



INFORMES TÉCNICOS MIDE UC / TECHNICAL REPORTS MIDE UC

Centro de Medición MIDE UC / Measurement Center MIDE UC

IT1403

Modelos de Evaluación Docente en el mundo e instrumentos utilizados para evaluar:
Ventajas, desventajas y consideraciones para su elección.

MARÍA ROSA GARCÍA

23 de abril de 2014



PONTIFICIA UNIVERSIDAD
CATÓLICA DE CHILE



ÍNDICE

1. EVALUACIÓN DOCENTE EN EL MUNDO.....	3
1.1 Nivel de centralización de la evaluación	4
1.2 Propósito de la evaluación	5
1.3 Instrumentos de evaluación	6
1.4 Marcos regulatorios o aproximaciones políticas.....	6
1.5 Estándares de referencia y aspectos evaluados	7
1.6 Frecuencia y obligatoriedad de la evaluación	8
1.7 Naturaleza de la evaluación y evaluadores	8
1.8 Uso de los resultados	9
2. CONSIDERACIONES GENERALES SOBRE LA ELECCIÓN DE INSTRUMENTOS DE EVALUACIÓN	11
3. REVISIÓN DE LOS PRINCIPALES INSTRUMENTOS USADOS EN EVALUACIONES DOCENTE Y SUS VENTAJAS Y DESVENTAJAS.....	12
3.1 Observación de clases	12
2.1.1 Descripción del instrumento.....	12
2.1.2 Uso del instrumento para fines de permanencia y ascenso.....	13
2.1.3 Ventajas y desventajas del instrumento.....	13
3.2 Entrevistas del directivo y profesor(a).....	15
2.2.1 Descripción del instrumento.....	15
2.2.2 Uso del instrumento para fines de permanencia y ascenso.....	15
2.2.3 Ventajas y desventajas del instrumento.....	15
3.3 Auto-evaluación del docente	17
2.3.1 Descripción del instrumento.....	17
2.3.2 Uso del instrumento para fines de permanencia y ascenso.....	18
2.3.3 Ventajas y desventajas del instrumento.....	18

3.4	Portafolio.....	19
2.4.1	Descripción del instrumento.....	19
2.4.2	Uso del instrumento para fines de permanencia y ascenso.....	21
2.4.3	Ventajas y desventajas del instrumento.....	21
3.5	Resultados de los estudiantes	23
2.5.1	Descripción del instrumento.....	23
2.5.2	Uso del instrumento para fines de permanencia y ascenso.....	24
2.5.3	Ventajas y desventajas del instrumento.....	25
3.6	Encuestas a estudiantes y padres	26
2.6.1	Descripción del instrumento.....	26
2.6.2	Uso del instrumento para fines de permanencia y ascenso.....	27
2.6.3	Ventajas y desventajas del instrumento.....	28
3.7	Pruebas para evaluar competencias.....	29
2.7.1	Descripción del instrumento.....	29
2.7.2	Uso del instrumento para fines de permanencia y ascenso.....	30
2.7.3	Ventajas y desventajas del instrumento.....	30
4.	CONCLUSIONES Y RECOMENDACIONES	32
5.	REFERENCIAS	36
6.	ANEXOS.....	40
6.1	Anexo N°1: Tipos de aproximaciones de evaluación docente en distintos países ..	40
6.2	Anexo N°2: Síntesis de las ventajas y desventajas de los distintos instrumentos de evaluación	41

1. RESUMEN

La evaluación de profesores cobra especial relevancia en los sistemas educativos que buscan mejorar los aprendizajes de los estudiantes, aun cuando difieren en múltiples aspectos relacionados con los propósitos, los marcos de evaluación, los estándares que utilizan, los instrumentos que escogen para evaluar y el uso de resultados y consecuencias que asignan. En particular, elegir los instrumentos de evaluación más adecuados para un sistema no es tarea fácil, dado que existen distintas ventajas y desventajas asociadas a cada uno, y cada vez más, los sistemas se orientan al uso de instrumentos diversos para evaluar el desempeño docente de forma más precisa, y también, aportar información que sirva a distintos propósitos. En este reporte, se discuten a nivel general los aspectos en los que difieren los distintos modelos de evaluación docente a nivel mundial, y luego, se presentan las características de los distintos instrumentos utilizados para evaluar el desempeño docente en sistemas de evaluación nacionales e internacionales, y se discuten las principales ventajas y desventajas de cada uno, y algunos criterios que pueden orientar su elección.

Autora: María Rosa García¹

Palabra(s) clave(s): Evaluación Docente, Instrumentos de evaluación, Carrera Docente.

¹ María Rosa García es Coordinadora de MIDE Internacional del Centro de Medición MIDE UC, Pontificia Universidad Católica de Chile. La correspondencia puede ser enviada a Vicuña Mackenna 4860, Edificio Centro de Medición MIDE UC, Macul, Santiago, Chile, o al correo mrgarci1@uc.cl.

2. EVALUACIÓN DOCENTE EN EL MUNDO

La evaluación de profesores ha sido un tema muy relevante a nivel mundial en la última década, y ha sido objeto de estudios de inédita magnitud a nivel internacional. Sin embargo, no ha estado exenta de controversias, en cuanto por una parte, se busca evaluar aquello que define el impacto en los aprendizajes de los estudiantes, y por otra, la presencia de diversos conflictos de interés entre los distintos actores en educación. Como consecuencia, existe el riesgo de que la evaluación termine siendo un ejercicio sin sentido, para cumplir con los requerimientos burocráticos que se solicitan (Danielson, 2001, Holland, 2005, Marshall, 2005 citados en OECD, 2009). Sin embargo, recientemente algunos países han comenzado a instalar los sistemas de evaluación como parte integral de las políticas escolares (Peterson, 2006; TDA 2007^a citados en OECD, 2009).

Los sistemas de evaluación docente en los distintos países han tomado múltiples formas. Hay diversos aspectos en los cuales difieren: el nivel de centralización o descentralización que tiene la evaluación, los propósitos que persigue, los instrumentos utilizados para evaluar, los marcos o aproximaciones políticas en los que se insertan evaluaciones docentes, los estándares en los que se basan, la frecuencia y obligatoriedad de la evaluación, los evaluadores que participan, y el uso de resultados y consecuencias que tienen (OECD, 2009). A continuación revisaremos a nivel general estos aspectos.

2.1 Nivel de centralización de la evaluación

Primeramente, si la evaluación se implementa de forma centralizada o descentralizada depende en gran medida del contexto gubernamental existente en cada país, y el nivel de autonomía que los colegios tienen. En aquellos países que no cuentan con políticas y marcos nacionales que guíen la evaluación docente, las autoridades locales y los colegios son quienes han desarrollado sistemas de evaluación docente con muy poco involucramiento del nivel central. En cambio, en el contexto de mayor preocupación pública por la calidad y equidad de los aprendizajes que alcanzan los estudiantes, se han desarrollado reformas para mejorar la calidad de la enseñanza, con un foco en la evaluación de los docentes de forma centralizada. Sin embargo, instalar sistemas centralizados de evaluación docente en sistemas educativos

con una larga tradición de autonomía puede ser complejo. En estos casos, las autoridades educativas deben considerar distintas opciones para balancear del mejor modo la guía centralizada con la autonomía local. Mientras los líderes locales normalmente tienen mucho mayor conocimiento de las necesidades específicas del colegio, involucrar al nivel central permite asegurar estándares y criterios relacionados con las metas educativas propuestas a nivel nacional (OECD, 2013).

2.2 Propósito de la evaluación

Otro aspecto en el que varían los sistemas se refiere a los propósitos de las evaluaciones docentes. Existen dos propósitos principales, uno *formativo* y otro *sumativo*. El propósito formativo, busca la mejora y el desarrollo profesional docente. Está orientado a identificar las fortalezas y debilidades docentes de forma que los directivos puedan tomar decisiones más informadas sobre qué actividades de desarrollo profesional resultan más idóneas para las necesidades docentes y las prioridades del contexto educativo en que se inserta. Es usualmente conducido por los colegios en forma individual y no siempre está regulado para implementarse en forma nacional. (OECD, 2013).

El propósito *sumativo*, se focaliza en responsabilizar a los profesores por su desempeño y conlleva en muchos casos, consecuencias para la carrera de los docentes como definir su progresión en esta, bonos salariales, o la posibilidad de tener sanciones asociadas al desempeño bajo. La evaluación con propósito sumativo, busca entregar información sintética sobre el desempeño de los docentes en relación a lo que se considera una “buena” enseñanza, y establecer incentivos para que los profesores se desempeñen en el mejor estándar. (OECD, 2013).

Ahora bien, la mayoría de los sistemas de evaluación docente utilizan los resultados de la evaluación para propósitos tanto formativos como sumativos. Sin embargo, combinar ambos focos no está exento de complejidades. Cuando la evaluación se orienta hacia la mejora de las prácticas, los profesores se muestran abiertos a revelar sus debilidades en la expectativa de que ello llevará a que se tomen decisiones más efectivas sobre sus necesidades de desarrollo y capacitación. Sin embargo, cuando los profesores son confrontados con las posibles consecuencias que la evaluación tendrá en su carrera o

remuneración, la inclinación a revelar sus debilidades se reduce, poniendo en peligro el foco formativo (OECD, 2013).

2.3 Instrumentos de evaluación

La revisión hecha por la OECD (2013) sobre los sistemas de evaluación que implementan diversos países mostró que los principales instrumentos que son utilizados para evaluar son la observación de clases, el portafolio, la auto-evaluación del profesor, la entrevista del directivo con el docente, los resultados de los estudiantes, las encuestas realizadas a estudiantes y padres, y las pruebas estandarizadas. Ahora bien, el propósito que se defina en el modelo de evaluación, incide en el tipo de instrumentos que resulta recomendable usar. Cuando el foco es formativo, se requiere de instrumentos que entreguen evidencia detallada acerca de la práctica docente que permita retroalimentar hacia la mejora, con estándares compartidos y conocidos que orienten hacia las buenas prácticas. En cambio, cuando el foco es sumativo, se releva en mayor medida la confiabilidad del instrumento y su capacidad de recoger y puntuar evidencia que se vea afectada lo menos posible por sus evaluadores, de forma que los resultados no resulten sesgados por las altas consecuencias que acarrearán.

2.4 Marcos regulatorios o aproximaciones políticas

Sumado a lo anterior, otro aspecto relevante refiere a las aproximaciones o marcos políticos o regulatorios en que se insertan las evaluaciones en relación a la carrera y profesión docente, los cuales son básicamente tres²:

El primero corresponde a *evaluaciones iniciales* de profesores principiantes, que tienen como fin su acreditación para el ejercicio de la docencia al cabo de su licenciatura o de un período de prueba.

Otro marco corresponde a las evaluaciones orientadas a *gestionar el desempeño*, aplicadas a profesores en ejercicio, con el fin de gestionar su carrera y el desarrollo

² En el anexo 1 se entrega un cuadro resumen que detalla el tipo de aproximación(es) o marco regulatorio que utiliza cada país según los datos de la OECD (2013).

profesional del docente. Aquí se incluyen evaluaciones que registran o certifican a los docentes en ejercicio, como también las evaluaciones regulares que definen la permanencia ejerciendo, y también las evaluaciones para definir la promoción o ascenso de los docentes.

Finalmente, existen las evaluaciones insertas en *esquemas de recompensa*, que tienen el objetivo explícito de identificar un número selecto de docentes de alto nivel de desempeño para reconocer sus competencias profesionales a través de premios, bonos o incrementos salariales.

2.5 Estándares de referencia y aspectos evaluados

Los sistemas de evaluación también requieren de estándares de referencia en los cuales basar la evaluación. Estos son requisito indispensable para desarrollar evaluaciones más sofisticadas que capten la complejidad de la tarea docente. Sin ellos, la evaluación trivializa la profesión docente (Danielson, 2011; Shulman, 2001). Los estándares proveen un lenguaje común para profesores y directivos sobre qué caracteriza una buena práctica docente (Sartain, Steoelinga, & Brown, 2011 citados en Coggshall, Rasmussen, Colton, Milton & Jacques, 2012).

Usualmente los estándares pueden contener: i) estándares de enseñanza con perfiles de competencias generales para todos los docentes y estándares de competencias específicas diferenciando por nivel o asignatura; ii) un conjunto de funciones incluidas descripciones de cargo para los profesores, y iii) un plan de desarrollo local del establecimiento. Ahora bien, el tipo de estándares de referencia que usan los distintos países debe ser consistente con el marco político en que se inserta la evaluación docente (OECD, 2013).

Varios sistemas de evaluación han sido influenciados por el Marco de Enseñanza elaborado por Danielson (2007 citado en OECD, 2013), el cual define las responsabilidades docentes agrupadas en cuatro áreas: planificación y preparación, instrucción, ambiente de la clase y responsabilidades profesionales. En algunos países los aspectos evaluados incluyen elementos más específicos, muchos de los cuales, aunque no todos, son cubiertos por estas cuatro categorías. Algunos países agregan a la evaluación la medición de los conocimientos disciplinarios y pedagógicos en distintas áreas. En algunos países no existen regulaciones

centrales asociadas a los aspectos que se deben evaluar, sino que en estos casos, son los evaluadores quienes determinan lo que se evaluará (OECD, 2013).

2.6 Frecuencia y obligatoriedad de la evaluación

Las evaluaciones pueden ser voluntarias u obligatorias. La frecuencia en que la evaluación se realiza también varía. Se puede realizar periódicamente en intervalos regulares, en etapas claves de la carrera, en ocasiones específicas como al finalizar el período de prueba o al renovar el contrato de trabajo. En general, las evaluaciones que se realizan al finalizar el período de prueba son obligatorias, al igual que las que se realizan en forma regular para gestionar el desempeño. Estas últimas suelen aplicarse regularmente cada un corto período de años. En cambio, las evaluaciones basadas en esquemas de recompensa suelen ser voluntarias. La frecuencia de la evaluación y su carácter obligatorio o voluntario debe ser consistente con el propósito que persigue la evaluación (OECD 2009, 2013).

2.7 Naturaleza de la evaluación y evaluadores

Otro de los aspectos en los que los sistemas difieren tiene relación con quién evalúa o quién tiene la responsabilidad por la evaluación docente, la cual puede ser externa o interna. Esta determinación está fuertemente influida por el propósito que persigue la evaluación, dado que no es trivial quien es el evaluador. Cuando es externa, los aspectos evaluados, los instrumentos utilizados y los criterios de evaluación son comunes entre los distintos colegios, y los evaluadores suelen ser externos al colegio en que trabaja el docente evaluado, lo cual ocurre principalmente cuando el propósito es sumativo. En cambio, en evaluaciones conducidas de forma interna, es el colegio el que diseña los instrumentos y define los criterios de evaluación, y se encarga de determinar los resultados obtenidos, usualmente con evaluadores que pertenecen al mismo colegio del evaluado. Aquí se incluyen los directivos, profesores pares y el propio docente. Así, las evaluaciones internas pueden ser más apropiadas cuando el propósito es formativo (OECD, 2009, 2013).

Las ventajas de las evaluaciones internas es que los procedimientos de evaluación pueden adaptarse mejor al contexto local del colegio y los evaluadores están familiarizados

con el contexto de trabajo del docente. En este sentido, usualmente logran el apoyo de los docentes y directivos (OECD, 2013).

En cambio, las evaluaciones externas resultan más pertinentes cuando se busca la responsabilización, y por tanto, se quiere alcanzar estándares comunes consistentes a lo largo de los distintos colegios. Asimismo, son especialmente útiles para colegios que no tienen una tradición y experiencia desarrollando evaluaciones docentes (OECD, 2013).

Ahora bien, al igual como sucede con los propósitos de las evaluaciones, hay países que utilizan evaluadores tanto internos como externos en la evaluación, conjugando sus beneficios y características, como es el caso de Australia, Austria, Chile, Francia, Israel, Luxemburgo, Portugal y Suiza (OECD, 2013). El uso de evaluadores múltiples es una práctica exitosa, en cuanto mejora la calidad de la evaluación cuando más de una persona participa en juzgar el desempeño docente (Peterson, 2000, Stronge and Tucker, 2003 citados en OECD, 2009). Esto se vuelve más importante en evaluaciones sumativas que asocian altas consecuencias a los resultados (Danielson & McGreal, 2000 citados en OECD, 2009).

2.8 Uso de los resultados

Otro aspecto crucial que define a los sistemas de evaluación tiene relación con el uso que se espera de sus resultados, lo cual determina la efectividad de la evaluación. Esto implica retroalimentar a quienes desarrollan servicios educativos, a los profesores y directivos; diseñar actividades de desarrollo profesional para mejorar la práctica docente; establecer premios y consecuencias que se desprenden de los resultados de la evaluación; y desarrollar canales a través de los cuales la información de la evaluación puede usarse para el desarrollo de políticas educativas (OECD, 2013).

Un uso formativo de los resultados implica que se establezcan vínculos claros entre la información de la evaluación y las oportunidades de desarrollo profesional docente. Un uso sumativo de los resultados puede implicar tomar decisiones asociadas a la carrera docente que indirectamente se asocian con consecuencias salariales, como por ejemplo, renovar un contrato, aprobar el período de prueba o apuntar a un profesor para un determinado cargo al interior del colegio. Asimismo, otro uso sumativo es entregar premios asociados al desempeño los que pueden significar el pago de un bono, beneficios asociados a la jornada

de trabajo, otorgar períodos sabáticos, oportunidades para desarrollar investigación o estudios de postgrado, entregar reconocimiento público, entre otros posibles premios (OECD, 2013).

Un último uso sumativo se relaciona con las sanciones asociadas a la obtención de desempeños bajo lo esperado. Hay países que luego de obtener un desempeño bajo lo esperado en un número definido de evaluaciones, sanciona a los profesores removiéndolos de sus funciones docentes, o simplemente terminando su contrato. De todas formas, comúnmente la identificación temprana de desempeños bajo el estándar esperado son utilizados para definir planes de apoyo para mejorar la práctica, pero cuando esto no ocurre, hay sistemas que definen que sean transferidos a otras funciones, otros los suspenden de sus funciones y en otros casos, son despedidos (OECD, 2013).

De este modo, todos los aspectos antes desarrollados inciden en el tipo de modelo u aproximación que tenga el sistema de evaluación docente de cada país. Desarrollar sistemas de evaluación docente puede ser costoso y difícil de implementar, pero es crítico lograr conjugar las demandas por una educación de calidad, con la mejora de las prácticas docentes a través del desarrollo profesional y el reconocimiento de los conocimientos, habilidades y competencias de los docentes. La expectativa es que involucrarse en prácticas reflexivas, analizar los métodos de enseñanza, y compartir experiencias con pares en los colegios, se convierta en parte de la rutina habitual de la práctica profesional docente, y así, estas sean de cada vez mayor calidad (OECD, 2013).



3. CONSIDERACIONES GENERALES SOBRE LA ELECCIÓN DE INSTRUMENTOS DE EVALUACIÓN

Los instrumentos utilizados para evaluar el desempeño docente en distintos países son diversos y la elección de los más idóneos depende de varios factores. En términos generales, puede decirse que la elección y desarrollo de instrumentos deberá tomar en cuenta las tradiciones específicas del país en el cual se aplicarán, junto con las características del sistema de educación respectivo y los recursos con los que se cuenta (Santiago, 2013b). Además, no todas las implicaciones políticas son igualmente relevantes en los distintos países. Para algunos, ciertas sugerencias ya están en funcionamiento, mientras que para otros, puede que no se ajusten considerando sus tradiciones y estructura social, económica y educacional. Así, el contexto particular de cada país determinará las prioridades al momento de desarrollar un sistema de evaluación docente (Santiago, Nusche, Shewbridge & Radinger, 2013a).

En segundo lugar, debe considerarse el propósito específico, ya sea formativo o sumativo, que tenga el sistema de evaluación (Santiago, 2013b). Algunos instrumentos, por ejemplo, tienen entre sus ventajas el otorgar información detallada y específica de la práctica docente, lo cual sin duda es relevante en un sistema cuyo foco es formativo y por tanto la retroalimentación es central. Otros en cambio, entregan una visión global sobre el nivel de logro lo que puede ser pertinente cuando el fin es sumativo.

En tercer lugar, tiene que considerarse qué aspectos consensuados de la práctica docente se quieren evaluar y bajo qué estándares (OECD, 2013). Finalmente, un cuarto aspecto de importancia, tiene que ver con las capacidades que los evaluadores requerirán al momento de la evaluación, más los apoyos que podrán o no ofrecerse desde el nivel central (Santiago et al., 2013a).

Así, todos estos elementos inciden en los instrumentos que se propongan para un sistema particular influyendo en los resultados obtenidos.

4. REVISIÓN DE LOS PRINCIPALES INSTRUMENTOS USADOS EN EVALUACIONES DOCENTE Y SUS VENTAJAS Y DESVENTAJAS

A continuación, se detallan los instrumentos y fuentes de información más importantes utilizados para evaluar el desempeño docente, según la revisión de la OECD (2013) realizada en 28 países. La presentación de cada instrumento se estructura en tres partes: primero se detallan sus principales características; luego se nombran los países que lo utilizan para gestionar el desempeño docente, es decir, que evalúan regularmente a sus docentes para definir su carrera y el desarrollo profesional³, y en tercer lugar, se analizan sus ventajas y desventajas.

4.1 Observación de clases

2.1.1 Descripción del instrumento

Muchos de los aspectos claves de la enseñanza se despliegan cuando los profesores interactúan con sus estudiantes en la sala de clases, y por ello, la observación de clases es una fuente crucial de información para evaluar (OECD, 2013). Consiste en registrar sistemática y objetivamente lo que sucede en el aula, para lo cual se utiliza una pauta de observación que puede ser más o menos estructurada.

La pauta para observar deberían focalizarse en dos aspectos: i) focalizar la atención del observador en aspectos específicos de la enseñanza, y ii) establecer niveles de logro de acuerdo a estándares claramente diferenciados (Bill & Melinda Gates Foundation, 2012).

La mayoría de las veces la observación de clases es realizada por el director(a) del colegio o algún miembro del equipo directivo, aunque hay países que graban la clase y luego

³ Las evaluaciones para la gestión del desempeño docente incluyen países que emplean evaluaciones para registrar o certificar docentes en ejercicio, como también las evaluaciones regulares que definen la permanencia, promoción o ascenso de los docentes. Excluyo por tanto, de forma deliberada, la mención a experiencias de países que emplean los instrumentos para acreditar docentes al término de su licenciatura o período de prueba, y para entregar bonos o recompensas para un grupo selecto según su desempeño.

la evalúa un observador entrenado, como es el caso de Chile (OECD, 2013). También es posible que la observación la realicen pares o profesores mentores (Amrein-Breadsley, 2014).

Algunos ejemplos de estos instrumentos son: Classroom Assessment Scoring System (CLASS), desarrollado por Pianta y Hamre, de la Universidad de Virginia; Framework for Teaching (FFT) desarrollado por Charlotte Danielson; Mathematical Quality of Instruction (MQI) desarrollado por Hill and Lowenberg Ball de la Universidad de Michigan; Protocol for Language Arts Teaching (PLATO) desarrollado por Pamela Grossman de la Universidad de Stanford; y Quality Science Teaching (QST) desarrollado por Raymond Pecheone también de la Universidad de Stanford (Bill & Melinda Gates Foundation, 2013).

2.1.2 Uso del instrumento para fines de permanencia y ascenso

La mayoría de los países que emplea la evaluación docente para gestionar el desempeño, definiendo con ello la permanencia de los docentes, utiliza observaciones de clases como instrumento. En el caso de Israel, Korea y Portugal por ejemplo, las utilizan específicamente para definir la promoción y ascenso docente (OECD, 2013).

2.1.3 Ventajas y desventajas del instrumento

La observación de clases es altamente recomendada y muy utilizada por los diversos países de acuerdo a la revisión de la OECD (2013). Entre sus ventajas se encuentra el hecho de que permite acceder de forma auténtica y directa a aspectos claves de la enseñanza, tal y como ocurren en la interacción con los estudiantes, otorgando una riqueza de información difícil de lograr mediante otros instrumentos. Además, permite identificar en forma precisa aspectos de la práctica que requieren mejorarse, facilitando así que puedan ser abordadas mediante actividades de desarrollo profesional docente apropiadas (Bill & Melinda Gates Foundation, 2012; Santiago et al., 2013a).

Por otra parte, estudios han demostrado que es un instrumento que logra predecir adecuadamente la efectividad docente respecto de los logros de aprendizaje que logran en sus estudiantes (Bill & Melinda Gates Foundation, 2012; Kane and Staiger, 2012, Kane et al., 2010, Milanowski, 2004 citados en Santiago et al., 2013a). Asimismo, otro de sus beneficios para sistemas que buscan potenciar el desarrollo docente, es que la utilización de múltiples

observaciones de clase junto con la entrega de retroalimentación detallada a los docentes, mejorar el desempeño de estos junto con su efectividad futura (Allen et al., 2011, Taylor & Tyler, 2011 citados en Bill & Melinda Gates Foundation, 2012; Milanowski, Kimball & White, 2004, Milanowski, 2004, Rockoff & Speroni, 2010 citados en Darling-Hammond, Amrein-Beardsley, Haertel & Rothstein, 2012a).

Ahora bien, entre sus desventajas o aspectos críticos a considerar se encuentra que requiere de observadores adecuadamente entrenados, de forma que exista una adecuada confiabilidad entre observadores, y excelentes pautas de registro y evaluación (Amrein-Beardsley, 2014; Bill & Melinda Gates Foundation, 2012; Kane and Staiger, 2012, Kane et al., 2010, Milanowski, 2004 citados en Santiago et al., 2013a). En esta línea, algunas de las dificultades que las pautas de observación han tenido es i) la inclusión de muchos ítems que muchas veces resultan redundantes entre sí o superfluos, ii) el uso de pautas que no siempre han sido validadas externamente o basadas en investigación, y iii) el uso de instrumentos que no están lo suficientemente anclados en un modelo teórico que lo sustente (Amrein-Beardsley, 2014).

Otra desventaja, es que para lograr una confiabilidad adecuada se requiere de más de una observación, lo que complejiza algunos modelos de implementación y uso de este instrumento. Ello debido a que un reciente estudio realizado por Bill and Melinda Gates Foundation (2013), encontró que existe alta variación en la práctica docente entre una clase y otra, y más aún entre un observador y otro, por lo que en sistemas de evaluación que tienen altas consecuencias, como definir la permanencia y el ascenso, se recomienda que más de un observador evalúe una clase, o si ello no es factible, utilizar la observación de más de una clase, de forma de aumentar la confiabilidad de la medida. Investigadores que critican la práctica de evaluar utilizando observaciones de clase, se refieren justamente a países en que las observaciones las realizan miembros del equipo directivo, quienes solo visitan al docente en forma anual, no obteniendo así evidencia auténtica de la práctica habitual que permita retroalimentar en forma constructiva a los docentes (Klinger et al., 2008, Daley and Kim, 2010, Danielson, 2011, Marshall, 2012, Papay, 2012 citados en OECD, 2013).

Un último aspecto a considerar cabe mencionarse para países en que la evaluación docente es más informal o no está regulada por marcos centralizados, en los cuales puede resultar difícil la inserción de los directivos en las aulas para evaluar la práctica docente mediante la observación. En estos casos, se recomienda comenzar a instalar la evaluación

docente mediante el uso de entrevista del profesor(a) con el directivo (Nusche et al., 2011a, 2011b, Shewbridge et al., 2011 citados en OECD, 2013). Otra alternativa, es la que ha seguido Chile, donde la clase del docente es grabada y luego corregida por profesores entrenados de acuerdo a estándares y procedimientos técnicos rigurosos, de forma de potenciar la confiabilidad del instrumento (Manzi, González & Sun, 2011).

4.2 Entrevistas del directivo y profesor(a)

2.2.1 Descripción del instrumento

Muchos modelos de evaluación docente requieren que el profesor establezca objetivos de desempeño a alcanzar en un determinado período de tiempo, en acuerdo con el directivo a cargo. La evaluación se realiza respecto del alcance de estos objetivos, a través de entrevistas individuales y discusiones reflexivas entre el evaluador, usualmente un miembro del equipo directivo, y el docente (OECD, 2013).

Para que las entrevistas individuales entre el docente y el directivo resulten efectivas, uno de los hallazgos centrales indica que estas reuniones deben basarse en evidencia, para reducir la subjetividad, y emplear un lenguaje común acerca de las prácticas docentes y sus posibilidades de mejora. En este sentido, se encontró que estas reuniones resultaban más reflexivas cuando se usaba, por ejemplo, el Marco para la Enseñanza de Charlotte Danielson, que cuando se usaban listas de cotejo de prácticas docentes (Sartain, Ray & Brown, 2011).

2.2.2 Uso del instrumento para fines de permanencia y ascenso

Este tipo de entrevistas son frecuentemente utilizadas entre los países, para evaluaciones orientadas a la gestión del desempeño. En Israel, es utilizada además con fines de promoción y ascenso. En Chile, esta entrevista es conducida por un profesor par, y además los directivos entregan una evaluación escrita (OECD, 2013).

2.2.3 Ventajas y desventajas del instrumento

Este instrumento resulta atractivo en cuanto permite asegurar la participación de un evaluador que está familiarizado con el trabajo del docente en el contexto particular en que

lo realiza, y así permite considerar factores contextuales en la evaluación del desempeño del docente (Santiago, Benavides, Danielson, Goe & Nusche, 2013b).

Asimismo, en sistemas de evaluación orientados al desarrollo de la carrera profesional, incorporar a los directivos tiene la ventaja de que se puede proveer una coordinación más fluida entre los resultados de la evaluación y los usos esperados, facilitando las promociones y ascensos dentro de la institución (Santiago et al., 2013b).

Además, se ha encontrado que las evaluaciones realizadas por los directivos respecto de las habilidades pedagógicas docentes logran predecir las ganancias futuras que los profesores lograrán en los aprendizajes de sus estudiantes. Específicamente, un estudio realizado en Estados Unidos mostró que los directivos lograban identificar acertadamente a aquellos docentes que producían las más altas y más bajas ganancias en los aprendizajes de sus estudiantes (20% superior e inferior). No obstante, eran menos hábiles en distinguir a los docentes que se encontraban en el medio de la distribución. Estos resultados sugerían que los programas que basan sus evaluaciones en los directivos y que recompensan o sancionan en función de incentivos económicos, debieran basarse en los docentes con más altos y más bajos niveles de desempeño (Jacob & Lefgren, 2006).

Dentro de sus desventajas o aspectos a considerar, se encuentra el entrenamiento que requieren los directivos tanto para evaluar como para retroalimentar a sus docentes. Respecto a la evaluación, la literatura argumenta extensamente sobre la necesidad de entrenar a los directivos en puntuar a sus profesores en función de la evidencia que recogen, en los criterios en los cuales deben basarse al momento de evaluar las buenas prácticas y en los niveles de desempeño que se corresponden con la calidad docente que se evalúa (OECD, 2013; Yariv, 2009).

Asimismo, los directivos debieran ser entrenados en cómo proveer retroalimentación constructiva y coaching que facilite la mejora en las prácticas docentes (OECD, 2009; Yariv, 2009). Se ha encontrado que la efectividad de las entrevistas entre docentes y directivos depende de las habilidades del directivo. Las entrevistas que requerían mejora se caracterizaban por ser instancias centradas fundamentalmente en los planteamientos del directivo dejando poco espacio de diálogo con el profesor, y en preguntas cerradas y de baja profundización, limitando el alcance de la retroalimentación (Sartain et al., 2011).

Estos mismos resultados se confirman en estudios que indagan sobre entrevistas o apoyos brindados por pares evaluadores que utilizan el rol de mentores de los profesores

evaluados. Se ha encontrado que la experticia de los profesores mentores influye en la calidad de estas reuniones, y que estas resultan más efectivas cuando se basan en evidencia de evaluaciones, y se centran en recomendaciones sobre las decisiones pedagógicas. Varios sistemas que utilizan esta aproximación han logrado no sólo mejorar las prácticas docentes, sino perfeccionar su efectividad en identificar profesores para permanecer en sus cargos y aquellos que requieren apoyo o bien, cuando es necesario, solicitar su retiro, pero requieren de un evaluador entrenado (NCTAF, 1996, Van Lier, 2008 citados en Darling-Hammond et al., 2012a).

Otra desventaja tiene relación con el tiempo que les implica a los directivos involucrarse en esta evaluación (OECD, 2009; Sartain et al., 2011). En esta línea, es importante que se considere reducir el trabajo administrativo de los directivos de forma de que cuenten con más tiempo para la evaluación docente, junto con tiempo para la retroalimentación y coaching a sus profesores (Marshall, 2005 citado en OECD, 2009).

4.3 Auto-evaluación del docente

2.3.1 Descripción del instrumento

La auto-evaluación implica que el docente evalúa su propio desempeño, reflexionando sobre cómo los factores personales, organizacionales e institucionales afectan su enseñanza. En general, los sistemas de evaluación docente utilizan este instrumento buscando que el docente analice sobre su práctica pedagógica y valore su propio desempeño profesional (OECD, 2013).

Una investigación reciente mostró que cuando la autoevaluación se basa en criterios y estándares objetivos, comprendidos por los docentes, estos lograban reportes reflexivos caracterizados por ser elaborados, concretos, específicos y detallados sobre su práctica. En cambio, cuando los docentes no utilizaban criterios de evaluación sus reportes eran limitados, en cuanto resultaban superficiales y generales (Van Diggelen, Den Brok & Beijaard, 2013).

En esta misma línea, el uso de este instrumento se fundamenta en los principios del aprendizaje de adultos, los cuales muestran que cuando las personas usan auto-evaluaciones sostienen más su aprendizaje y desarrollo profesional que cuando algo o alguien externo les señala los aspectos a mejorar de su práctica (Danielson & McGreal, 2000).

Ahora bien, entre los aspectos que deben definirse al diseñar este instrumento se encuentran cómo se corregirá (p.e si se usa una rúbrica), quién lo evaluará, cuándo y con qué escala (Amrein-Beardsley, 2014).

2.3.2 Uso del instrumento para fines de permanencia y ascenso

Este instrumento es utilizado en la mayoría de los países para evaluaciones orientadas a la gestión del desempeño, y en el caso de Estonia e Israel es utilizada con fines de promoción y ascenso (OECD, 2013).

2.3.3 Ventajas y desventajas del instrumento

En general, los beneficios que fundamentan el uso de este instrumento se relacionan con que la auto evaluación propicia que los profesores comprendan qué caracteriza una buena práctica docente (Samuels & Betts, 2007 citados en Van Diggelen et al., 2013), estimula la reflexión y el auto monitoreo de las capacidades de los docentes (Crooks, 1998 citado en Van Diggelen et al., 2013), y prepara a los profesores para el aprendizaje permanente (Boud, 1995 citado en Van Diggelen et al., 2013).

Sumado a lo anterior, otra ventaja es que puede ser fácilmente desarrollada e implementada, requiriendo poco del limitado tiempo disponible de los profesores (OECD, 2013), y que todos los profesores pueden entregar estos reportes para ser evaluados. Asimismo, valora el juicio de los docentes más que cualquier otra medida (Amrein-Beardsley, 2014).

Asimismo, otro beneficio es que permite que los docentes cuenten con un instrumento en que pueden entregar información contextual o reflexionar sobre su propio proceso o resultados de la evaluación, sobre todo en sistemas que utilizan múltiples instrumentos. Ahora, el riesgo que debe cautelarse es que al incluir auto-reportes y reflexiones puede que los docentes incluyan muchas excusas en relación a su desempeño, aun cuando el beneficio es lograr comprender explicaciones que en el contexto, pueden resultar atingentes y necesarias (Amrein-Beardsley, 2014).

No obstante, este instrumento se desaconseja en sistemas de altas consecuencias (Santiago et al, 2013b). Ello debido a que los beneficios recién expuestos son difícilmente

respaldadas por evidencia empírica (Ross & Bruce, 2007), y a que los docentes tienden a evaluarse en el más alto estándar para no perjudicar su evaluación (Amrein-Beardsley, 2014; Santiago et al, 2013b). Esto último puede verse en el Sistema de Evaluación Docente chileno, donde un 99% de los profesores en los distintos períodos se han evaluado en los dos niveles de desempeño más altos, ubicándose alcanzando el estándar esperado (Manzi et al., 2011).

De esta forma, para que este instrumento tenga valor para los docentes y para su profesión, es necesario que los profesores puedan conducir su proceso de auto evaluación en privado, sin que altas consecuencias medien sus resultados. De lo contrario, es muy difícil que los profesores, aun cuando fueran precisos en su evaluación, lo que de todas formas no está asegurado, sean honestos (Santiago et al, 2013b).

Asimismo, esta baja validez del instrumento incide en que existan estudios que muestran una baja relación con el logro de aprendizajes en los estudiantes. Una investigación realizada por el Ministerio de Educación de Chile y el PNUD, muestra que la auto evaluación provee poca información para identificar buenas prácticas docentes, y además muestra una baja correlación con los resultados de los estudiantes (Alvarado, Cabezas, Falck & Ortega, 2012).

En síntesis, hay estudios cuyos resultados demuestran que la autoevaluación promueve mejoras en las prácticas docentes, sobre todo cuando además se combina con otras técnicas como coaching de pares u observación de clases de agentes externos (Ross & Bruce, 2007). No obstante, en evaluaciones con altas consecuencias no se encuentran estos mismos resultados, resultando en un instrumento pobre para indagar sobre las prácticas docentes (Santiago et al., 2013b).

4.4 Portafolio

2.4.1 Descripción del instrumento

El Portafolio (o carpeta) deriva de las colecciones de modelos, fotógrafos, arquitectos, diseñadores o artistas, con las que se pretende demostrar la calidad o nivel alcanzado en la profesión a través de la integración de un repertorio de trabajos más representativos o sobresalientes (Díaz & Pérez, 2010). Desde este contexto se deriva para su uso en educación, y particularmente en la evaluación docente.

En la literatura se encuentran diversas definiciones sobre lo que es un portafolio docente. En términos generales, puede definirse como una colección de trabajos o producciones elaboradas por el profesor –de manera individual o colectiva-, que está enfocada en la planificación, implementación o evaluación del proceso de enseñanza-aprendizaje de sus alumnos, y que ha sido realizada en el transcurso de un ciclo o curso escolar (Díaz & Pérez, 2010).

En esta línea, es un instrumento significativo en cuanto permite recopilar evidencia directa de aspectos claves de la práctica docente. Puede incluir una diversidad de cosas hechas por el profesor en el aula u otro espacio relacionado, como planificaciones de clases o de unidades de aprendizaje, recursos de aprendizaje, muestras del trabajo de los estudiantes y comentarios de este, cuestionarios de auto reporte y reflexiones sobre la práctica (OECD, 2013). Sumado a esto, puede incluir documentos que avalen experiencias claves de índole formativa, de perfeccionamientos y creatividad en la docencia, junto con observaciones o evaluaciones que ha recibido respecto de su labor como profesor, cartas de recomendación, premios, testimonios de tutores u otros integrantes de la comunidad educativa (Díaz & Pérez, 2010).

En el caso de Estados Unidos, particularmente de la certificación docente que ofrece el National Board for Teaching Standards (en adelante NBPTS), el portafolio incluye además de la evidencia directa de la práctica del docente, un conjunto de ejercicios o tareas de enseñanza que los profesores deben realizar en un centro de evaluación. En estos centros, los docentes deben demostrar, con juegos de rol, discusiones grupales o ejercicios de situación real, su manejo de conocimiento disciplinario y pedagógico. Esta forma de evaluar ha sido una alternativa a las pruebas estandarizadas de aprendizaje que usualmente se han usado para medir este ámbito (Pearlman, 2008).

En la última década cada vez más han dejado de usarse los portafolios físicos o impresos en papel, dando paso a los portafolios electrónicos –e portfolio o e-folio-. Estos aprovechan al máximo posible el entorno multimedia y las tecnologías de comunicación e información que los docentes utilizan, permitiendo la presentación de sonidos, gráficos, videos, hipertexto, demostraciones por computador, fotografías, foros, etc. (Díaz & Pérez, 2010).

El portafolio se utiliza en evaluaciones con distintos fines, lo cual determina el contenido que es solicitado en el Portafolio. Si la evaluación tiene como fin el desarrollo

profesional, debiera orientarse a obtener una honesta reflexión del docente sobre sus prácticas, de forma de poder determinar un plan de desarrollo. En cambio, si la evaluación tiene un fin orientado a determinar la progresión en su carrera profesional, el portafolio debiera permitir al profesor entregar evidencia de cómo sus prácticas favorecen el aprendizaje de los estudiantes (Santiago et al., 2013b).

Finalmente es importante notar que no sólo es un instrumento de evaluación per se, sino que es una herramienta para favorecer un enfoque reflexivo de la práctica docente (OECD, 2013).

2.4.2 Uso del instrumento para fines de permanencia y ascenso

El portafolio es comúnmente utilizado en los distintos países para todos los tipos de evaluaciones. Como se mencionó anteriormente, en Estados Unidos, el NBPTS ofrece una certificación para los profesores mediante la presentación de un portafolio. En Chile, también se utiliza este instrumento como parte de la evaluación regular orientada a la gestión del desempeño, y también como parte de una evaluación de recompensa (OECD, 2013).

2.4.3 Ventajas y desventajas del instrumento

Los portafolios docentes resultan especialmente atractivos en cuanto es un instrumento idóneo para aproximarse a prácticas docentes auténticas a través de la recolección de diversa evidencia del trabajo del profesor, la cual puede ser combinada de diversos modos de acuerdo al propósito específico del sistema de evaluación (Santiago, 2013b).

Asimismo, es un instrumento adecuado para propiciar el involucramiento de los docentes en la evaluación, en cuanto son ellos quienes deben preparar y decidir la evidencia que consideran más pertinente presentar acerca de su trabajo en cada ámbito evaluado (Attinello et al., 2006, Pecheone et al., 2005, Van der Shaaf et al., 2008 citados en Santiago et al., 2013b; Tucker, Stronge, Gareis & Beers, 2003). En este sentido, se argumenta que a diferencia de otros instrumentos, el portafolio propicia el desarrollo del juicio crítico profesional de los docentes, propiciando así la reflexión y su desarrollo profesional (Airasian & Gullickson, 1997, Wolf, 1999 citados en Tucker et al., 2003).

Asimismo, otra de sus ventajas es que permite evaluar ámbitos difícilmente evaluables a través de otros instrumentos. Un ejemplo de esto es el estudio de Tucker y otros (2003), quienes encuentran que el portafolio es especialmente útil para medir el desempeño docente en las áreas de evaluación y responsabilidades profesionales, las que eran difícilmente observables en clases. Sumado a esto, permite recoger información de todos los docentes independiente de la disciplina que enseñan (Amrein-Beardsley, 2014).

Por otra parte, este instrumento resulta beneficioso para potenciar el desarrollo docente. Estudios de profesores que participan de la certificación del NBPTS muestran que los profesores se vuelven más conscientes de sus decisiones pedagógicas y cambian las prácticas que reportan como resultado de esta mayor capacidad de darse cuenta y de las prácticas que requiere la evaluación (Athanases, 1994, Buday & Kelly, 1996, Sato, Wei & Darling-Hammond, 2008 citados en Darling-Hammond, Newton & Cheung, 2012b). Asimismo, un estudio acerca las percepciones de los profesores sobre sus competencias pedagógicas antes y después de completar el portafolio para el NBPTS, mostró que los profesores reportaban aumentos estadísticamente significativos en su desempeño en cada área evaluada (planificación, diseño e implementación de la enseñanza, manejo de la clase, evaluación del aprendizaje, uso del conocimiento disciplinar y participación en comunidades de aprendizaje) (Tracz et al., 2005 citado en Darling-Hammond et al., 2012b). Los profesores comentaban que analizar los videos de sus clases y el trabajo de los estudiantes potenció su capacidad de organizar la enseñanza y las tareas de aprendizaje, cómo analizar el aprendizaje de los estudiantes y cómo modificar su enseñanza cuando era necesario (Darling-Hammond et al., 2012b).

Sumado a lo anterior, estudios avalan su validez en cuanto se ha asociado con los resultados que logran los estudiantes. Un reciente estudio realizado por Wilson, Hallam, Pecheone & Moss (2014) con docentes principiantes, mostró que el portafolio elaborado por los profesores logra predecir las ganancias en los resultados que obtendrán en sus estudiantes, en cuanto los puntajes obtenidos en el portafolio lograban distinguir entre profesores que eran más y menos exitosos en lograr aprendizajes en sus alumnos. Estos resultados se confirman en varios estudios realizados con el portafolio del programa NBPTS, en donde se encuentra que las ganancias en los resultados de los estudiantes de profesores que obtuvieron la certificación son significativamente mayores que las de aquellos que no se certificaron o que nunca realizaron el proceso de evaluación (Cavaluzzo, 2004, Goldhaber &

Anthony, 2005, Clotfelter, Ladd, & Vigdor, 2006, 2007, Cantrell, Fullerton, Kane, & Staiger, 2007, Vandevort, Amrein-Beardsley, & Berliner, 2004 citados en Darling- Hammond et al., 2012b).

Entre sus limitaciones se encuentra que, al igual que en las observaciones de clase, se requiere contar con evaluadores entrenados que corrijan esta evidencia, sobre todo cuando se utiliza con fines de determinar la progresión en su carrera profesional (Díaz & Pérez, 2010; Santiago et al., 2013b), y además se requieren de bastante tiempo de corrección, y de rúbricas adecuadamente validadas (Amrein-Beardsley, 2014).

Otra dificultad es que en general, los docentes presentan distintos artefactos que tienen diferentes propósitos, y que al ser corregidos resulta difícil compararlos y se prestan para confusión (Amrein-Beardsley, 2014).

Además, se debe considerar que los docentes cuenten con tiempo suficiente para completarlo (Santiago et al., 2013b). En algunos contextos, los docentes consideran que elaborar el portafolio les quita tiempo importante de su labor, por lo cual, los sistemas que utilizan este instrumento debieran propiciar que los profesores puedan preparar su portafolio, recolectando material que naturalmente utilizarían como parte de su trabajo. Por ejemplo, la planificación de clases podría describir una unidad que esté actualmente enseñando (OECD, 2013). De lo contrario, se desencadena una percepción general de sobrecarga laboral producto de la evaluación, e intensas reacciones emocionales negativas por parte de los docentes evaluados (Taut, Santelices & Manzi, 2011).

4.5 Resultados de los estudiantes

2.5.1 Descripción del instrumento

Debido a que el éxito de un sistema educacional se basa en los resultados de aprendizaje de los estudiantes, modelos de evaluación docente han incorporado como una medida de la efectividad del profesor(a), la medición de la contribución del este a los logros de aprendizaje de sus estudiantes. Para esto, han utilizado modelos de valor agregado que, mediante modelos estadísticos, miden los cambios en los puntajes obtenidos por los estudiantes en pruebas estandarizadas de aprendizaje a lo largo del tiempo, considerando las

características de los estudiantes y otros factores que influyen en sus logros (Darling-Hammond et al., 2012a).

Estas medidas de valor agregado en general se obtienen de exámenes que se aplican al término de cursos, o de pruebas estatales, o de pruebas que se han desarrollado especialmente para estos propósitos, que se administran dos veces durante un año escolar y que buscan controlar en los resultados -a través de sofisticadas técnicas estadísticas-, el efecto de profesores anteriores, y de otras variables que inciden en los logros de aprendizaje que muestran los estudiantes y que no dependen directamente del profesor (Baker, 2011, Harris & Anderson, 2013, Nelson, 2011, Newton et al., 2010, Papay, 2011 citados en Amrein-Beardsley, 2014).

En algunos países, la mayoría de los profesores que obtiene los más altos desempeños en el sistema nacional de evaluación docente no obtiene, a un nivel agregado, altos resultados en los aprendizajes de sus estudiantes. Esta falta de convergencia es la que sustenta la importancia de usar los resultados de los estudiantes como parte de las fuentes de evaluación del desempeño docente (OECD, 2013).

Según la Encuesta Internacional sobre Enseñanza y Aprendizaje (en adelante TALIS) realizada por la OECD (2009), un 65% de los profesores reportó que consideraba que los resultados de los estudiantes debía ser una parte importante de la retroalimentación sobre su desempeño.

Los sistemas que incluyen como parte de su evaluación el resultado de los estudiantes buscan incentivar a los docentes a orientarse a que todos sus alumnos logren los aprendizajes esperados de acuerdo a los estándares definidos según el currículum nacional (OECD, 2013).

2.5.2 Uso del instrumento para fines de permanencia y ascenso

La mayoría de los países no utiliza los resultados de los estudiantes como parte de sus esquemas formales de evaluación. México utiliza los resultados de los estudiantes como parte de la evaluación de los docentes asociada a la gestión del desempeño. En Chile, los resultados de los estudiantes medidos en pruebas de aprendizaje estandarizadas se utilizan para evaluar grupos de docentes, de un mismo establecimiento, como parte del Sistema Nacional de Evaluación del Desempeño (SNED). Estados Unidos, Inglaterra, Escocia y

Singapur reportan utilizar los resultados de los estudiantes en alguna de las etapas del proceso global de evaluación docente (OECD, 2013).

2.5.3 Ventajas y desventajas del instrumento

Uno de los aspectos beneficios del uso de los resultados de los estudiantes para la evaluación del desempeño docente, es que traslada la discusión sobre la efectividad docente hacia el aprendizaje de los estudiantes, que es el objetivo central de la enseñanza. Sumado a ello, introduce una medida cuantitativa objetiva sobre el desempeño docente (Braun, 2005, citado en OECD, 2013), y permite contar con indicadores del progreso a lo largo del tiempo (Amrein-Breadsley, 2014).

Ahora bien, investigaciones recientes han mostrado que las medidas de valor agregado son inestables, variando en un mismo docente en mediciones de un año y otro, y entre grupos de estudiantes, y según el modelo estadístico utilizado. Sumado a ello, la investigación ha revelado que es difícil identificar la contribución específica de un profesor(a) en el aprendizaje de un estudiante, dado que el aprendizaje está influido por múltiples factores⁴, entre los cuales está la labor de otros docentes, lo que resulta difícil de medir y controlar de forma precisa mediante esta técnica (Baker et al., 2010; Darling-Hammond et al., 2012a; Isoré, 2009 citado en OECD, 2013; Papay, 2011).

Por ello, existe amplio consenso en la literatura sobre dos puntos. Primero, no se recomienda utilizar estas medidas en sistemas de evaluación docente con altas consecuencias, que toman decisiones a nivel individual y que evalúan poblaciones disímiles de colegios o estudiantes (Baker et al., 2010; Darling-Hammond et al., 2012a; Isoré, 2009 citado en OECD, 2013; Kane & Staiger, 2002, McCaffrey et al., 2003, CAESL, 2004, Braun, 2005, Ingvarson et al., 2007, citados en OECD, 2013; Papay, 2011). Segundo, usar el resultado de los estudiantes como instrumento de evaluación resulta más relevante en evaluaciones de un establecimiento en su globalidad, más que para evaluaciones de desempeño individual docente (OECD, 2013).

⁴ Entre los factores que influyen se encuentran las propias habilidades del estudiante, sus expectativas, motivación y conducta; el soporte que los estudiantes reciben de sus familias y la influencia de su grupo de pares; la organización del colegio, sus recursos y clima institucional; la estructura y contenido curricular, entre otros.

Sumado a estas dificultades estadísticas, se han encontrado algunas consecuencias no deseadas que atentan contra el uso de los resultados de los estudiantes para evaluar la docencia. Primeramente, estudios han mostrado que se estrecha y sobre simplifica el currículum para enfocarse especialmente en las asignaturas que son medidas con las pruebas estandarizadas. Además, desalienta a los profesores a trabajar con estudiantes con necesidades educativas, dado que ello influirá negativamente en su desempeño, y además desincentiva la colaboración entre docentes (Baker, 2010; Scherrer, 2011).

Por otra parte, otra de las desventajas de este instrumento es que no permite contar con información específica sobre los aspectos destacados y deficitarios de la práctica docente, por lo que no ayuda a que los profesores mejoren su práctica, ni orienta a los directivos sobre cómo fortalecer el desarrollo profesional de estos (Bill & Melinda Gates Foundation, 2012; Santiago, 2013a).

Otra dificultad recae en que tiene una baja "validez aparente", en cuanto resulta difícil explicar a los profesores y directivos de los colegios cómo funciona la técnica del valor agregado y justificar su uso. Sumado a esto, son mediciones costosas, y además sólo se pueden utilizar para evaluar a los profesores de las asignaturas que habitualmente se miden en evaluaciones estandarizadas (Santiago, 2013a).

4.6 Encuestas a estudiantes y padres

2.6.1 Descripción del instrumento

Las encuestas son cuestionarios previamente diseñados que obtienen información a través de preguntas, comúnmente cerradas, que advierten acerca de cómo los estudiantes, padres y otros actores que están en permanente contacto con el docente perciben su desempeño (Peterson, 2000, Peterson et al., 2000, 2003, Jacob & Lefgren, 2005 citados en OECD, 2013).

En educación superior estas encuestas han sido utilizadas por décadas para evaluar la satisfacción de los estudiantes en relación a la enseñanza de los docentes, su calidad instruccional, la percepción sobre su efectividad, entre otros (Amrein-Beardsley, 2014).

Un ejemplo de una encuesta utilizada para evaluar desempeño docente corresponde a la denominada "Tripod", la cual fue desarrollada por Ronald Ferguson y Rob Ramsdell,

investigadores de Harvard, y administrada por el departamento de Educación de Cambridge. Esta encuesta evalúa el grado en que los estudiantes experimentaban involucramiento, exigencia y apoyo en su desarrollo intelectual a través del ambiente de la clase que es propiciado. Esta encuesta pregunta a los estudiantes si ellos están de acuerdo o en desacuerdo con afirmaciones específicas⁵ sobre la práctica del profesor(a), de forma que los profesores puedan mejorar el uso del tiempo de clase, la calidad de los comentarios que entregan en las tareas, sus prácticas pedagógicas y las relaciones que establecen con los estudiantes (Bill & Melinda Gates Foundation, 2010).

2.6.2 Uso del instrumento para fines de permanencia y ascenso

Las encuestas que completan estudiantes y padres son escasamente usadas en las evaluaciones formales que implementan los distintos países para evaluar el desempeño docente, a pesar de que en educación superior se utilizan ampliamente (Bill & Melinda Gates Foundation, 2010; OECD, 2013).

Los resultados de la Encuesta TALIS (OECD, 2013), muestran que la mayoría de los docentes percibe que la retroalimentación de los estudiantes y de los padres es un aspecto importante que es considerado en sus evaluaciones. Asimismo, en algunos países es común que los mismos profesores diseñen encuestas para obtener retroalimentación de su práctica y del progreso del aprendizaje de los estudiantes.

En evaluaciones que buscan gestionar el desempeño, México y Canadá utilizan las encuestas de los estudiantes en toda ocasión, mientras que Nueva Zelanda, República Eslovaca y Estados Unidos las utilizan en algunos casos. En Polonia se considera la opinión del centro de padres sobre el desempeño docente para la promoción y ascenso docente. En Korea se utilizan encuestas completadas por padres y estudiantes como parte del sistema regular de evaluación docente con fines de gestión del desempeño (OECD, 2013).

⁵ Ejemplos de afirmaciones son las siguientes: "Mi profesor sabe cuándo la clase le está entendiendo y cuando no", "Mi profesor usa varias y buenas maneras de explicar cada tópico que revisamos en clase", y "Cuando entrego mis tareas mi profesor me entrega retroalimentación útil que me ayuda a mejorar".

2.6.3 Ventajas y desventajas del instrumento

Uno de los aspectos más atractivos de las encuestas es que permiten incluir una perspectiva de actores relevantes que de otro modo tiende a quedar fuera (Peterson et al., 2003 citado en Santiago et al., 2013b), y que pueden otorgar información contextual rica de la práctica docente, y cercana en el caso de los docentes. Asimismo, cuando se utilizan preguntas abiertas se puede recoger información útil para ser utilizada con fines formativos, y retroalimentar a los profesores (Amrein-Breadsley, 2014).

Otra ventaja es que mediante las encuestas es posible recoger percepciones de prácticamente todos los docentes independiente de la disciplina que enseñan, exceptuando los de educación parvularia y tal vez, primeros cursos de primaria (Baeder, 2010, Harris, 2011, Kyriakides, 2005, Rothstein, 2011, Toch y Rothman, 2008 citados en Amrein-Breadsley, 2014; Bill y Melinda Gates Foundation, 2013).

Otra ventaja es que pueden alcanzar confiabilidades altas (de 0.8 y más), que además son estables para un mismo profesor durante un mismo año (Bill & Melinda Gates Foundation, 2010).

Además, en el caso de las encuestas que completan los estudiantes, se ha visto que predicen los logros de aprendizaje que estos tienen. Específicamente, un estudio realizado por la Fundación Bill y Melinda Gates (2013), mostró que las percepciones de los estudiantes sobre las fortalezas y debilidades de la práctica de los docentes son consistentes a lo largo de los distintos estudiantes que enseña cada profesor, y se relacionan con sus logros de aprendizaje. Los resultados de este estudio avalan el uso de encuestas de alumnos como una medida complementaria a otros instrumentos de evaluación docente, como la observación de clases, siempre que las encuestas soliciten a los estudiantes entregar retroalimentación específica sobre aspectos de la práctica docente como el uso del tiempo de clases, la calidad de la retroalimentación entregada en sus tareas, la relación que establece con ellos, entre otros.

Otras ventajas de este instrumento tienen relación con que sus costos tienden a ser bajos en comparación a otros instrumentos. Además, su validez aparente es alta, en cuanto los distintos usuarios (padres, estudiantes y directores) comprenden la relación de lo medido con la enseñanza y su relevancia. Por último, las encuestas se pueden extender a cursos y asignaturas que comúnmente no son evaluados (Bill & Melinda Gates Foundation, 2010).

Dentro de las dificultades que deben considerarse asociados al uso de este instrumento, se señala que la consistencia observada en sus resultados en ocasiones se debe al efecto halo o efecto de error de puntuación que está presente. Este tiene relación con que quienes responden las encuestas suelen dar respuestas holísticas para casi todos los ítems del instrumento, sin focalizarse o discriminar su puntuación respecto al contenido particular de cada ítem (Amrein-Beardsley, 2014).

Otra limitación que se ha descrito tiene relación con que las puntuaciones de las encuestas resultan susceptibles de ser influidas por fuentes que aportan varianza irrelevante al constructo, como i) variables asociadas a quienes responden (p.e edad, género, nacionalidad, etnia, lenguaje, comprensión de la relevancia y fines de la encuesta etc.); ii) variables asociadas al profesor (p.e género, edad, nacionalidad, etnia, lenguaje, atractivo físico, carisma etc.); iii) variables del curso (p.e tamaño, disciplina etc.); y iv) características del instrumento (p.e modo de administración, escalas usadas, etc.) (Block, 1998, Davies, Hirschberg, Lye, Johnston & McDonald, 2007, Gamliel & Davidovitz, 2005, Liu & Teddlie, 2005, Sedlmeier, 2006, Trout, 1997, Williams & Ceci, 1997 citados en Amrein-Breadsley, 2014).

Otras desventajas que amenazan la validez del instrumento tienen relación con la dificultad para recoger una adecuada tasa de respuesta por parte de los padres, y con que la construcción de las encuestas en ocasiones no se funda en un modelo teórico a su base (Amrein-Breadsley, 2014).

Finalmente, específicamente respecto de las encuestas que completan los padres, hay quienes aconsejan que se utilicen para evaluaciones globales del establecimiento educativo más que para resultados individuales, dado que ellos muchas veces no tienen conocimiento de la práctica habitual y además, suelen valorar aspectos que no siempre tienen relación directa con la calidad de la enseñanza para favorecer aprendizajes en los estudiantes (Isoré, 2009, citado en OECD, 2013).

4.7 Pruebas para evaluar competencias

2.7.1 Descripción del instrumento

En algunos países, los profesores son evaluados mediante pruebas estandarizadas que miden sus competencias generales y específicas en diferentes áreas. Comúnmente, se

utilizan pruebas de conocimientos disciplinarios y pedagógicos, para lo cual se requiere evaluar mediante pruebas diferenciadas a los docentes de distintas asignaturas (OECD, 2013).

Muchas de estas pruebas están fundamentalmente basadas en preguntas de opción múltiple, validadas a través de pruebas de campo, y algunas preguntas abiertas validadas a través de entrevistas cognitivas (Centro Microdatos, 2012).

2.7.2 Uso del instrumento para fines de permanencia y ascenso

La utilización de estas pruebas es relativamente escasa entre los distintos países examinados por la OECD. México y Portugal utilizan pruebas de evaluación de conocimientos disciplinarios y pedagógicos en las evaluaciones regulares que buscan gestionar el desempeño. En Polonia se utilizan pruebas de conocimientos disciplinarios y pedagógicos para evaluar la promoción y ascenso (OECD, 2013).

2.7.3 Ventajas y desventajas del instrumento

Una de las ventajas de las pruebas estandarizadas, que años atrás constituyó el fundamento para usarlas para evaluar la docencia, es que permiten difundir a los docentes los estándares requeridos para el ejercicio profesional del profesorado. Específicamente, daban a conocer el conocimiento disciplinario y pedagógico que se espera que los docentes manejen en cada área de aprendizaje, esperando así que estos sean logrados por los estudiantes en mayor medida (Darling-Hammond et al., 2012b).

Sumado a ello, otra de sus ventajas es la alta confiabilidad que es posible alcanzar y garantizar en el uso de este instrumento (Centro Microdatos, 2012), las que en las distintas pruebas están en torno a .85 (Rodríguez et al., 2013), lo cual es sobre todo beneficioso en sistemas con propósitos sumativos, que tienen altas consecuencias para los docentes.

Ahora bien, la gran crítica que se ha realizado a este instrumento es que, basar el juicio sobre la competencia docente sólo sobre la base del resultado en una prueba estandarizada, de preguntas de opción múltiple, sobre conocimiento disciplinar y pedagógico, es simplificar y reducir la complejidad del rol crítico de los profesores. En este sentido, el hecho de que esta evaluación deje fuera la demostración de las habilidades pedagógicas

relevantes en el ejercicio de la práctica, como la capacidad para facilitar discusiones, motivar a los estudiantes, manejar la convivencia, entre otras, es lo que se critica fuertemente de acuerdo a los nuevos modelos de evaluación docente (Arnaut & Giorguli, 2010; Pecheone & Chung 2006).

Por otra parte, otra desventaja de las pruebas estandarizadas que evalúan a los docentes, es que son limitadas en términos de la retroalimentación en cuanto no permite enfocarse en la mejora de las prácticas. Ello debido a que utilizan preguntas de selección múltiple que hacen difícil obtener información detallada de los aspectos logrados y por lograr de los profesores. Evidentemente, esta es una desventaja si el propósito de la evaluación es propiciar el desarrollo profesional docente mediante la mejora continua de las habilidades y competencias pedagógicas del profesorado (Bravo, Falck, González, Manzi & Peirano, 2008; Darling Hammond et al., 2012b; Santiago et al., 2013b).

Una última desventaja de este instrumento es que estudios encuentran una baja predicción del resultado en las pruebas que rinden los docentes, en los aprendizajes que logran en sus estudiantes (Wilson et al., 2014). Un estudio reciente que evaluó cuán predictivo del desempeño docente resultaba el portafolio y las pruebas de aprendizajes, mostró que los resultados de portafolio agregaban información que no estaba contenida en las pruebas estandarizadas rendidas por los profesores, y que el portafolio resultaba ser un mejor predictor de las contribuciones de los profesores a los logros de aprendizaje de sus estudiantes, que las pruebas estandarizadas (Wilson et al., 2014). Estudios anteriores confirmaban la misma tendencia, en cuanto las tradicionales pruebas estandarizadas para medir conocimientos y habilidades de los docentes fallaron al intentar probar una relación significativa entre los puntajes obtenidos y las ganancias provocadas en los aprendizajes de los estudiantes (Andrews, Blackmon & Mackey, 1980, Ayers & Qualls, 1979, Haney, Madaus, & Kreitzer, 1986, Wilson et al., 2007 citados en Darling-Hammond et al., 2012b).

En esta misma línea, cuando se ha estudiado la validez convergente de sistemas de evaluación que utilizan múltiples instrumentos en comparación con aquellos que utilizan pruebas de conocimiento disciplinario, se ha encontrado una relación débil que indicaría una convergencia baja. En cambio, esta convergencia es moderada al comparar los resultados en dos sistemas que utilizan portafolio (Taut et al., 2011).

5. CONCLUSIONES Y RECOMENDACIONES

La evaluación de la práctica y el desempeño docente comenzó a implementarse hace décadas, como resultado de las demandas por lograr una educación de calidad, pero la preocupación por los modelos desarrollados sigue presente hoy. Es un tema controvertido, en el cual las investigaciones han encontrado evidencia empírica mixta con respecto a sus efectos en el aprendizaje de los estudiantes, las que varían entre instrumentos, y con conflictos de interés entre los principales actores educativos (OECD, 2009).

Existe una amplia variedad de modelos y aproximaciones distintas para evaluar la profesión docente a lo largo de los distintos países. Hay diferencias en el rol asignado a los implementadores, en los propósitos que los sistemas de evaluación tienen, en los métodos e instrumentos que utilizan para evaluar, así como también en los estándares y criterios en base a los cuales definen lo que consideran una "buena" enseñanza, y en las consecuencias o recompensas que los sistemas establecen (OECD, 2009).

Ahora bien, existe amplio consenso sobre la necesidad de que los profesores se involucren en el desarrollo del proceso de evaluación. Un sistema de evaluación efectivo, justo y confiable requiere que los profesores acepten y se apropien del sistema (OECD, 2009). Asimismo, el involucramiento de los directivos en las evaluaciones docentes es crítico, no sólo por el conocimiento que estos tienen del contexto en que se desempeña el docente, sino por la posibilidad que su involucramiento abre en la instalación de su rol como líderes pedagógicos (Clayton, 2013).

Asimismo, se ha encontrado que los sistemas de evaluación resultan más efectivos cuando cuentan con evaluadores entrenados, cuando la retroalimentación es precisa y frecuente, y cuando tienen procesos de apoyo y guía hacia la mejora para los profesores, por ejemplo, a través del apoyo de pares mentores (Darling-Hammond et al., 2012a).

Cada vez más países muestran interés por implementar sistemas de evaluación docente comprensivos, donde se enfatiza la importancia de incluir una variedad de evaluadores, criterios e instrumentos de evaluación de forma de reflejar la complejidad de la buena enseñanza (Darling-Hammond et al., 2012a; OECD, 2009). Además, cuando los resultados de la evaluación se utilizarán para decisiones de altas consecuencias como definir la permanencia y ascenso del docente, se recomienda utilizar múltiples instrumentos y

fuentes de información (Bill & Melinda Gates Foundation, 2012; Sykes & Winchell, 2010 citados en OECD, 2013), de forma que los evaluadores midan diferentes conocimientos y habilidades de las competencias de los docentes (Goe et al., 2008, Peterson, 1987, Rockoff and Speroni, 2011, citados en OECD, 2013). Combinar distintas medidas ofrece varias ventajas: aumenta el poder predictivo de la medida, aumenta la confiabilidad y perfecciona el potencial diagnóstico que permitirá a los profesores entender cómo mejorar sus prácticas (Bill & Melinda Gates Foundation, 2012).

Cómo escoger los instrumentos de evaluación que se incorporarán en un sistema de evaluación no es tarea fácil, y existen varios factores que debieran considerarse, como las características del sistema educacional del país respectivo y su contexto político, los propósitos de la evaluación, los estándares que se quieren medir, los evaluadores y sus capacidades para implementar la evaluación, y las ventajas y desventajas que cada uno tiene, entre otros (OECD, 2013; Santiago et al., 2013a; Santiago, 2013b).

Los instrumentos más utilizados para evaluar a los profesores en distintos países, de acuerdo a la revisión hecha por la OECD (2013), son la observación de clases, la entrevista con el docente, la autoevaluación del profesor(a) y el portafolio. Esta misma organización recomienda utilizar el siguiente conjunto de instrumentos, para desarrollar sistemas de evaluación orientados a definir la progresión en la carrera docente: i) portafolio docente que incluya una auto evaluación reflexiva cualitativa por parte del docente, y que sea corregido por evaluadores entrenados; ii) observación de clases realizada por evaluadores acreditados (que pueden ser pares o directivos de otros colegios); y iii) reporte de los directivos para favorecer el vínculo con el colegio del docente (Santiago et al., 2013b).

En esta misma línea, el análisis sobre las ventajas y desventajas de los distintos instrumentos fundamenta la recomendación del uso de algunos más que otros, lo que se discutirá a continuación⁶.

La observación de clases y el portafolio aparecen a su vez altamente recomendados de acuerdo a la revisión de ventajas y desventajas de los distintos instrumentos de evaluación antes presentada, y sobre todo para evaluaciones de permanencia y ascenso. Ambos tienen la ventaja de acercarse a las prácticas auténticas del docente tal como ocurren en la sala de

⁶ En el anexo 2 puede revisarse una síntesis de las principales ventajas y desventajas de los distintos instrumentos de evaluación analizados.

clases, de permitir obtener información específica para orientar mejoras, y de haber mostrado buena predicción de la efectividad del docente en los logros de aprendizaje que propicia en sus alumnos (Bill & Melinda Gates Foundation, 2012; Cavaluzzo, 2004, Goldhaber & Anthony, 2005, Clotfelter, Ladd, & Vigdor, 2006, 2007, Cantrell, Fullerton, Kane, & Staiger, 2007, Vandevort, Amrein-Beardsley, & Berliner, 2004 citados en Darling- Hammond et al., 2012b; Santiago et al., 2013a; Santiago, 2013b; Wilson et al., 2014), todas ventajas que la literatura sugiere priorizar al momento de escoger instrumentos de evaluación docente (Santiago et al., 2013b).

Por su parte, la entrevista con el docente se vuelve central en modelos de evaluación que buscan fortalecer el liderazgo pedagógico de los directivos, de forma de propiciar que en las instituciones locales se instalen capacidades tanto de evaluación, como de promoción del desarrollo profesional. En sistemas que se orientan a conjugar la evaluación con la carrera docente, este instrumento se recomienda en cuanto se vuelve más crítico promover un rol activo por parte de los directivos, de forma que puedan coordinar de mejor forma los esfuerzos de promoción, ascenso y gestión de roles y cargos al interior de los establecimientos, para favorecer ambientes de colaboración y desarrollo profesional entre los distintos profesores (Santiago et al., 2013b).

En el caso de las encuestas, la evidencia revisada mostraría que las completadas por padres no se recomiendan para usos a nivel individual, pero las completadas por estudiantes pueden aportar una mirada que enriquece la evaluación, sobre todo si se la combina con otros instrumentos. Estas han mostrado predecir los futuros logros de aprendizaje que los docentes tendrán con sus alumnos, mostrar estabilidad a lo largo del año, y altas confiabilidades respaldadas por la cantidad de estudiantes que evalúan a un mismo docente (Bill & Melinda Gates Foundation, 2010). De todas formas, en sistemas con fines sumativos, deben manejarse con precaución considerando que los estudiantes no son expertos en enseñanza y que están involucrados emocionalmente con el trabajo del profesor (Santiago et al., 2013b).

Respecto a los demás instrumentos, tanto la autoevaluación como el uso de los resultados de los estudiantes no se recomiendan en sistemas de altas consecuencias a nivel individual (Baker et al., 2010; Darling-Hammond et al., 2012a; Isoré, 2009 citado en OECD, 2013; Kane & Staiger, 2002, McCaffrey et al., 2003, CAESL, 2004, Braun, 2005, Ingvarson et al., 2007, citados en OECD, 2013; Papay, 2011; Santiago et al, 2013b). En el caso de la

autoevaluación, sus beneficios en términos de la reflexión y auto-monitoreo de las propias prácticas dejan de percibirse cuando los docentes saben que su evaluación incidirá de forma importante en las consecuencias que esta tenga para ellos (Alvarado et al., 2012; Manzi et al., 2011; Santiago et al., 2013b). En el caso de los resultados de los estudiantes, las medidas de valor agregado hasta ahora desarrolladas son bastante inestables (Baker et al., 2010; Darling-Hammond et al., 2012a; Isoré, 2009 citado en OECD, 2013; Papay, 2011) y además, no entregan información sobre las prácticas docentes que permita retroalimentar y conducir a la mejora (Bravo et al., 2008; Bill & Melinda Gates Foundation, 2012; Santiago, 2013a). Esto último es también uno de los motivos que hace poco recomendadas a las pruebas estandarizadas (Darling Hammond et al., 2012b; Santiago et al., 2013b). Sumado a ello, las pruebas han mostrado una baja predicción de la efectividad docente (Andrews, Blackmon & Mackey, 1980, Ayers & Qualls, 1979, Haney, Madaus, & Kreitzer, 1986, Wilson et al., 2007 citados en Darling-Hammond et al., 2012b; Wilson et al., 2014).

Ahora bien, dado que la evidencia sobre el aprendizaje de los estudiantes es fundamental para definir la efectividad docente, los profesores pueden mostrar evidencia a favor del progreso de sus alumnos, por ejemplo, a través de portafolios, o también de la entrevista con el directivo, en que pueden acordar metas específicas del aprendizaje a lograr por sus cursos o grupos de alumnos. A este respecto es importante que las metas que se establezcan sean alcanzables, y que se sitúen más en las contribuciones que el docente realiza para alcanzar los aprendizajes de sus estudiantes (Santiago et al., 2013a).



6. REFERENCIAS

- Alvarado, M., Cabezas, G., Falck, D. & Ortega, M.E. (2012). *La evaluación docente y sus instrumentos: Discriminación del desempeño docente y asociación con los resultados de los estudiantes*. Artículo de investigación, Centro de Estudios Ministerio de Educación de Chile y Programa de Desarrollo de las Naciones Unidas [PNUD], Santiago.
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education. Critical perspectives on tests and assessment-based accountability*. New York: Routledge.
- Arnaut, A. & Giorguli, S. (2010). *Los grandes problemas de México. Cap. VII. Educación*. México, D.F.: El Colegio de México.
- Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., Ravitch, D., Rothstein, R., Shavelson, R., Shepard, L. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.
- Bill & Melinda Gates Foundation. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: Author.
- Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Author.
- Bill & Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Author.
- Bravo, D., Falck, D., González, R., Manzi, J. & Peirano, C. (2008). *La relación entre la evaluación docente y el rendimiento de los alumnos: Evidencia para el caso de Chile*. Extraído el 07 de abril de 2014 de http://www.microdatos.cl/docto_publicaciones/Evaluacion%20docentes_rendimiento%20escolar.pdf
- Centro Microdatos. (2012). *Evaluación de desempeño docente: Asignación de excelencia pedagógica (AEP), evaluación docente, AVDI e inicia*. Extraído el 16 de abril de 2014 de http://www.microdatos.cl/doctos_noticias/moduloII docentes.pdf

- Clayton, C. (2013). *Understanding current reforms to evaluate teachers: A literature review on teacher evaluation across the career span*. Nueva York: Pace University.
- Cogshall, J., Rasmussen, C., Colton, A., Milton, J. & Jacques, C. (2012). *Generating teaching effectiveness: The role of job-embedded professional learning in teacher evaluation*. Washington D.C.: National Comprehensive Center for Teacher Quality.
- Danielson, C. & McGreal, T. (2000). *Teacher Evaluation to Enhance Professional Practice*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Danielson, C. (2011). *Competencias docentes: desarrollo, apoyo y evaluación*. Santiago: PREAL.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. & Rothstein, J. (2012a). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), p. 8-15.
- Darling-Hammond, L., Newton, S. & Cheung, R. (2012b). *Developing and assessing beginning teacher effectiveness: The potential of performance assessments*. Stanford Center for Opportunity Policy in Education, extraído el 11 de abril de 2014 de <https://edpolicy.stanford.edu/publications/pubs/657>
- Díaz, F. & Pérez, M. (2010). El portafolio docente a escrutinio: sus posibilidades y restricciones en la formación y evaluación del profesorado. *Observar 4*, p. 6-27. ISSN: 1988-5105.
- Isoré, M. (2010). *Evaluación docente: prácticas vigentes en los países de la OCDE y una revisión de la literatura*. Santiago: PREAL.
- Jacob, B. & Lefgren, L. (2006). When principals rate teachers. *Education Next*, 6 (2), p. 59-69.
- Manzi, J., González, R. & Sun, Y. (eds.). (2011). *La Evaluación Docente en Chile*. Pontificia Universidad Católica de Chile, Facultad de Ciencias Sociales, Escuela de Psicología, Centro de Medición MIDE UC, Santiago, Chile.
- Organización para la Cooperación y el Desarrollo Económico [OECD]. (2009), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264072992-en>.
- Organización para la Cooperación y el Desarrollo Económico [OECD]. (2013). *Teachers for the 21st century: Using evaluation to improve teaching*. OECD Publishing.

- Papay, J. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Education Research Journal*, 48(1), p.163-193. DOI: 10.3102/0002831210362589.
- Pearlman, M. (2008). The design architecture of NBPTS certification assessments. En Stake, R., Kushner, S., Ingvarson, L. & Hattie, J. (Eds.). *Advances in Program Evaluation Vol. 11: Assessing Teachers for Professional Certification. The First Decade of the National Board for Professional Teaching Standards* (p.x
- Pearlman, M. (2008). The design architecture of NBPTS certification assessments, in Ingvarson, L. & Hattie, J. (Eds.). *Assessing Teachers for Professional Certification: the first decade of the national board for professional teaching standards, advances in program evaluation*, Vol. 11, p. 55-92. Oxford: Elsevier Ltd.
- Pecheone, R. & Chung, R. (2006). Evidence in teacher education: The performance assessment for California teachers (PACT). *Journal of Teacher Education*, 57(1), p.22-36.
- Rodríguez, B., Carreño, X., Muñoz, V., Ochsenius, H., Mahías, P., Bosch, A. (2013). *¿Cuánto saben de matemática los docentes que la enseñan y cómo se relaciona ese saber con sus prácticas de enseñanza?* Proyecto FONIDE N°: F611150. Extraído el 20 de abril de 2014 de www.fonide.cl
- Ross, J.A. & Bruce, C.D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23(2), 146-159.
- Santiago, P. (2013a, noviembre). *Evaluación docente en Chile: Fortalezas y desafíos*. Ponencia presentada en el Seminario Evaluación Docente en Chile: Un enfoque internacional, Santiago, Chile.
- Santiago, P. (2013b, noviembre). *Evaluación docente en Chile: Las recomendaciones de la OECD*. Ponencia presentada en el Seminario Evaluación Docente en Chile: Un enfoque internacional, Santiago, Chile.
- Santiago, P., Nusche, D., Shewbridge, C. & Radinger, T. (2013a). *Synergies for better learning. An international perspective on evaluation and assessment*. OECD Publishing.
- Santiago, P., Benavides, F., Danielson, C., Goe, L. & Nusche, D. (2013b). *Teacher evaluation in Chile 2013*. OECD Reviews of evaluation and assessment in education. OECD Publishing.

- Sartain, L., Ray, S. & Brown, E. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences and district implementation*. Consortium on Chicago School Research. Chicago: University of Chicago, Urban Education Institute.
- Sherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin*, 95 (2), p. 122-140.
- Shulman, L. (2001). Conocimiento y enseñanza. *Estudios Públicos*, 83, p.163-196.
- Taut, S., Santelices, V. & Manzi, J. (2011). Estudios de validez de la evaluación docente. En Manzi, J., González, R. & Sun, Y. (Eds.). *La Evaluación Docente en Chile (p.155 –175)*. Pontificia Universidad Católica de Chile, Facultad de Ciencias Sociales, Escuela de Psicología, Centro de Medición MIDE UC, Santiago, Chile.
- Tucker, P., Stronge, J., Gareis, C. & Beers, C. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference? *Educational Administration Quarterly*, 39 (5), p. 572-602.
- Van Diggelen, M., Den Brok, P. & Beijaard, D. (2013). Teachers' use of a self-assessment procedure: the role of criteria, standards, feedback and reflection. *Teachers and Teaching: Theory and practice*, 19 (2), p. 115-134.
- Wilson, M. Hallam, P.J., Pecheone, R. & Moss, P.A. (2014). Evaluating the validity of portfolio assessments for licensure decisions. *Education Policy Analysis Archives*, 22 (6). Extraído el 09 de abril de 2014 de <http://dx.doi.org/10.14507/epaa.v22n6.2014>.
- Yariv, E. (2009). Principals' *informal* methods for appraising poor-performing teachers. *Educational Assessment, Evaluation and Accountability*, 21, p. 283-298. DOI 10.1007/s11092-009-9081-3.



7. ANEXOS

7.1 Anexo N°1: Tipos de aproximaciones de evaluación docente en distintos países

Figura N°1: Tipos de aproximaciones políticas asociadas a la evaluación docente a lo largo de los países (OECD, 2013)

		Australia	Austria	Bélgica	Canadá	Chile	República Checa	Dinamarca	Estonia	Finlandia	Francia	Hungría	Islandia	Irlanda	Israel	Italia	Korea	Luxemburgo	México	Holanda	Nueva Zelanda	Noruega	Polonia	Portugal	República Eslovaca	Eslovenia	España	Suiza	Reino Unido
Evaluación de recompensa																													
Gestión del desempeño	Promoción																												
	Evaluación regular																												
	Registro																												
Evaluación inicial																													

7.2 Anexo N°2: Múltiples medidas de evaluación y sus ventajas y desventajas

Tabla N°1: Múltiples medidas de evaluación y sus ventajas y desventajas

Instrumento	Ventajas	Desventajas
Observación de clases	<ul style="list-style-type: none"> • Información auténtica de la práctica docente, capturando aspectos que "importan" en términos la efectividad docente (p.e relación entre docente y estudiantes). • Justo en cuanto todos los docentes pueden ser observados • Entrega información útil para retroalimentar la práctica • Predice efectividad docente • Educadores tienen experiencia usando modelos observacionales • La evaluación ocurre en el tiempo real, durante un día de clases 	<ul style="list-style-type: none"> • Requiere de entrenamiento de los observadores • Toma tiempo para lograr una evaluación objetiva (baja confiabilidad en observaciones únicas) • Instrumentos usados no siempre están validados externamente o basados en investigación • Instrumentos en ocasiones no están fundados en un modelo teórico
Portafolio	<ul style="list-style-type: none"> • Información auténtica de la práctica docente y muestras del trabajo de los estudiantes • Favorece el involucramiento y la reflexión docente • Potencia el desarrollo profesional docente • Predice efectividad docente • Justo en cuanto todos los docentes pueden entregar esta información independiente de la disciplina que enseñan 	<ul style="list-style-type: none"> • Requiere de entrenamiento de evaluadores y de rúbricas validadas • Diferentes artefactos entregados por diferentes profesores puede incrementar la confusión • Alto tiempo requerido por docentes y posible estrés • Se necesita bastante tiempo para corregir
Entrevistas directivo-profesor	<ul style="list-style-type: none"> • Permite considerar información contextual y facilitar los usos esperados • Recoge percepción de actores claves en liderar mejoras a nivel institucional • Predice efectividad docente 	<ul style="list-style-type: none"> • Requiere entrenamiento de directivos en observar y retroalimentar • Tiempo requerido por directivos
Encuestas a estudiantes y padres	<ul style="list-style-type: none"> • Información contextual rica de proveniencia directa de personas cercanas como los estudiantes • Bajo costo • Alta validez aparente • Predice efectividad docente • Alta confiabilidad • Justa en cuanto se puede obtener información de todos los docentes 	<ul style="list-style-type: none"> • Efecto halo influye en consistencia interna y confiabilidad • Baja tasa de respuesta de padres puede amenazar la validez • Inválida para ciertas poblaciones (p.e estudiantes de educación parvularia o niveles iniciales) • Susceptible a varianza irrelevante de constructo

	<p>independiente de la disciplina que enseñan</p> <ul style="list-style-type: none"> • Preguntas abiertas puede proveer información rica para usos formativos • Brinda a los estudiantes y padres la oportunidad de opinar sobre el proceso educacional 	<ul style="list-style-type: none"> • En el caso de las encuestas de padres, no se recomienda para usar para resultados individuales • Instrumentos en ocasiones no están fundados en un modelo teórico
Autoevaluación	<ul style="list-style-type: none"> • Valora el juicio de los docentes más que cualquier otra medida • Favorece el auto-monitoreo y la reflexión docente • Permite que los docentes conozcan lo que caracteriza las buenas prácticas • Requiere de poco tiempo • Justo en cuanto todos los docentes pueden entregar estos reportes • Permite explicar la información recogida a través de otros instrumentos y contextualizarla 	<ul style="list-style-type: none"> • En sistemas de altas consecuencias validez es baja • Alto potencial para distorsionar o inflar los puntajes • Baja predicción de efectividad docente
Pruebas estandarizadas	<ul style="list-style-type: none"> • Facilita que los docentes conozcan estándares • Alta confiabilidad 	<ul style="list-style-type: none"> • Limitada capacidad de retroalimentar en base a los resultados • No mide las habilidades pedagógicas en la práctica • Baja predicción de efectividad docente
Resultados de los estudiantes	<ul style="list-style-type: none"> • Centra efectividad docente en objetivo central de la enseñanza • Permite contar con indicadores del progreso a lo largo del tiempo • Mejor para evaluar a nivel escuela o para realizar análisis institucionales 	<ul style="list-style-type: none"> • Limitada capacidad de retroalimentar en base a los resultados • Inestabilidad de las medidas de valor agregado • Se aplica sólo a docentes de asignaturas comúnmente evaluadas en pruebas estandarizadas • Alto costo • Dificultad para comprenderse y usarse, es decir, baja validez aparente

