de la competencia demostrada.
5. El programa de enseñanza facilita el desarrollo y evaluación de las competencias especificadas.

El enfoque general propuesto por Elam (1971) se etiquetaba como basado en el desempeño, el cual estaba compuesto por el enfoque de competencias y el Sistema Keller de instrucción personalizada, el cual fue muy popular en las décadas de los setenta y los ochenta. Este enfoque sigue vigente pero ya no es tan popular; entre sus limitaciones se encuentran que los estudiantes bajo este sistema no pueden avanzar a su ritmo debido a las restricciones del periodo escolar y que no avanzan al ritmo esperado pues postergan (*procrastination*) la entrega de los trabajos (Eyre, 2007).

De acuerdo con Edelfelt y Raths (1998), en años posteriores se realizaron otros informes importantes. En 1972 por primera vez aparece la formación de docentes con base en competencias. El Comité sobre Prioridades Nacionales en la Formación Docente dio a conocer un reporte denominado “El poder de la formación docente basada en competencias”. En 1974 la American Association of Colleges for Teacher Education creó la Comisión para el Estudio de la Profesión Docente, la cual publicó un informe dos años después en el que estableció 52 recomendaciones que eran esencialmente estándares para la formación docente.

EL PRESENTE Y NUEVAS TENDENCIAS EN LOS ESTÁNDARES DE EVALUACIÓN DOCENTE. LOS PRIMEROS ESTÁNDARES PARA LOS ESTUDIANTES Y LOS DOCENTES

El preámbulo formal y más reciente que dio pie a la creación de estándares docentes fue el informe “*Nation at risk*” (“Nación en riesgo”). En 1981 en el Departamento de Educación de Estados Unidos se creó la Comisión Nacional sobre Excelencia en Educación (The National Commission on Excellence in Education, 1983) conformada por profesores universitarios, juntas de padres de familia, docentes y otros miembros de la comunidad educativa. El informe preparado por esa comisión hizo notar las debilidades del currículum, la declinación de los puntajes en los test de aprovechamiento, el insuficiente tiempo destinado a la enseñanza, el poco tiempo que los escolares dedicaban a su aprendizaje fuera de la escuela, así como la débil preparación de los maestros (The National Commission on Excellence in Education, 1983). El informe recomendó, por primera vez, que se elaboraran estándares más exigentes para evaluar a los estudiantes graduados de bachillerato y una jornada de clase más larga. Asimismo, fue una de las primeras ocasiones en que Estados Unidos empleó datos de otros países para comparar su desempeño escolar.

Los estándares originalmente se crearon para proponer qué conocimiento debía enseñarse, es decir, estándares de conocimientos o curriculares. En 1989 se crearon los primeros estándares de esta clase; fueron promovidos por el Consejo Nacional de Profesores de Matemáticas (NCTM, por sus siglas en inglés). Su finalidad fue mejorar la calidad de la educación en matemáticas (National Council of Teachers of Mathematics, 2014). Estos estándares causaron mucha controversia en Estados Unidos; animaban el uso de la calculadora por los estudiantes y la solución de problemas abiertos, en lugar de ejercicios repetitivos (University of Rochester, 2006). Por otro lado, en la década pasada comenzaron a surgir las críticas sobre el uso de los estándares. El empleo de docenas de estándares en la evaluación del aprendizaje fue criticado porque sacrificaba la profundidad en la enseñanza en pos de la cobertura e impedía que los docentes emplearan sus propios criterios acerca de lo que se debía enfatizar y qué

suprimir de los extensos libros de texto (Malone y Nelson, 2006). En cuanto a los estándares para evaluar a los docentes, dada la inercia histórica que existió en el siglo xx, los primeros creados fueron para evaluar y certificar a los nuevos docentes y también eran excesivos; por ejemplo, los estándares de calidad del programa y eficacia para los programas de preparación profesional de la Universidad del estado de California estaban constituidos por 32 estándares (Commission on Teacher Credentialing State of California, 1992).

La Interstate New Teacher Assessment and Support Consortium (INTASC), creada por el Consejo de Jefes Oficiales Escolares Estatales (Council of Chief State School Officers, ccsso), que agrupa a los titulares de los 50 departamentos de educación estatales y del Distrito de Columbia, desarrolló los primeros estándares para evaluar a los nuevos docentes y certificar a los docentes activos (Interstate New Teacher Assessment and Support Consortium, 1992). La comisión tuvo la participación de Linda Darling-Hammond. Los diez estándares diseñados para evaluar al docente están agrupados en tres dimensiones: *conocimiento*, *desempeños* y *disposiciones* bajo los cuales los nuevos docentes son certificados. Estas mismas dimensiones de los estándares siguen vigentes en la última edición (Interstate New Teacher Assessment and Support Consortium, 2010) que no sólo está dedicada a evaluar a los nuevos docentes sino también a los experimentados. A mitad de la década de los noventa una comisión semejante creó los primeros estándares para evaluar a los líderes escolares y la evaluación sólo se compone de seis estándares (Interstate School Leaders Licensure Consortium, 1996).

EL PRESENTE

Los nuevos estándares de evaluación docente, tanto en Estados Unidos como en otros países anglosajones, se caracterizan por ser breves y generales. No obstante, al menos en Estados Unidos, los docentes son evaluados tomando en alta consideración el desempeño académico de sus alumnos, medido a través de las pruebas estandarizadas; los resultados se emplean como medida de rendición de cuentas por parte de los docentes y de sus escuelas. Esto ha sido fuertemente influido

por las políticas creadas en Estados Unidos, las llamadas reformas basadas en estándares. Dichas políticas se han plasmado en “Ningún niño será ignorado” (*No child left behind*) de la administración del presidente George Bush y durante la del presidente Obama “La carrera hacia la cima” (*Race to the top*). Estas políticas miden el desempeño básicamente a través de las pruebas de rendimiento, lo que se ha denominado en ese país como Sistema de Evaluación del Valor Agregado (*Value-Added Assessment System*). Este sistema de evaluación recompensa monetariamente a los docentes por los resultados obtenidos en las pruebas estandarizadas de rendimiento. Los docentes en un principio rechazaron este tipo de recompensas, pero ahora son aceptadas por la mayoría de ellos (Dillon, 2007). El sistema de evaluación docente con base en los resultados ha sido cuestionado por dos razones: por su validez y por las consecuencias que tiene en la escuela.

A pesar del empleo de los resultados del desempeño de los estudiantes en la evaluación docente, no hay evidencias suficientes de que tenga una adecuada validez (Shepard, 2012). De acuerdo con Furham (2003) el uso de los test como instrumentos de rendición de cuentas parte de los siguientes presupuestos:

1. El desempeño o logro de los estudiantes es un valor u objetivo clave de la escuela así como la construcción de la rendición de cuentas alrededor de éste.
2. El desempeño es medido con exactitud y autenticidad por los instrumentos empleados en la evaluación.
3. Las consecuencias de la evaluación motivan al personal y a los estudiantes.
4. El resultado es el mejoramiento de la docencia y niveles de desempeño más altos.
5. Las consecuencias desafortunadas o no buscadas son mínimas.

La realidad es que aún no existe suficiente investigación que ampare la validez de estos presupuestos. Shepard (2012) logra en su investigación poner en balance las bondades y limitaciones del empleo del valor agregado como instrumento de evaluación docente. Entre sus aciertos, señala la autora, está que puede distinguir el desempe-

ño de los docentes novatos de los experimentados. Asimismo, esta investigadora cita estudios que indican que las escuelas con mejores docentes, que cuentan con habilidades y conocimientos previos, logran responder a esas políticas. La presión por mejorar los resultados puede estimular el desempeño académico en ese tipo de escuelas.

No obstante, existen otro tipo de evidencias que apuntan a que esta política de evaluación docente no es del todo válida. La autora citada (Shepard, 2012) señala que un factor que atenta contra la validez del empleo del valor agregado en la evaluación docente es que no es posible agrupar a los estudiantes con base en sus características académicas; son grupos heterogéneos de estudiantes y en ocasiones los efectos de la labor de un docente con alto desempeño y comprometido se manifiestan en el siguiente curso escolar, pero al nuevo docente se le atribuyen las ganancias en el aprendizaje y no al docente que realmente las generó, es decir, el docente anterior.

En cuanto a las consecuencias del empleo del valor agregado en la evaluación docente, un estudio reciente realizado en Arizona (Collins, 2014) señala que la presión y competencia generada por este tipo de evaluación reduce la moral y la colaboración entre los docentes, y se convierte en un estímulo para hacer trampa o enseñar para responder los exámenes. Este fenómeno no sólo se da en Estados Unidos. En México fue notable ver cómo los estados del país que típicamente ocupaban los últimos lugares en la Evaluación Nacional del Logro Académico en Centros Escolares (Enlace), de repente lograron escalar a los primeros lugares en desempeño. Otro de los efectos es que reduce el currículum a lo que simplemente se evalúa en los test (Figazzolo, 2013).

MÉTODOS E INSTRUMENTOS DE EVALUACIÓN DOCENTE

Es probable que los primeros métodos e instrumentos para evaluar el desempeño docente tengan su origen en los países anglosajones. Los antecedentes más remotos conocidos datan de la época colonial de Estados Unidos. Lortie (1975), en su célebre estudio de la sociología de la profesión docente, señala que la evaluación docente

se realizaba por la junta escolar y los padres de familia a través de las recitaciones públicas de los alumnos. Posteriormente, en tiempos más recientes, los docentes eran evaluados a través de la exposición de los trabajos de los estudiantes.

Los sistemas educativos descentralizados, como Estados Unidos, confieren la facultad a los estados y distritos escolares de contratar a su personal docente. Para ello se auxilian de las estrategias de la administración contemporánea de los recursos humanos. A los estudiantes de administración educativa se les enseña que el proceso está conformado por un ciclo que consta de las siguientes fases (Seyfarth, 2008): 1) planeación, en dos fases: estimación de la matrícula del siguiente año y estimación del personal que será requerido para satisfacer las necesidades; 2) reclutamiento de los candidatos a docentes; 3) selección e inducción al puesto; 5) nombramiento; 6) desarrollo profesional, y 7) evaluación. Los profesores de la materia proveen de consejos prácticos, como el valorar al candidato desde el momento que llena su solicitud con letra manuscrita y determinar la calidad de su redacción y buena letra.

A partir de la evaluación docente con base en estándares existe un progreso notable en el desarrollo de instrumentos de evaluación del docente. Sin embargo, los avances no son concluyentes. Una de sus debilidades, como lo muestran diversos estudios, es que no poseen una adecuada validez discriminante. Por ejemplo, en varios estados de la Unión Americana 99 por ciento de los docentes obtienen una evaluación satisfactoria (Shakman *et al.*, 2012).

DISEÑO DE UN SISTEMA DE EVALUACIÓN DOCENTE BASADO EN ESTÁNDARES

Después del diseño de una política de evaluación docente y previa o simultáneamente a la selección y diseño de estándares e instrumentos de evaluación, es conveniente diseñar estrategias de comunicación con los actores involucrados para difundir, retroalimentar y ganar la aceptación de la innovación. Es decir, una cultura de evaluación docente que no se crea de la noche a la mañana. Al respecto, Cogshall (2007) propone establecer un marco de comunicación para medir

la calidad y eficacia docente. Este marco está integrado por cuatro componentes: 1) planeación de la comunicación, 2) clarificación de objetivos, 3) términos que definen la calidad del docente y 4) instrumentos de medición y recursos.

La planeación de la comunicación tiene como objetivo buscar los canales más adecuados para comunicar y difundir la evaluación e involucrar a los actores clave en ese proceso. Esto se puede lograr por medio de conferencias, foros, reuniones en el cabildo municipal, con las juntas de padres de familia y profesores, así como comunicados a los medios de comunicación, entre otros. Otro aspecto esencial, según la autora (Coggshall, 2007), es que cuando se ha logrado la difusión y comunicación es importante realizar una investigación sobre lo que es la calidad docente y alcanzar un consenso sobre lo que significa; por ejemplo, qué datos e información del docente y los alumnos se debe recolectar para medirla.

El segundo componente, clarificación de objetivos, se enfoca a analizar y lograr una comprensión de lo que significa la calidad y eficacia docente, así como los objetivos y propósitos para medirla. Dichos propósitos, refiere la autora (Coggshall, 2007), se derivan también de la clarificación de los propósitos de las políticas de evaluación docente. Asimismo, la autora sugiere preguntas en varios ámbitos, federal, estatal y local, como las siguientes:

- ¿Los docentes de calidad se distribuyen equitativamente en todos los estudiantes de cada estado, municipios y zonas escolares?
- ¿Los programas de preparación de docentes están produciendo docentes de alta calidad?
- ¿Se están contratando docentes de alta calidad?
- ¿Los programas de desarrollo docente mejoran eficazmente la calidad de los docentes?
- ¿El sistema propuesto es capaz de identificar a los docentes de bajo desempeño y suministrar recursos para mejorar?

Estas preguntas y otras similares pueden contribuir a clarificar los objetivos de la evaluación docente y a planear otros aspectos y condiciones que requieren los docentes y las escuelas para mejorar su calidad.

El tercer componente lo constituyen los términos que definen la calidad del docente. La eficacia docente es un componente multidimensional. Coggschall (2007) propone que la eficacia docente sea una de las dimensiones centrales de la calidad. En el cuadro 1 se pueden apreciar las dimensiones de la calidad docente, sus definiciones y los instrumentos que la autora propone. De acuerdo con ella, el primer paso es la definición de esas dimensiones con base en las dimensiones analizadas arriba. Asimismo, este estudio propone una lista amplia y diversa de instrumentos para valorar las siete dimensiones de la calidad del docente que se pueden evaluar. Es una síntesis completa y sucinta de la evaluación multidimensional de la calidad docente.

CUADRO 1

Dimensiones de la calidad docente

Dimensiones	Definición	Medición e instrumentos
Eficacia del docente	En la investigación sobre calidad docente, y en muchas comunidades estudiantiles de las políticas educativas, la palabra eficaz connota algún impacto directo –o efecto– en los resultados. En el caso de los docentes, este término usualmente se define como la contribución a los logros de los estudiantes en las pruebas, no obstante es posible medir otros resultados valiosos tales como: tasas de graduación del bachillerato; motivación del estudiante; creencias académicas de eficacia u otro resultado social, intelectual o de comportamiento. Entonces los docentes altamente eficaces pueden ser definidos como aquellos que muestran evidencia de producir resultados sobresalientes (como sean definidos o medidos).	<ul style="list-style-type: none"> - Rendimiento de los estudiantes (incluyendo métodos de valor agregado y modelos de crecimiento de los estudiantes). - Tasas de deserción. - Trabajo documentado de los estudiantes. - Compromiso, afecto y persistencia de los estudiantes.
Cualificaciones del docente	Los docentes cualificados poseen credenciales que certifican que han completado o aprobado exitosamente un programa de preparación docente (con frecuencia nacionalmente acreditados); han demostrado su buen perfil (usualmente con carta de no antecedentes penales) y tienen un grado de licenciatura. Además los estados certificadores casi siempre requieren que presenten un examen de contenido y pedagogía para ser certificados.	<ul style="list-style-type: none"> - Grados académicos. - Cursos o revisión de sus certificados escolares. - Certificación y requerimientos para la obtención de la cédula profesional. - Prestigio y estatus del programa de formación docente.
Cualificaciones del docente	Los docentes altamente cualificados como son definidos en el acta "Ningún niño será ignorado" (NCLB, por sus siglas en inglés), son quienes están totalmente certificados por el Estado, tienen un grado de licenciatura y demuestran que tienen <i>expertise</i> en el contenido de las materias que realmente enseñan.	<ul style="list-style-type: none"> - Puntajes obtenidos en la certificación.

Cuadro 1 (continuación)

Dimensiones	Definición	Medición e instrumentos
Docentes expertos	Los docentes expertos tienen un conocimiento amplio y profundo de la materia que enseñan y de la manera de enseñar ese contenido. Los docentes expertos también tienen conocimiento de cómo aprenden los estudiantes así como de un amplio rango de métodos de enseñanza efectivos para ayudar a todos los estudiantes a aprender. Los docentes expertos son también culturalmente competentes en el contexto en el cual enseñan.	<ul style="list-style-type: none"> - Exámenes (incluyendo test de opción múltiple o test con respuestas abiertas, conocimiento didáctico o conocimiento docente). - Portafolios docentes. - Observación en el aula. - Evaluación del desarrollo profesional.
Capacidad del docente	Los docentes con capacidad para tener éxito demuestran una habilidad para aprovechar su contexto profesional para enseñar mejor. Por ejemplo, están bien preparados y comprometidos con un aprendizaje continuo, son reflexivos y organizados, demuestran habilidad verbal, son capaces de analizar su práctica docente y articular y refinar sus filosofías de la docencia.	<ul style="list-style-type: none"> - Entrevistas al docente. - Revisión de sus cualificaciones. - Estudios de casos o viñetas. - Observaciones en el aula. - Test de inteligencia, test de habilidad verbal. - Declaraciones sobre la filosofía de la docencia.
Carácter del docente	Los docentes de carácter tienen ciertos rasgos y disposiciones que se han observado están relacionados con la calidad en la docencia: sensibilidad, calidez, entusiasmo, pasión, creatividad, persistencia, afectividad, compromiso, sentido de autoeficacia y autenticidad.	<ul style="list-style-type: none"> - Entrevistas al docente. - Referencias profesionales y personales. - Encuestas y entrevistas a los estudiantes. - Observaciones en el aula. - Declaraciones de filosofía sobre la docencia. - Revisión de sus antecedentes penales.
Desempeño del docente	Docentes con alto desempeño son aquellos en cuyas acciones se observa que cumplen o exceden altos estándares de práctica docente. Docentes con alto desempeño demuestran el conocimiento y habilidades para proporcionar docencia de alta calidad a todos sus estudiantes. Estos docentes probablemente producirán resultados sobresalientes en el aprendizaje de los estudiantes pero puede darse el caso de que no demuestren evidencia suficiente, válida y confiable de los resultados de aprendizaje de los estudiantes.	<ul style="list-style-type: none"> - Portafolios docentes. - Protocolos de observación realizados por el director, pares o especialistas. - Evaluación con base en test estructurados. - Muestras del trabajo del docente. - Evaluación por medio del video. - Encuestas y entrevistas a los estudiantes. - Encuestas y entrevistas a los padres. - Análisis de casos de enseñanza. - Diarios del docente. - Encuestas sobre el conocimiento del currículum oficial.

Cuadro 1 (continuación)

Dimensiones	Definición	Medición e instrumentos
Éxito del docente	El término docentes exitosos podría significar docentes que son "altamente eficaces" en producir estudiantes exitosos (como se han definidos estos) por lo que ellos probablemente producirán crecimiento y éxito en los estudiantes.	Mediciones combinadas de eficacia y desempeño.

Fuente: Coggshall, 2007: 8-11.

Sin embargo, la lista de instrumentos propuestos por Coggshall es muy amplia; en la práctica, la evaluación docente basada en estándares se limita a menos instrumentos. La síntesis de investigación realizada por Goe, Bell y Little (2008) señala que los instrumentos más empleados en esta tarea se reducen a los siguientes: 1) observación, 2) observación por el director, 3) muestras de trabajo, 4) portafolios, 5) autoevaluación, 6) encuestas a padres y alumnos, y 7) valor agregado/resultados de las pruebas. Los cuadros 2 a 8 describen de manera independiente los instrumentos más empleados en la evaluación docente. De acuerdo con la amplia síntesis de investigación realizada por Goe, Bell y Little, (2008) además de su descripción, fortalezas y reservas, incluyen como un instrumento de observación al director. Probablemente las autoras incluyeron al director como instrumento debido a que en los países anglosajones la evaluación docente es fundamentalmente responsabilidad del director escolar. No obstante, se respetó la clasificación de las autoras ya que pueden aportar información sobre el director como instrumento de evaluación docente.

En los mencionados cuadros se puede apreciar que el estado del diseño y uso de instrumentos para la evaluación docente es aún una tarea incompleta, que existe poca investigación sobre la validez y confiabilidad de los instrumentos empleados y que todos éstos tienen ventajas y limitaciones. De acuerdo con la revisión de Shakman *et al.* (2012), de los siete instrumentos o métodos señalados por Goe, Bell y Little, el principal instrumento que se emplea en la evaluación docente, al menos en Estados Unidos, es la observación por parte del director y los resultados de los test. La obser-

vación usualmente se realiza tomando como referencia, o criterios, los estándares de evaluación. Otro instrumento frecuentemente empleado es la muestra de trabajos docentes (*teaching artifacts*); la diferencia con los portafolios es que estos últimos son preparados deliberadamente, en cambio la muestra de los trabajos son como ocurren en el aula y su entorno. En cuanto a los portafolios docentes, en Estados Unidos se emplean sobre todo para certificar a los docentes, no es común que se empleen como evaluación de los docentes ya certificados.

CUADRO 2

Instrumentos y técnicas empleados con mayor frecuencia en la evaluación docente. La observación

Descripción	Investigación	Fortalezas	Reservas
Empleada para medir procesos observables en el aula, incluyendo prácticas del docente, aspectos holísticos de la enseñanza e interacciones entre docentes y estudiantes. Puede medir aspectos globales de la docencia o específicos de la práctica docente de la materia o del contexto.	Algunos protocolos –de observación– ampliamente investigados han encontrado relación con el desempeño de los estudiantes, pero las relaciones son en algunos casos modestas. Los hallazgos de la investigación y su validez son muy dependientes de los instrumentos empleados. Hay carencia de investigación sobre los protocolos de observación para la evaluación docente.	<ul style="list-style-type: none"> - Proporciona información rica sobre los comportamientos y actividades en el aula. - Los actores involucrados la consideran como una medida directa y justa. - Dependiendo del protocolo, puede ser empleada en varias materias, grados y contextos. - Puede proveer de información útil tanto para la evaluación formativa como sumativa. 	<ul style="list-style-type: none"> - Se debe poner cuidado en la elección o diseño de un protocolo confiable y capacitar y sintonizar a los observadores. - La observación en el aula es cara debido al costo de los observadores, la capacitación intensiva y la sintonización entre los observadores; es un costo adicional pero necesario para su validez. - Este método evalúa comportamientos del aula observables pero no es útil para evaluar creencias, sentimientos, intenciones o actividades fuera del aula.

Fuente: Goe, Bell y Little, 2008.

CUADRO 3

Instrumentos y técnicas más empleados en la evaluación docente. La evaluación del director

Descripción	Investigación	Fortalezas	Reservas
<ul style="list-style-type: none">- Está basada en la observación, puede ser estructurada o no estructurada; sus usos y procedimientos varían ampliamente.- Es empleada comúnmente para dar a los docentes novatos definitividad o despedirlos.	Los estudios que han comparado las evaluaciones subjetivas del director encontraron resultados contradictorios. Existe poca evidencia sobre la validez de las evaluaciones que se llevan en la escuela, pero la evidencia que existe indica que la capacitación del director es limitada y ocasional, lo cual disminuye la validez de sus evaluaciones.	<ul style="list-style-type: none">- Puede representar una perspectiva útil pues está basada en el conocimiento que tiene el director de la escuela y el contexto.- Es generalmente viable y puede ser un componente útil en un sistema empleado para realizar juicios sumativos y proporciona retroalimentación formativa.	<ul style="list-style-type: none">- Los instrumentos de evaluación usados sin una adecuada capacitación o consideración del propósito que persigue disminuirán su validez.- Los directores pueden no estar cualificados para evaluar a los docentes en materias o contextos muy especializados.

Fuente: Goe, Bell y Little, 2008.

CUADRO 4

Instrumentos y técnicas más empleados en la evaluación docente. Muestras del trabajo docente en el aula

Descripción	Investigación	Fortalezas	Reservas
Los protocolos estructurados son empleados para analizar las muestras de trabajo docente con el fin de determinar la calidad de la enseñanza en un aula. Pueden incluir planeación de las unidades didácticas, tareas que asigna el docente, evaluaciones, rúbricas y trabajo de los estudiantes.	La investigación exploratoria ha relacionado la calificación de las muestras de trabajo docente con valoraciones de la práctica docente. La calidad de los trabajos de los estudiantes y sus ganancias en el aprendizaje. Se requiere de más investigación para establecer la confiabilidad en las calificaciones y determinar la cantidad ideal de trabajo de la muestra. Existe carencia de investigación en el uso de muestras estructuradas de trabajo en la práctica.	<ul style="list-style-type: none">- Si se emplea un protocolo válido, observadores capacitados, y las muestras son variadas en su calidad, puede ser una medida útil de la calidad de la docencia.- Es práctico y viable ya que las muestras ya han sido creadas para el trabajo en el aula.	<ul style="list-style-type: none">- Se requiere de más validez e investigación.- Evaluadores capacitados y con conocimiento pueden ser costosos, pero son necesarios para asegurar su validez.- Este método puede ser prometedor en términos de confiabilidad y validez entre una observación completa y métodos menos directos como las autoevaluaciones docentes.

Fuente: Goe, Bell y Little, 2008.

CUADRO 5

Instrumentos y técnicas más empleados en la evaluación docente. Portafolios

Descripción	Investigación	Fortalezas	Reservas
Es usado para documentar un amplio rango de comportamientos y responsabilidades docentes. Se ha empleado mucho en los programas de formación de docentes para la evaluación de los desempeños de los candidatos a docentes y nuevos docentes.	La investigación sobre la validez y confiabilidad está en proceso y se han manifestado preocupaciones acerca de su consistencia/estabilidad en la calificación de este instrumento. Existe carencia de investigación que relacione los portafolios con el desempeño de los estudiantes.	<ul style="list-style-type: none"> - Es holística y puede medir aspectos de la docencia que no son fácilmente observables en el aula. - Puede ser empleada con docentes de todos los campos. - Proporciona altos niveles de credibilidad entre los actores. - Es una buena herramienta para la reflexión y mejoramiento del docente. 	<ul style="list-style-type: none"> - Este método consume mucho tiempo tanto de los docentes como de los evaluadores; los evaluadores deben tener conocimiento del contenido. - La estabilidad de los puntajes puede no ser suficiente para emplearlo en evaluaciones sumativas. - Los portafolios son difíciles de estandarizar (compararlos con otros docentes o escuelas). - Los portafolios representan el trabajo sobresaliente del docente, pero no las actividades cotidianas del aula.

Fuente: Goe, Bell y Little, 2008.

CUADRO 6

Instrumentos y técnicas más empleados en la evaluación docente. Autoevaluación del docente

Descripción	Investigación	Fortalezas	Reservas
Los informes de lo que los docentes hacen en el aula pueden ser evaluados a través de encuestas, diarios sobre la enseñanza y entrevistas. Puede variar ampliamente en lo que se enfocan y el nivel de detalle.	Los estudios sobre la validez de los instrumentos de autoevaluación presentan resultados no concluyentes y diversos. Instrumentos muy detallados sobre la práctica puede que sean más aptos para capturar las prácticas docentes reales, pero puede ser difícil establecer la confiabilidad o puede resultar en instrumentos muy restringidos.	<ul style="list-style-type: none"> - Puede medir factores no observables que pueden afectar la docencia tales como: conocimiento, intenciones, expectativas y creencias. - Proporciona una perspectiva singular del docente. - Es muy viable y eficiente en costo; puede recolectar una gran cantidad de información en una sola vez. 	<ul style="list-style-type: none"> - La validez y confiabilidad de las autoevaluaciones no está completamente establecida y depende del instrumento empleado. - El empleo o la creación de un instrumento bien desarrollado y validado decrecerá su costo-eficiencia, pero incrementará la precisión de los hallazgos. - Este método no debe ser empleado como medida principal o única en la evaluación docente.

Fuente: Goe, Bell y Little, 2008.

CUADRO 7

Instrumentos y técnicas más empleados en la evaluación docente. Encuestas a los estudiantes

Descripción	Investigación	Fortalezas	Reservas
Se emplea para recoger las opiniones o juicios de los estudiantes acerca de la práctica docente como parte de la evaluación docente y para proveer de información acerca de la docencia, como es percibida por los estudiantes.	Varios estudios han mostrado que las evaluaciones de los docentes por los estudiantes pueden ser útiles en proporcionar información acerca de la docencia; pueden ser tan válidas como los juicios realizados por estudiantes universitarios y otros grupos y, en algunos casos, se pueden correlacionar con mediciones del desempeño escolar. La validez depende de los instrumentos empleados y su administración y se recomienda solamente para un uso formativo.	<ul style="list-style-type: none"> - Suministra la perspectiva del estudiante quien tiene la mayor experiencia con los docentes. - Puede suministrar información formativa para colaborar con los docentes a mejorar la práctica y de alguna manera los conectará con los estudiantes. - Los estudiantes son tan capaces como los adultos en suministrar evaluaciones precisas del desempeño docente. 	<ul style="list-style-type: none"> - Las evaluaciones de los docentes por los estudiantes no han sido validadas para su empleo en evaluaciones sumativas y no deberían emplearse como un instrumento único o principal en la evaluación docente. - Los estudiantes no pueden proporcionar información de aspectos de la docencia tales como: el conocimiento que el docente tiene del contenido, el cumplimiento del currículum y las actividades profesionales.

Fuente: Goe, Bell y Little, 2008.

CUADRO 8

Instrumentos y técnicas más empleados en la evaluación docente. Modelo de valor agregado

Descripción	Investigación	Fortalezas	Reservas
Es empleado para determinar la contribución de los docentes a las ganancias de los estudiantes en las pruebas estandarizadas.	Poco se sabe acerca de la validez del uso de los puntajes de valor agregado para identificar <i>la docencia eficaz</i> ; pero la investigación que emplea modelos de valor agregado sugiere que los docentes difieren notablemente en su contribución a las ganancias de los estudiantes en esas pruebas.	<ul style="list-style-type: none"> - Suministra una manera de evaluar la contribución de los docentes al aprendizaje de los estudiantes, lo que no hacen la mayoría de los otros instrumentos. - No requiere visitas al aula debido a que los datos del docente pueden ser analizados a distancia. 	<ul style="list-style-type: none"> - Los modelos no son capaces de diferenciar los efectos del docente de los efectos del aula. - Se asume que existe una alineación vertical en las pruebas (esto es que las pruebas miden lo mismo de grado a grado).

Cuadro 8 (continuación)

Descripción	Investigación	Fortalezas	Reservas
Puede también ser empleado como una herramienta de investigación (ejemplo, determinar la distribución de docentes “eficaces” de acuerdo con las características de los estudiantes o escuelas).	Sin embargo, al correlacionar el valor agregado con las cualificaciones de los docentes, características o prácticas ha producido resultados confusos y pocos hallazgos significativos. Entonces es obvio que los docentes varían en su eficacia pero que las razones de esto no son conocidas.	<ul style="list-style-type: none"> - Implica poca carga a nivel de aula o escuela porque la mayoría de datos ya han sido recolectados si se hacen evaluaciones estatales o nacionales. - Puede ser útil para identificar docentes sobresalientes cuyas aulas pueden servir como “laboratorios de aprendizaje”, así como para docentes en dificultades y con necesidades de apoyo. 	<ul style="list-style-type: none"> - Los puntajes de valor agregado no son útiles para propósitos formativos porque los docentes no aprenden nada acerca de cómo sus prácticas contribuyeron (o impidieron) al aprendizaje del estudiante. - Los modelos de valor agregado son controvertidos porque sólo miden las contribuciones de los docentes a las ganancias de los estudiantes en las pruebas estandarizadas.

Fuente: Goe, Bell y Little, 2008.

Como puede apreciarse en los cuadros presentados, todos los instrumentos tienen ventajas y limitaciones. Sin duda, la observación es el instrumento más empleado, junto con las muestras de trabajo docente y los modelos de valor agregado. Existe todavía un largo camino por recorrer para determinar la validez y confiabilidad de los instrumentos los cuales, en su mayoría, tienen un cariz formativo más que sumativo.

LAS NUEVAS TENDENCIAS EN EL DISEÑO E IMPLEMENTACIÓN DE ESTÁNDARES EN LA EVALUACIÓN DOCENTE

La evaluación docente y la rendición de cuentas, tanto de los docentes como de las escuelas, es una política promovida por la OCDE (OECD, 2013b) que se está extendiendo tanto a países miembros como no miembros. La creencia básica que guía esta política es que los docentes son uno de los factores más importantes para elevar la calidad de los resultados educativos. Esta sección tiene su base en los análisis previos realizados en este trabajo, en el enfoque integral de formación y evaluación docente que postula Darling-Hammond.

Se complementarán con ejemplos de diseño y evaluación basada en estándares de algunos países anglosajones, los cuales son pioneros en la evaluación basada en estándares.

Linda Darling-Hammond, de la Universidad de Stanford, es sin duda la autora más influyente, desde la década de los años noventa, que ha propugnado por un currículum basado en estándares y la evaluación docente basada en el mismo enfoque. Sus posturas sobre la formación y evaluación docente están traspasando las fronteras estadounidenses. La SEP publicó a principios de la década pasada uno de sus libros más conocidos, *El derecho de aprender* (Darling-Hammond, 2002), originalmente publicado en 1997. Chile ha puesto mucha atención en los estudios de esta autora; además de que fue invitada al país y los medios destacaron su visita, se le publicó un libro (Darling-Hammond, 2012a) y un artículo en una revista de ese país (Darling-Hammond, 2012b). El reconocimiento a su labor ya cruzó el Atlántico, la Universidad de Alicante le otorgó el doctorado Honoris Causa (Universidad de Alicante, 2012). Dada su trayectoria y sus ideas plasmadas en sus diversas publicaciones, es posible afirmar que Darling-Hammond representa las nuevas tendencias en relación con los estándares curriculares y la evaluación docente. En este apartado se destacarán las principales posturas y recomendaciones que fueron consultadas en diversos trabajos de la autora (Darling-Hammond, 2002; Darling-Hammond y Loewenberg-Ball, 1998; Darling-Hammond y Baratz-Snowden, 2007; Darling-Hammond y Rothman, 2011; Darling-Hammond, 2012c). Las políticas y recomendaciones que sugiere esta investigadora no sólo se basan en sus abundantes estudios de la realidad educativa estadounidense, sino también en el análisis comparativo que ha hecho entre los sistemas educativos más exitosos, como Singapur, Finlandia y Ontario (Darling-Hammond y Rothman, 2011).

De esos trabajos se seleccionaron las siguientes recomendaciones:

1. Tomar el sistema de formación, certificación, evaluación y formación continua en su conjunto.
2. Un sistema fuerte de reclutamiento, preparación y certificación de nuevos docentes.

3. Creación de un sistema de estándares curriculares.
4. Diseño e implementación de un conjunto coherente de estándares de evaluación docente.
5. Un sistema de desarrollo profesional continuo.

Los análisis realizados sobre estos aspectos son exploratorios; se resumirán las políticas y estrategias que sugiere Darling-Hammond y se ilustrarán con ejemplos derivados del diseño e implementación de estándares en otros países anglosajones, especialmente Australia.

Tomar el sistema de formación, certificación, evaluación y formación continua en su conjunto

Darling-Hammond y Rothman (2011) concluyen que las lecciones derivadas de los sistemas educativos con mejor desempeño en las pruebas internacionales, como Singapur, Finlandia y la provincia de Ontario, Canadá, indican que para mejorar la calidad de la educación se tienen que establecer políticas múltiples (no una sola) que aborden el sistema en su conjunto; esto es, el reclutamiento, preparación, inducción (al puesto), certificación, un programa de aprendizaje continuo y desarrollo profesional, y evaluación del desempeño. La adopción de políticas que aborden el sistema parcialmente tendrán un éxito limitado o poco significativo. Darling-Hammond y Loewenberg-Ball (1998) demuestran con datos del desempeño escolar de Estados Unidos en matemáticas, lectura y ciencias que los docentes que reciben preparación *ad hoc* previa, como es el caso de los dedicados a la lectura, tienen mejor desempeño que aquellos que fueron habilitados como docentes sin tener preparación previa, como ocurre en el campo de la enseñanza de las matemáticas. Para Darling-Hammond la preparación de los nuevos docentes es un aspecto determinante en el desempeño escolar y por ello debe asumirse con mucha seriedad.

Un sistema fuerte de reclutamiento, preparación y certificación de nuevos docentes

Los sistemas educativos a los que refieren los autores citados tienen un sistema interconectado de reclutamiento, formación, certificación evaluación y desarrollo profesional. Sin un sistema fuerte de atracción y preparación de docentes, así como su inducción en los primeros años a través de mentores/tutores, junto con un sistema sólido de certificación y las demás fases aludidas arriba, no es posible crear un sistema sólido de evaluación docente. Linda Darling-Hammond señala que la evaluación de los nuevos docentes, básicamente por medio de pruebas de opción múltiple, es insuficiente; *la verdadera evaluación debe realizarse en la práctica profesional, en el aula, donde se aprecia la integración de la teoría con la práctica*. La enseñanza es una profesión cuyo desempeño sólo se puede apreciar en su práctica, no sólo con pruebas de conocimiento. Darling-Hammond (2012c) advierte que los test de conocimiento no son buenos predictores del desempeño profesional docente.

Los nuevos docentes se evalúan conforme a los estándares de evaluación que sean apropiados, que estén basados en la investigación y el juicio de expertos y otros agentes educativos. Por ejemplo, el sistema de evaluación docente de la provincia de Ontario, Canadá, no evalúa a los docentes en todas las dimensiones e indicadores que conforman los estándares. Los nuevos docentes sólo son evaluados en 3 de las 5 dimensiones y en 8 de los 17 indicadores de competencia que conforman este sistema (Ontario Ministry of Education, 2010).

Creación de un sistema de estándares básicos curriculares

Los estándares de enseñanza o parámetros curriculares –como se les denomina en Brasil (Casimiro Lopes, 2008)– son un ingrediente necesario para determinar y diseñar qué tipo de estándares de evaluación docente se deben diseñar. Darling-Hammond y Loewenberg-Ball (1998) ilustran su postura citando la frase del gato de Cheshire de *Alicia en el País de las Maravillas*: “si no sabes a dónde vas cual-

quier camino te llevará allí”. Darling-Hammond (2012c) postula que, como ocurre actualmente en Estados Unidos, se deben definir estándares de enseñanza o curriculares. La autora recomienda que esos estándares sean un factor clave para establecer una clara concepción de los objetivos de aprendizaje y sobre las expectativas centrales del currículum. La autora afirma que los países con alto desempeño en las pruebas internacionales, como Finlandia, Japón, Corea y Singapur, han producido un conjunto de estándares, apoyos curriculares y evaluaciones que proveen un contexto para la evaluación y desarrollo profesional de los docentes.

Estos estándares, naturalmente, deben estar articulados con la evaluación de los docentes (Darling-Hammond y Loewenberg-Ball, 1998). Para Darling-Hammond (2002) los estándares curriculares, elevados y rigurosos, también garantizan una rendición de cuentas del profesorado; señala que se debe construir un consenso político en torno a las metas escolares para adoptar estándares de alta calidad educativa, que vaya acompañado con acciones que promuevan la equidad.

Diseño e implementación de un conjunto coherente de estándares de evaluación docente

La eficacia docente es una tarea esencial para elevar la calidad de la educación, como lo señala la investigación revisada líneas atrás. La concepción clara fundamentada tanto en la investigación internacional como nacional es un factor esencial para construir un sistema de evaluación del desempeño docente basado en estándares. Este aspecto es necesario, pero no suficiente, además se requiere, como lo señala Coggshall (2007), una concepción clara de los objetivos de la evaluación docente y un plan de comunicación para difundir e involucrar a la sociedad, así como a actores involucrados en el proceso.

Darling-Hammond (2012c: iii-iv) considera que un sistema de evaluación docente basado en estándares comprende los siguientes factores:

1. La evaluación docente debe estar basada en estándares profesionales de docencia y debe ser suficientemente sofisticada para

evaluar la docencia a través de un *continuum* que va desde los docentes nóveles hasta los expertos. Es decir, los estándares deben distinguir diferentes tipos de desempeño docente.

2. La evaluación debe incluir evidencias múltiples de la práctica del docente, aprendizaje de los estudiantes y contribuciones profesionales. No basta sólo evaluar el desempeño docente, es necesario también recolectar evidencias del desarrollo de los alumnos en aquellas facetas que se consideren importantes, como las pruebas nacionales y otras evidencias del progreso de los estudiantes.
3. Los evaluadores deben tener conocimiento de la docencia y estar adecuadamente capacitados en el sistema de evaluación.
4. La evaluación debe estar acompañada de una retroalimentación útil y conectada con oportunidades de desarrollo profesional como aprendizaje formal, colaboración entre pares, observación y tutoría.
5. El sistema de evaluación debe valorar y animar la colaboración de los docentes tanto en el diseño de estándares como en los criterios empleados para evaluar el trabajo docente.
6. Los docentes expertos deben ser parte en la revisión y asesoría del proceso de revisión de aquellos docentes que necesitan ayuda adicional para cumplir con los estándares de la evaluación.
7. Comités de funcionarios y docentes deben supervisar la evaluación.

EL DISEÑO DE UN SISTEMA DE EVALUACIÓN DOCENTE: EL EJEMPLO AUSTRALIANO

Las estrategias recomendadas por Darling-Hammond (2012c) y Coggshall (2007) guardan parecido con las que Australia ha implementado para diseñar su sistema de evaluación docente. El sistema de diseño e implementación de estándares australianos es un caso ejemplar y único. Antes, los territorios que conforman el país tenían su propio sistema de evaluación docente; ahora se está construyendo un sistema nacional de evaluación, bajo el liderazgo del Australian Institute for Teaching and School Leadership.

En primer lugar, Australia estableció cuáles son los objetivos principales que persigue en la educación: informa que el desempeño de su sistema educativo se compara adecuadamente con el de otras naciones (se encuentra entre los 20 mejores en los resultados de PISA), pero que aspira a ser uno de los mejores sistemas educativos del mundo. Las metas principales son dos: 1) la educación australiana promueve la equidad y la excelencia, y 2) todos los jóvenes estudiantes son aprendices exitosos, individuos seguros y creativos, y ciudadanos activos e informados. El sistema australiano reconoce que los docentes son un pieza clave para llevar a cabo esas metas (Australian Institute for Teaching and School Leadership, 2012); su papel justifica la creación de estándares que guíen el desarrollo profesional y el compromiso como medios para facilitar la mejora de la calidad docente. Ésta fue la manera como se diseñaron:

Los estándares y sus indicadores representan un análisis de la práctica contemporánea eficaz de los docentes de toda Australia. Su desarrollo incluye la síntesis de los indicadores de conocimiento, práctica y compromiso profesional de los docentes empleados por las autoridades, profesionales asociados, oficinas de acreditación y registro, y asociaciones profesionales. Se procuró comprender lo que los docentes requieren en diferentes etapas de su carrera profesional. Se hizo una amplia validación del proceso que involucró casi a 6000 docentes para asegurar que cada indicador estaba conformado por las características de la profesión. Los estándares apoyan la Declaración de Melbourne, la cual describe las aspiraciones de los jóvenes australianos para la próxima década (Australian Institute for Teaching and School Leadership, 2011: 1).

Como lo informa la cita, los estándares australianos están organizados en sólo tres dimensiones: conocimiento profesional, práctica profesional y compromiso profesional. Generalmente, como es el caso australiano, los sistemas educativos anglosajones emplean nombres breves para las dimensiones en que se agrupan los estándares. También vale la pena destacar que estas dimensiones están conformadas sólo por siete estándares que describen el desempeño de los docentes de acuerdo con su categoría: graduado, certificado, experimentado y guía o líder.

Además, ese instituto realizó una amplia revisión de la literatura internacional sobre el tema; una parte de esa bibliografía se cita al final del documento como evidencia de la validez de su marco conceptual. El marco conceptual que sirvió de base para diseñar los estándares parte de la creencia de que se requiere crear una cultura fuerte de desempeño y desarrollo docente y no se crea de la noche a la mañana, requiere tiempo hacerlo; el marco pone atención al proceso y no sólo a los resultados. Se señala que el marco de desempeño y desarrollo de los docentes australianos se caracteriza por (Australian Institute for Teaching and School Leadership, 2012):

1. Enfocarse a los resultados de los estudiantes. Este instituto advierte que la mejora de la calidad de la docencia no es un fin en sí mismo, que la docencia debe estar ligada a un efecto en los resultados de los estudiantes, que incluyen aprendizaje, compromiso con el aprendizaje y bienestar. Esto, desde luego, no sólo se refiere a medir el desempeño de los estudiantes exclusivamente a través de pruebas objetivas.
2. Una clara comprensión de la eficacia docente. Este marco indica que para enfocarse a la mejora de la docencia se debe tener una clara visión de lo que es una docencia eficaz.
3. Liderazgo. Este marco conceptual destaca el papel primordial que juega el liderazgo escolar en la mejora del desempeño escolar y señala la importancia de los estándares profesionales para directores, que han sido creados conjuntamente con los correspondientes a los docentes.
4. Flexibilidad. Se indica que las escuelas son diferentes y necesitan responder a contextos e historias singulares. Las escuelas varían ampliamente en sus enfoques de desempeño y desarrollo docente. Se advierte, por lo tanto, que una implementación eficaz toma en consideración el estado inicial de desempeño de esas escuelas y que la secuencia de cambio será distinta en diferentes contextos.
5. Coherencia. El documento indica que es importante que el desempeño docente vaya de la mano con el desarrollo profesional de los docentes, así que aquél deberá ser consistente con los planes de la escuela.

Los cinco presupuestos anteriores se llevan a cabo bajo un ciclo de desempeño y desarrollo continuo de acción conformado por las siguientes etapas: 1) reflexión y establecimiento de objetivos; 2) práctica profesional y aprendizaje y 3) retroalimentación y evaluación. Este ciclo es el que corresponde a la evaluación. Primero, el docente establece tanto con los evaluadores como con el director sus objetivos de desempeño; segundo, su práctica docente va acompañada de planes de desarrollo profesional y, por último, se da una retroalimentación al docente, y ambos (evaluador y docente) evalúan sus resultados (Australian Institute for Teaching and School Leadership, 2012).

Por otro lado, aunque el proceso de construcción e implementación de estándares nacionales inició en 2010 y éstos se dieron a conocer en 2011, en 2014 aún no se implementaban definitivamente. Debido a que uno de los grandes objetivos que tiene el proceso es crear una cultura de evaluación, Australia está difundiendo el proceso de evaluación mediante múltiples canales, incluyendo YouTube (bajo el nombre *Australian professional standards for teachers*). Además, en el sitio oficial existen videos que ilustran el desempeño de los docentes de acuerdo con los cuatro niveles: graduado, certificado, experimentado y guía (*lead*). El sitio ofrece otras herramientas para ayudar a los docentes a conocer y aplicar los estándares.

Asimismo, recientemente Australia inició un estudio de casos sobre la implementación exitosa de los estándares. Los convocados para que documenten las evidencias son las escuelas, individuos y consorcios educativos de ese país (University of Melbourne, 2014); en suma, la implementación de estándares estaba en proceso. Después de cuatro años de haberse iniciado parece que a Australia se aplica el refrán “despacio que voy de prisa”. Los australianos toman con mucha seriedad la creación de una cultura de evaluación y la evaluación de los docentes.

SUMARIO Y CONCLUSIONES

El propósito de este capítulo fue realizar un análisis de la evolución de la investigación sobre eficacia docente y su relación con los estándares

de evaluación. La globalización económica que comenzó a ocurrir durante la década de los ochenta ha traído consigo la competencia por los mercados entre los países del orbe. Puesto que existe la creencia de que hay una relación entre la robustez de la economía y la calidad educativa, la competencia económica se ha trasladado a la escuela. A partir de los resultados de PISA y estudios concurrentes, la evaluación de los docentes se está propagando entre los sistemas educativos de diferentes partes del mundo. El diseño de políticas de evaluación docente actualmente se lleva a cabo por medio de estándares de desempeño. Los estándares se pueden conceptualizar como los conocimientos y habilidades que un docente debe ser capaz de demostrar.

Otro factor que ha impulsado la evaluación docente basada en estándares es el cúmulo y avances de la investigación sobre la eficacia docente. Los estándares docentes descansan en los hallazgos de la investigación sobre eficacia docente. La investigación sobre este tema ya se ha consolidado durante más de un siglo y ha atravesado las etapas que señala Walberg (1986): identificación de resultados valiosos, identificación de factores plausibles, observación de factores y efectos, replicación de covariaciones, síntesis de la investigación y adopción y evaluación.

Durante la última década la investigación retornó, de alguna manera, a sus inicios, su objeto de estudio cambió de la eficacia de la escuela a la eficacia del docente. La investigación sobre eficacia docente parece haber alcanzado consenso sobre lo que constituye un docente eficaz. La eficacia está relacionada con tres grandes dimensiones de la actuación del docente: el conocimiento del currículum, conocimiento, interacción y evaluación de los estudiantes, y la eficacia de las estrategias docentes empleadas en el aula. A partir de estas dimensiones varios sistemas educativos han construido los estándares contemporáneos de evaluación docente a los que se ha agregado el desarrollo y compromiso profesional; es decir, una preocupación activa de los docentes por desarrollarse profesionalmente y la colaboración del docente con colegas y otros grupos o instancias de la comunidad escolar.

Los estándares de evaluación docente comenzaron en Estados Unidos; su inicio no tiene una fecha marcada. Se considera que des-

de mediados del siglo XIX ese país mostró preocupación por definir las normas y expectativas de la formación de docentes en las escuelas normales. El preámbulo formal y más reciente que dio pie a la creación de estándares docentes fue el informe “Nación en riesgo”. El informe preparado por una comisión (The National Commission on Excellence in Education) hizo notar las debilidades del sistema educativo y recomendó, por primera vez, estándares más exigentes para el currículum. En Estados Unidos los docentes son evaluados tomando en consideración el desempeño académico de sus alumnos, calculado por medio de las pruebas estandarizadas; los resultados se emplean como medida de rendición de cuentas por parte de los docentes y sus escuelas. A pesar del empleo de los resultados de los estudiantes en los test en la evaluación docente, no hay evidencias suficientes de que tenga una adecuada validez.

Coggshall (2007) propone establecer un marco de comunicación para medir la calidad y eficacia docente que se puede aplicar al diseño e implementación de estándares de evaluación. Este marco está integrado por cuatro componentes: planeación de la comunicación, clarificación de objetivos, términos que definen la calidad del docente e instrumentos de medición y recursos. Los instrumentos más empleados en la evaluación docente por medio de estándares son: la observación, observación por el director, muestras de trabajo, portafolios, autoevaluación del docente, encuestas a padres y alumnos, y valor agregado. El instrumento más empleado es la observación, seguido de las medidas de valor agregado y muestras de trabajo. Todos los instrumentos tienen ventajas y limitaciones en la valoración del desempeño docente.

Para el diseño de un sistema de evaluación docente basado en estándares Darling-Hammond recomienda que el diseño de estándares de evaluación se haga dentro de un marco más amplio que considere los siguientes componentes: 1) tomar el sistema de formación, certificación, evaluación y formación continua en su conjunto; 2) un sistema fuerte de reclutamiento, preparación y certificación de nuevos docentes; 3) creación de un sistema de estándares curriculares; 4) diseño e implementación de un conjunto coherente de estándares de evaluación docente, y 5) un sistema de desarrollo profesional

continuo. La estrategia de diseño e implementación de estándares de evaluación docente en Australia tiene parecido con las recomendaciones que Darling-Hammond hace al respecto.

REFERENCIAS

- Australian Institute for Teaching and School Leadership (2011), *Australian professional standards for teachers*, Carlton South, Australia, <<http://www.aitsl.edu.au/professional-growth/australian-teacher-performance-and-development-framework>>, consultado el 10 de abril de 2014.
- Australian Institute for Teaching and School Leadership (2012), *Australian teacher performance and development framework*, Carlton South, Australia, <<http://www.aitsl.edu.au/professional-growth/australian-teacher-performance-and-development-framework>>, consultado el 10 de abril de 2014.
- Barr, A. S. (1948), “The Measurement and prediction of teaching efficiency: A summary of investigations”, *The journal of experimental education*, vol. 16, núm.4, pp. 203-283.
- Brophy, J. y T. Good (1986), “Teacher behavior and student achievement”, en M. C. Wittrock (ed.), *Handbook of research on teaching*, 3a. ed., Nueva York, MacMillan, pp. 328-375.
- Bruns, B. y J. Luque (2014), *Profesores excelentes. Cómo mejorar el aprendizaje en América Latina y el Caribe* (resumen), Washington, Banco Mundial, <<http://openknowledge.worldbank.org/bitstream/handle/10986/20488/Spanish-excellent-teachers-report.pdf?sequence=5>>, consultado el 15 de diciembre de 2015.
- Casimiro Lopes, A. (2008), “Articulaciones en las políticas de currículo”, *Revista Perfiles Educativos*, vol. 30, núm. 120, pp. 63-78.
- Cogshall, J. (2007), *Communication framework for measuring teacher quality and effectiveness: Bringing coherence to the conversation*, Washington, National Comprehensive Center for Teacher Quality, <<https://eric.ed.gov/?id=ED543771>>, consultado el 1 de agosto de 2014.
- Coleman, J. (1990), *Equality and achievement in education*, Boulder, Westview Press.
- Collins, C. (2014), “Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system (EVAAS®)”,

- Education policy analysis archives*, vol. 22, núm. 98, <<http://dx.doi.org/10.14507/epaa.v22.1594>>.
- Commission on Teacher Credentialing, State of California (1992), *Standards of program quality and effectiveness for professional teacher preparation programs for multiple a single subject teaching credential with a (bilingual) crosscultural, language and academic development (CLAD/ BICLAD) emphasis*, Sacramento.
- Darling-Hammond, L. (2002), *El derecho de aprender. Crear buenas escuelas para todos*, México, SEP.
- Darling-Hammond, L. (2007), “A good teacher in every classroom: Preparing the highly qualified teachers our children deserve”, *Educational Horizons*, vol. 65, núm. 2, pp. 111-132, <<http://eric.ed.gov/?id=EJ750647>>.
- Darling-Hammond, L. (2012a), *Educación con calidad y equidad. Los dilemas del siglo XXI*, Santiago de Chile, Centro de Innovación en Educación de Fundación Chile, <http://ww2.educarchile.cl/UserFiles/P0001/File/CR_Articulos/Home/Educacionconcalidadyequidad_LDarlingH.pdf>, consultado el 20 de mayo de 2014.
- Darling-Hammond, L. (2012b), “Desarrollo de un enfoque sistémico para evaluar la docencia y fomentar una enseñanza eficaz”, *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, vol. 49, núm. 2, pp.1-20.
- Darling-Hammond, L. (2012c), *Creating a comprehensive system for evaluating and supporting effective teaching*, Stanford University-Center for Opportunity Policy in Education, <<https://edpolicy.stanford.edu/sites/default/files/publications/creating-comprehensive-system-evaluating-and-supporting-effective-teaching.pdf>>, consultado el 20 de mayo de 2014.
- Darling-Hammond, L. y J. Baratz-Snowden (2007), “A good teacher in every classroom: Preparing the highly qualified teachers our children deserve”, *Educational Horizons*, vol. 85, núm. 2, pp. 111-132.
- Darling-Hammond, L. y D. Loewenberg-Ball (1998), *Teaching for high standards: What policy makers need to know and be able to do*, Filadelfia, The Consortium for Policy Research in Education, <<http://eric.ed.gov/?id=ED426491>>.
- Darling-Hammond, L. y R. Rothman (2011), *Teacher and leader effectiveness in high-performing education systems*, Stanford University-Center for Opportunity Policy in Education, <<http://edpolicy.stanford.edu/>>

- sites/default/files/publications/teacher-and-leader-effectiveness-high-performing-education-systems.pdf>, consultado el 30 de mayo de 2014.
- Diario Oficial de la Federación* (2010), “Acuerdo Número 540 por el que se actualiza el diverso número 181 por el que se establecen el plan y los programas de estudio para la educación primaria”, 20 de agosto, <http://dof.gob.mx/nota_detalle_popup.php?codigo=5156090>, consultado el 16 de abril de 2014.
- Dillon, S. (2007), “Long reviled merit pay gains among teachers”, *The New York Times*, 21 de junio, <http://www.nytimes.com/2007/06/18/education/18pay.html?pagewanted=all&_r=0>, consultado el 30 de septiembre de 2015.
- Dunkin, M. J. y B. J. Biddle (1974), *The study of teaching*, Nueva York, University Press of America.
- Edelfelt, R. A. y J. D. Rath (1998), *A Brief history of standards in teacher education*, <<https://eric.ed.gov/?id=ED461627>>.
- Edmonds, R. (1979), “Effective schools for the urban poor”, *Educational Leadership*, vol. 37, núm. 1, pp. 15-24.
- Elam, S. (1971), *Performance based teacher education. What is the state of the art?*, American Association of Colleges for Teacher Education, Washington, <<https://eric.ed.gov/?id=ED058166>>.
- Eyre, H. (2007), “Keller’s personalized system of instruction: Was it a fleeting fancy or is there a revival on the horizon?”, *The Behavior Analyst Today*, vol. 8, núm. 3, pp. 317-323, <<http://eric.ed.gov/?id=EJ800986>>.
- Figazolli, L. (2013), *The use and misuse of teacher appraisal. An overview of cases in the developed world*, Bruselas, Education International Research Institute, <<https://download.ei-ie.org/Docs/WebDepot/TeacherAppraisal.pdf>>, consultado el 15 de mayo de 2016.
- Furham, S. H. (2003), *Redesigning accountability systems for education. Reporting on Issues and Research in Education Policy and Finance*, University of Pennsylvania-Graduate School of Education, <<http://files.eric.ed.gov/fulltext/ED498338.pdf>>.
- Glass, G. V. (1976), “Primary, secondary, and meta-analysis of research”, *Educational Researcher*, vol. 5, núm. 10, pp. 3-8.
- Goe, L., C. Bell y O. Little (2008), *Approaches to evaluating teacher effectiveness: A research synthesis*, <<http://files.eric.ed.gov/fulltext/ED521228.pdf>>.

- Hanushek, E., P. Peterson y L. Woessmann (2012), *Achievement growth: International and U.S. State trends in Student performance*, Harvard's Program on Education Policy and Governance/Education, Next Taubman Center for State and Local Government, Harvard Kennedy School, PEPG Report 12-03, <https://www.hks.harvard.edu/pepg/PDF/Papers/PEPG12-03_CatchingUp.pdf>, consultado el 19 de mayo de 2016.
- Hopkins, D., A. Harris, L. Stoll y T. Mackay (2011), *School and system improvement: State of the art review*, International Congress of School Effectiveness and School Improvement, Limassol, Chipre, enero, <http://www.icsei.net/2016/icsei2011/State_of_the_art/State_of_the_art_Session_C.pdf>, consultado el 19 de mayo de 2016.
- Interstate New Teacher Assessment and Support Consortium (1992), *Model standards for beginning teacher licensing, assessment and development: A resource for state dialogue*, Washington, Council of Chief State School Officers, <<http://programs.ccsso.org/content/pdfs/corestrd.pdf>>, consultado el 10 de enero de 2013.
- Interstate New Teacher Assessment and Support Consortium (2010), *Model core teaching standards: A resource for state dialogue* (Draft for public comment), Washington, Council of Chief State School Officers, <http://www.ccsso.org/Documents/2011/InTASC_Model_Core_Teaching_Standards_2011.pdf>, consultado el 10 de marzo de 2016.
- Interstate School Leaders Licensure Consortium (1996), *Standards for school leaders*, Washington, Council of Chief State School Officers, <http://soe.unc.edu/academics/requirements/standards/ISLLC_Standards.pdf>, consultado el 5 de mayo de 2014.
- Isoré, M. (2010), *Evaluación docente: prácticas vigentes en los países de la OCDE y una revisión de la literatura*, Santiago, PREAL, <<https://www.oas.org/cotep/GetAttach.aspx?lang=en&cId=13&aid=25>>, consultado el 1 de abril de 2014.
- Klette, K. (2007), "Trends in research on teaching and learning in schools: didactics meets classroom studies", *European Educational Research Journal*, vol. 6, núm. 2, <doi:10.2304/eeerj.2007.6.2.147>.
- Ko, J. y P. Sammons, (2013), *Effective teaching: a review of research and evidence*, Reading, CFBT Education Trust, <<http://eric.ed.gov/?id=ED546794>>.
- Lee, M. (2007), *History of American educational standards*, <www.csun.edu/~ml727939/coursework/610/HistoryofEducation.doc>, consultado el 5 de octubre de 2014.

- Lortie, D. C. (1975), *School teacher. A sociological study*, Chicago, The University of Chicago Press.
- Malone, B. G. y J. S. Nelson (2006), “Standards based reform: Panacea for the twenty-first century”, *Educational Horizons*, vol. 84, núm. 2, pp. 121-128.
- Medley, D. M. (1972), “Early history of research on teacher behavior”, *International Review of Education*, vol. 18, núm. 1, pp. 430-439.
- Medley, D. M. (1977), *Teacher competence and teacher effectiveness. A review of process-product research*, <<https://eric.ed.gov/?id=ED143629>>.
- Mosadeghrad, A. (2014), “Why TQM programmes fail? A pathology approach”, *The TQM Journal*, vol. 26, núm. 2, pp. 160-187, <https://www.researchgate.net/publication/263364760_Why_TQM_programs_fail_A_pathology_approach>, consultado el 1 de abril de 2016.
- Murillo, F. J. (2004), “Un marco comprensivo de la mejora de la eficacia escolar”, *Revista Mexicana de Investigación Educativa*, vol. 9, núm. 21, pp. 319-359.
- National Council of Teachers of Mathematics (2014), *Principles to Actions: Executive Summary*, <https://www.nctm.org/uploadedFiles/Standards_and_Positions/PtAExecutiveSummary.pdf>, consultado el 10 de marzo de 2016.
- Organization for Economic Co-operation and Development (OECD) (2013a), *Lessons from PISA 2012 for the United States, strong performers and successful reformers in education*, París, OECD, <<https://www.oecd.org/pisa/46623978.pdf>>, consultado el 1 de abril de 2014.
- OECD (2013b), *Teachers for the 21st century. Using evaluation to improve teaching*, París, OECD, <<http://www.oecd.org/site/eduistp13/TS2013%20Background%20Report.pdf>>, consultado el 1 de abril de 2014.
- Ontario Ministry of Education (2010), *Teacher performance appraisal. Technical requirements*, <http://www.edu.gov.on.ca/eng/teacher/pdfs/TPA_Manual_English_september2010l.pdf>, consultado el 22 de junio de 2015.
- Pinar, W. F. (2004), *What is curriculum theory?*, Nueva York, Routledge.
- Scheerens, J. (2000), *Improving school effectiveness*, París, UNESCO, <<http://doc.utwente.nl/92592/>>, consultado el 22 de junio de 2015.
- Secretaría de Educación Pública (SEP) (2011), *Plan de estudios 2011, Educación básica*, México, SEP.
- Seyfarth, J. (2008), *Human resource leadership for effective schools*, 5^a ed., Boston, Pearson.

- Shakman, K., J. Riordan, M. T. Sánchez, M. T. DeMeo, K. Cook, R. Fournier y J. Brett (2012), *An examination of performance based teacher evaluation systems in five states*, (Issues & Answers Report, REL 2012, núm. 129), Washington, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast and Islands, <<http://files.eric.ed.gov/fulltext/ED529800.pdf>>, consultado el 9 de agosto de 2015.
- Shepard, L. A. (2012), “Evaluating the use of tests to measure teacher effectiveness: Validity as a theory-of-action framework”, Trabajo presentado en la reunión anual del National Council on Measurement in Education, Vancouver, abril, <<https://www.cde.state.co.us/educatoreffectiveness/scencmereport>>, consultado el 9 de agosto de 2015.
- 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*, 3a. ed., Nueva York, MacMillan, pp. 6-36.
- The National Commission on Excellence in Education (1983), *A Nation at risk: The imperative for educational reform*, Washington, The National Commission on Excellence in Education.
- Universidad de Alicante (2012), *Doctora Honoris Causa: Linda Darling-Hammond*, Universidad de Alicante, <<http://web.ua.es/es/protocolo/eventos/honoris/darling-hammond-linda-2012/doctora-honoris-causa-linda-darling-hammond.html>>, consultado el 25 de enero de 2015.
- University of Melbourne (2014), *2014 Case studies. Lesson from practice*, Australian Institute for Teaching and School Leadership/University of Melbourne, <<http://www.aitsl.edu.au/research-and-evaluation/evaluation-of-the-standards/activities/case-studies>>, consultado el 4 de febrero de 2016.
- University of Rochester (2006), *Five documents in the Math wars, 1989-2006—Mathematics*, <http://www.math.rochester.edu/people/faculty/rarm/debate_appendix.html>, consultado el 4 de marzo de 2015.
- Walberg, H. J. (1986), “Synthesis of research on teaching”, en M. C. Wittrock (ed.), *Handbook of research on teaching*, 3a. ed., Nueva York, MacMillan, pp. 214-229.
- Weber, G. (1971), *Inner city can be taught to read: Four successful schools*, (CBE Occasional Papers, 18), <<https://eric.ed.gov/?id=ED057125>>.
- Wittrock, M. C. (ed.) (1986), *Handbook of Research on Teaching*, 3a. ed., Nueva York, MacMillan.