

Sobre/Contra el SIMCE

Dr. Félix Angulo

Investigador Principal

Línea "Evaluación y currículum para la inclusión"
Centro de Investigación para la Educación Inclusiva

I argued that not only should tests be evaluated in terms of their measurement properties, but that testing applications should be evaluated in terms of their potential social consequences. Samuel Messick (1979)¹

Las pruebas SIMCE² son lo que, en la literatura especializada, se denominan High-Stake Testing y corresponden a tests estandarizados nacionales, que pueden ser interpretados, perfectamente, como test de alto impacto. Su alto impacto se basa en que es una prueba censal, por lo que se administra a todo el alumnado del curso seleccionado en todas las escuelas de Chile y, por ello -al abarcar a *la* población-, supone serias repercusiones para los docentes, los centros escolares y el propio alumnado, que pueden llevar -a través de sanciones-, al cierre de los centros y la ruptura de las relaciones sociales y afectivas entretejidas por el alumnado de la escuela y el consecuente descrédito de los y las docentes. Las sanciones están establecidas en la Ley de Aseguramiento de la Calidad (Ley No. 20.529, Ley SAC) y la Ley de Subvención Escolar Preferencial (Ley No. 20.248, Ley SEP, 2008). En la Ley de Aseguramiento de la Calidad, en su artículo 2º, añade la confluencia de las pruebas externas de carácter censal (el SIMCE) y la rendición de cuentas, añadiendo consecuencias jurídicas y un régimen de sanciones por el incumplimiento de lo establecido y estipulado.

Una prueba psicométrica con tales repercusiones e impactos en el mismo sistema educativo y en la vida de miles de alumnos y alumnas chilenas se convierte en un instrumento con tal poder, que cumple la ley de Campbell (Campbell, 1979), quien estableció el efecto corruptor de los indicadores cuantitativos, enunciado de la siguiente manera:

¹ "Sostengo que no solamente se debería evaluar la capacidad de medición de las pruebas, sino también las potenciales consecuencias sociales de su aplicación". Samuel Messick (1979).

² SIMCE es el acrónimo de Sistema de Medición de la Calidad de la Educación. Las pruebas miden materias como Lenguaje y Comunicación,

"The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social process it is intended to monitor". (Campbell, 1979, p.85; ver además Berliner & Nichols, 2007; Nichols & Berliner, 2005). Pero antes de seguir por esta línea, en la que claramente podemos incluir a las pruebas SIMCE, quisiera detenerme en dos ideas que considero importantes.

Los tests no suponen evaluación

Existen tres conceptos que en castellano se los nombra normalmente con el mismo término, pero, aunque aparezcan relacionados, tienen implicaciones epistemológicas muy distintas. Uno de ellos es el concepto de evaluación (Angulo, 1994, 2019). La evaluación se centra en la calidad de un servicio, un programa social, una escuela o un proyecto curricular, por ejemplo; la calidad viene apreciada a través del análisis de sus componentes y participantes³. La evaluación pretende nutrir el juicio de valor sobre lo evaluado, atendiendo a su complejidad sistémica. Si quisiéramos evaluar una escuela tendríamos que tener en cuenta y analizar su historia, su contexto socio-cultural, sus prácticas pedagógicas, las interrelaciones entre docentes y apoderados, las opiniones del alumnado y, además -y sólo si fuera necesario-, los resultados o logros del alumnado en algunas materias.

En inglés se emplea el concepto de *assessment* justamente para esto último: para denotar los procesos de evaluación del alumnado, tanto formativos como sumativos. En este caso se pueden emplear una infinidad de métodos disponibles (Angelo & Cross, 1993; Greenstein, 2010; Wiggins, 1998).

Matemáticas, Ciencias Naturales, Historia, Geografía y Ciencias Sociales, Inglés y Educación Física. Dichas pruebas suelen aplicarse a 2º, 4º, 6º y 8º Básicos y II y III Medio.

³ Sobre Evaluación ver, por ejemplo: Chelimsky y Shadish (2009); House (1986, 2000); y Ryan y Cousin (2009).

El tercer concepto es el de *testing*, que hace referencia a la aplicación de pruebas estandarizadas. Pensemos que evaluar al alumnado o llevar a cabo un proceso de *assessment*, puede implicar o no la utilización de un test. La evaluación del alumnado puede perfectamente llevarse a cabo sin su empleo. De alguna manera no trivial, en la medida en que la evaluación del alumnado, o el *assessment*, se oriente hacia el concepto de evaluación e intente abarcar la complejidad misma del proceso educativo del docente y del aprendizaje del alumnado, no de manera puntual o meramente terminal (resultados, *attainements* u *outputs*) se amplifica su calidad y profundidad como proceso y en ese sentido su validez formativa. Pero si, por el contrario, el *assessment* se remite y se circunscribe a la aplicación de una o unas pruebas estandarizadas o test, el proceso mismo se degrada, cuando no se desvirtúa (Figura 1). Podemos concluir que querer evaluar un sistema educativo o la labor educativa de un centro de escolar meramente por la aplicación de test, supone claramente una simplificación y, por ello, un despropósito y un grave error epistemológico.



Figura 1
Apreciar la calidad educativa
Fuente: Elaboración propia.

Desgraciadamente es en esta simplificación, reducción y error epistemológico en donde podemos situar a los test SIMCE. Dicho de otra manera, no se puede juzgar a un centro escolar con equidad y validez simplemente aplicando las pruebas SIMCE.

Los tests no representan o pueden representar al contenido curricular

Pero aceptemos que lo único que nos interese *medir* es el logro y los aprendizajes del alumnado en distintas materias. Centrémonos en el curriculum de las materias. Como cualquier educador sabe, el campo de las matemáticas, de la historia, de la geografía, de la biología o de cualquier otra disciplina no siempre tiene límites precisos puesto que junto a los solapamientos disciplinares, todos los campos de conocimiento han ido creciendo considerablemente en los últimos años.

Pues bien, el curriculum nacional realiza una selección de todo un campo disciplinar, en razón de lo que se creen oportuno que 'aprendan' o adquieran las nuevas generaciones y en razón del nivel de desarrollo cognitivo del alumnado. No se puede enseñar logaritmos en primero de básica, como tampoco abordar en profundidad la Ecuación de Schrödinger⁴ al final de la educación media. A esta primera y necesaria selección del contenido curricular, se le añade otra, clave para lo que quiero mostrar. La segunda reducción tiene que ver con el contenido que se selecciona para ser incluido en la prueba estandarizada que aplicamos.

Dicho de otra manera, si el curriculum nacional selecciona el contenido de una materia, las pruebas estandarizadas vuelven a reducir dicho contenido a una serie de preguntas o problemas, los mismos para todo el alumnado. La Figura 2 intenta expresar este razonamiento.

El conocimiento que se mide es distinto y menor -en cantidad- que el conocimiento que se pretende enseñar en la escuela, que es a su vez una selección del conocimiento disciplinar. Así pues, si un alumno o una alumna ha superado la puntuación media en una prueba de matemáticas, sólo podemos afirmar que ha superado la prueba de matemáticas, no que haya adquirido el conocimiento matemático seleccionado para su curso y edad y mucho menos que sepa de matemáticas.

clave en la física cuántica.

4 La Ecuación de Schrödinger describe la evolución temporal de una partícula subatómica masiva de naturaleza ondulatoria y no relativista y es



Figura 2

Del conocimiento medido a lo que es el conocimiento disciplinar

Fuente: Elaboración propia.

Los tests estandarizados, ha enfatizado Zhao (2016, p.6), “miden el grado en que el estudiantado cumple las expectativas de los que han fabricado la prueba o test”. La única salida a esta situación reduccionista es que la prueba posea validez del contenido y del constructo cuando menos. La pregunta sería: ¿la prueba es representativa del contenido curricular? Una pregunta que no puede distraernos de otra igualmente importante: ¿el contenido curricular nacional es representativo del contenido curricular disciplinar? Las respuestas a estas preguntas nos colocan de lleno en los problemas de validez de las pruebas SIMCE (Florez Petour, 2013) y de cualquier otra prueba estandarizada (Nichols & Berliner, 2005).

Validez del contenido y del constructo⁵

La validez del contenido (también llamada content-related evidence of validity) de un instrumento hace mención a si dicho instrumento mide lo que se propone medir. Dicho de otra manera: los ítems del test han de ser vistos como “indicadores confiables o fidedignos del dominio de conocimiento que se pretende medir (assess)” (Nichols &

Berliner, 2005, p.112). Aplicando esta idea al SIMCE nos encontramos que -tal como ha analizado Florez Petour *in extenso*- dichas pruebas pretenden medir un constructo que denominan *calidad* de la educación, pero lo hacen a través de medir una selección de contenidos curriculares (Florez Petour 2013, p.59).

Esta posición presenta algunos problemas de partida. En primer lugar, en los documentos analizados por dicha autora no existe una definición clara y contundente del constructo *calidad* (Florez Petour 2013)⁶. En segundo lugar, la supuesta *validez del constructo* se hace depender de la *validez del contenido*; porque el constructo *calidad* se define por el contenido curricular. Pero de esta manera, se elude enfrentarse a un concepto que resulta enormemente complejo de medir (si no imposible), reduciendo la *calidad* de las escuelas y, por proyección, del sistema educativo, a la *validez del contenido* que se supone que expresa dicha *calidad* (Figura 3).

No sólo la *calidad* se transforma en un asunto de medir con precisión el *currículum*; sino que la *validez del constructo* parece depender de la *validez del contenido*; algo que epistemológicamente podría ser muy cuestionable (Messick, 1998). Tengamos en cuenta que, con este desplazamiento, lo que hacemos es trasladar en el fondo el planteamiento ineludiblemente político e ideológico de lo que sea *calidad* en educación, al planteamiento también político de qué contenido curricular es valioso para ser adquirido por las nuevas generaciones (Angulo, 2015a, 2015b; Apple, 1986, 1993, 2002).

De todas maneras, no podemos escapar a este círculo. Nichols y Berliner (2008, p.112) afirman que la *validez del constructo* es una cuestión mucho más complicada y difícil de cumplir que la del *contenido*, pues supone, al menos, dar respuesta a la pregunta ¿está midiendo el test que usamos las características subyacentes que esperamos medir o está midiendo otra cosa? Se entiende pues que el constructo real

⁵ En este apartado he seguido las penetrantes argumentaciones de Nichols y Berliner (2005), y Florez Petour (2013). Por otro lado, este escrito no acepta la idea de una validez unitaria y menos que dicha unicidad de la validez de los instrumentos sea únicamente la validez del constructo (Florez Petour, 2013; Messick, 1989, 1998). Nichols y Berliner (2005) emplean el término *criterio* (criterion-referenced evidence of validity) en lugar del de

contenido. Hemos preferido usar aquí el de contenido porque da más sentido al hecho de que los criterios son en realidad contenidos curriculares.

⁶ Algo parecido, es decir la ausencia de una definición clara, ocurre con el constructo de *literacy* (Alfabetización) de las pruebas PISA (Carabaña, 2015) y podríamos decir lo mismo del constructo *inteligencia* (Angulo, 2020). Parece que esta indefinición del constructo es una característica de estas pruebas psicométricas.

no sea 'calidad' y que la respuesta directa a qué se entiende por calidad en el SIMCE, sea justamente los contenidos curriculares de las diversas disciplinas: matemáticas, lenguaje y comunicación, historia, geografía, etc. Como hemos visto más arriba, este desplazamiento no suaviza los problemas, sino que aporta los suyos propios, puesto que las pruebas estandarizadas como el SIMCE miden un conjunto enormemente limitado de contenidos; miden, podríamos añadir, una parte sólo de aquello que puede medirse del contenido curricular, olvidando además -como la mayoría de las pruebas estandarizadas vigentes- los aspectos menos cognitivos y más cercanos con la creatividad, la imaginación, la curiosidad, el pensamiento crítico y divergente, la resolución de problemas, etc. (Emler et al., 2019; Zhao, 2016).

El SIMCE, como otras pruebas estandarizadas muestra una acusada tendencia a simplificar lo medible a aquello que puede ser medido y con ello a homogeneizar al estudiantado en las materias medidas (Zhao, 2016, p.5). Como señala Florez Petour

el constructo se reduce más todavía cuando en la mayoría de los documentos disponibles se indica que el SIMCE solamente incluye aquellos ámbitos de aprendizaje que pueden evaluarse por medio de una prueba de lápiz y papel... lo que en realidad se evalúa en el SIMCE es finalmente una versión

bastante reducida, rutinaria y básica del currículum de cada área⁷. (Florez Petour, 2013, p.60).

De esta manera, se transita el camino técnicamente más fácil para asegurarse la supuesta robustez psicométrica a costa de ofrecer una versión disminuida, parcial y reducida no sólo del contenido cultural, sino de las capacidades cognitivas del alumnado. Como advirtió Messick (1998, p.41): "no se puede restringir el significado de la validez del constructo a la exactitud de las puntuaciones inferidas". Esta problemática puede complicarse todavía mucho más y en una dirección no trivial. Aceptemos que SIMCE mide el constructo 'matemáticas'. ¿Qué ocurre cuando a al alumnado de un centro se le adiestra en la preparación del test de matemáticas? Algo realmente posible dado que se trata de test de lápiz y papel. En ese momento estamos frente a otro 'constructo': el 'conocimiento curricular de matemáticas en alumnado que ha sido adiestrado'. Denominemos Cp al constructo principal ('matemáticas'), al segundo podremos denominarlo Cp1, de tal manera que $Cp \neq Cp1$. Podríamos incluso continuar incluyendo otros constructos, como por ejemplo 'conocimiento curricular de matemáticas en alumnado que ese día no ha desayunado' (Cp2) o 'conocimiento curricular de matemáticas en alumnado que acaba de recuperarse de una gripe' (Cp3), etc. $Cp \neq \{Cp1, Cp2, Cp3...Cpn\}$.

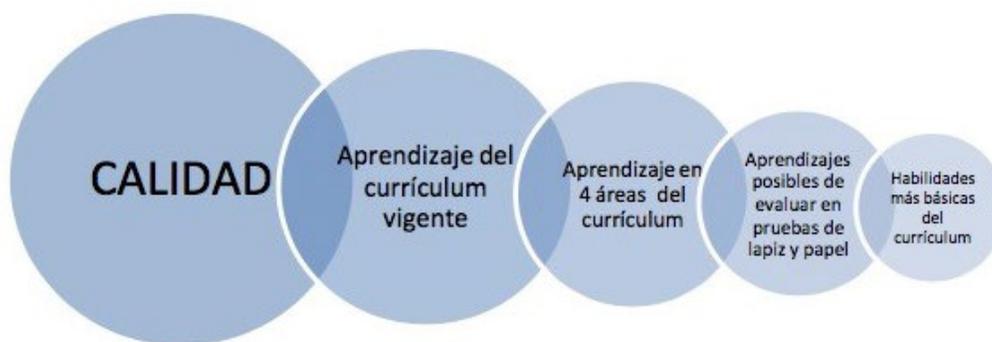


Figura 3

Reducción de los límites del constructo a evaluar en SIMCE en base a evidencia

Fuente: Florez Petour (2013, p.63).

⁷ Florez Petour (2013, p.61) añade que "las preguntas que tienden a comportarse mejor en el proceso de validación psicométrica son aquellas

de formulación más simple y que están ligadas a habilidades más básicas, lo que refuerza la existencia de un filtro hacia lo más elemental de cada área".

El mismo test, la misma prueba elaborada para ser válida con respecto al constructo primero, no lo sería con respecto al constructo segundo, ni a los otros. Esto quiere decir que las 'variables extrañas intervinientes' son parte del constructo y no externas al mismo. Cuantas más variables extrañas converjan, más cambiante y más lábil es el constructo. Pensar que podemos controlarlas o limitarlas es una fantasía psicométrica. La vida real, la cultura, los acontecimientos sociales y políticos son parte del entramado en el que la educación, los centros escolares, el profesorado y el alumnado (con su aprendizaje) se mueve y vive.

Por ello comparar los resultados sería como comparar constructos sensiblemente distintos; no podríamos de ninguna manera, como hace el SIMCE, jerarquizar centros de enseñanza en relación a sus resultados en matemáticas si en algunos de ellos, repito, se ha adiestrado al alumnado, o sus alumnos están mal alimentados, o carecen de contención afectiva en sus casas, o alguno de sus apoderados está en la cárcel. Esta comparación carecería de equidad y sería injusta y ninguna argucia psicométrica podría hacerle frente o mitigarla. Pero aquí nos adentramos en el terreno, siempre olvidado, en las pruebas estandarizadas y los test, de la *validez consecucional*.

En conclusión, las pruebas SIMCE son un pálido reflejo del conocimiento cultural y del conocimiento curricular. Intercambian la validez del constructo por validez del contenido y asumen una visión simplista y equívoca con respecto a un constructo que no queda claramente definido si no es por substitución de lo que supuestamente sería el elemento a considerar en la validez del contenido, i.e. los aprendizajes medibles.

Aproximación a la validez consecucional

If consequences are not part of the validation process, many sources of invalidity will remain unexposed to the detriment of the science and practice of educational and psychological measurement (Messick, 1998, p.43).⁸

La validez consecucional, aun siendo uno de los

criterios de validez menos aceptados en el campo de la psicometría (Messick, 1998), es clave para todo test puesto que las consecuencias y decisiones asociadas con las puntuaciones, el uso social intencionado y las repercusiones de su uso son una cuestión vital y social y políticamente trascendentales (Messick, 1979; Nichols & Berliner, 2005).

La pregunta por el uso y propósito de un test debe responderse tanto en base a evidencia como en base a imperativos éticos, es decir, se debe evaluar no solamente si el test sirve para un propósito determinado, sino también si la prueba debe o no utilizarse con dicho propósito. Esta dimensión, entonces, implica una evaluación de las posibles consecuencias de una prueba, aun cuando se trata de una tarea sumamente compleja, sobre todo en relación con sus efectos no intencionados (Florez Petour, 2013, p.23).

La validez consecucional está asociada a la Ley de Campbell, como lo está todo test de alta repercusión (high-stakes test) como la prueba SIMCE, pero por ello, está sujeta a imperativos éticos y políticos que aquí tampoco pueden solaparse por la técnica psicométrica. Zhao (2018) y Emler et al (2018) utiliza el concepto de *side effects o efectos secundarios*, que es muy similar al de efectos consecucionales o consecuencias de un test. Para estos autores los *side effects* son efectos negativos que las pruebas de alto impacto tienen, en razón de que distorsionan la educación.

Una distorsión que sucede a diversos niveles:

- Haciendo creer que los resultados en las pruebas representan la calidad de la educación.
- Cuando el curriculum se convierte en la preparación para el test, reduciéndolo y estrechándolo; con lo que se distorsiona tanto el sentido de las materias como el de la cultura misma.
- Cuando la instrucción se convierte en instrucción y preparación para el test, privando al alumnado de la adquisición de otras habilidades y conocimientos mucho más importantes para su futuro y el de la sociedad en la que viven (Emler et al, 2019; Nichols & Berliner, 2005; Zhao 2018).

⁸ "Si las consecuencias no forman parte del proceso de validación, muchas fuentes de invalidez permanecerán ocultas en detrimento de la ciencia y

práctica de la medición educacional y psicológica" (Messick, 1998, p.43).

Consecuencias políticas de las pruebas SIMCE

Pero estas consecuencias de las pruebas estandarizadas son todavía mucho más graves si tenemos en cuenta las pruebas SIMCE. Como ya hemos indicado, dado que dichas pruebas están engarzadas con los sistemas de Aseguramiento de la Calidad y de Aseguramiento de la Subvención Escolar, tenemos la tormenta perfecta. En este caso el impacto de las pruebas afecta a la vida del alumnado y del profesorado además de la vida del propio centro. La amenaza de cierre o la retirada de subvención, por el no cumplimiento de la supuesta calidad, tiene unas repercusiones enormes y terribles.

Este es un terreno especialmente político porque necesitaríamos preguntarnos qué sentido último tienen este tipo de pruebas en el sistema educativo chileno. ¿Por qué han permanecido tanto tiempo y a través de todos los gobiernos de la democracia? Hay una respuesta clave pero incompleta y tiene que ver con la perpetuación de las élites tecnocráticas y su alianza con las élites políticas que se amparan en una cierta meritocracia, que en realidad no es más que *darwinismo social* (Chorover, 1982; Gould, 1984). El sistema, en su interpretación política, parece querer mostrar que permite seleccionar aquellos centros y alumnado cuyo esfuerzo va a permitir que puedan situarse social y económicamente donde sus méritos le lleven. Pero esta creencia está lejos de ser cierta⁹. En Chile el ascensor social se encuentra paralizado según la OECD (2018), pues se necesitan seis generaciones para ascender de clase. Esta parálisis, en la que también se incluye varios países latinoamericanos, se centra en los extremos

socioeconómicos: son los más ricos y los más pobres los más propensos a permanecer como tales¹⁰. No obstante, la probabilidad de que los hijos e hijas de los cuadros directivos se incorporen a puestos de trabajo como el de sus padres y permanezcan en sus clases sociales de origen, es mucho mayor que la posibilidad de que los hijos de padres obreros o clases vulnerables puedan ascender (OECD, 2018)¹¹.

En este contexto una prueba con el impacto político del SIMCE, conlleva validar una meritocracia ficticia, que extiende y confirma año tras año la división de clases sociales a través de la educación, con el agravante de lo que Sandel (2020) denomina *soberbia meritocrática*:

las meritocracias generan unas actitudes morales reprobables entre los que llegan a la cima. Cuando estamos convencidos de que hemos triunfado sólo gracias a nuestro esfuerzo, es poco probable que nos sintamos en deuda con nuestros conciudadanos. El énfasis implacable en que hay que ascender y prosperar empuja a los triunfadores a emborracharse con su éxito y despreciar a quienes no tienen esas credenciales meritocráticas (Sandel, 2020).

El SIMCE existe esencialmente como un recordatorio de que en la sociedad chilena cada uno ocupa el lugar que le corresponde como si, utilizando los conceptos de Esposito (2003, 2005)¹², estuviéramos levantando una muralla inmunológica inmunitas-entre las élites y el resto de las clases sociales. No olvidemos, y ésta sería una segunda razón para su permanencia, que el mismo SIMCE es un emprendimiento y un ámbito de lucro, como otros tantos que existen en Chile (Mönckeberg, 2015).

9 Ver los siguientes estudios internacionales -entre otros- cuyas conclusiones desmontan esta falacia: Berliner (2006); Lee & Burkam (2002); y Tranter (2012).

10 Estos resultados confirman la idea del Palma Ratio, según la cual la desigualdad medida por el índice GINI no es sensible al hecho de la estabilidad de las clases y grupos intermedios. La *Ratio de Palma* propone apreciar la desigualdad comparando la distancia, por así decir, entre los extremos, país por país: las clases populares y vulnerables y las clases corporativas y ricas (Palma, 2011). Por otro lado, Chile es uno de los países con mayor concentración de la riqueza y el segundo país de la OCDE con la mayor brecha de ingresos entre los extremos socioeconómicos: 10% más rico y 10% más pobre (OCDE, 2019). Como han mostrado Pérez y Sandoval (2020) en Chile el 1% más rico captura cerca del 17% de los ingresos fiscales, mientras que el 10% más rico percibe más del 50% de todos los ingresos. Sobre la *desigualdad acumulada* ver Jackson (2015).

11 Uno de los factores en los que esta diferencia podría encontrarse

es en el vocabulario. Como Hart y Risley (2003) han demostrado el contexto socio-cultural determina poderosamente la potencia y calidad del lenguaje, hasta el extremo de que a los 4 años la diferencia de vocabulario entre clases sociales puede llegar a ser de 30 millones de palabras.

12 En Esposito, *communitas* e *immunitas* están inter-penetrados uno en el otro. Si *communitas* supone voluntad de donación al otro, la *immunitas* exonera de tal donación; si *communitas* remite a algo 'colectivo' y abierto, la *immunitas* remite a lo particular, como algo que 'se sustrae a la condición común'; si *communitas* señala a la ruptura de barreras protectoras de la identidad individual, la *immunitas* es un intento de reconstrucción de las mismas, de restitución de muros y separaciones; si la *communitas* conlleva, al salir al otro a restituir la donación, el peligro del contagio, la *immunitas* nos asegura indirectamente la protección individual, la defensa contra todo elemento externo que amenace o nos amenace con su contagio (Esposito, 2009, p.17; Redon Angulo & Vallejos, 2015).

Como ha informado Ciper Chile, el SIMCE desde 2013 ha movido \$29M de presupuesto público, permitiendo además que empresas externas -según contraloría- ejerzan las labores asignadas a la Agencia de Calidad, lo que se denomina 'traspaso de potestades públicas' como medio de lucro. Empresas coludidas y relacionadas entre sí y con lazos con la Agencia de Calidad como CGS Logística, Seminarium Certificación Limitada, Ubix Tecnología Limitada y TATA Consultancy Services (TCS), participan o han participado de alguna manera en la aplicación, supervisión o interpretación de las pruebas SIMCE. En un país tan fuertemente neoliberal como Chile en donde la pseudo-meritocracia, el emprendimiento y el lucro constituyen el eje de la

subjetividad, no debería resultarnos extraño su permanencia durante tantas décadas, ni que las élites políticas se hayan mostrado tan condescendientes con esta prueba¹³; tanto que ni siquiera hayan pensado seriamente en convertir el SIMCE en una prueba muestral *diagnóstica*, algo que, psicométricamente, no depende de la transformación de los instrumentos¹⁴, sino de un adecuado proceso de muestreo. Pero en este caso, la información que generaría no podría ser utilizada para el control inmunitario, que es lo que en el fondo se pretende (Falabella, 2020). Las pruebas SIMCE no tienen por función servir para la toma de decisiones políticas o para mejora de la educación sino para para estigmatizar, discriminar y controlar a escuelas, docentes e individuos.

13 La única crítica institucionalizada ha sido el movimiento Alto al SIMCE. Ver <http://www.altoalsimce.cl/>

14 Aunque los instrumentos tendrían que ser mejorados considerablemente.

Referencias

- Angelo, T. A. & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. Jossey-Bass.
- Angulo Rasco, J. F. (1994). ¿A qué llamamos evaluación? Las distintas acepciones del término evaluación o por qué no todos los conceptos significan lo mismo. En J. F. Angulo Rasco & N. Blanco, (Comps.), *Teoría y desarrollo del currículum* (pp. 283-296). Aljibe.
- Angulo Rasco J. F. (2015a). Un enfoque naturalista sobre los contenidos. En J. Gimeno Sacristán, (Comp.), *Los contenidos: Una reflexión necesaria* (pp. 103-115). Morata.
- Angulo Rasco, J. F. (2015b). In search of the lost curriculum. In J. Paraskeva & S. Steinberg (Eds.), *Curriculum: Decanonizing the field* (pp. 137-155). Peter Lang.
- Angulo Rasco, J. F. (2019). *Evaluación y estándares en educación: Legitimidad y equidad*. Serie Por una Educación Inclusiva, 2, Centro de Investigación para la Educación Inclusiva.
- Angulo Rasco, J. F. (2020). Inteligencia. En R. Espinoza Lolas & J. F. Angulo Rasco, (Coords.), *Conceptos para pensar la educación capitalista*. Terra Ignota. (En Prensa).
- Apple, M. W. (1986). *Ideología y currículo*. Akal.
- Apple, M. W. (1993). *Official knowledge: Democratic education in a conservative age*. Routledge.
- Apple, M. W. (2002). *Educar "como Dios manda": Mercados, niveles, religión y desigualdad*. Paidós.
- Berliner, D. C. (2006). Our impoverished view of educational reform. *Teachers College Record*, 108, 949-995.
- Berliner, D. C. & Nichols, S. L. (2007). High-stakes testing is putting the nation at risk. *Education Week*, 12/03/.
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, 2, 67-90. [https://doi.org/10.1016/0149-7189\(79\)90048-X](https://doi.org/10.1016/0149-7189(79)90048-X)
- Carabaña, J. (2015). *La inutilidad de PISA para las escuelas*. Catarata.
- Chelmsky, E. & Shadish, W. R. (1997). *Evaluation for the 21 Century: A handbook*. SAGE.
- Chorover, S. L. (1982). *Del génesis al genocidio: La socio biología en cuestión*. H. Blume Ediciones.
- Emler, T. E., Zhao Yong, D., Jiayi, Yin, D., & Wang, Y. (2019). Side effects of large-scale assessment in education. *ECNU Review of Education*, 2(3), 279-296. <https://doi.org/10.1177/2096531119878964>
- Esposito, R. (2003). *Communitas: Origen y destino de la comunidad*. Amorrortu.
- Esposito, R. (2005). *Immunitas: Protección y negación de la vida*. Amorrortu.
- Falabella, A. (2020). The seduction of hyper-surveillance: standards, testing and accountability. *Educational Administration Quarterly*. <https://doi.org/10.1177/0013161X20912299>
- Florez Petour, T. (2013). *Análisis crítico de la validez del Sistema de Medición de la Calidad de la Educación (SIMCE)*. Consejo Nacional de Educación (Chile).
- Gould, S. J. (1984). *La falsa medida del hombre*. Antoni Bosch.
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. ASCD.
- Hart, B. & Risley, T. R. (2003). The early catastrophe: The 30-million-word gap. *American Educator*, 27(1), 4-9.
- House, E. R. (Ed.). (1986). *New directions in educational evaluation*. The Falmer Press.
- House, E. R. (2000). *Evaluación, ética y poder*. Morata.
- Jackson, M. I. (2015). Cumulative inequality in child health and academic achievement. *Journal of Health and Social Behavior*, 56(2), 262-280.
- Lee, V. E. & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Institut.
- Messick, S. (1979). Test validity and the ethics of assessment. *ETS Research Report Series*, 1, 1-43.
- Messick, S. (1989). Validity. In: R. L. Linn (Ed.). *Educational measurement (3rd ed.)*, (pp. 13-103). American Council on Education/Macmillan.
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45, 35-44.
- Mönckeberg, M. O. (2015). *El saqueo de los grupos económicos al Estado Chileno*. DeBolsillo.
- Nichols, S. L. & Berliner, D. C. (2005). *The inevitable corruption of indicators and educators through high-stakes testing*. The Great Lakes Center for Education Research & Practice.
- OECD. (2018). *A broken social elevator? How to promote social mobility*. OECD Publishing.
- OECD. (2019). *Society at a glance 2019: OECD social indicators*. OECD Publishing.
- Palma, J. G. (2011). Homogeneous middles vs. heterogeneous tails, and the end of the 'inverted-u': It's all about the share of the rich. *Development and*

Change, 42(1), 87-153.

Pérez, R. & Sandoval, D. (2020). La geografía de la desigualdad y el poder. CIPER Chile 26/02/2020.

<https://ciperchile.cl/2020/02/26/la-geografia-de-la-desigualdad-y-del-poder/>

Redon, S., Angulo Rasco, J. F., Vallejos, N. (2015). El sentido de lo común como experiencia de construcción democrática: Estudio de casos en escuelas en contextos de pobreza en Chile. *Archivos Analíticos de Políticas Educativas*, 23, 13.

Ryan, K. E. & Cousins, J. B. (2009). *The SAGE international handbook of educational evaluation*. SAGE.

Sandel, M. (2020). ¿Estamos todos juntos en esto? El País, mayo 01. <https://elpais.com/opinion/2020-05-01/estamos-todos-juntos-en-esto.html>

Tranter, D. (2012). Unequal schooling: How the school curriculum keeps students from low socio-economic backgrounds out of university. *International Journal of Inclusive Education*, 16(9), 901-916.

Zhao, Y. (2016). The danger of misguiding outcomes: Lessons from Easter Island. In Y. Zhao (Ed.), *Counting what counts: Reframing education outcomes* (pp. 1-12). Solution Tree Press.

Zhao, Y. (2018). *What works may hurt: Side effects in education*. Teacher College Press.

Centro Eduinclusiva

Buscamos desarrollar investigación transdisciplinaria sobre los factores individuales, organizativos y de política necesarios para lograr una educación inclusiva de calidad; apoyar mediante la propuesta y la transferencia de intervenciones y estrategias basadas en la evidencia para una educación inclusiva; y formar capital humano avanzado para la investigación en educación inclusiva.

Serie Policy Briefs

El Área de Incidencia en Política Pública del Centro Eduinclusiva desarrolla una serie de Policy Briefs, los cuales hemos denominado “Propuestas para Políticas Educativas”, los cuales tienen por objetivo aportar al debate público de los distintos actores sociales en pos de la transformación del sistema educativo hacia la inclusión y la calidad.



Av. El Bosque 1290, Viña del Mar
Campus Sausalito PUCV
ciecomunicaciones@pucv.cl
32 - 237 2575