

# Las evaluaciones educacionales estandarizadas desde la experiencia de los actores<sup>1</sup>

Johana Contreras<sup>2</sup>

ORCID: 0000-0001-6081-0476

Álex Torres<sup>2</sup>

ORCID: 0000-0003-4501-9453

## Resumen

Este artículo presenta un estudio que aborda, desde la sociología de la experiencia, la problemática de la relación entre los actores del mundo escolar y los dispositivos de evaluación estandarizada. Pese a las críticas que estos instrumentos han suscitado en la sociedad chilena, existen directivos de instituciones educacionales dispuestos a implementarlos voluntariamente. El objetivo de este estudio fue indagar en las razones y sentidos que ellos le dan a un programa de evaluaciones de bajas consecuencias desarrollado por una universidad de prestigio en Chile y orientado a medir el desempeño y progreso de estudiantes en relación con el currículum nacional. El diseño contempló la ejecución de una investigación cualitativa en base a entrevistas semiestructuradas a directivos de establecimientos y administradores de redes educacionales públicas y privadas que utilizan el programa de evaluación, seleccionados de acuerdo a criterios de muestro intencionado (N=33). La información fue analizada en base a lineamientos de la *Grounded Theory* y, posteriormente, interpretada desde la sociología de la experiencia. Los resultados revelaron la presencia de lógicas de la acción heterogéneas (integración, estratégica y subjetivación) subyacentes a las decisiones y usos de las evaluaciones, cuyas posibilidades de articulación recaen en los actores y dependen en gran medida de las condiciones organizacionales de los establecimientos, dando lugar a distintas experiencias en relación a las evaluaciones estandarizadas. Este estudio identificó tres: la experiencia academicista, de adhesión relativa y la subutilización de las evaluaciones. Se concluye señalando las implicancias de este estudio para la investigación, política y práctica educativa.

## Palabras clave

Evaluación estandarizada – Experiencia social – Usos de la evaluación.

**1-** Disponibilidad de datos: el conjunto de datos de este estudio no está disponible públicamente, debido a consideraciones de confidencialidad hacia los participantes. La solicitud de acceso puede ser hecha directamente a los autores al mail: jtcontre@uc.cl

**2-** Pontificia Universidad Católica de Chile, Santiago, Chile. Contacto: jtcontre@uc.cl; atorresp@uc.cl



<https://doi.org/10.1590/S1678-4634202349248451es>

This content is licensed under a Creative Commons attribution-type BY 4.0.

## ***Standardized educational evaluations from the actors' experience***

### **Abstract**

*This article presents a study that approaches, from the sociology of experience, the problematic of the relationship between the actors of the school context and the standardized evaluation devices. Despite the criticisms that these instruments have raised in Chilean society, there are school principals and local administrators willing to implement them voluntarily. The objective of this study was to investigate the reasons and meanings that they give to the use of a low-stakes evaluation program developed by a prestigious university in Chile, oriented to measure the performance and progress of students in relation to the national curriculum. The design contemplated the execution of a qualitative research based on semi-structured interviews with school principals and administrators of public and private educational networks that use the evaluation program, selected according to purposive sampling criteria (N=33). The information was analyzed based on Grounded Theory guidelines for open coding and, subsequently, interpreted from the sociology of experience. The results revealed the presence of heterogeneous logics of action (integration, strategic and subjectivation) underlying the decisions and uses of the assessment, whose possibilities of articulation lie with the actors and depend to a large extent on the organizational conditions of the establishments, giving rise to different experiences in relation to standardized evaluations. This study identified three types: the academicist experience, relative adherence, and the underutilization of assessments. It concludes by pointing out the implications of this study for educational research, policy and practice.*

### **Keywords**

*Standardized assessment – Social experience – Uses of assessment.*

---

### **Introducción**

La evaluación de desempeños escolares existe desde fases tempranas del desarrollo de los sistemas educativos y con el nacimiento de los sistemas escolares modernos adquiere un rol esencial en la selección meritocrática (BOURDIEU; PASSERON, 2001; DURU-BELLAT, 2009; STOBART, 2010). Sin evaluación y calificación no es posible realizar, durante la trayectoria escolar, las distinciones que culminarán como diferencias de certificaciones (MERLE, 2018; PERRENOUD, 2008).

Cada sistema escolar ha dispuesto de mecanismos de evaluación, siendo el más común las calificaciones de los docentes. Sin embargo, a medida que se fue desarrollando

la medición psicológica y educacional, se fueron expandiendo las evaluaciones estandarizadas (MONARCA, 2015; RAMOS-ZINCKE, 2018; SMITH, 2017), es decir, las pruebas dirigidas a medir los logros cognitivos de los alumnos cuyo diseño, aplicación y corrección son uniformes (MONS, 2009).

En Chile, este tipo de pruebas no es reciente (RAMOS-ZINCKE; FALABELLA, 2020), aunque también ha adquirido una importancia mayor en las últimas décadas. Esto ocurre en el contexto del modelo de mercado en la educación, el que requería medir la efectividad de los establecimientos escolares para informar la elección de escuela de las familias. Para ello se crea el Sistema Nacional de Medición de la Calidad de la Educación (SIMCE), un conjunto de pruebas de desempeño cuyos resultados por escuela comienzan a publicarse durante los años 1990 (BRAVO, 2011; FALABELLA, 2016).

A partir de leyes promulgadas entre 2009 y 2011, las evaluaciones estandarizadas suman un rol regulador al constituirse en dispositivos con altas consecuencias, bajo el modelo de rendición de cuentas (FLÓREZ; ROZAS, 2020; LOEB; BYUN, 2019; OZGA, 2020; SUSPITSYNA, 2010). Los resultados de los estudiantes se asocian a incentivos y sanciones a docentes e instituciones, que van desde beneficios económicos hasta el cierre de escuelas (DE LA VEGA; PICAZO, 2016; ELACQUA *et al.*, 2016; FALABELLA, 2016).

Dentro de este conjunto de políticas educativas se crean las Asesorías Técnicas Educativas (ATE), servicios ofrecidos por entidades privadas para la capacitación, asesoría pedagógica y evaluación de los establecimientos. Entre esas instituciones se sitúan las universidades (ASESORIAS PARA EL DESARROLLO, 2014).

Este trabajo se concentra en un programa de evaluaciones estandarizadas (Sistema de Evaluación del Progreso del Aprendizaje, SEPA) que no conlleva altas consecuencias ni fines de rendición de cuentas, sino que, junto a otras iniciativas privadas, conforman una oferta de servicios a disposición de los establecimientos escolares para el monitoreo de los aprendizajes de sus estudiantes. SEPA consiste en un conjunto de pruebas estandarizadas de lenguaje y matemáticas, alineadas con el currículum nacional, contratadas voluntariamente por escuelas, administraciones locales y programas de intervención educacional y que se aplican en alrededor de 200 escuelas a lo largo del país.

Este tipo de evaluaciones ha recibido poca atención en la literatura nacional e internacional, respecto de los sistemas de altas consecuencias. Además, en Chile frente al excesivo uso de estos sistemas, ha emergido un movimiento crítico de estas evaluaciones en el mundo académico y en la sociedad civil (PINO; OYARZÚN; SALINAS, 2016). En este contexto es relevante analizar el rol de las evaluaciones estandarizadas de bajas consecuencias insertas en un régimen de rendición de cuentas y los sentidos que les atribuyen los actores que deciden voluntariamente implementarlas. Describir esas experiencias de los actores permitirá ampliar el conocimiento respecto de las maneras en que ellos articulan las demandas y presiones heterogéneas del sistema educativo y enfrentan los desafíos que este les impone.

## El rol político de las evaluaciones estandarizadas

Desde las perspectivas sociológicas se postula que las evaluaciones estandarizadas constituyen instrumentos de una nueva regulación política de los sistemas educativos (DEMAILLY, 2006; DUBET, 2016; MONARCA, 2015; MONS, 2009; OZGA, 2020). De acuerdo con Ball (2003), las evaluaciones a gran escala son un elemento clave en las tecnologías políticas de mercado, gerencialismo y performatividad en las que reposan las reformas educativas a nivel global, que tienden a alinear el funcionamiento de los servicios públicos al del sector privado. Las evaluaciones son artefactos centrales en esta regulación debido a que esta consiste en una vigilancia a distancia que deja en manos de los actores el conducirse según *lo que funciona* (BALL, 2003).

En el campo de la educación, los llamados *governmentality studies* se han interesado en cómo las prácticas y discursos de la política se internalizan y encarnan a nivel del self (FIMYAR, 2006). Enfatizan en las interrelaciones entre el gobierno de sí y las tecnologías que moldean la conducta de individuos y colectivos y no en lo que los individuos gobernados efectivamente enuncian o hacen (BRÖCKLING; KRASMANN; LEMKE, 2010). Sobre la evaluación, se señala que, apoyada en argumentos técnicos y científicos, constituye un instrumento de gobierno de individuos asociado a definiciones de aprendizaje, calidad de la educación o incluso una herramienta para la producción de individuos conformes a las políticas neoliberales en educación (RAMOS-ZINCKE, 2018; SUSPITSYNA, 2010).

Estos trabajos suelen enfocarse en las evaluaciones de altas consecuencias o con fines de rendición de cuentas. Sin embargo, también permiten comprender el rol que cumplen evaluaciones como las que se abordan en este estudio, que participan del entramado de discursos y tecnologías políticas basadas en la medición del desempeño y que, de hecho, son desarrolladas por instituciones que gozan de legitimidad técnica y científica (OZGA, 2020; RAMOS-ZINCKE, 2018).

Ahora bien, en los análisis del rol de político de las evaluaciones predomina un énfasis en el punto de vista macrosocial, siendo menos estudiada la forma en que estos instrumentos son percibidos, apropiados y cuestionados por sus destinatarios (BRÖCKLING; KRASMANN; LEMKE, 2010). Si bien existe una literatura abundante sobre los usos de las evaluaciones (BEAVER; WEINBAUM, 2015; BROWN; ZHANG, 2017; EBBELER *et al.*, 2017; SCHILDKAMP, 2019; VANLOMMEL; SCHILDKAMP, 2019), esta se mantiene dissociada del rol de regulación y de las relaciones de poder en juego, centrándose en aspectos técnicos de validez o de la toma de decisiones basada en datos. Actualmente, los estudios comienzan a integrar las perspectivas micro y macrosocial; como plantea Ball (2016) a partir de Foucault (2001), en este tipo de tecnologías políticas el principal lugar de lucha y rechazo es la subjetividad de los individuos.

En Chile, algunos trabajos han estudiado los sentidos, significados y cómo las evaluaciones han penetrado en las culturas escolares (FALABELLA, 2016, 2020; PARRA; MATUS, 2016). No obstante, estos estudios no abordan las evaluaciones con bajas consecuencias. Precisamente esa es una contribución de este estudio, el cual, reconociendo el rol político de las evaluaciones educacionales, privilegia la experiencia de los actores del mundo escolar, para comprender las razones y sentidos que le dan al uso de evaluaciones externas.

## Las evaluaciones estandarizadas desde la perspectiva de los actores

Estudiar las evaluaciones estandarizadas desde la perspectiva de los actores supone una postura ascendente desde el *actor* hacia el *sistema*, sin desconocer las fuerzas macroestructurales. Es el caso de la sociología de la experiencia (DUBET, 1994, 2020) y del individuo (MARTUCCELLI, 2007, 2019), que se sitúan desde la experiencia de los actores con el fin de comprender los sentidos que dan a las orientaciones y mandatos que provienen desde la política educativa. Adicionalmente, esta perspectiva concibe diversas formas de ejercicio del poder, algunas actuando como presiones heterogéneas y difusas que se viven como exigencias individuales llamando a la iniciativa personal (MARTUCCELLI, 2007).

Dubet (1994) propone la noción de experiencia, entendida como la actividad del actor que se orienta a combinar y articular las lógicas heterogéneas del sistema social o escolar y a darles sentido. Los actores se enfrentan a un mundo que no está organizado y a una institución escolar no integrada, que se les presenta como una serie de pruebas, presiones y prescripciones cuya resolución recae en el propio individuo (MARTUCCELLI; DE SINGLY, 2018; SANTIAGO, 2016).

Los individuos responden en base a tres lógicas que corresponden a tipos-ideales de orientación a la acción. En primer lugar, la lógica de integración reúne las situaciones en las cuales el actor se define a sí mismo y explica sus acciones a partir de categorías, valores, roles sociales, que reflejan su pertenencia a un grupo o comunidad. En segundo lugar, la lógica estratégica se manifiesta cuando el actor concibe el mundo social como un mercado competitivo y su identidad como un estatus o una posición en relación con otros, y su actuar se orienta a situarse en las mejores posiciones de las jerarquías. Finalmente, la subjetivación emerge cuando el actor se define como un sujeto que mantiene una distancia crítica respecto de las otras lógicas, como alguien capaz de desafiar al orden social. La identidad es construida como un compromiso respecto de ideales y modelos culturales transformadores, se vive como un proceso creativo, de resistencia o de emancipación (DUBET, 1994; DUBET; MARTUCCELLI, 1998).

Desde esta perspectiva teórica, no se asume a priori que la decisión de contratar evaluaciones estandarizadas sea el resultado de una adhesión al modelo de rendición de cuentas. Ese modelo no es entendido como un sistema integrado de determinaciones provenientes de un único organismo claramente identificable, sino como una serie de prescripciones diversas e incluso contradictorias (MARTUCCELLI, 2007). Por otro lado, los actores no producen una respuesta única y predecible de interiorización, sino que existen diversas experiencias posibles, pudiendo incluir situaciones de crítica, resistencia o de atribución de sentido consciente y deliberado (BALL, 2016; DEMAILLY, 2003; DUBET, 2016). De esta manera, este estudio se pregunta ¿cuáles son los sentidos que los actores, en particular los directivos de escuelas, confieren a esas evaluaciones estandarizadas de bajas consecuencias? Se estudian así, las razones que motivan a directivos de institucionales educacionales a contratar voluntariamente estas evaluaciones, a pesar de la crítica y malestar asociado a la sobrevaluación derivada de la rendición de cuentas que predomina en el sistema educativo chileno actual.

La investigación se realizó sobre el Sistema de Evaluación del Progreso del Aprendizaje (SEPA), una prueba de bajas consecuencias desarrollada desde el año 2007 por una universidad chilena de prestigio. Este sistema comprende el diseño de las pruebas, su administración y entrega de resultados en una plataforma online de acceso restringido. Los resultados incluyen el estado del aprendizaje al finalizar un año escolar, el progreso entre años consecutivos, y, cuando se disponen de dos o más mediciones, el valor agregado de la escuela sobre esos estudiantes (aporte que realiza el establecimiento al aprendizaje de sus estudiantes por sobre otros factores).

Actualmente, se evalúa a más de doscientos establecimientos constituidos por más de 50.000 estudiantes en las asignaturas de Lenguaje y Matemática, a los que el programa les ofrece acompañamiento en el uso y gestión de los resultados de las evaluaciones, transmitiendo los usos promovidos y no deseados, desde un enfoque formativo y de la toma de decisiones basada en datos (ABARZÚA; CONTRERAS, 2019). Adicionalmente, si bien los responsables de la prueba recaban rutinariamente información de parte de sus usuarios, se estima necesario producir evidencia sistemática y de estándares científicos a favor o en contra de la validez de SEPA (por ejemplo, HEIN; TAUT, 2010), agenda institucional en la que se inserta el presente trabajo.

## **Metodología**

Para indagar en profundidad en las experiencias y sentidos que los directivos de escuelas otorgan a la evaluación SEPA se optó por una aproximación cualitativa, dada su pertinencia para trabajar con la subjetividad de los actores sociales, considerando además elementos contextuales y macrosociales (FLICK, 2004; GOBO, 2005). El alcance del estudio es descriptivo, permitiendo exponer las distintas configuraciones presentes en los actores.

La estrategia de producción de datos consistió en entrevistas semiestructuradas en las que se abordaron diversos temas en torno al uso y las significaciones de las evaluaciones estandarizadas. Estas entrevistas fueron realizadas de manera individual a aquellos actores que tomaron la decisión de contratar e implementar la prueba SEPA.

A partir de la definición anterior se realizó un muestreo determinado a priori (FLICK, 2004), delimitado por el entrecruzamiento de tres características relevantes en el sistema educacional chileno y SEPA: tipo de contratante, dependencia administrativa-financiera de las escuelas y ubicación geográfica. La primera característica refiere a si el participante es una autoridad central que administra una red de establecimientos, si administra un establecimiento o si es parte de un programa de investigación/intervención educacional. En Chile existen tres tipos de escuelas según su dependencia: escuelas públicas, privadas y subvencionadas (administración privada con financiamiento público), las que corresponden al 51%, 21% y 28% de los usuarios de SEPA respectivamente. Finalmente, la ubicación geográfica es relevante en un país altamente concentrado en su capital, por lo que se procuró obtener la participación de la capital y de las regiones. La muestra final de 33 participantes se distribuyó de la siguiente forma (ver tabla 1):

**Tabla 1-** Usuarios de 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.

Esta investigación obtuvo la aprobación del Comité de ética de la Pontificia Universidad Católica de Chile, lo que implica que se garantiza el respeto de los derechos de los participantes, tales como los consentimientos informados y las autorizaciones de las instituciones. Además, dado que el estudio se llevó a cabo por la misma universidad que implementa SEPA, el trabajo de campo fue realizado por un equipo externo con el fin de favorecer la honestidad de los participantes y la apertura a las experiencias por parte de los entrevistadores.

## Análisis de datos

En primera instancia, el material fue analizado de forma inductiva siguiendo los lineamientos de la codificación abierta propuesta por la *Grounded Theory* (STRAUSS; CORBIN, 2002) y utilizando el programa Max.QDA versión 2018.2. A nivel descriptivo, se identificaron los elementos constitutivos de la experiencia y sentidos del uso de la evaluación SEPA, tales como: los motivos de contratación y las valoraciones de esta.

En un segundo momento, y siguiendo las orientaciones metodológicas de la sociología de la experiencia (DUBET, 1994), se exploró al interior de las categorías de la etapa previa, la forma que adoptan las tres lógicas de la acción de la experiencia (estratégica, integración y subjetivación) en el marco del uso de las evaluaciones SEPA. Esto contempló un análisis de contenido de carácter abductivo (KENNEDY; THORNBERG, 2018), mediante una codificación

más analítica (GIBBS, 2012), vinculando las interpretaciones iniciales con los tipos puros que constituyen las tres lógicas de la acción. Finalmente, describir las experiencias implica analizar la actividad o trabajo que realizan los actores al combinar esas lógicas heterogéneas, y a veces contradictorias, para dar sentido a sus acciones respecto de las evaluaciones.

## Resultados

En este apartado se presentan los resultados de la investigación distinguiendo los dos pasos analíticos para la descripción de las experiencias. En primer lugar, se exponen los discursos en torno a las lógicas de la acción describiendo la forma particular que asumen estos tipos ideales en el uso de evaluaciones estandarizadas estudiadas. En la segunda parte se presentan las experiencias de los actores que, a diferencia de las lógicas de la acción, son articulaciones empíricas y ancladas a contextos específicos (DUBET, 1994; DUBET; MARTUCELLI, 1998).

### Tres lógicas de la acción en el uso de evaluaciones externas

#### a. El uso estratégico de las evaluaciones externas

En esta categoría se sitúan los discursos que conciben las evaluaciones estandarizadas como una solución pragmática a las necesidades de la institución. Propios de esta lógica son los análisis de costo/beneficio al tomar la decisión de contratar la evaluación, y valorar el *prestigio* de sus creadores como una garantía de calidad (confiable y válida).

Al situarse desde la lógica estratégica, el actor se percibe inserto en un mercado competitivo y aspira a que su escuela ocupe posiciones aventajadas. Las evaluaciones son valoradas porque predicen y mejoran los resultados en otras evaluaciones, porque sus resultados pueden satisfacer las aprehensiones de las familias y autoridades, o porque son números que permiten competir con otras instituciones, clasificar a grupos e individuos, exigirles rendición de cuentas o atribuirles sanciones y recompensas.

En algunos casos el uso de las evaluaciones no logra ser el resultado de un cálculo racional, sino que son más bien estrategias de sobrevivencia, soluciones de compromiso o formas de lidiar con las presiones derivadas del sistema escolar. En estos casos, la evaluación externa permite responder a las exigencias de mejora de la administración central y no perder financiamiento, disponiendo de mediciones que legitimen la gestión o que demuestren la necesidad de cambios de las prácticas pedagógicas. Si a las evaluaciones se les asignan consecuencias directas para los individuos o instituciones, es probable que emerjan estrategias como el engaño, el entrenamiento excesivo para la prueba, la manipulación de las condiciones de aplicación o de los resultados.

#### b. Adhesión a un modelo evaluativo

Esta lógica de integración agrupa aquellos discursos en que la evaluación se concibe acorde a una cultura dominante, a la que los actores adhieren, ya sea por deseo de mantenerse incluidos en una comunidad mayor o porque interiorizaron su relevancia.

Son posibles dos configuraciones. Por una parte, los discursos que adhieren al modelo de rendición de cuentas; por otra, los que valoran los beneficios de las evaluaciones formativas como SEPA, manifestando un discurso alternativo respecto de las evaluaciones de altas consecuencias.

Dado que los participantes corresponden a contratantes de un programa evaluativo que dispone de instancias de acompañamiento e informa sus propósitos a los usuarios, la adhesión al modelo de la evaluación formativa fue predominante en los relatos de los participantes. Por ejemplo, los entrevistados generalmente señalan con firmeza que los resultados de las evaluaciones no se deben utilizar para desvincular docentes o para calificar estudiantes. Se muestran, por tanto, contrarios a los usos punitivos de la evaluación. Por supuesto, la adhesión a este modelo no está exenta de matices, distancias y contradicciones que se perciben en las entrevistas. Por ejemplo, se afirma que la evaluación no debe servir para desvincular a docentes, pero se realiza una responsabilización directa del profesor por los resultados de sus estudiantes.

### **c. Distancia respecto de los modelos evaluativos dominantes**

La lógica de subjetivación corresponde a aquellos discursos en los cuales el actor se sitúa desde la distancia crítica hacia las dos lógicas precedentes. No se interiorizan plenamente los propósitos del sistema de evaluación o de otro modelo dominante, pero tampoco es una racionalidad puramente estratégica. Es visible en la crítica, en la emancipación o la orientación a principios transformadores que defienden el bien, la verdad, lo justo y cuestionan desde ahí los modelos evaluativos disponibles.

Dado que los entrevistados presentaban un alto grado de adhesión al programa, los relatos susceptibles de situarse en la lógica de la subjetivación fueron tenues. En términos de los motivos de contratación, se pudo apreciar cuando los actores se presentaban como sujetos de su decisión, afirmando con convicción sus buenas razones para implementar el sistema de evaluación, razones sentidas como construidas por ellos, ni impuestas ni irreflexivas. Estos actores presentan una convicción de que esas evaluaciones sirven efectivamente para tomar decisiones que permiten la mejora del aprendizaje. Mejorar la calidad de la educación pasaría por asegurarse de que los estudiantes están aprendiendo, y calidad de la educación o aprendizaje no es sinónimo de posición en un ranking o de un puntaje, sino una apropiación de los contenidos.

En otros casos, se valora la evaluación por la posibilidad de disponer de información individual para desarrollar una pedagogía personalizada; o porque a partir de la evaluación se generan reflexiones profundas e instancias de trabajo colaborativo.

La distancia crítica se pudo apreciar en los relatos que cuestionaban aspectos de este sistema de evaluación en particular, por ejemplo, la necesidad de un mayor acompañamiento en el uso de los datos para tomar decisiones e idear intervenciones, la falta de flexibilidad en los tiempos y condiciones de aplicación, la dificultad de comprensión de algunos reportes de resultados. Adicionalmente, se relativiza el poder de la evaluación afirmando que la prueba es solo una prueba, un solo punto de vista que, además, conlleva error y, por tanto, no es siempre reflejo de la realidad que pretende medir. En ese sentido, si bien se percibe como necesaria, no es suficiente y debe ser complementada con otra información.

En algunos casos se aprecian disidencias o desobediencias sutiles, por ejemplo, cuando se reconoce que, pese a que SEPA desaconseja ciertas acciones, se realizan porque el usuario estima que es apropiado en su contexto y para sus propósitos.

La distancia de mayor importancia en el discurso de los entrevistados es la situación en la cual los resultados no se utilizan o “no se aprovechan”. Algunos directivos señalan que los profesores no se han involucrado lo suficiente e incluso ellos mismos, no siempre se implican desde el rol de directivos. El caso más elocuente de esta resistencia pasiva es el de una dirección de educación municipal que toma la decisión de implementar las evaluaciones a nivel central sin involucrar a los directivos y docentes en el proceso. Las respuestas de los establecimientos son variadas y, entre ellas, aparecen las escuelas que aplican las evaluaciones y no las aprovechan porque no se les han ofrecido las condiciones necesarias. Adicionalmente, se observaron casos en los cuales la administración local manifestaba una baja valoración de las evaluaciones, al punto de desear discontinuarlas, sin embargo, los directivos de las escuelas las consideraban de gran utilidad y señalaban realizar un trabajo pedagógico valioso a partir de ellas.

## **Experiencias en torno al uso de las evaluaciones externas**

La experiencia consiste en el trabajo que realizan los actores para integrar esas lógicas heterogéneas y, a veces, contradictorias y dar un sentido a sus acciones. Así, los actores se enfrentan a *pruebas* que son similares en su estructura y origen, pero que se distribuyen de manera desigual según los contextos escolares. En este sentido, es posible identificar contextos que son favorables a la aparición de experiencias articuladas y otros donde se acumulan las condiciones desfavorables para la integración de la experiencia (DUBET, 1994; MARTUCCELLI; DE SINGLY, 2018; SANTIAGO, 2016). De acuerdo con los datos analizados, fue posible identificar tres situaciones que, sin ser exhaustivas, permiten describir tres experiencias empíricas en torno a las evaluaciones estandarizadas con fines formativos.

En primer lugar, la experiencia academicista, reúne condiciones favorables al uso de evaluaciones externas y estandarizadas, pues se trata de culturas escolares orientadas al logro de resultados, que valoran la eficiencia, la validez, la confiabilidad de los instrumentos de medición. Valoran la mirada externa que les permite legitimarse frente a las familias y las autoridades públicas o privadas. En general, los actores adhieren al modelo del uso de la evaluación para la mejora educativa (medida a través del desempeño en evaluaciones). En los casos de mayor integración de las lógicas de la acción, la mirada externa que aporta la evaluación es percibida como una oportunidad para auto observarse, reflexionar sobre las propias prácticas y concebir, en ambientes colaborativos, estrategias de enseñanza orientadas a suplir las áreas deficientes. En los casos donde predomina el modelo de rendición de cuentas, la evaluación sirve como entrenamiento o predicción de las evaluaciones con altas consecuencias. Esta variante de la experiencia presenta una supremacía de la estrategia y la integración, dejando poco espacio a la crítica y a concepciones más profundas del aprendizaje. La siguiente cita ilustra este tipo de experiencias:

Yo siempre digo, le digo a los profesores “yo estoy seguro que ustedes cuando están en una asado en alguna parte, siempre dicen (E: [risas]) ‘no, mi colegio es de alto rendimiento, tenemos sobre...’”, el año pasado obtuvimos 350 puntos en el SIMCE en matemática acá po, 350 puntos del SIMCE, superamos al Lastarria, superamos al liceo 7, somos el mejor de la comuna, le ganamos a todos los colegios... a todos los colegios privados, viene la UC a trabajar con nosotros y aquí todos hablan ahí no más del liceo, pero externamente hay un... hay un tema de identidad que es muy... que es muy potente... (director, establecimiento público).

En segundo lugar, la experiencia de adhesión relativa presenta tantas condiciones favorables como desfavorables a la integración de las lógicas de la acción. Describe aquellos contextos y actores que son críticos a las evaluaciones nacionales y que deciden contratar evaluaciones privadas porque valoran sus fines formativos (en oposición a los sumativos), su carácter individualizado (en oposición a los resultados agregados), su foco en lo pedagógico (en oposición a la gestión y rendición de cuentas). En estos casos se aprecian decisiones estratégicas de contratación, incluyendo análisis del mercado de ATEs, por ejemplo. No obstante, suele ser la situación de establecimientos que contratan los servicios con recursos públicos, lo que significa que deben rendir cuentas y obtener buenos resultados en las evaluaciones nacionales para mantener el financiamiento. Esta condición los obliga a integrarse al sistema de rendición de cuentas, pese a que lo cuestionen. Son contextos en los cuales se vive la tensión entre la presión de las evaluaciones nacionales y la adhesión a modelos evaluativos más formativos, entre la necesidad de dar cumplimiento a las exigencias ministeriales y visiones más creativas de la educación y el aprendizaje:

[...] en la escuela no había una prueba así que pudiera reportar de manera tan minuciosa el informe general del curso, de la escuela, por alumno, por habilidades, en esas dos áreas ¿ya? y así que, de acuerdo al plan de mejoramiento, consideramos que la prueba SEPA nos podía ayudar bastante, ahora igual al principio los profesores la tomaron como una prueba estandarizada más, parecida al SIMCE, la relacionaron con cobertura, tiene que ver con cobertura, si eso es claro, pero el propósito es saber efectivamente dentro de... de los estándares de aprendizaje que están propuestos en los planes y programas o bases curriculares, cuánto de aquello los estudiantes están aprendiendo o no, porque igual es algo que nosotros tenemos que dar cuenta de eso ¿ya? sí, las escuelas, muy pocas escuelas tienen planes y programas propios, por ejemplo, entonces creo que... pero ahora, ahora anda bien el tema, o sea los profesores ya lo están utilizando como una herramienta de análisis, más que de, como si se evalúa en su trabajo... (Director establecimiento municipal).

En tercer lugar, la experiencia de subutilización de la evaluación consiste en aquella situación en la cual los actores adhieren a los discursos dominantes sobre la importancia de las evaluaciones estandarizadas (ya sea los de rendición de cuentas o formativos), en algunos casos valoran las características técnicas de la evaluación, sin embargo, en la práctica no las utilizan o no les sacan provecho. En general, esa subutilización de la evaluación resulta de la ausencia de condiciones organizacionales (especialmente

de tiempo) y de competencias para comprender la información y desprender medidas concretas a partir de ella.

[...] y por eso me asaltan dudas de si continuar o no, no por la falla del instrumento, no por debilidad del instrumento, sino por nuestra capacidad de gestión respecto de lo que nos entrega SEPA, entonces eh... y también por eso, eso me lleva también a ser reticente a veces cuando me proponen otros instrumentos, entonces sumar otro instrumento, sumar otro y si no... no tenemos resultados ¿para qué?, si a lo mejor tener pocos o uno o dos instrumentos al respecto que... pero utilizarlos bien utilizarlos bien y no picar por aquí y por allá, eh... esperando que alguno nos haga el milagro, nos haga el milagro ¿cierto? y que sea capaz de, a lo mejor superar nuestras propias debilidades o subsidiar nuestras debilidades y entregarnos buenos resultados a pesar de nosotros mismos y eso no va a ocurrir... (administrador municipal de establecimientos).

La subutilización también puede ser la experiencia de quienes no lograron apreciar las ventajas de la evaluación en el tiempo porque estas solo les mostraban sistemáticamente los resultados deficientes de los estudiantes y no se produjeron cambios producto de su utilización, como lo esperaban los usuarios. En este estudio, al tratarse de un programa específico y dadas las limitaciones de la muestra, no se observaron tensiones tan fuertes como las reportadas en otros trabajos (CONTRERAS, 2014; FALABELLA, 2016; 2020) que llevaran a una desilusión con la evaluación y a su descrédito, por ejemplo, bajo el argumento de que la prioridad de la educación no es el rendimiento escolar sino la formación de personas, como ocurre en algunos contextos de vulnerabilidad social donde predomina la desesperanza ante el estancamiento de los resultados deficientes.

## Discusión

Este trabajo indagó en las razones por las que los directivos de instituciones educacionales contratan voluntariamente las evaluaciones SEPA, aun cuando en Chile las altas consecuencias asociadas a las evaluaciones estandarizadas han generalizado la crítica y la desconfianza (PINO; OYARZÚN; SALINAS, 2016). Con el fin de no asumir a priori estas valoraciones y sentidos, se utilizó un enfoque que incluye los condicionamientos sociales y las vivencias de los individuos bajo el concepto de experiencia social, de tal manera que, sin desconocer la vertiente política y reguladora de las evaluaciones, busca comprender cómo los actores les dan sentido (DUBET, 1994; MARTUCCELLI; DE SINGLY, 2018; SANTIAGO, 2016).

No porque un programa de evaluación sea de bajas consecuencias carece de un rol político, es más, estos programas forman parte del entramado de instituciones técnicas que sustentan las concepciones sobre educación, aprendizaje y evaluación, que se presentan como verdades a los actores del mundo escolar. Regulan las conductas de los individuos a través de clasificaciones, cifras y distinciones, participando de los discursos dominantes sobre *lo que funciona*, cómo se mide y cómo se rinde cuenta de ello (BALL, 2003; RAMOS-ZINCKE, 2018).

Fue posible identificar lógicas distintas a las promovidas por el programa, orientadas a la mejora escolar, y otras propias de un sistema de rendición de cuentas.

Si bien no fueron predominantes, indican, al igual que estudios en otros contextos, que las evaluaciones no siempre se utilizan de acuerdo con los propósitos para los que fueron diseñadas (ALKIN; KING, 2016; SUSSMAN; WILSON, 2019). Así, las consecuencias dependen más del uso real que de la definición de la evaluación. Por esto, una distinción entre altas y bajas consecuencias que no considere el uso de las evaluaciones no tiene sentido (TANNENBAUM; KANE, 2019).

Se distinguieron las tres lógicas de la acción en el uso de las evaluaciones SEPA: el uso estratégico, la adhesión a un modelo evaluativo y la distancia crítica respecto de los modelos dominantes. Luego, se identificaron tres experiencias en relación con esas lógicas. La experiencia academicista, donde las condiciones parecen reunidas para una integración de las lógicas, con una presencia importante del uso estratégico de las evaluaciones y, en algunos casos, con prácticas asociadas a la lógica de rendición de cuentas. La experiencia de adhesión relativa a las evaluaciones, en las cuales existen condiciones facilitadoras y obstaculizadoras para dar sentido a las evaluaciones externas. Se revela la coexistencia y tensión entre modelos de altas consecuencias individuales y colectivas con instancias asociadas a modelos de *accountability* inteligente, es decir, basados en la colaboración y confianza entre equipos, relaciones horizontales entre servicios, liderazgo pedagógico y una fuerte orientación a la evaluación para el aprendizaje (TOLO *et al.*, 2020). Algunos de estos elementos de las evaluaciones para el aprendizaje también actúan como ideales en la experiencia de subutilización de la evaluación, en la que se toma la decisión inicial de contratar servicios de evaluaciones externas para generar procesos reflexivos que conduzcan a la mejora de las prácticas pedagógicas y, por ende, de los aprendizajes. Sin embargo, no son las lógicas de rendición de cuentas las que dificultan estos procesos, sino la falta de condiciones materiales, profesionales y organizacionales en los establecimientos educacionales.

Esos hallazgos permiten concluir que, efectivamente existen situaciones en las que se aprecia una congruencia entre las dos facetas de la gubernamentalidad: las técnicas de gestión de las poblaciones y la racionalidad que las subyace, por un lado, y la autoconducción o técnicas de sí de los sujetos individuales que han interiorizado las normas sociales (SUSPITSYNA, 2010). No obstante, los resultados revelan la existencia de lógicas estratégicas y de subjetivación que tensionan la interiorización de los beneficios de las evaluaciones estandarizadas por parte de actores y demuestran que la cultura de la auditoría parece más una propuesta que una realización. Se observan situaciones en las cuales los actores se muestran como sujetos de sus decisiones, adhiriendo o resistiendo, y otras en las cuales prueban tácticas y negociaciones para responder a las presiones de las cuales son completamente conscientes.

En términos de la política y práctica educativa, se identifican concepciones de la evaluación orientadas a usos pedagógicos y de mejora escolar, lo que permite un cierto grado de optimismo frente a un extenso período (alrededor de tres décadas) de dominio de las evaluaciones de altas consecuencias, asociadas en primer lugar a un modelo de mercado y, luego, a un sistema estatal de aseguramiento de la calidad basado en la responsabilización por desempeño escolar (FALABELLA; DE LA VEGA, 2016; PARCERISA; FALABELLA, 2017). Si bien prevalecen los usos de evaluaciones con fines de rendición de cuentas, el discurso de la Agencia de Calidad, organismo estatal a cargo de las evaluaciones,

ha ido incorporando modelos formativos y más centrados en la mejora escolar (AGENCIA DE CALIDAD DE LA EDUCACIÓN, 2018). De hecho, la evaluación para el aprendizaje es el modelo evaluativo que predomina en la más reciente política para el fortalecimiento de la evaluación en el aula (CHILE, 2018).

Esas pueden ser señales de un avance hacia formas de *accountability* inteligente (TOLO *et al.*, 2020), las que son incipientes como se observa en los discursos de los actores de este estudio. Para avanzar en esta línea será necesario intervenir en dos aspectos de la práctica educativa. Por un lado, en la instalación de condiciones concretas para el uso de datos. Entre ellas destacan, las competencias para el uso de datos en los equipos pedagógicos, las condiciones materiales y tiempos necesarios para procesos reflexivos, los espacios de trabajo colaborativo, el liderazgo distribuido (BROWN; ZHANG, 2017; SCHILDKAMP, 2019; SCHILDKAMP *et al.*, 2016; TOLO *et al.*, 2020). Por otro lado, desde los niveles centrales serán necesarias formas más horizontales y democráticas de responsabilización, pues mientras persistan las altas consecuencias individuales y colectivas, los modelos alternativos entrarán necesariamente en tensión con ellos, haciendo difícil la experiencia de los actores en torno a las evaluaciones. En Chile, este proceso será paulatino, pues no solo basta con cambiar el enfoque de las políticas. Los modelos evaluativos y de responsabilización son inseparables de las culturas escolares y lo que determina en gran medida el paso de una cultura de la competencia a una de rendición de cuentas inteligente, es la confianza en las instituciones centrales y locales y al interior de los equipos pedagógicos (FALABELLA, 2020; MALIN *et al.*, 2020; TOLO *et al.*, 2020).

Dado que la finalidad de esta investigación fue recoger la experiencia de un grupo particular de participantes (cargos directivos) y debido a la metodología y muestra utilizadas no es posible realizar generalizaciones a la población de usuarios del programa. No obstante, se logra generar evidencia relevante para la agenda de validación del programa acerca de los usos y su alineamiento a los propósitos de este. Así, constituye una primera etapa de una agenda de validación más amplia (ABARZÚA; CONTRERAS, 2019). Desde la perspectiva de la transferibilidad (FLICK, 2004) y en la medida que los resultados de esta investigación proponen la identificación de tipos de experiencia y no una mera descripción, sus hallazgos ofrecen una forma de comprender este fenómeno en contextos de similares características. Por último, un riesgo de los análisis efectuados y de la producción de tipologías es la simplificación de la realidad, de modo que deben interpretarse como manifestaciones particulares y no como una descripción exhaustiva de esta.

En virtud de estas limitaciones, investigaciones futuras deberán orientarse a identificar los usos, sentidos y valoraciones que dan otros actores a las evaluaciones estandarizadas, siendo de especial relevancia la perspectiva de los docentes. Adicionalmente, sería de gran interés científico y social, estudiar las condiciones que favorecen y obstaculizan estos usos en cada contexto, con especial énfasis en aquellos que permiten la transición hacia culturas de *accountability* inteligente y democrático.

## Referencias

- ABARZÚA, Andrea; CONTRERAS, Johana. Sistema de Evaluación de Progreso del Aprendizaje, SEPA: evidencia de su confiabilidad y validez. *In*: MANZI, Jorge; GARCÍA, María; TAUT, Sandy (org.). **Validez de evaluaciones educacionales en Chile y Latinoamérica**. Santiago de Chile: UC, 2019. p. 127-156.
- AGENCIA DE CALIDAD DE LA EDUCACIÓN. **Nuevo sistema nacional de evaluación de aprendizajes**: la evaluación al servicio de los aprendizajes. Santiago de Chile: Agencia de Calidad de la Educación, 2018.
- ALKIN, Marvin; KING, Jean. The historical development of evaluation use. **American Journal of Evaluation**, Los Angeles, v. 37, n. 4, p. 568-579, 2016. <https://doi.org/10.1177/1098214016665164>
- ASESORIAS PARA EL DESARROLLO. **Estudio sobre el funcionamiento de los servicios de Asistencia Técnica Educativa (ATE) chileno**: informe final. Santiago de Chile: Ministerio de Educación: PNUD, 2014.
- BALL, Stephen John. Subjectivity as a site of struggle: refusing neoliberalism? **British Journal of Sociology of Education**, Oxford, v. 37, n. 8, p. 1129–1146, 2016. <https://doi.org/10.1080/01425692.2015.1044072>.
- BALL, Stephen John. The teacher's soul and the terrors of performativity. **Journal of Education Policy**, London, v. 18, n. 2, p. 215-228, 2003. <https://doi.org/10.1080/0268093022000043065>
- BEAVER, Jessica; WEINBAUM, Elliot. State test data and school improvement efforts. **Educational Policy**, Los Altos, v. 29, n. 3, p. 478-503, 2015. <https://doi.org/10.1177/0895904813510774>.
- BOURDIEU, Pierre; PASSERON, Jean-Claude. **La reproducción**: elementos para una teoría del sistema de enseñanza. Madrid: Popular, 2001.
- BRAVO, Juan. Simce: Pasado, presente y futuro del sistema nacional de evaluación. **Estudios Públicos**, Santiago de Chile, n. 123, p. 189-211, 2011. <https://doi.org/10.38178/cep.vi123.336>
- BRÖCKLING, Ulrich; KRASMANN, Susanne; LEMKE, Thomas. From Foucault's lectures at the Collège de France to studies of governmentality: an introduction. *In*: BRÖCKLING, Ulrich; KRASMANN, Susanne; LEMKE, Thomas (org.). **Governmentality**. New York: Routledge, 2010. p. 9-41. <https://doi.org/10.4324/9780203846476-4>
- BROWN, Chris; ZHANG, Dell. How can school leaders establish evidence-informed schools: an analysis of the effectiveness of potential school policy levers. **Educational Management Administration and Leadership**, London, v. 45, n. 3, p. 382-401, 2017. <https://doi.org/10.1177/1741143215617946>.
- CHILE. Ministerio de Educación. **Política para el fortalecimiento de la evaluación en el aula**. Santiago de Chile: Mineduc, 2018.
- CONTRERAS, Johana. **Les modes de production des inégalités, les cultures scolaires et les expériences des élèves**: une comparaison des systèmes éducatifs français et chilien. 2014. 581 p. Tesis (Doctorado en Sociología) – École Doctorale: Sociétés, politique, santé publique, Université de Bordeaux, Bordeaux, 2014. Disponible en: <https://tel.archives-ouvertes.fr/tel-01152627> Acceso en: 20 oct. 2020.

DE LA VEGA, Luis; PICAZO, María Inés. La accountability de la calidad y equidad de una política educativa: el caso de la subvención escolar preferencial para la infancia vulnerable en Chile. **Reforma y Democracia**, Caracas, v. 2016, n. 65, p. 193-224, jun. 2016.

DEMAILLY, Lise. En Europe: l'évaluation contre la crise des systèmes scolaires, l'évaluation en crise. **Education et Societes**, Paris, v. 17, n. 1, p. 105-120, 2006. <https://doi.org/10.3917/es.017.0105>

DEMAILLY, Lise. L'évaluation de l'action éducative comme apprentissage et négociation. **Revue Francaise de Pedagogie**, Paris, n. 142, p. 115-130, 2003. <https://doi.org/10.3406/rfp.2003.29>

DUBET, François. Les instruments et l'institution: le cas de l'école. **Sociologie du Travail**, Paris, v. 58, n. 4, p. 381-389, 2016. <https://doi.org/10.1016/j.soctra.2016.09.019>

DUBET, François. **Sociologie de l'expérience**. Paris: Seuil, 1994.

DUBET, François. The return of society. **European Journal of Social Theory**, London, p. 1-19, 2020. <https://doi.org/10.1177/1368431020950541>

DUBET, François; MARTUCCELLI, Danilo. **En la escuela**: sociología de la experiencia escolar. Buenos Aires: Losada, 1998.

DURU-BELLAT, Marie. **Le mérite contre la justice**. Paris: Presses des Sciences Po, 2009.

EBBELER, Johanna et al. The effects of a data use intervention on educators' satisfaction and data literacy. **Educational Assessment, Evaluation and Accountability**, Dordrecht, v. 29, n. 1, p. 83-105, 2017. <http://dx.doi.org/10.1007/s11092-016-9251-z>

ELACQUA, Gregory *et al.* Short-run effects of accountability pressures on teacher policies and practices in the voucher system in Santiago, Chile. **School Effectiveness and School Improvement**, Abingdon, v. 27, n. 3, p. 385-405, 2016. <https://doi.org/10.1080/09243453.2015.1086383>

FALABELLA, Alejandra. ¿Qué aseguran las políticas de aseguramiento de la calidad? Un estudio de casos en distintos contextos escolares. **Estudios Pedagógicos**, Valdivia, v. 42, n. 1, p. 107-126, 2016. <https://doi.org/10.4067/S0718-07052016000100007>

FALABELLA, Alejandra. The ethics of competition: accountability policy enactment in Chilean schools' everyday life. **Journal of Education Policy**, London, v. 35, n. 1, p. 23-45, 2020. <https://doi.org/10.1080/02680939.2019.1635272>

FALABELLA, Alejandra; DE LA VEGA, Luis. Políticas de responsabilización por desempeño escolar: un debate a partir de la literatura internacional y el caso chileno. **Estudios Pedagógicos**, Valdivia, v. 42, p. 395-413, 2016.

FIMYAR, Olena. Using governmentality as a conceptual tool in education policy research. **Educate**, London, v. 1, n. 1, p. 3-18, 2006.

FLICK, Uwe. **Introducción a la investigación cualitativa**. Madrid: Morata, 2004. <https://doi.org/M-15.306-2007>

FLÓREZ, María Teresa; ROZAS, Tamara. Accountability from a social justice perspective: Criticism and proposals. **Journal of Educational Change**, Berkeley, v. 21, n. 1, p. 157-182, 2020. <https://doi.org/10.1007/s10833-019-09361-3>

FOUCAULT, Michel. Le sujet et le pouvoir. *In*: DEFERT, Daniel (org.). **Dits et écrits II 1976-1988**. Paris: Gallimard, 2001. p. 1041-1062.

GIBBS, Graham. **El análisis de datos en investigación cualitativa**. Madrid: Morata, 2012.

GOBO, Giampietro. The renaissance of qualitative methods. **Forum Qualitative Sozialforschung**, Berlin, v. 6, n. 3, p. 1-12, 2005. <https://doi.org/10.17169/fqs-6.3.5>

HEIN, Andreas; TAUT, Sandy. El uso de información evaluativa externa con fines formativos: el caso de establecimientos educacionales chilenos participantes de SEPA. **Revista Iberoamericana de Evaluación Educativa**, [S. l.], v. 3, n. 2, 160-181, 2010.

KENNEDY, Brianna; THORNBERG, Robert. Deduction, induction, and abduction. *In*: FLICK, Uwe (org.). **The SAGE handbook of qualitative data collection**. London: SAGE, 2018. p. 49-64. <https://doi.org/10.4135/9781526416070.n4>

LOEB, Susanna; BYUN, Erika. Testing, accountability, and school improvement. **Annals of the American Academy of Political and Social Science**, Thousand Oaks, v. 683, n. 1, p. 94-109, 2019. <https://doi.org/10.1177/0002716219839929>

MALIN, Joel *et al.* World-wide barriers and enablers to achieving evidence-informed practice in education: what can be learnt from Spain, England, the United States, and Germany? **Humanities and Social Sciences Communications**, London, v. 7, n. 1, p. 1-14, 2020. <https://doi.org/10.1057/s41599-020-00587-8>

MARTUCCELLI, Danilo. **Cambio de rumbo: la sociedad a escala del individuo**. Santiago de Chile: LOM, 2007.

MARTUCCELLI, Danilo. Variantes del individualismo. **Estudios Sociológicos**, Ciudad de México, v. 37, n. 109, p. 7-37, 2019.

MARTUCCELLI, Danilo; DE SINGLY, F. **L'individu et ses sociologies**. Paris: Armand Colin, 2018.

MERLE, Pierre. **Les pratiques d'évaluation scolaire**. Paris: PUF, 2018.

MONARCA, Héctor. **Evaluaciones externas: mecanismos para la configuración de representaciones y prácticas en educación**. Buenos Aires: Miño y Dávila, 2015.

MONS, Nathalie. Effets théoriques et réels des politiques d'évaluation standardisée. **Revue Française de Pédagogie**, Paris, n. 169, p. 99-140, 2009. <https://doi.org/10.4000/rfp.1531>

OZGA, Jenny. The politics of accountability. **Journal of Educational Change**, Berkeley, v. 21, n. 1, p. 19-35, 2020. <https://doi.org/10.1007/s10833-019-09354-2>

PARCERISA, Lluís; FALABELLA, Alejandra. La consolidación del Estado evaluador a través de políticas de rendición de cuentas: trayectoria, producción y tensiones en el sistema educativo chileno. **Education Policy Analysis Archives**, Tempe, v. 25, p. 1-24, 2017. <https://doi.org/10.14507/epaa.25.3177>

PARRA, Victoria; MATUS, Gladys. 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**, Santiago de Chile, n. 45, p. 207-250, 2016. <https://doi.org/10.4067/s0718-45652016000200007>

PERRENOUD, Philippe. **La evaluación de los alumnos**. Buenos Aires: Colihue, 2008.

PINO, Mauricio; OYARZÚN, Gonzalo; SALINAS, Iván. Crítica a la rendición de cuentas: narrativa de resistencia al sistema de evaluación en Chile. **Cadernos Cedes**, São Paulo, v. 36, n. 100, p. 337-354, 2016. <https://doi.org/10.1590/CC0101-32622016171362>

RAMOS-ZINCKE, Claudio. Dispositivo de evaluación y gubernamentalidad del sistema educacional: entretejimiento de ciencia social y poder. **Cinta de Moebio**, Santiago de Chile, v. 61, p. 41-55, 2018. <https://doi.org/10.4067/S0717-554X2017000100041>

RAMOS-ZINCKE, Claudio; FALABELLA, Alejandra. Dispositivo de evaluación educacional y gubernamentalidad en Chile: los orígenes (1844-1970). **Cuhso**, Temuco, v. 2, n. 30, p. 1-30, 2020. <https://doi.org/10.7770/0719-2789.2019.cuhso.03.a07>

SANTIAGO, José. Repensar la acción social tras la crisis de la idea de sociedad: las promesas de las nuevas sociologías del individuo. *In*: GATTI, Gabriel; TEJERINA, Benjamín (org.). **Pensar la agencia en la crisis**. Madrid: Centro de Investigaciones Sociológicas, 2016. p. 103-121.

SCHILDKAMP, Kim. Data-based decision-making for school improvement: research insights and gaps. **Educational Research**, Berkshire, v. 61, n. 3, p. 257-273, 2019. <https://doi.org/10.1080/00131881.2019.1625716>

SCHILDKAMP, Kim *et al.* Factors promoting and hindering data-based decision making in schools. **School Effectiveness and School Improvement**, Abingdon, v. 28, n. 2, p. 242-258, 2016. <https://doi.org/10.1080/09243453.2016.1256901>

SMITH, William. National testing policies and educator based testing for accountability. **OECD Journal**, Paris, v. 2017, n. 1, p. 131-149, 2017. [https://doi.org/10.1787/eco\\_studies-2016-5jg1jxftj4r3](https://doi.org/10.1787/eco_studies-2016-5jg1jxftj4r3)

STOBART, Gordon. **Tiempos de pruebas: los usos y abusos de la evaluación**. Madrid: Morata, 2010.

STRAUSS, Anselm; CORBIN, Juliet. **Bases de la investigación cualitativa: técnicas y procedimientos para desarrollar la teoría fundamentada**. Medellín: Universidad de Antioquia, 2002.

SUSPITSYNA, Tatiana. Accountability in american education as a rhetoric and a technology of governmentality. **Journal of Education Policy**, London, v. 25, n. 5, p. 567-586, 2010. <https://doi.org/10.1080/02680930903548411>

SUSSMAN, Joshua; WILSON, Mark. The use and validity of standardized achievement tests for evaluating new curricular interventions in mathematics and science. **American Journal of Evaluation**, New York, v. 40, n. 2, p. 190-213, 2019. <https://doi.org/10.1177/1098214018767313>

TANNENBAUM, Richard J.; KANE, Michael T. Stakes in testing: tot a simple dichotomy but a profile of consequences that guides needed evidence of measurement quality. **ETS Research Report Series**, Princeton, v. 2019, n. 1, p. 1-16, 2019. <https://doi.org/10.1002/ets2.12255>

TOLO, Astrid et al. Intelligent accountability in schools: a study of how school leaders work with the implementation of assessment for learning. **Journal of Educational Change**, Berkeley, v. 21, n. 1, p. 59-82, 2020. <https://doi.org/10.1007/s10833-019-09359-x>

VANLOMMEL, Kristin; SCHILDKAMP, Kim. How do teachers make sense of data in the context of high-stakes decision making? **American Educational Research Journal**, Thousand Oaks, v. 56, n. 3, p. 792-821, 2019. <https://doi.org/10.3102/0002831218803891>

*Recibido en: 05.02.2021*

*Revisado en: 28.03.2022*

*Aprobado en: 19.04.2022*

**Editora:** Shirley Silva

**Johana Contreras** es psicóloga educacional de la Pontificia Universidad Católica de Chile, doctora y máster en Sociología de la Universidad de Bordeaux. Actualmente es investigadora del centro de medición MIDE UC.

**Álex Torres** es psicólogo y magíster en Psicología Social Comunitaria de la Pontificia Universidad Católica de Chile. Actualmente es instructor adjunto en dicha universidad y en la Universidad Alberto Hurtado y miembro de la Unidad de Análisis de MIDE UC.