

PERSPECTIVAS DE DIRECTIVOS EDUCACIONALES SOBRE EL USO DE PRUEBAS ESTANDARIZADAS

El caso de una evaluación de bajas consecuencias

JOHANA CONTRERAS / ÁLEX TORRES

Resumen:

Las evaluaciones estandarizadas de aprendizaje se han extendido en el ámbito escolar, sin embargo, no existe suficiente claridad sobre cómo y bajo qué condiciones son utilizadas. Este artículo presenta una investigación que aborda los usos de un programa de evaluación de bajas consecuencias implementado en alrededor de 200 escuelas de Chile. Empleando un enfoque cualitativo, se realizaron 33 entrevistas a personal directivo de escuelas, de redes escolares y de programas educacionales, contratantes de estas evaluaciones. Los resultados permiten describir la utilización de este tipo de evaluación según los perfiles de las y los usuarios y las condiciones que la favorecen u obstaculizan, revelándose la existencia de usos pedagógicos y de mejora escolar, así como situaciones de subutilización. Finalmente, se discuten las implicaciones para las prácticas y política escolar.

Abstract:

The use of standardized achievement tests has expanded throughout the school setting, yet clarity is lacking on the form of usage and the conditions under which they are administered. This article presents a study of a low-stakes assessment program implemented in approximately 200 schools in Chile. Based on a qualitative focus, thirty-three interviews were conducted with the directors of schools, school networks, and educational programs that contract such assessments. The results permit a description of the use of standardized tests according to user profiles and the conditions that favor or hinder their usage, revealing the existence of pedagogical and school improvement uses, as well as situations of underutilization. Lastly, the implications for school policy and practices are discussed.

Palabras clave: evaluación de estudiantes; uso de la información; validez de las pruebas; investigación cualitativa.

Keywords: student assessment; use of information; test validity; qualitative research.

Johana Contreras: investigadora de la Pontificia Universidad Católica de Chile, Centro de Medición MIDE UC. Vicuña Mackenna 4860, Macul, Santiago, Chile. CE: jtcontre@uc.cl / <https://orcid.org/0000-0001-6081-0476>

Álex Torres investigador de la Pontificia Universidad Católica de Chile, Centro de Medición MIDE UC. Macul, Santiago, Chile. CE: atorresp@uc.cl / <https://orcid.org/0000-0003-4501-9453>

Introducción

Las evaluaciones estandarizadas de aprendizaje han experimentado una expansión importante durante las últimas décadas, transitando desde la certificación individual hacia un rol de rendición de cuentas y regulación de los sistemas escolares (Braun y Singer, 2019; Chalhoub-Deville, 2016; Maroy y Dutercq, 2017; Martínez-Rizo, 2016; Mons, 2009; Smith, 2014).

En Chile, debido a la implementación de políticas de rendición de cuentas, los estudios de los usos de las evaluaciones se han centrado en dispositivos de altas consecuencias, demostrando la existencia de impactos negativos en las prácticas de enseñanza y culturas escolares (Manzi, Bogolasky, Gutiérrez, Grau *et al.*, 2014; Elacqua, Martínez, Santos y Urbina, 2016; Falabella, 2016, 2020; Parra y Matus, 2016; Taut, Cortés, Sebastian y Preiss, 2009). Por ejemplo, se ha constatado la reducción del currículum de las disciplinas evaluadas, el entrenamiento para las evaluaciones y la preeminencia de una cultura de competición. Estos hallazgos coinciden con los trabajos desarrollados en otros contextos, en particular, en Estados Unidos (Koretz, 2017; Loeb y Byun, 2019).

Junto a ello, también se ha comenzado a promover el uso de las evaluaciones para la mejora escolar. Desde esta perspectiva, las evaluaciones estandarizadas no solo debiesen ser utilizadas por quienes diseñan y evalúan las políticas públicas, sino también, por los niveles intermedios y locales de toma de decisión, incluidos los equipos de gestión así como las y los docentes¹ (Datnow y Park, 2018; Ebbeler, Poortman, Schildkamp y Pieters, 2017; Mandinach y Gummer, 2016; Schildkamp, Poortman, Luyten y Ebbeler, 2016).

Este artículo aborda la validez de los usos y las consecuencias de evaluaciones estandarizadas *de bajas consecuencias*, es decir, diseñadas para entregar información que oriente la mejora escolar sin que sus resultados conlleven incentivos o sanciones para docentes y escuelas (AERA, APA y NCME, 2018; Tannenbaum y Kane, 2019). Este aspecto de la validez ha recibido menor atención en Chile y en el contexto internacional.

Esta investigación se desarrolló en el marco del Sistema de Evaluación del Progreso del Aprendizaje (SEPA), un conjunto de pruebas estandarizadas de lenguaje y matemáticas alineadas con el currículum nacional, que son contratadas voluntariamente por escuelas, administraciones locales y programas de intervención educacional en alrededor de 200 establecimientos educacionales a lo largo del país. El objetivo de este trabajo fue describir

y comprender los usos que dan a las evaluaciones SEPA los directivos y líderes de instituciones educacionales que las implementan.

Antecedentes

Usos de las evaluaciones estandarizadas de aprendizaje

Las evaluaciones estandarizadas de aprendizaje, cuyo diseño, aplicación y corrección son uniformes, se dirigen a medir los logros cognitivos de los alumnos y constituyen evidencia sobre la cual se toman decisiones (Mons, 2009). Sin embargo, se ha constatado una subutilización de esa información o una discrepancia entre los propósitos para las cuales fueron diseñadas y cómo son usadas (Alkin y King, 2016; Alkin y Taut, 2002).

Por esta razón, los reales usos y consecuencias de las evaluaciones constituyen una fuente de evidencia de validación de programas evaluativos, promovida ampliamente por los estándares internacionales de la medición educacional y la investigación asociada al tema (AERA, APA y NCME, 2018; Brewer, Knoeppl y Lindle, 2015; Hitchcock, Onwuegbuzie y Khoshaim, 2015; Kane, 2016; Lane, 2014; Loeb y Byun, 2019; Messick, 1980).

En la literatura se encuentran descripciones que pueden agruparse en tipologías según sus características y sus propósitos. Respecto de sus características, se identifican seis usos (Alkin y King, 2016, 2017; Alkin y Taut, 2002).

En primer lugar, el *uso instrumental* es aquel donde los resultados de las evaluaciones desembocan en acciones o decisiones dirigidas a resolver un problema. El *uso conceptual* influye o cambia el pensamiento y conocimiento del usuario sobre el objeto de la evaluación, sin conllevar una acción específica. El tipo *simbólico* ocurre cuando la evaluación solo se emplea como un acto político o una forma de obtener estatus. El tipo *legitimante* corresponde a su utilización para justificar decisiones o acciones emprendidas (Alkin y King, 2016; Alkin y Taut, 2002; Hein y Taut, 2010). El *uso de descubrimiento* refiere al empleo de hallazgos para ampliar el conocimiento en un campo disciplinar, más allá del programa evaluado. Finalmente, el *uso de proceso* (Patton, 1997) se refiere a los efectos que tiene la implementación misma de la evaluación en los usuarios y no sus hallazgos. Trabajos recientes integran este tipo como un eje transversal a los ya mencionados, de modo que lo conceptual e instrumental pueden describir tanto al proceso evaluativo como a sus hallazgos (Alkin y King, 2016; Alkin y Taut, 2002). De acuerdo con esta distinción, este trabajo

se concentra en los usos de los resultados de las evaluaciones y no de su proceso.

Por otro lado, Alkin y King (2017) proponen situar los usos de las evaluaciones en dos continuos: uno que va del “uso” al “no uso” y otro desde el “mal uso” hasta el “no mal uso”, incluyendo además un eje que distingue entre su carácter pasivo y activo. El uso y mal uso se sitúan en el polo activo y el no uso en el pasivo. Esta contribución es importante para esta investigación, por cuanto permite describir situaciones intermedias y la subutilización de las evaluaciones que, como se verá más adelante, representan la experiencia de diversos entrevistados.

Otra clasificación pertinente para este estudio es la que identifica los usos para fines de rendición de cuentas, para la mejora escolar y los usos pedagógicos (Bolhuis, 2019; Ebbeler, Poortman, Schildkamp y Pieters, 2016; Jimerson, Garry, Poortman y Schildkamp, 2020; Schildkamp, 2019; Schildkamp *et al.*, 2016). El primero se produce cuando los resultados de la evaluación sirven para demostrar a una audiencia externa la efectividad de acciones emprendidas por la escuela o los docentes. La utilización de datos para la mejora escolar, en cambio, analiza las acciones emprendidas con el fin de determinar las brechas respecto de las metas del establecimiento y determinar estrategias para reducirlas. Este aspecto corresponde más típicamente al ámbito de la gestión escolar e incluye las decisiones de desarrollo profesional docente, asignación de recursos, creación de planes de mejoramiento escolar, etc. El uso pedagógico se sitúa a nivel de aula y se refiere a la toma de decisiones orientada a modificar la planeación y las estrategias pedagógicas. Algunas acciones pedagógicas basadas en datos pueden ser, entre otras: definir metas de aprendizaje, determinar el dominio de conocimientos y habilidades o definir estrategias según grupos de estudiantes (Beaver y Weinbaum, 2015; Datnow, Lockton y Weddle, 2020; Ebbeler *et al.*, 2016; Farley-Ripple, Jennings y Butram, 2019; Gelderblom, Schildkamp, Pieters y Ehren, 2016; Prenger y Schildkamp, 2018).

Facilitadores y obstaculizadores en el uso de las evaluaciones

El estudio de los usos de las evaluaciones educacionales es prácticamente inseparable de los factores que inciden en ellos, lo que se manifiesta en numerosos trabajos publicados, sin embargo, este aspecto ha recibido poca atención en Chile y América Latina (Datnow y Park, 2018; Dunn, Airola, Lo y Garrison, 2013; Prenger y Schildkamp, 2018; Schildkamp, 2019;

Schildkamp *et al.*, 2016; Schildkamp, Smit y Blossing, 2017; Van Gasse, Vanlommel, Vanhoof y Van Petegem, 2017).

Además, la evidencia muestra que, pese al creciente volumen de datos disponibles, los equipos pedagógicos siguen tomando decisiones basadas en la intuición más que en los hallazgos de investigaciones y evaluaciones (Ebbeler *et al.*, 2017; Gelderblom *et al.*, 2016; Prenger y Schildkamp, 2018; Van Gasse *et al.*, 2017).

Según Schildkamp *et al.* (2016), los factores que favorecen la utilización de datos se pueden clasificar en tres tipos: características del dato, cualidades del usuario y condiciones organizacionales. Las primeras incluyen aspectos como qué tan oportuna es la entrega de resultados, su accesibilidad, usabilidad, confiabilidad o precisión.

Respecto de las cualidades del usuario, se han estudiado las competencias y habilidades para la comprensión e interpretación de datos, destacando la alfabetización estadística como un factor clave y un aspecto a fomentar en la formación inicial y el desarrollo profesional docente (Beaver y Weinbaum, 2015; Datnow, Lockton y Weddle, 2020; Mandinach y Gummer, 2016; Mandinach y Jimerson, 2016). Otros estudios abordan la autoeficacia respecto del uso de datos, es decir, las creencias relacionadas con las propias habilidades para interpretar y desprender acciones a partir de los datos o elementos, como las actitudes y la ansiedad frente a su utilización (Bolhuis, 2019; Dunn *et al.*, 2013; Gelderblom *et al.*, 2016; Jimerson, 2016; Kippers, Poortman, Schildkamp y Visscher, 2018; Prenger y Schildkamp, 2018).

Finalmente, dentro de las condiciones organizacionales destacan el liderazgo de los directivos y el soporte para el uso de datos, la visión compartida y normas institucionales, el tiempo y los espacios disponibles para el trabajo en torno a los datos (Coburn y Turner, 2011; Schildkamp *et al.*, 2016; Van Gasse *et al.*, 2017). Adicionalmente, la colaboración en los equipos es una de las condiciones organizacionales que ha demostrado unánimemente ser un facilitador en la utilización de datos (Abrams, Varier y Mehdi, 2020; Datnow, Lockton y Weddle, 2020; Jimerson *et al.*, 2020; Mandinach y Jimerson, 2016; Schildkamp *et al.*, 2016).

Un obstaculizador que mencionan algunos trabajos es el contexto de evaluaciones con fines de rendición de cuentas. Por ejemplo, Schildkamp (2019) afirma que existen tensiones entre los propósitos de mejora escolar y de rendición de cuentas; este último uso muchas veces genera un énfasis en el rendimiento, en detrimento de un enfoque formativo. Centrarse en

la rendición de cuentas también puede reforzar el pensamiento de déficit, es decir, los equipos pedagógicos atribuyen los resultados insatisfactorios de los estudiantes a sus características personales o sociodemográficas (Lasater, Bengtson y Albiladi, 2020).

Los hallazgos sobre los usos para rendición de cuentas son relevantes para esta investigación. Aunque el programa evaluativo estudiado sea de bajas consecuencias, se desarrolla en un sistema educativo donde predominan las políticas de rendición de cuentas y, por ende, se inserta en culturas escolares orientadas a los resultados y a la competición (Elacqua, Martínez, Santos y Urbina, 2016; Falabella, 2016, 2020; Parra y Matus, 2016).

En función de estos antecedentes, el presente estudio se dirigió a describir los usos que declaran líderes escolares y coordinadores de programas educacionales que deciden implementar un sistema estandarizado de evaluación de los estudiantes. Este objetivo se desglosó en tres preguntas: ¿Cómo se utilizan los resultados de las evaluaciones SEPA?, ¿difieren estos usos de acuerdo con los niveles de toma de decisión de sus usuarios? y, ¿qué condiciones favorecen u obstaculizan el uso efectivo de las evaluaciones SEPA?

El contexto del estudio es el Sistema de Evaluación del Progreso del Aprendizaje, desarrollado desde 2007 por un centro perteneciente a una universidad de gran prestigio en el país. Las evaluaciones son contratadas por instituciones educacionales públicas y privadas, incluyendo redes o administraciones escolares, establecimientos educacionales independientes, organizaciones públicas y no gubernamentales. Anualmente, se evalúan alrededor de 50,000 estudiantes pertenecientes a más de 200 establecimientos del país. Los servicios incluyen el diseño, administración de las pruebas y entrega de resultados, estos últimos accesibles de forma confidencial desde una plataforma en línea, distinguiendo cuatro perfiles de resultados: administrador de red de establecimientos, equipo directivo, profesor y estudiante.

El programa SEPA entrega resultados del estado del aprendizaje al finalizar un año escolar, del progreso entre años consecutivos y, a los establecimientos que disponen de dos o más mediciones, se les entrega un índice de valor agregado (aporte neto que realiza el establecimiento al aprendizaje de sus estudiantes, por sobre otros factores). Por último, el programa ofrece un acompañamiento en el uso y la gestión de los resulta-

dos de las evaluaciones, que incluye reuniones y talleres con los usuarios del programa, con el fin de transmitir los usos promovidos y no deseados, desde un enfoque formativo y de la toma de decisiones basada en datos (Abarzúa y Contreras, 2019).

Metodología

Para abordar la pregunta de investigación desde los significados y valoraciones que expresan los actores escolares (Newby, 2014), se concibió un estudio cualitativo, de alcance descriptivo, centrado en la utilización que distintas instituciones dan a la evaluación estandarizada SEPA, así como en las situaciones que favorecen o dificultan esos usos.

Instrumento y participantes

La recolección de datos se llevó a cabo a través de entrevistas semiestructuradas individuales orientadas a indagar sobre las valoraciones de las evaluaciones, la comprensión de los datos, así como su utilización y consecuencias en distintos contextos de aplicación. Las entrevistas se realizaron a autoridades municipales, coordinadores de programas educativos, directores y responsables pedagógicos de escuelas, dado que estos actores generalmente toman la decisión de contratar e implementar el programa y, por lo tanto, inciden en el tipo de uso que se dará a la evaluación.

Al interior de esta definición inicial de casos se realizó un muestreo determinado *a priori* (Flick, 2004), con base en características identificadas como relevantes de acuerdo con la configuración del programa SEPA y con el sistema educacional chileno, distribuyendo los casos dentro de la matriz que conforman dichas características:

Contratantes: Difieren en su complejidad y en los usos posibles que dan a las evaluaciones. Los perfiles presentes son una autoridad central que administra una red de centros, establecimientos individuales o programas de investigación/intervención en el sistema educativo.

Dependencia del establecimiento: Refiere a la administración y financiamiento de la escuela. En el sistema chileno existen escuelas públicas, privadas y subvencionadas (administración privada con financiamiento público). Al realizar este estudio, de los 234 establecimientos que utilizaron el programa SEPA, 51% eran públicos, 28% particulares subvencionados y 21% privados.

Ubicación geográfica: Chile se encuentra altamente centralizado en su capital, generando distinciones entre esta región y las restantes del país. Por esta razón, se incluyeron casos de ambas situaciones geográficas.

La tabla 1 muestra las anteriores características de acuerdo con los entrevistados.

TABLA 1
Usuarios del programa SEPA

Tipo contratante	Dependencia	Ubicación	Entrevistados
Administrador red de establecimientos	Pública	Capital	5
		Regiones	5
	Privada	Capital	1
		Regiones	4
Establecimiento individual	Pública	Capital	2
		Regiones	4
	Privada	Capital	2
		Regiones	2
	Subvencionada	Capital	4
		Regiones	2
Investigación/intervención	—	Capital	2
Total			33

Fuente: elaboración propia.

Dado que esta investigación fue desarrollada por la misma institución que implementa el programa, con el fin de favorecer la honestidad de los participantes y la apertura a las experiencias por parte de los entrevistadores, el trabajo de campo fue realizado por un equipo externo a la institución, el cual también participó del proceso de construcción de los guiones de entrevistas, así como en los análisis inductivos del material.

Adicionalmente, el estudio se sometió a la revisión del Comité de Ética de la Pontificia Universidad Católica de Chile, obteniendo su aprobación, lo que supuso contar con los elementos que garantizan la voluntariedad

en la participación y el respeto de los derechos de los participantes, tales como los consentimientos informados de los entrevistados y las autorizaciones de las instituciones.

Análisis de datos

Los datos, constituidos por transcripciones íntegras de cada entrevista, fueron sometidos a análisis cualitativos de contenido. En un primer momento el material fue abordado descriptivamente, generando conceptos y categorías (Dey, 1993) según los procedimientos de la *codificación abierta* propuesta por Strauss y Corbin (2002) y utilizando el programa Max.QDA versión 2018.2. En este proceso se identificaron elementos relevantes en la experiencia con el SEPA: definiciones del programa, componentes y finalidades, resultados esperados a corto y mediano plazos, comprensión de los datos y reportes, valoración del programa y usos de las pruebas, entre otros. Este artículo se focaliza específicamente en aquellos contenidos referidos al uso de esta evaluación.

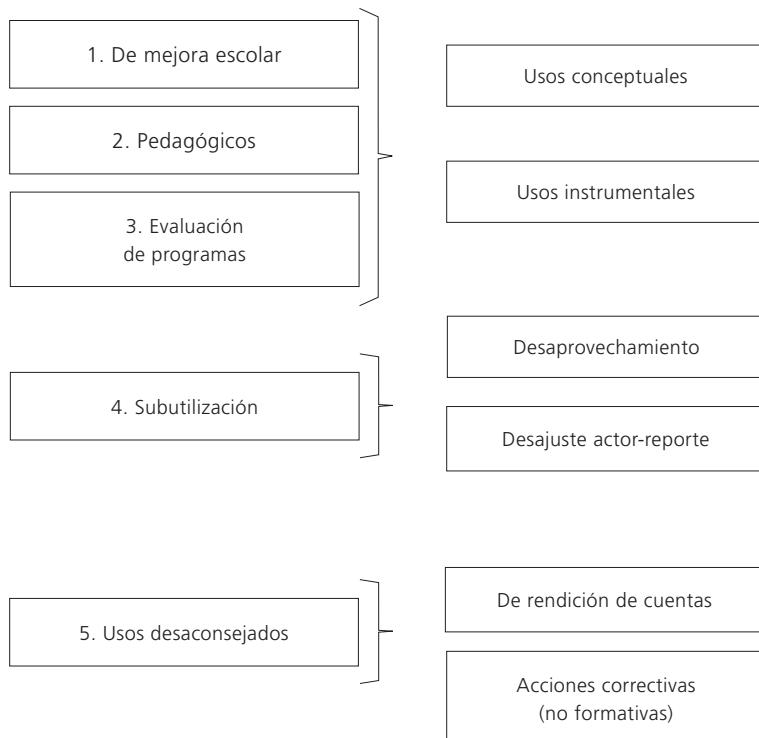
Posteriormente, se realizó una fase interpretativa que permitió establecer relaciones entre los datos y los antecedentes conceptuales mediante un proceso iterativo entre inducción y deducción, denominado *abducción* (Kennedy y Thornberg, 2018). Sin tener una hipótesis basada en una teoría, se disponía de literatura sobre clasificaciones de usos y definiciones de sus variantes. El proceso interpretativo consistió en revisar el listado de usos emergentes a partir de esas categorizaciones preexistentes. Como resultado, surgió una clasificación nueva basada en los datos obtenidos en este estudio que, cuando coincidieron con usos reportados en trabajos precedentes, mantuvieron las denominaciones más comúnmente utilizadas en la literatura. Una vez construida esta clasificación se procedió a hacer un análisis comparativo de las similitudes y diferencias de los usos entre los tres perfiles de actores y a identificar, descriptivamente, los obstáculos y facilitadores mencionados globalmente por los actores.

Resultados

Los resultados están organizados según las tres preguntas que guiaron la investigación: ¿cómo se utilizan los resultados de las evaluaciones SEPA? (figura 1), ¿difieren estos usos de acuerdo con los niveles de toma de decisión de sus usuarios? y, ¿qué condiciones favorecen u obstaculizan el uso efectivo de las evaluaciones SEPA?

FIGURA 1

Usos del programa SEPA declarados por equipos directivos de instituciones educacionales



Fuente: elaboración propia.

Usos de las evaluaciones SEPA

Usos orientados a la mejora escolar

Debido a que los participantes ejercen roles de dirección de instituciones educacionales, se esperaba que mencionaran usar las evaluaciones para la gestión escolar. Efectivamente, se relevan procesos reflexivos y acciones orientadas a la mejora de las escuelas. Por ejemplo, un uso de mejora escolar conceptual declarado fue la triangulación de información con el fin de monitorear la cobertura curricular. Adicionalmente, esta triangulación se realiza para aumentar el conocimiento sobre los estudiantes e identificar los factores que explican sus resultados de aprendizaje. Asimismo, se señala la importancia de basar las decisiones en más de una

fuente de información, siendo los resultados del programa SEPA, uno entre varios datos posibles.

Como usos conceptuales de mejora escolar también se pueden mencionar las situaciones en las cuales se realizan análisis de los datos a nivel de la escuela o de equipos de docentes según las asignaturas o etapas escolares. Estas sesiones incluyen desde simples comparaciones entre cursos y niveles, hasta instancias más reflexivas.

Los usos de mejora escolar instrumental basados en el sistema de evaluación hacen referencia a la definición de metas y contenidos en las planificaciones, programas de reforzamiento de contenidos o habilidades, asignación de recursos, definición de prioridades de apoyo externo o material didáctico, cambios en la estructura organizacional, entre otros:

Si necesito material tecnológico, si necesitamos docentes de apoyo, si necesitamos asistente de aula, tendría que ser tendiente a ir mirando cuáles son las debilidades que tenemos nosotros, entonces las acciones que debieran planificarse, debieran estar en absoluta correlación con los resultados que estamos teniendo, y que nos están diciendo con nombre y apellido “aquí está la falla” (Caso 34, directivo de escuela pública).

Usos pedagógicos

En las entrevistas también se hizo referencia a usos pedagógicos conceptuales e instrumentales. Con respecto al primero, destacaron los procesos de análisis de datos a nivel de aula, ya sea para identificar las fortalezas y debilidades del alumnado o para generar subgrupos en función de los resultados. También, se menciona la utilización de la evaluación como diagnóstico, con el fin de conocer el dominio de contenidos y habilidades del estudiantado al inicio del año escolar. Adicionalmente, los entrevistados señalan que se genera reflexión pedagógica a partir de los resultados, es decir, los docentes analizan y cuestionan sus prácticas, orientados a mejorar lo que no esté funcionando:

Lo usamos básicamente para analizarlo nosotros mismos, o sea jamás cuestionamos a un chiquillo, por ejemplo, no. Es para analizar nuestras prácticas pedagógicas, para ver cómo mejoramos, esas son como las ideas, o sea, analizamos estos resultados, [...], que nos sirve para tomar decisiones, entonces ¿cómo están nuestras prácticas pedagógicas? (Caso 22, directivo de escuela pública).

Los usos pedagógicos instrumentales se relacionan generalmente con la planificación de las clases. Los resultados se utilizan para establecer metas de aprendizaje, seleccionar contenidos y habilidades a desarrollar o seleccionar estrategias pedagógicas. En algunos casos, estas acciones se realizan basándose en agrupamientos de alumnos en función de sus resultados en SEPA y otras evaluaciones internas o externas.

Usos de evaluación de programas

En menor medida, los entrevistados mencionaron utilizar el programa SEPA para la evaluación de programas educativos. En general se refirieron a usos conceptuales, distinguiéndose los de tipo formativo, como aumentar el conocimiento de la población objetivo para retroalimentar y mejorar la intervención. Además, se mencionaron usos sumativos, referidos a la evaluación de la efectividad de programas a través de los resultados de aprendizaje de los estudiantes:

[...] Lo que permite [SEPA] es medir como no solo el estado sino como el progreso de los estudiantes..., este mismo grupo de escuelas lo vamos a volver a tomar en noviembre de este año, para ver... y a los mismos estudiantes que la aplicamos antes, o a los mismos cursos, para ver si es que se movilizaron o no los aprendizajes y cuánto se movilizaron (Caso 12, directivo de programa de intervención educativa).

Subutilización de los resultados de las evaluaciones

Esta categoría comprende las situaciones en las que, si bien existe la intención de usar las evaluaciones, diversas condiciones contextuales lo obstaculizan, tales como la falta de tiempo, materiales o competencias en los equipos pedagógicos. Los usuarios declaran no obtener el “máximo provecho” de una herramienta que les parece útil. Por ejemplo, no se logra generar un ciclo completo de uso de datos que vaya desde la interpretación a la planificación de acciones y a su evaluación. A veces se reciben los resultados y luego no se sabe bien qué hacer con esa información.

Cuando a uno le entregan un auto, por ejemplo, un último modelo ¿ya? qué lindo el auto, precioso, mire y le dicen: “esto te lleva para allá, te lleva para acá, tiene... todo este gasto en combustible”, qué se yo “ya, toma, aquí tienes las llaves”, y uno después no sabe cómo... cómo manejarse, cómo subirse al

auto, cómo encenderlo, un poco eso entiendo yo que es lo que pasa con la SEPA (Caso 32, sostenedor de escuelas públicas).

Otra situación que refleja una subutilización de la información es el *desajuste actor-reporte*, es decir, el informe disponible para un tipo de actor no es utilizado por este. Por ejemplo, cuando los directivos entregan los resultados a los docentes sin trabajarlos a nivel de escuela, o cuando los administradores de educación pública no utilizan los reportes de sus escuelas a nivel central y traspasan los datos directamente a los establecimientos escolares.

Usos desaconsejados

Esta última categoría reúne a las prácticas que no van en el sentido pedagógico ni de mejora escolar, pudiendo conllevar riesgos o consecuencias indeseadas, pero que no son alteraciones deliberadas de los procesos o resultados evaluativos –lo que sería un mal uso–. Este tipo de usos desaconsejados fueron minoritarios en el discurso de los participantes.

En primer lugar, se identifican usos de la evaluación que se asocian a culturas escolares de rendición de cuentas. Algunos de estos no necesariamente constituyen una práctica indeseada, sino que simplemente se plantean como una forma de transparencia y de responsabilidad frente a las decisiones y acciones emprendidas. Por ejemplo, se rinde cuentas de los resultados a instancias centrales y a financiadores. Sin embargo, también se mencionan aquellos menos alineados con lo pedagógico, como el entrenamiento para pruebas de altas consecuencias o la responsabilización de los docentes:

La usamos como insumo también para ver la dotación docente al año siguiente, porque si nos arroja que este curso en todos sus períodos que se le aplicó [...] vemos que no hay avance, mire, esta colega, que la teníamos por ejemplo en primero, le teníamos hartas expectativas, pero el instrumento externo no nos arroja buenos resultados, entonces quiere decir que ahí hay algo de formación, algo de la docente que no está dando los resultados, no ha sacado a los niños adelante como debe ser, ese es un tema, y el otro es que nos va diciendo cómo está usando... si está usando o no los planes y programas y qué estrategias está usando en clases. (Caso 33, directivo de escuela pública).

Por último, se mencionan prácticas que, si bien tienen la intención de orientarse a la mejora escolar, son correctivas más que formativas y podrían tener efectos negativos en los actores. Por ejemplo, pedir a los docentes que analicen los resultados de SEPA durante sus vacaciones u obligar a los estudiantes a asistir a talleres de las asignaturas evaluadas en su tiempo libre. Un taller de reforzamiento puede ofrecer oportunidades adicionales de aprendizaje a los estudiantes que más lo requieren y, en ese sentido, ser una estrategia orientada a la mejora, sin embargo, desde la perspectiva de los estudiantes, la obligación a asistir a clases después de una extensa jornada escolar puede ser percibida como punitiva. Debido a que este estudio no recoge la opinión de los estudiantes no es factible afirmar cuál es su postura, aunque es posible señalar que representa riesgos y que la distinción entre un uso formativo o correctivo estará dada por la manera en que sea implementado y presentado a los estudiantes.

Usos según el nivel de toma de decisión de los contratantes

Entendiendo que existe diversidad de usuarios del programa SEPA, tanto en su complejidad como en el nivel de toma de decisiones en que se sitúan, se indagó si existían usos diferenciados entre los tres perfiles de contratantes: administradores de servicios educativos, equipos directivos y programas educativos de intervención.

La utilización orientada a la gestión y mejora escolar es mencionada por los tres tipos de actores, aunque con características propias. Por ejemplo, solo los sostenedores y los programas señalan usos, generalmente conceptuales, que comparan de forma simultánea varios establecimientos para identificar sus avances, mientras que los directivos buscan comparar subgrupos (cursos, alumnos rezagados, etc.) dentro de sus establecimientos.

A su vez, los tres perfiles señalan explícitamente usos de carácter pedagógico, tanto instrumentales como conceptuales, sin embargo, difiere la profundidad e involucramiento en ellos. En concreto, los equipos directivos describen detalladamente este tipo de usos, mientras que los representantes de programas educativos, si bien los mencionan, no los ubican dentro de sus intereses principales. Por su parte, los administradores refieren este tipo de usos con mayor lejanía a las acciones específicas de cada escuela y, en varios casos, bajo una delegación en sus equipos directivos.

Un uso efectivamente diferenciador es el de rendición de cuentas. Mientras que es una práctica implementada por los programas educativos

con el fin de acreditar sus logros, es mencionada excepcionalmente por los equipos directivos de las escuelas que se definen a sí mismas como de alto desempeño. Respecto de los administradores, contrario a lo que se podría esperar, no se utiliza el programa SEPA para pedir rendición de cuentas a las escuelas. Esta situación solo se planteó en un caso como algo deseable y no factual.

Finalmente, los administradores de escuelas públicas son los actores que más abiertamente reconocen que subutilizan el programa SEPA en su rol de agencias centrales, al no establecer estrategias o usos propios a sus funciones. En efecto, señalan que no es una herramienta aprovechable en su nivel organizacional y que prefieren delegar a las escuelas la responsabilidad de dar buen uso a las evaluaciones.

[...] como unidad técnico pedagógica, dirección del DAEM [Departamento de Administración de Educación Municipal], hemos sido débiles en ese aspecto o no hemos tenido la assertividad como para hacerlo, como para generar o definir, diseñar un mecanismo que nos permita retroalimentarnos y poder evaluar nosotros a los colegios, en ese aspecto. (Caso 32, Sostenedor Público, 36).

Condiciones que favorecen u obstaculizan el uso efectivo de las evaluaciones SEPA
A pesar de que el objetivo inicial del estudio no consistió en identificar las condiciones que facilitan u obstaculizan el uso de las evaluaciones, este fue un aspecto que emergió espontáneamente en las entrevistas, especialmente cuando los participantes explicaban la subutilización del programa SEPA. Las respuestas se agruparon en tres condiciones: organizacionales, individuales y de los datos.

A nivel organizacional, una barrera importante para el uso efectivo de los datos es la falta de tiempo de los equipos y la alta carga administrativa. En estos casos se reconoce la importancia y el deseo de realizar análisis de la información, sin embargo, los espacios para reunirse son limitados y la posibilidad de los líderes escolares de dar un apoyo pedagógico se ve relegada frente a las tareas administrativas:

Tenemos que decir la verdad, no la explotamos todo lo que la deberíamos debido a la falta de tiempo ¿ya? en cuanto al profesorado. Nosotros tratamos de aprovechar al máximo el tiempo, pero el ministerio de educación [...] nos bombardea con tanto trabajo administrativo, que en este momento la parte

técnica muchas veces queda relegada a segunda instancia, entonces no es que los profesores no tengan los medios, falta muchas veces el tiempo (Caso 23, sostenedor de escuelas privadas).

Como aspectos organizacionales que facilitan el uso de las evaluaciones se mencionan la existencia de una cultura de uso de datos, la instalación de esta práctica en la rutina de la organización, la existencia de espacios y el trabajo colaborativo en torno a los datos:

[...] yo te diría que esos son como los tres mínimos que se use la información..., o sea que se entienda que la información es para ocuparla, que [...] haya una lectura un poco más compleja de los resultados, en términos de que uno puede analizar no solamente qué áreas... qué está bien, qué está mal, dónde estamos fuerte, dónde..., sino que se pueda vincular con las prácticas y con las prácticas institucionales, y que existan espacios donde esa toma de decisiones se haga a nivel institucional (Caso 9, sostenedor de escuelas públicas).

Las características individuales de los usuarios son evocadas en menor medida. Solo se mencionan como obstaculizadores la falta de competencias para la interpretación y el análisis de los datos en algunos docentes, así como una cultura del profesorado poco habituado a concebir acciones concretas a partir de los resultados de las evaluaciones:

Nos ha costado más hacer lo que viene después, como que efectivamente usen los resultados para tomar decisiones pedagógicas y tengan un plan de acción, como lo que viene después de levantar evidencias (ajá), pero eso nos ha costado con todos los instrumentos, no, es como más cultural de los profesores mismos, creemos nosotros (Caso 10, directivo de programa de intervención educativa).

Finalmente, algunas características del diseño, reportes de resultados y acompañamiento de los usuarios del programa SEPA actuarían como obstaculizadores o facilitadores de su uso efectivo. Las condiciones desfavorables son, por ejemplo, que los resultados no se entregan de forma oportuna y a veces están retrasados respecto de los procesos de planificación escolar. También se menciona la dificultad de analizar los resultados porque es demasiada información para procesar o porque algunos reportes son de más

difícil comprensión (por ejemplo, el progreso o el valor agregado). Ahora bien, estas opiniones son minoritarias y lo que predomina es una valoración generalizada del tipo de información que entregan las evaluaciones, de su confiabilidad, validez, nivel de detalle y accesibilidad.

Por último, algunos participantes consideran que el acompañamiento que realiza el programa mediante reuniones y charlas les ha facilitado la comprensión y el uso de datos. No obstante, algunos entrevistados no lo han utilizado por falta de tiempo o planificación y, en otros casos, el uso de videoconferencias de acompañamiento, dada la lejanía territorial, no les fue del todo satisfactorio.

Discusión y conclusiones

Este trabajo permitió describir cómo utilizan una evaluación estandarizada de aprendizaje, de bajas consecuencias, los líderes escolares y directivos de servicios y programas educacionales chilenos. Además, se presentaron algunas diferencias en los usos declarados según el nivel de toma de decisión (escuela, administración de escuelas y programas de intervención) y, por último, se identificaron las condiciones facilitadoras y obstaculizadoras de la utilización efectiva de las evaluaciones.

La metodología cualitativa escogida, si bien se adaptó al objetivo de comprender los usos de las evaluaciones y sus contextos, presenta algunas limitaciones. Por un lado, la posibilidad de generalización de los hallazgos a otros grupos y poblaciones es limitada. Por otro lado, para estudiar los usos de las evaluaciones al interior del aula es esencial acceder a la visión de los docentes, sin embargo, en este estudio solo fue posible incluir a los directivos de instituciones educativas, quienes nos ofrecen un acercamiento inicial a estos usos pedagógicos. Por esta razón, actualmente se está desarrollando una segunda investigación que aborda específicamente los usos en el aula desde la perspectiva docente.

Más allá de esas limitaciones, este estudio ofrece importantes hallazgos sobre los usos de las evaluaciones estandarizadas. Primero, fue posible evidenciar que los actores entrevistados utilizan las evaluaciones de formas variadas, predominando usos pedagógicos y de mejora escolar que son altamente promovidos en la literatura especializada y por el programa SEPA. Ahora bien, pese a que se evocaron procesos reflexivos y de toma de decisiones a partir de los datos, también se expresó la dificultad para comprometerse en un ciclo completo de uso de datos.

La información disponible plantea que el uso de datos y evaluaciones que efectivamente propicia la mejora escolar no se reduce a la acción aislada de aplicar un instrumento de evaluación y analizar sus resultados, sino que debe formar parte de un proceso con diversos pasos, incluyendo la selección de un propósito, recolección de datos, interpretación, acción y evaluación de las iniciativas implementadas (Coburn y Turner, 2011; Kippers *et al.*, 2018; Schildkamp, 2019; Schildkamp, Smit y Blossig, 2017). Al respecto, los entrevistados describen que, por distintas razones, las evaluaciones se limitan a una recolección de información desconectada de un propósito de mejora y que, en muchos casos, finaliza con el análisis de resultados sin conducir a acciones concretas. Por lo tanto, se observa la dificultad para hacer análisis profundos de la información, realizar procesos reflexivos sistemáticos, seguimiento de las acciones o involucrar a toda la comunidad escolar.

Dentro de este escenario, se destaca el caso de los administradores de educación pública como los actores que menos utilizaban las evaluaciones, siendo común traspasar esta responsabilidad a las escuelas sin mediar análisis ni seguimiento desde el nivel central. Además, es relevante señalar que, si bien predomina una cultura de rendición de cuentas en el sistema escolar chileno, no se menciona este tipo de uso del programa SEPA entre las escuelas y las administraciones públicas de educación.

Esto se podría explicar porque el SEPA es un programa de bajas consecuencias y sus usuarios han integrado su discurso sobre el uso formativo de las evaluaciones. Sin embargo, también puede deberse a cambios que trascienden al programa y que están mostrando una tendencia, al menos discursivamente, hacia una mayor valoración de las evaluaciones para la mejora en detrimento de la lógica de rendición de cuentas. En Chile esta tendencia podría derivar del descontento frente al uso excesivo de evaluaciones de altas consecuencias, que llegó hasta la constitución de un movimiento de intelectuales y de la sociedad civil que exigía el “alto al Simce”,² debido a los efectos nocivos que estaría provocando esta lógica evaluativa en las comunidades escolares (Pino-Yancovic, Oyarzún y Salinas, 2016). Adicionalmente, desde la Agencia de Calidad, responsable de las evaluaciones nacionales, se ha promovido paulatinamente un uso más pedagógico de las evaluaciones y se han ampliado los indicadores de calidad con el fin de mitigar algunos efectos no deseados (Agencia de Calidad de la Educación, 2018).

No obstante lo anterior, se identifican prácticas y usos del programa SEPA que se asocian a esta cultura de rendición de cuentas, como la responsabilización de docentes o la preparación de evaluaciones con altas consecuencias. Si bien parecen ser usos minoritarios, representan un desafío que no se debe desestimar debido a que una evaluación de bajas consecuencias puede transformarse en una de altas consecuencias si los usuarios le atribuyen incentivos o sanciones. En efecto, el carácter de “bajas” y “altas” consecuencias no debiese entenderse como una propiedad intrínseca de la evaluación, sino como un aspecto a indagar en el uso que le dan los actores del mundo escolar (Tannenbaum y Kane, 2019).

La subutilización también representa un desafío importante para la política y práctica educativas. Dado que el uso de datos para fines pedagógicos y de mejora escolar tiene efectos positivos en las instituciones y los aprendizajes de los estudiantes (Carlson, Borman y Robinson, 2011; Ebbeler *et al.*, 2016, 2017; Van Geel, Keuning, Visscher y Fox, 2016) es importante comprender bajo qué condiciones esos usos son posibles y qué factores los obstaculizan, con el fin de maximizar los recursos y los efectos deseados.

Si bien este estudio no tuvo como objetivo inicial identificar las condiciones que facilitan y obstaculizan el uso de las evaluaciones, algunas fueron mencionadas espontáneamente en las entrevistas. Así, coincidiendo con lo observado en otros contextos, condiciones organizacionales como una cultura de uso de datos, la instalación de esta práctica en la rutina de la organización, la existencia de espacios y el trabajo colaborativo fueron aspectos considerados cruciales para el uso efectivo de las evaluaciones. En menor medida se mencionaron características del usuario, como las competencias en estadística y la capacidad para pasar de los datos a la toma de decisiones pedagógicas. Respecto de las características del dato, se mencionaron términos como la confiabilidad, validez, objetividad, pero por, sobre todo, lo que parece facilitar el uso pedagógico es el detalle de la información, que sea oportuna y entregue resultados a nivel individual.

Los entrevistados que señalaron una subutilización de las evaluaciones SEPA y el deseo de lograr un mejor aprovechamiento de este instrumento consideraron que se debía principalmente a la falta de una o algunas de las condiciones organizacionales mencionadas. Por ejemplo, los equipos directivos explicaron que la sobrecarga de trabajo administrativo les impedía liderar un proceso reflexivo más profundo en torno a los resultados

de las evaluaciones; del mismo modo, en las administraciones municipales de educación se hizo referencia a la falta de un acompañamiento a las escuelas en el uso de datos. Como se explicó en los antecedentes de este trabajo, la evidencia muestra que la subutilización de las evaluaciones no es una excepción y que, incluso en países donde el uso de datos está más instalado en los equipos pedagógicos, persisten las dificultades para tomar decisiones basadas en hallazgos de investigaciones y evaluaciones (Ebbeler *et al.*, 2017; Gelderblom *et al.*, 2016; Prenger y Schildkamp, 2018; Van Gasse *et al.*, 2017).

Por otro lado, los trabajos que recomiendan un ciclo completo de uso de datos (Coburn y Turner, 2011; Kippers *et al.*, 2018; Schildkamp, 2019; Schildkamp, Smit y Blossing, 2017) habitualmente se han realizado en sistemas escolares o programas donde existe una estructura organizacional que lo permite, ya sea un profesional especializado o un equipo de uso de datos. En Chile, pese a la relevancia que han adquirido las evaluaciones estandarizadas, no existen instancias como estas ni la capacitación adecuada de los docentes.

Estos hallazgos tienen implicancias importantes para la política y práctica educativas. Por un lado, relevan la necesidad de instalar competencias en el uso de datos en los equipos pedagógicos generando instancias de formación inicial y continua. Si no es posible disponer de profesionales especializados en el uso de datos en el corto plazo, es esencial que los equipos pedagógicos, que ya son usuarios de las evaluaciones, accedan a una formación de calidad sobre este tema. De hecho, el programa SEPA ofrece un acompañamiento a sus usuarios en el que promueve el uso formativo de la evaluación en el marco de un ciclo completo de uso de datos. Sin embargo, no es usual que los programas evaluativos brinden instancias formativas a los usuarios, a pesar de que los estándares internacionales sugieren que los programas se responsabilicen por los usos y consecuencias de las evaluaciones que diseñan (AERA, APA y NCME, 2018).

En Chile no se dispone de evidencia respecto de la formación inicial y continua en el uso de evaluaciones. No obstante, el Sistema Nacional de Evaluación Docente, basado en los estándares oficiales de desempeño docente a los cuales debiesen alinearse los programas de formación inicial, demuestra que el indicador *reflexión a partir de las evaluaciones* es un aspecto de la práctica que sistemáticamente aparece menos desarrollado en los docentes evaluados (Centro de Perfeccio-

namiento, Experimentación e Investigaciones Pedagógicas, 2016; Sun, Correa, Zapata y Carrasco, 2011).

Por otro lado, sería relevante intervenir para generar las condiciones organizacionales necesarias para el uso de datos. Además de los tiempos, se requiere instalar procesos reflexivos colectivos que se orienten a analizar los datos desde una mirada pedagógica y no desde el pensamiento del déficit, que se focaliza en las falencias y las explica invocando factores extraescolares (Lasater, Bengtson y Albiladi, 2020). Adicionalmente, sería beneficioso que los equipos pedagógicos dispongan de instancias formativas sobre el ciclo completo en el uso de datos, lo que no solo aportaría al uso de las evaluaciones, sino a los procesos de gestión curricular en general.

Algunas políticas educativas recientes podrían tener efectos positivos en este sentido. Por ejemplo, la Ley de Desarrollo Profesional Docente (de 2016) reduce las horas lectivas de los docentes y promueve ampliamente el trabajo colaborativo, lo que constituye una oportunidad para instalar formalmente el uso formativo de las evaluaciones y disminuir el énfasis en la rendición de cuentas. Por otro lado, la creación de centros de liderazgo escolar se vislumbra como una posibilidad de influir en la política pública y en las escuelas y que podría reforzar las competencias de liderazgo necesarias para una orientación hacia la mejora escolar (Pino-Yancovic y Ahumada, 2020).

Notas

¹ En adelante, en este artículo se usará el masculino con el único objetivo de hacer más fluida la lectura, sin menoscabo de género.

² Simce: Sistema de la Medición de la Calidad de la Educación, prueba estandarizada de altas consecuencias del Estado Chileno.

Referencias

- Abarzúa, Andrea y Contreras, Johana (2019). “Sistema de Evaluación de Progreso del Aprendizaje, SEPA: Evidencia de su confiabilidad y validez”, en J. Manzi, M. García y S. Taut (eds.), *Validez de evaluaciones educacionales en Chile y Latinoamérica, Santiago, Chile*: Ediciones UC, pp. 127-156.
- Abrams, Lisa; Varier, Divya y Mehdi, Tareque (2020). “The intersection of school context and teachers’ data use practice: Implications for an integrated approach to capacity building”, *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2020.100868>
- AERA, APA y NCME (2018). *Estándares para pruebas educativas y psicológicas*, Washington, D.C.: American Educational Research Association/American Psychological Association/National Council for Measurement in Education.
- Agencia de Calidad de la Educación (2018). *Nuevo sistema nacional de evaluación de aprendizajes. La evaluación al servicio de los aprendizajes*, Santiago, Chile: Agencia

- de Calidad de la Educación. Disponible en: http://archivos.agenciaeducacion.cl/Sistema_Nacional_de_Evaluacion_17abr.pdf
- Alkin, Marvin y Taut, Sandy (2002). "Unbundling evaluation use", *Studies in Educational Evaluation*, vol. 29, núm. 1, pp. 1-12. [https://doi.org/10.1016/S0191-491X\(03\)90001-0](https://doi.org/10.1016/S0191-491X(03)90001-0)
- Alkin, Marvin y King, Jean (2016). "The historical development of evaluation use", *The American Journal of Evaluation*, vol. 37, núm. 4, pp. 568-579. <https://doi.org/10.1177/1098214016665164>
- Alkin, Marvin y King, Jean (2017). "Definitions of evaluation use and misuse, evaluation influence, and factors affecting use", *American Journal of Evaluation*, vol. 38, núm. 3, pp. 434-450. <https://doi.org/10.1177/1098214017717015>
- Beaver, Jessica y Weinbaum, Elliot (2015). "State test data and school improvement efforts", *Educational Policy*, vol. 29, núm. 3, pp. 478-503. <https://doi.org/10.1177/0895904813510774>
- Bolhuis, Erik (2019). "The development of data use, data skills, and positive attitude towards data use in a data team intervention for teacher educators", *Studies in Educational Evaluation*, vol. 60, pp. 99-108. <https://doi.org/10.1016/j.stueduc.2018.12.002>
- Braun, Henry y Singer, Judith (2019). "Assessment for monitoring of education systems: International comparisons", *Annals of the American Academy of Political and Social Science*, vol. 683, núm. 1, pp. 75-92. <https://doi.org/10.1177/0002716219843804>
- Brewer, Curtis; Knoepfle, Robert y Lindle, Jane (2015). "Consequential validity of accountability policy: public understanding of assessments", *Educational Policy*, vol. 29, núm. 5, pp. 711-745. <https://doi.org/10.1177/0895904813518099>
- Carlson, Deven; Borman, Geoffrey y Robinson, Michelle (2011). "A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement", *Educational Evaluation and Policy Analysis*, vol. 33, núm. 3, pp. 378-398. <https://doi.org/10.3102/0162373711412765>
- Centro de Perfeccionamiento, Experimentación e Investigaciones Pedagógicas (2016). *Resultados de evaluación docente 2015*, Santiago, Chile: Ministerio de Educación. Disponible en: https://www.docentemas.cl/docs/Resultados_Evaluacion_Docente_2015.pdf
- Chalhoub-Deville, Micheline (2016). "Validity theory: Reform policies, accountability testing, and consequences", *Language Testing*, vol. 33, núm. 4, pp. 453-472. <https://doi.org/10.1177/0265532215593312>
- Coburn, Cynthia y Turner, Erica (2011). "Research on data use: A framework and analysis", *Measurement: Interdisciplinary Research and Perspectives*, vol. 9, núm. 4, pp. 173-206. <https://doi.org/10.1080/15366367.2011.626729>
- Datnow, Amanda y Park, Vicki (2018). "Opening or closing doors for students? Equity and data use in schools", *Journal of Educational Change*, vol. 19, núm. 2, pp. 131-152. <https://doi.org/10.1007/s10833-018-9323-6>
- Datnow, Amanda; Lockton, Marie y Weddle, Hayley (2020). "Capacity building to bridge data use and instructional improvement through evidence on student thinking", *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2020.100869>

- Dey, Ian (1993). *Qualitative data analysis. A user-friendly guide for social scientists*, Londres: Routledge.
- Dunn, Keree; Airola, Denise; Lo, Wen-Juo y Garrison, Mickey (2013). "What teachers think about what they can do with data: Development and validation of the data driven decision-making efficacy and anxiety inventory", *Contemporary Educational Psychology*, vol. 30, núm. 1, pp. 87-98. <https://doi.org/10.1016/j.cedpsych.2012.11.002>
- Ebbeler, Johanna; Poortman, Cindy; Schildkamp, Kim y Pieters, Jules (2016). "Effects of a data use intervention on educators' use of knowledge and skills", *Studies in Educational Evaluation*, vol. 48, pp. 19-31. <https://doi.org/10.1016/j.stueduc.2015.11.002>
- Ebbeler, Johanna; Poortman, Cindy; Schildkamp, Kim y Pieters, Jules (2017). "The effects of a data use intervention on educators' satisfaction and data literacy", *Educational Assessment, Evaluation and Accountability*, vol. 29, núm. 1, pp. 83-105. <https://doi.org/10.1007/s11092-016-9251-z>
- Elacqua, Gregory; Martínez, Matías; Santos, Humberto y Urbina, Daniela (2016). "Short-run effects of accountability pressures on teacher policies and practices in the voucher system in Santiago, Chile", *School Effectiveness and School Improvement*, vol. 27, núm. 3, pp. 385-405. <https://doi.org/10.1080/09243453.2015.1086383>
- Falabella, Alejandra (2016). "¿Qué aseguran las políticas de aseguramiento de la calidad? Un estudio de casos en distintos contextos escolares", *Estudios Pedagógicos*, vol. 42, núm. 1, pp. 107-126. <https://doi.org/10.4067/S0718-07052016000100007>
- Falabella, Alejandra (2020). "The ethics of competition: accountability policy enactment in Chilean schools' everyday life", *Journal of Education Policy*, vol. 35, núm. 1, pp. 23-45. <https://doi.org/10.1080/02680939.2019.1635272>
- Farley-Ripple, Elizabeth; Jennings, Austin y Buttram, Joan (2019). "Toward a framework for classifying teachers' use of assessment data", *AERA Open*, vol. 5, núm. 4. <https://doi.org/10.1177/2332858419883571>
- Flick, Uwe (2004). *Introducción a la investigación cualitativa*, Madrid: Ediciones Morata.
- Gelderblom, Gert; Schildkamp, Kim; Pieters, Jules y Ehren, Melanie (2016). "Data-based decision making for instructional improvement in primary education", *International Journal of Educational Research*, vol. 80, pp. 1-14.
- Hein, Andreas y Taut, Sandy (2010). "El uso de información evaluativa externa con fines formativos: el caso de establecimientos educacionales chilenos participantes de Sepa", *Revista Iberoamericana de Evaluación Educativa*, vol. 3, núm. 2, pp. 160-181.
- Hitchcock, John; Onwuegbuzie, Anthony y Khoshaim, Heba (2015). "Examining the consequential validity of standardized examinations via public perceptions: a review of mixed methods survey design considerations", *International Journal of Multiple Research Approaches*, vol. 9, núm. 1, pp. 24-39. <https://doi.org/10.1080/18340806.2015.1076757>
- Jimerson, Jo (2016). "How are we approaching data-informed practice? Development of the Survey of Data Use and Professional Learning", *Educational Assessment, Evaluation and Accountability*, vol. 28, núm. 1, pp. 61-87. <https://doi.org/10.1007/s11092-015-9222-9>

- Jimerson, Jo; Garry, Vanessa; Poortman, Cindy y Schildkamp, Kim (2020). "Implementation of a collaborative data use model in a United States context", *Studies in Educational Evaluation*. <https://doi.org/10.1016/j.stueduc.2020.100866>
- Kane, Michaelk (2016). "Explicating validity", *Assessment in Education: Principles, Policy and Practice*, vol. 23, núm. 2, pp. 198-211. <https://doi.org/10.1080/0969594X.2015.1060192>
- Kennedy, Brianna y Thornberg, Robert (2018). "Deduction, induction, and abduction", en U. Flick (ed.), *The SAGE Handbook of Qualitative Data Collection*, Londres: SAGE Publications.
- Kippers, Wilma; Poortman, Cindy; Schildkamp, Kim y Visscher, Adrie (2018). "Data literacy: What do educators learn and struggle with during a data use intervention?", *Studies in Educational Evaluation*, vol. 56, pp. 21-31. <https://doi.org/10.1016/j.stueduc.2017.11.001>
- Koretz, Daniel (2017). *The testing charade. Pretending to make schools better*, Chicago: University of Chicago Press.
- Lane, Suzanne (2014). "Evidencia de validez basada en las consecuencias del uso del test", *Psicothema*, vol. 26, núm. 1, pp. 127-135. <https://doi.org/10.7334/psicothema2013.258>
- Lasater, Kara; Bengtson, Ed y Albiladi, Waheeb (2020). "Data use for equity?: How data practices incite deficit thinking in schools", *Studies in Educational Evaluation*, vol. 69 <https://doi.org/10.1016/j.stueduc.2020.100845>
- Loeb, Susanna y Byun, Erika (2019). "Testing, accountability, and school improvement", *The ANNALS of the American Academy of Political and Social Science*, vol. 683, núm. 1, pp. 94-109. <https://doi.org/10.1177/0002716219839929>
- Mandinach, Ellen y Gummer, Edith (2016). "What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions", *Teaching and Teacher Education*, vol. 60, pp. 366-376. <https://doi.org/10.1016/j.tate.2016.07.011>
- Mandinach, Ellen y Jimerson, Jo (2016). "Teachers learning how to use data: A synthesis of the issues and what is known", *Teaching and Teacher Education*, vol. 60, pp. 452-457. <https://doi.org/10.1016/j.tate.2016.07.009>
- Manzi, Jorge; Bogolasky, Francisca; Gutiérrez, Gabriel; Grau, Valeska y Volante, Paulo (2014). *Ánalisis sobre valoraciones, comprensión y uso del SIMCE por parte de directores escolares de establecimientos subvencionados*, Santiago, Chile: Fondo de Investigación y Desarrollo en Educación.
- Maroy, Christian y Dutercq, Yves (2017). "Le travail et le professionnalisme enseignants face aux politiques de responsabilisation", en C. Maroy y Y. Dutercq (eds.) *Professionnalisme Enseignant et Politiques de Responsabilisation*, París: Deboek Superieur.
- Martínez-Rizo, Felipe (2016). "Impacto de las pruebas en gran escala en contextos de débil tradición técnica: experiencia de México y el Grupo Iberoamericano de PISA", *RELIEVE*, vol. 22, núm. 1, pp. 1-12. <https://doi.org/10.7203/relieve.22.1.8244>
- Messick, Samuel (1980). "Test Validity and the Ethics of Assessment", *American Psychologist*, vol. 35, núm. 11, pp. 1012-0127.
- Mons, Nathalie (2009). "Effets théoriques et réels des politiques d'évaluation standardisée", *Revue Française de Pédagogie*, vol. 169. <https://doi.org/10.4000/rfp.1531>

- Newby, Peter (2014). *Research methods for education, second edition*, Londres: Taylor and Francis Group.
- Parra, Victoria y Matus, Gladys (2016). “Usos de datos y mejora escolar: una aproximación a los sentidos y prácticas educativas subyacentes a los procesos de toma de decisiones”, *Calidad en la Educación*, vol. 45, pp. 207-250. <https://doi.org/10.4067/s0718-45652016000200007>
- Patton, Michael (1997). *Utilization-focused evaluation: The new century text*, 3^a ed., Thousand Oaks: Sage Publications.
- Pino-Yancovic, Mauricio; Oyarzún, Gonzalo y Salinas, Iván (2016). “Crítica a la rendición de cuentas: narrativa de resistencia al sistema de evaluación en Chile”, *Cadernos CEDES*, vol. 36, núm. 100, pp. 337-354. <https://doi.org/10.1590/CC0101-32622016171362>
- Pino-Yancovic, Mauricio y Ahumada, Luis (2020). “Collaborative inquiry networks: The challenge to promote network leadership capacities in Chile”, *School Leadership and Management*, vol. 40, núms. 2-3, pp. 221-242.
- Prenger, Rilana y Schildkamp, Kim (2018). “Data-based decision making for teacher and student learning: a psychological perspective on the role of the teacher”, *Educational Psychology*, vol. 38, núm. 6, pp. 734-752. <https://doi.org/10.1080/01443410.2018.1426834>
- Schildkamp, Kim (2019). “Data-based decision-making for school improvement: Research insights and gaps”, *Educational Research*, vol. 61, núm. 3, pp. 257-273. <https://doi.org/10.1080/00131881.2019.1625716>
- Schildkamp, Kim; Poortman, Cindy; Luyten, Hans y Ebbeler, Johanna (2016). “Factors promoting and hindering data-based decision making in schools”, *School Effectiveness and School Improvement*, vol. 289, núm. 2, pp. 242-258. <https://doi.org/10.1080/09243453.2016.1256901>
- Schildkamp, Kim; Smit, Maaike y Blossing, Ulf (2017). “Professional development in the use of data: From data to knowledge in data teams”, *Scandinavian Journal of Educational Research*, vol. 63, núm. 3, pp. 393-411. <https://doi.org/10.1080/00313831.2017.1376350>
- Smith, William (2014). “The global transformation toward testing accountability”, *Education Policy Analysis Archives*, vol. 22, núm. 116, pp. 1-34. <https://doi.org/10.14507/epaa.v22.1571>
- Strauss, Anselm y Corbin, Juliet (2002). *Bases de la investigación cualitativa. Técnicas y procedimientos para desarrollar la teoría fundamentada*, Antioquia: Editorial Universidad de Antioquia.
- Sun, Yulan; Correa, Mónica; Zapata, Álvaro y Carrasco, Diego (2011). “Resultados: qué dice la evaluación docente acerca de la enseñanza en Chile”, en J. Manzi, R. González y Y. Sun (eds.), *La evaluación docente en Chile*, Santiago: Centro de Medición MIDE UC. Disponible en: https://www.mideuc.cl/libroed/pdf/La_Evaluacion_Docente_en_Chile.pdf
- Tannenbaum, Richard y Kane, Michael (2019). “Stakes in testing: Not a simple dichotomy but a profile of consequences that guides needed evidence of measurement quality”, *ETS Research Report Series*, vol. 2019, núm. 1, pp. 1-16. <https://doi.org/10.1002/ets2.12255>

- Taut, Sandy; Cortés, Flavio; Sebastian, Christian y Preiss, David (2009). "Evaluating school and parent reports of the national student achievement testing system (SIMCE) in Chile: Access, comprehension, and use", *Evaluation and Program Planning*, vol. 32, núm. 2, pp. 129-137. <https://doi.org/10.1016/j.evalprogplan.2008.10.004>
- Van Gasse, Roos; Vanlommel, Kristin; Vanhoof Jan y Van Petegem, Peter (2017). "The impact of collaboration on teachers' individual data use", *School Effectiveness and School Improvement*, vol. 28, núm. 3, pp. 489-504. <https://doi.org/10.1080/09243453.2017.1321555>
- Van Geel, Marieke; Keuning, Trynke; Visscher, Adrie y Fox, Jean-Paul (2016). "Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools", *American Educational Research Journal*, vol. 53, núm. 2, pp. 360-394. <https://doi.org/10.3102/0002831216637346>

Artículo recibido: 7 de diciembre de 2020

Dictaminado: 26 de mayo de 2021

Segunda versión: 21 de junio de 2021

Tercera versión: 25 de junio de 2021

Aceptado: 1 de julio de 2021