

**Las evaluaciones internacionales a gran escala
y la regulación global de los sistemas educativos:
un análisis integrativo ***

**Avaliações internacionais em larga escala
e o regulamento global de sistemas educacionais:
uma análise integrativa**

***International Large-Scale Assessments
and the Global Regulation of Educational Systems:
An Integrative Analysis***

Claudio Ramos Zincke **

Desde la década de 1990, se han expandido a través del mundo las evaluaciones internacionales a gran escala del desempeño educacional. Aunque con orígenes diferentes, ellas han ido dando forma a un dispositivo técnico que se conecta y superpone a los gobiernos nacionales e incide, de formas diversas, en la regulación global y nacional de la educación. Este artículo se focaliza en responder a dos preguntas: 1) cuál es la fuente de su fuerza expansiva y regulatoria; y 2) qué es lo que estos mecanismos evaluativos producen. Las respuestas aportadas confluyen en mostrar que estos sistemas de evaluación internacional constituyen un dispositivo de gubernamentalidad global que está rediseñando el funcionamiento de los sistemas de educación a través del mundo, estableciendo nuevos estándares normativos y redefiniendo las nociones de calidad, abriendo los sistemas educacionales a una revisión y discusión en un marco de referencia global. Tal dispositivo se sustenta en una estructura operativa basada en redes internacionales, en una labor investigativa a escala de *big science* y orientada por una creciente preocupación práctica por el diseño de políticas educacionales. En las últimas dos décadas, este dispositivo ha adquirido una gran potencialidad adaptativa que continuamente está ampliando su fuerza regulatoria y performativa.

159

Palabras clave: evaluaciones internacionales a gran escala; regulación global; *big science*; gubernamentalidad

* Recepción del artículo: 17/04/2020. Entrega de la evaluación final: 25/11/2020. El artículo pasó por una instancia de corrección y reevaluación.

** Sociólogo. Profesor titular del Departamento de Sociología de la Universidad Alberto Hurtado, Chile. Correo electrónico: cramos@uc.cl.

Desde a década de 1990, as avaliações internacionais em larga escala do desempenho educacional se expandiram em todo o mundo. Embora com origens diferentes, eles têm moldado um dispositivo técnico que conecta e se sobrepõe aos governos nacionais e influencia, de diferentes maneiras, a regulamentação global e nacional da educação. O artigo concentra-se em responder a duas perguntas: 1) qual é a fonte de sua força expansiva e regulatória; e 2) que esses mecanismos avaliativos produzem. As respostas fornecidas convergem ao mostrar que esses sistemas internacionais de avaliação constituem um dispositivo de governamentalidade global que está redesenhando a operação dos sistemas educacionais em todo o mundo, estabelecendo novos padrões normativos e redefinindo as noções de qualidade, abrindo os sistemas educacionais a revisão e discussão em um quadro global de referência. Esse dispositivo é sustentado por uma estrutura operacional baseada em redes internacionais, por trabalhos de investigação científica em grande escala e orientada por uma crescente preocupação prática pelo desenho de políticas educacionais. Nas últimas duas décadas, esse dispositivo adquiriu um grande potencial adaptativo que está continuamente expandindo sua força regulatória e performativa.

Palavras-chave: avaliações internacionais em larga escala; regulamentação global; *big science*; governamentalidade

Since the 1990s, international large-scale assessments of educational performance have expanded throughout the world. Although with different origins, they have given shape to a technical device, both connected with national governments and overlapping them, that affects in various ways the global and national regulations of education. The article focuses on answering two questions: 1) what is the source of their expansive and regulatory force; and 2) what do these evaluation mechanisms produce. The answers provided converge in showing that these international evaluation systems constitute a global governmentality device that is redesigning education systems' functioning throughout the world, establishing new normative standards and redefining the notions of quality, opening educational systems to review and discussion in a global frame of reference. Such a device is grounded on an operational structure based on international networks, on research to a big science scale, and guided by a growing practical concern for designing educational policies. In the last two decades, this device has acquired great adaptive potential that is continuously expanding its regulatory and performative force.

Keywords: international large-scale assessments; global regulation; *big science*; governmentality

Introducción

Las pruebas internacionales que evalúan el desempeño escolar —PISA y TIMSS, entre las más conocidas— se han convertido en parte del panorama educacional global. Sus resultados, entregados periódicamente, son esperados, debatidos y usados de diferentes maneras en los sistemas educacionales. Sus alcances han llegado decisivamente a América Latina conectándola a una nueva forma de regulación global. Pese a su relevancia, en esta región tales pruebas no han sido objeto de la investigación y el debate que han tenido en los países centrales.¹ Para orientar o impulsar más investigación y debate en la región se requieren trabajos que sistematicen la vasta acumulación de trabajos sobre estas pruebas, sobre las cuales cuesta tener una visión de conjunto, y permitan tener un cuadro parsimonioso de aspectos centrales del debate, desde un punto de vista científico social. Tal es el fin de este artículo: discernir y ordenar elementos centrales de la operación, forma institucional y efectos de las pruebas internacionales, a través de una sistematización de la literatura. El foco del artículo es la regulación global que se genera a través de estas formas de evaluación, pero observada desde lo que es pertinente para América Latina. Desde esta perspectiva, algunas evaluaciones aplicadas solo en la región, como la prueba ERCE, adquieren importancia y otras, como la prueba SACMEQ aplicada en África o la PASEC, usada para los países de habla francesa, no son consideradas.

Un precedente de estos estudios comparativos internacionales tomó lugar en los años 1930, a través de la International Examinations Inquiry (IEI), en la que participaron nueve países, todos del Primer Mundo, entre ellos Estados Unidos, Inglaterra, Francia y Alemania. Su foco investigativo fue “la determinación del modo más efectivo de examinar a los estudiantes para su ingreso a la escuela secundaria” (Lawn, 2008, p. 7) y los problemas principales abordados tuvieron que ver con el establecimiento de métodos estandarizados de medición, la comunicación académica internacional y la comprensión comparativa de la diversidad nacional. Su financiamiento fundamentalmente provino de la Carnegie Foundation. La IEI fue un proyecto científico que operó internacionalmente, integrando a unos cien investigadores eminentes de Estados Unidos y Europa, quienes intentaron definir problemas comunes generando proyectos asociados y publicaciones, aunque no un proyecto investigativo o evaluativo común. El proyecto expresaba confianza en los esfuerzos de expertos científicos realizados sin control gubernamental ni conexiones con los gobiernos. En su interior convivieron diferentes enfoques y finalmente la Segunda Guerra Mundial impidió su continuación (Lawn, 2008).

Las evaluaciones internacionales a gran escala (EIGE) del desempeño educacional entran a escena, aunque todavía de modo preparatorio, con un proyecto piloto llevado a cabo entre 1959 y 1961 por un nuevo organismo, la International Association for the Evaluation of Educational Achievement (IEA), apoyada por la Unesco. En él se evaluaron los logros educacionales en varios dominios cognitivos (comprensión

1. Países “centrales” en el sentido que le asigna Wallerstein.

lectora, matemáticas, ciencias, lenguaje y habilidad no verbal) en nueve países (Bélgica, Inglaterra, Francia, Finlandia, Israel, Grecia, Suiza, Estados Unidos y Yugoslavia), usando muestras intencionales. Sirvió para probar la factibilidad de los estudios comparativos internacionales del desempeño educativo (Keeves, 2011).

Luego de ese ensayo inicial, los tests educacionales a gran escala propiamente comienzan en 1964, con el *First International Mathematics Study* (FIMSS), aplicado por la IEA en 12 países. En esta prueba se emplearon muestras probabilísticas y se plantearon hipótesis sobre la incidencia de la organización de la escuela y de diversos factores sociales y económicos sobre los logros en esta disciplina escolar (Keeves, 2011, p. 7), lógica explicativa que continuará en las pruebas futuras.

Las EIGE han sido objeto de variedad de análisis que buscan hacer sentido de lo que representan. Se ha considerado, especialmente en referencia a PISA, que juegan un rol en el avance de “un nuevo modo de gobernanza educacional global” (Meyer y Benavot, 2013), dando forma a un *global eye* (Novoa *et al.*, 2003), y que constituyen “la expresión de una matriz cognitiva de gobierno” (Carvalho, 2018) y de un *soft power* (Sellar y Lingard, 2013), en el cual el conocimiento científico técnico juega un rol crucial. Se puede entender a estos sistemas de evaluación como parte de la modernidad reflexiva, constituyendo un medio global que ayuda a las sociedades a revisar el ajuste de sus sistemas educacionales a los veloces cambios en los procesos sociales y culturales, y en su propio entramado institucional. En esa perspectiva, diversos enfoques conciben a las EIGE como una nueva herramienta de gobierno, hablando los autores, así, de gobernanza por indicadores (Gorur, 2017), *network governance* (Ball y Junemann, 2012), gobierno por números (Grek, 2009), gobernanza por comparación (Materns y Nieman, 2010) y gobierno a través de “vigilancia mutua” (Carvalho, 2018, p. 112). Lingard y Sellar (2016) hablan de una gobernanza epistemológica que conecta lo cognitivo y lo normativo. Se las considera, además, como parte de la economía global del conocimiento (Exly, Bran y Bell, 2011), contribuyendo a la operación de *policy networks* (Ball y Junemann, 2012).

En concordancia con varios elementos de esas interpretaciones, en este artículo se aborda a las EIGE como un ensamblaje internacional de elementos heterogéneos, el cual monta un proceso investigativo que integra elementos cognitivos, normativos y de poder, generando datos comparativos, y se analizan las derivaciones performativas de su operación. Aunque una visión positivista convencional aún siga diciendo que las pruebas internacionales miden algo que estaba allí esperando ser medido y revelado, desde los enfoques de la teoría del actor-red y de la performatividad de la ciencia, la actividad evaluativa es vista como una operación productiva (Callon, 1998; MacKenzie, 2006; Muniesa, 2014).² Los procedimientos de medición, con sus conceptualizaciones, diseños metodológicos, instrumentos y análisis, seleccionan particulares elementos (y no otros), los articulan de particulares formas y los someten a particulares interpretaciones. Cuando esta medición además está articulada con

2. Hay algunos autores que han estudiado las pruebas desde estos enfoques. La más destacada es Gorur (2011, 2016a, 2018), quien ha centrado su atención particularmente en la prueba PISA.

instituciones estatales, los resultados de la medición son usados prácticamente como manifestaciones de una realidad unívoca, guiando “con evidencia fáctica” las decisiones. El aparataje científico técnico, en su articulación estatal, se convierte así en un dispositivo con gran fuerza performativa.

El objetivo aquí es discernir algunos elementos centrales de esa red productiva que sustenta y provee la infraestructura de estas pruebas, y luego explorar esa labor “productiva”. Dado el carácter internacional del ensamblaje y dada la complejidad del tejido, que incluye variado tipo de entidades, lo que se puede mostrar en estas breves páginas es solo una versión estilizada de las conexiones, dejando fuera varias de ellas. Lo que se busca aquí es visualizar el cuadro general que se oculta tras la proliferación de información circulante.

Específicamente, este artículo busca responder dos preguntas sobre estas pruebas: 1) ¿cuál es la fuente o los factores explicativos de su fuerza expansiva y de la capacidad regulatoria que están exhibiendo?; y 2) ¿qué es lo que ellas producen?, ¿cuál es su efecto performativo? Estas interrogantes están planteadas no con respecto a una prueba en particular, sino que, con respecto al conjunto de ellas, pues nuestro planteamiento es que lo que se configura y los efectos que se producen derivan de su operación combinada.

La metodología del estudio se basó en una revisión de los informes técnicos y las publicaciones de la IEA, Unesco y OCDE, en referencia a sus pruebas internacionales, y en la revisión de más de 200 textos de la literatura especializada sobre las EIGE, seleccionados de acuerdo con un muestreo estratégico que buscó reflejar la diversidad de contenido dentro de tales textos, a fin de obtener una representatividad cualitativa. Los textos fueron objeto de un trabajo de análisis e interpretación guiado por las preguntas planteadas.

163

Primero se hace una breve revisión sobre el crecimiento histórico de las EIGE y una caracterización básica de las principales pruebas. Segundo, se explican diversas facetas del fenómeno de las EIGE que permiten entender su fuerza y alcance, las cuales están relacionados con su proceso productivo y estructura organizativa. Tercero, se describe lo que, de acuerdo con las investigaciones y nuestra interpretación, estas pruebas generan. Esto remite tanto a sus productos directos como a una cadena de efectos que inciden en los sistemas educacionales. Finalmente, se extraen conclusiones sobre el carácter del dispositivo evaluativo en marcha y sobre el alcance y la relevancia de sus efectos.

1. La expansión mundial de las evaluaciones educacionales internacionales a gran escala (EIGE)

Las EIGE tienen una historia que comienza con los tempranos esfuerzos de la IEI en la década de 1930, que se ven detenidos por la Segunda Guerra Mundial. Se reactivan en los años 1960, avanzando lentamente durante los 1970 y 1980. Como se observa en la siguiente tabla, su despliegue global comienza en los años 1990, masificándose desde los 2000. De América Latina, Chile es el primer país que participa

en ellas, involucrándose tempranamente, a fines de los años 1960, en la *Six Subject Survey*, una evaluación pionera de la IEA. En 1998, la oficina regional de la Unesco crea una prueba para esta región y con ello comienza la participación generalizada de los países latinoamericanos. La primera prueba PISA, de la OCDE, en el 2000, marca un salto global en la repercusión de las EIGE, al instalar una modalidad más sistemáticamente conectada al diseño de políticas y a la difusión en los medios de comunicación masiva. En 2009 ya hay 65 países que aplican la prueba PISA (ocho de América Latina) y en 2011 son 63 los que aplican TIMSS (dos de América Latina). El 2015, 72 países participan en PISA. Nada más considerando el período entre 2009 y 2011, hay más de 1.500.000 estudiantes rindiendo pruebas internacionales. Hoy, estas pruebas son fuente básica de información sobre su realidad educativa para los ministerios de educación del mundo y para las audiencias nacionales.

Tabla 1. Evaluaciones internacionales a gran escala (EIGE), ordenadas por año de aplicación, con países participantes y cantidad de evaluados

ORGANISMO A CARGO	PRUEBA	AÑO	NÚMERO DE PAÍSES	PAÍSES LATINOAMERICANOS PARTICIPANTES	NÚMERO DE INDIVIDUOS EVALUADOS (muestra global aproximada)
IEA	FIMSS	1964	12	-	133.000
IEA	SSS	1970	19	CH	250.000
IEA	SIMS	1982	10	-	124.000
IEA	SISS	1984	19	-	262.200
ETS	IAEP	1991	20	BR	175.000
OCDE	IALS	1994	8	CH	68.800
IEA	TIMSS	1995	45	CO	280.000
OCDE	IALS	1998	22	CH	68.800
UNESCO	ERCE-1	1998	11	AR, BO, BR, CH, CO, CU, HO, MX, PAR, RD, VE	54.500
IEA	CIVED	1999	28	CH	140.000
IEA	TIMSS	1999	38	CH	180.000
OCDE	PISA	2000	43	AR, BR, CH, ME, PE	223.000
IEA	PIRLS-1	2001	34	AR (Buenos Aires), CH	319.000
OCDE	PISA	2003	41	BR, MX, UR	276.000
IEA	TIMSS	2003	49	AR, CH,	355.000
OCDE	PISA	2006	57	AR, BR, CH, CO, MEX, UR	400.000
UNESCO	ERCE-2	2006	16	AR, BR, CH, CO, CR, CU, EC, ES, GU, MX, NI, PAN, PAR, PE RD, UR	196.000
IEA	PIRLS	2006	41	CH	215.000
IEA	TIMSS	2007	59	CH, CO, SA	425.000
OCDE	TALIS	2008	24	CH	70.000
IEA	TEDS-M	2008	17	CH	22.000
IEA	ICCS	2009	38	CH, MX, PAR, RD	140.000
OCDE	PISA	2009	65	AR, BR, CH, CO, MX PE, UR, TT	470.000
IEA	TIMSS	2011	63	CH, HO	600.000

ORGANISMO A CARGO	PRUEBA	AÑO	NÚMERO DE PAÍSES	PAÍSES LATINOAMERICANOS PARTICIPANTES	NÚMERO DE INDIVIDUOS EVALUADOS (muestra global aproximada)
IEA	PIRLS	2011	48	CH, CO, HO	325.000
OCDE	PISA	2012	65	AR, BR, CH, MX, PE, CO, CR	510.000
OCDE	TALIS	2013	33	BR, CH, MX	170.000
UNESCO	ERCE-3	2013	15	AR, BR, CH, CO, CR, EC, GU, HO, MX, NI, PAN, PAR, PE, RD, UR	117.000
IEA	ICILS	2013	18	AR, CH	60.000
IEA	TIMSS	2015	57	AR, CH	580.000
OCDE	PISA	2015	72	AR, BR, CH, CO, MX, PE, CR, RD, TT, UR	540.000
OCDE	PIAAC	2015	25+88 (2 rounds)	CH	215.000
IEA	PIRLS	2016	49	AR, CH	300.000
IEA	ICCS	2016	24	CH, CO, CR, RD, MX, PE	94.000
OCDE	PISA	2018	80	AR, BR, CH, CO, CR, EC, PAR, PE, RD, TT, UR	500.000
OECE	TALIS	2018	49	AR, BR, CH, CO, CR, EC, MX, PE, PAR, PE, RD, TT, UR	210.000
IEA	TIMSS	2019	70	CH	
UNESCO	ERCE-4	2019	18	AR, BO, BR, CH, CO, CR, CU, EC, ES, GU, HO, MX, NI, PAN, PAR, PE RD, UR	

165

Fuente: documentos técnicos de IEA, ETS, OCDE y Unesco.

En la actualidad, las EIGE se llevan a cabo en más de 130 países, alcanzando alrededor del 90% de los 146 países con más de dos millones de habitantes. Países de todas las tradiciones culturales y con diversas formas de organización económica y política se han sumado a su aplicación.

En la siguiente tabla se muestran los contenidos abordados en estas pruebas y las poblaciones que han sido el foco de las evaluaciones.

Tabla 2. Contenidos y población de estudio de principales pruebas internacionales

PRUEBA	CONTENIDOS								POBLACIÓN EVALUADA
	Lectura/ literacy	Matemáticas/ numeracy	Ciencias	Educación cívica	Solución de problemas	Dominio de TICs	Enseñanza	Otros	
SSS	X	X	X	X				Literatura Inglés & francés como segundo idioma	Estudiantes de 10 y 14 años, y último año Secundaria
TIMSS		X	X					Conocimientos avanzados matemática y física	4° y 8° Básico
IALS	X	X							Adultos 16 a 65 años
ERCE	X	X							3° y 6° Básico
CIVED				X					14 años y último año Ed. Sec.
PISA	X	X	X		Módulos opcionales	Opcional		Alfabetización financiera, bienestar, competencia global	Alumnos de 15 años
PIRLS	X								4° básico
TALIS							X		Profesores de aula
TEDS-M							X		Futuros profesores
ICILS					Opcional, solución computacional	X			8° Básico
PIAAC	X	X			S.P. en ambientes enriquecidos tecnológicamente				Adultos 16 a 65 años
ICCS				X					8° Básico (aprox. 14 años)

Fuente: documentos técnicos de IEA, OCDE y Unesco.

El crecimiento y propagación de las pruebas internacionales consolida y crea nueva demanda por su realización. Se hace casi ineludible para los países valerse de una herramienta de uso cada vez más generalizado, la cual, además, es incentivada o demandada por organismos internacionales.³ Esto lleva no solo a que los países se sumen, sino también a que incluyan más pruebas, con los mismos o con nuevos focos de evaluación. Las pruebas, además, entregan información sobre diferentes aspectos, habilidades para la vida (PISA), *curriculum* (TIMSS y ERCE), permitiendo comparaciones en el contexto mundial o regional. De tal modo, se ha instalado sólidamente una “cultura evaluativa global” (Ydesen, 2019).

No es una característica solo de años recientes la llegada a los países de América Latina de discursos y prácticas educacionales internacionales. Ya en el siglo XIX circularon por la región modelos educacionales ingleses, franceses y alemanes, y en

3. Puede hipotetizarse que esto ocurre por un proceso de isomorfismo en que inciden factores tanto normativos como miméticos y coactivos.

el XX enfoques norteamericanos. Sin embargo, tales discursos y prácticas tenían una circulación menos sistemática, sin una estructura internacional bien montada, que sostuviera un flujo continuo y multidireccional. No existía nada parecido al mecanismo actual de circulación de conocimiento. Por ejemplo, la amplia discusión educacional que tuvo lugar a principios del siglo XX respecto a las ideas de Dewey y de la escuela nueva, con enfoques provenientes de Europa y Estados Unidos, se basaba en publicaciones dispersas y en viajeros esporádicos. ¿Qué es lo que ocurre ahora en términos de circulación internacional de estas pruebas?

2. Dimensiones del fenómeno de las EIGE que explican su fuerza expansiva y reguladora

Frente a la pregunta sobre qué es lo que permite entender y explicar el crecimiento e impacto que han tenido las EIGE, consideramos que hay tres factores o dimensiones del fenómeno que son cruciales como parte de la respuesta:

- (1) Su peculiar arquitectura organizativa que conecta organismos internacionales, gobiernos nacionales, centros de investigación y redes de expertos a través del mundo.
- (2) Su basamento científico, que estructura el procedimiento central de constitución y operación de estas pruebas.
- (3) El tipo de resultados que generan, su forma de analizar y formatear los resultados, y las características del proceso de entrega y difusión de ellos.

167

A continuación, se atiende a cada dimensión y a cómo ella contribuye a explicar la fuerza de las EIGE.

2.1. Estructura operativa basada en redes internacionales

Las pruebas internacionales pioneras diseñadas por la IEA en los años 1960 contaron con el apoyo de Unesco y la Ford Foundation, estableciendo vínculos más bien formales con los gobiernos. La modalidad de elaboración de pruebas que se generaliza desde los años 1990 incluye la participación no solo financiera de los gobiernos, sino también sustantiva, a través de una continua colaboración técnica entre los expertos de los organismos internacionales y de los países.

Varios organismos internacionales desarrollaron iniciativas en materia de evaluaciones internacionales. No obstante, desde los años 1990, los organismos dominantes son tres: OCDE, IEA y Unesco.⁴

4. ETS también aplica, tempranamente, pruebas internacionales, en 1988 y 1991, pero que no continuó.

2.1.1. IEA (*International Association for the Evaluation of Educational Achievement*)

La IEA, creada en 1958, en el Institute for Education, en Hamburgo, bajo el alero de la Unesco, es el organismo más antiguo de los tres en lo que a las EIGE se refiere. Montó el primer gran dispositivo de evaluación internacional en la segunda mitad de los años 1960 (IEA, 2018b). Sus fundadores, gente como Torsten Husen, de la Universidad de Estocolmo, Benjamin Bloom, de la Universidad de Chicago, y Robert Thorndike, del Teachers College de la Universidad de Columbia, quienes trabajaron colaborativamente con muchos otros, eran académicos destacados en el ámbito educacional y su orientación era hacia la investigación (Keeves, 2011). Les interesaba detectar la variabilidad educacional internacional e identificar los factores que explican tal variabilidad. En palabras de Husen (1975, p. 7), “el mundo puede ser concebido como un gran laboratorio educacional donde diferentes prácticas son experimentadas, en términos de organización de la escuela, contenido curricular y métodos de instrucción”. Se buscaba obtener hallazgos generalizables que contribuyeran a la política educativa de los países y, en particular, ayudaran a modernizar los países menos desarrollados.

En la actualidad la IEA cuenta con 68 países miembros y entre sus pruebas más destacadas se cuentan: TIMSS, desde 1994 (que primero fue FIMSS, en 1964, y luego SIMSS, en 1982); PIRLS (2001); CIVED (1999), luego denominada ICCS (desde 2009); y TEDS (2008). De ellas, TIMSS y PIRLS son curriculares, mientras que CIVED e ICCS están referidas a la formación cívica.

2.1.2. Unesco (*United Nations Educational Scientific and Cultural Organization*)

Establecida en 1945, ha promovido la educación con una orientación normativa asociada a la paz y los derechos humanos. Este organismo ha buscado ser líder en producción y diseminación de conocimiento educacional (Menashy y Manion, 2016). En la década de 1950, asume la tarea de recopilación y estandarización de estadísticas educacionales con perspectiva mundial. En 1952 se funda el Instituto de la Unesco para la Educación (IUE), que aborda el estudio comparado de la educación. En 1952 la Unesco publica el *World Handbook of Educational Organization and Statistics*, con datos de 57 países, cuyas comparaciones se ven dificultadas por las grandes diferencias existentes en cuanto a denominaciones y clasificaciones de la realidad educacional.⁵ Entre 1966 y 1974, Unesco trabaja en la elaboración de la International Standard Classification of Education (ISCED), finalmente aprobada en 1978. Esta clasificación fue diseñada para armonizar internacionalmente las mediciones e inicialmente pensada “para la planificación educacional [de los países] y la promoción mundial de la educación” (Schneider, 2013, p. 366), acorde con la orientación de este organismo en esos años. La Unesco ha sido decisiva en la difusión y legitimación de la lógica de la estandarización educacional internacional y de la comparación entre países, ayudando a proveer la infraestructura necesaria para la producción de datos (Gorur, 2018).⁶ Un objetivo destacado que tiene la generación

5. Desde 1921 hay esfuerzos para generar estadísticas educacionales en un marco comparativo, pero con falta de estandarización que obstaculiza el análisis (Gorur, 2018).

6. En los años 1980, por la pérdida de apoyo financiero que experimenta la Unesco, asociada a la creciente politización del organismo, ella es desplazada por el Banco Mundial que, con mucho mayor presupuesto, asumió el liderazgo mundial en el ámbito educacional (Mundy, 2010).

de este tipo de información es la creación de programas de ayuda internacional para los países que no tienen los recursos para implementar mejoras en la educación. Sin datos comparativos, a nivel internacional, agencias como Unesco, Unicef o el Banco Mundial realizarían su labor a ciegas.

En 1994 se crea, en la Oficina Regional de Educación de la Unesco para América Latina, con sede en Santiago de Chile (Orealc-Unesco), un Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE). Su organización fue diseñada como una red de unidades dedicadas a la medición y evaluación de los sistemas educacionales de América Latina, coordinada por un equipo en la oficina regional (Unesco-Orealc, 2015). Esta iniciativa surge en un período en que “muchos países estaban iniciando procesos de reforma educacional sin contar con información adecuada y suficiente, con escasa masa de recursos críticos en medición de la calidad de la educación” (Unesco-Orealc, 2014, p. 9). De tal modo, el LLECE “tiene como objetivo aportar conocimientos, evidencia empírica y asistencia técnica a sus países miembros a través de evaluaciones de aprendizaje de carácter internacional y comparativo entre países”. Esto permitiría “retroalimentar a los sistemas educativos en general, pero también a sus actores relevantes tanto a nivel del aula y escuela como a nivel político, para fortalecer la toma de decisiones educativas, dada la entrega de resultados de logros de aprendizaje y factores asociados a estos últimos” (Unesco-Orealc, 2015, p. 3). También señala, desde 2008, el objetivo de contribuir a la formación de opiniones, a la circulación y difusión de ideas y al debate informado sobre aprendizajes y procesos educacionales (Unesco-Orealc, 2008, p. 19).

Las evaluaciones de la Unesco-Orealc, denominadas ERCE (Estudio Regional Comparativo en Educación), son de índole curricular y buscan atender a las peculiaridades de la región. Para eso, se realiza un extenso y participativo proceso de revisión de los currículos de los países participantes, el cual es coordinado por una de las instituciones colaboradoras: el Instituto Colombiano para el Fomento de la Educación Superior (ICFES). Así, una peculiaridad de los ERCE es atender a las “raíces curriculares propias” y considerar características contextuales de relevancia para la región (condición urbana/rural, tipo de ciudades, tipo de administración del establecimiento y otros factores asociados) (Unesco-Orealc, 1998, p. 7). Ese análisis curricular es, de hecho, un gran esfuerzo colectivo que toma un par de años, entre revisión de documentos, consultas y logro de acuerdos. Según sus coordinadores, constituye un “esfuerzo cooperativo sin precedentes en la región” (Unesco-Orealc, 1998, p. 17). El ERCE lleva ya tres versiones (1998, 2006 y 2013) y está en proceso la cuarta (2019), siendo una prueba de lenta preparación y desarrollo. En cada caso ha tomado más de cinco años, entre el diseño y la entrega final de resultados. Estas evaluaciones han recibido financiamiento del Banco Mundial, del BID y de la Fundación Ford (Unesco-Orealc, 2008, p. 15).

Producto del análisis de los “factores asociados” y de la identificación de algunas condicionantes particularmente influyentes (y que son modificables), los autores del ERCE hablan de un “modelo latinoamericano” de evaluación.

2.1.3. OCDE (*Organización para la Cooperación y el Desarrollo Económicos*)

La OCDE nace en 1961, como heredera de la organización que se había encargado, desde 1948, de hacer funcionar el Plan Marshall para la reconstrucción luego de la Segunda Guerra Mundial. Es una organización que genera y analiza información, con una directa orientación a la revisión y diseño de políticas. Para tales fines, organiza y analiza información estadística de diversas procedencias (Unesco, Eurostat, etc.) y genera sus propios datos. Es un lugar de interconexión entre expertos en diseño de políticas, con una perspectiva inicialmente europea, pero que gradualmente se ha ido haciendo internacional y global. En 1968 la OCDE crea el Centre for Educational Research and Innovation (CERI), con el cual buscaba una interacción más estrecha entre investigación, gestión educacional y diseño de políticas, y que será fundamental para el trabajo posterior en el área (Burgi y Trohler, 2018).

En los años 1980, Estados Unidos y Francia presionan a la OCDE para que produzca mejores datos y que sean más comparables. Estados Unidos se encontraba acicateada por la competencia tecnológica, componente de la Guerra Fría; el gobierno de izquierda de Francia, por su parte, estaba preocupado por las oportunidades educacionales para niños desaventajados socioeconómicamente (Fulge *et al.*, 2016, pp. 458-459). En respuesta, desde 1991, el Directorate for Education, Employment, Labour and Social Affairs comenzó a integrar indicadores respecto a los sistemas educacionales y a publicar síntesis globales y nacionales. Expresión fundamental de esto es el informe *Education at a Glance*, elaborado por el CERI y reconocido internacionalmente, que comenzó a publicarse en 1992. En su afán por generar indicadores más apropiados, en 1995 decidió el desarrollo del Programme for International Student Assessment (PISA), evaluación no curricular sobre la preparación de los jóvenes para enfrentar la sociedad actual, comenzando a aplicar la prueba en el 2000. Poco después, como reflejo de la importancia que adquiriría el trabajo en este campo, se estableció un Education Directorate autónomo (Sellar y Lingard, 2013, p. 190). Gradualmente esta agencia, impulsada por sus propios analistas, ha ido constituyendo su propia agenda de generación de datos (Fulge *et al.*, 2016) y se ha convertido en un productor central de conocimientos orientado al diseño de políticas, de gran relevancia especialmente en el espacio europeo (Sellar y Lingard, 2013; Grek, 2009, 2018).

En el 2018 la OCDE cuenta con 37 países miembros, entre ellos México y Chile, con Colombia en camino a serlo, y sus principales pruebas son PISA, IALS, PIAAC y TALIS.

2.1.4. Consorcios

Las EIGE de las últimas dos décadas han sido realizadas a través de consorcios internacionales con centros y personas especialistas en las diferentes materias en juego. En la tabla siguiente se registra una nómina de integrantes de estos consorcios, en algunas de las pruebas realizadas.

Tabla 3. Integrantes de los consorcios a cargo de la elaboración de pruebas internacionales

IEA	OCDE	UNESCO
<p>CIVED (1999)</p> <p>IEA, American Institutes for Research (AIR), Western Statistics (WESTAT), Estados Unidos, Humboldt University of Berlin (the International Coordinating Center)</p>	<p>PISA (2000)</p> <p>Australian Council for Educational Research (ACER), National Institute for Educational Measurement (CITO) de Holanda, Service de Pédagogie Expérimentale (Université de Liège, Belgica) WESTAT, Educational Testing Service (ETS), Estados Unidos, National Institute for Educational Research (NIER), Japón</p>	<p>ERCE (2006)</p> <p>Instituto Colombiano para el Fomento de la Educación Superior (ICFES), Australian Council for Educational Research (ACER)</p> <p>ETS (revisión y certificación de las pruebas), IEA (apoyo técnico)</p>
<p>TIMSS (1999)</p> <p>Boston College, IEA Secretariat (Amsterdam), Centro de Procesamiento de Datos de la IEA (Hamburgo-Alemania), Statistics Canada (Ottawa-Canada), Educational Testing Service (ETS), Estados Unidos</p>	<p>PISA (2006)</p> <p>ACER, CITO, Unité d'analyse des systèmes et des pratiques d'enseignement (aSPe) de la Université de Liège (Belgica), ETS, NIER, WESTAT, International Language School (ILS), Universidad de Oslo, Noruega, Leibniz Institute for Science and Mathematics Education (IPN) de la Universidad de Kiel, Alemania</p>	<p>ERCE (2013)</p> <p>Instituto Colombiano para el Fomento de la Educación Superior (ICFES), MIDE UC, Centro de Política Comparada Educativa (CPCE-UDP)</p>
<p>ICCS (2009)</p> <p>Australian Council for Educational Research (ACER), National Foundation for Educational Research (NFER), Laboratorio di Pedagogia Sperimentale de la Universidad Roma Tre University (LPS).</p>	<p>PISA (2015)</p> <p>ETS, Pearson, DIPF, aSPe; Center for Educational Technology (CET) de Israel, CRP-HT, The Educational Measurement and Research Center (EMACS) de la Universidad de Luxemburgo, GESIS – Leibniz-Institute for the Social Sciences de Alemania, WESTAT, ACER (como subcontratista de WESTAT para operaciones y muestreo).</p>	

Fuente: reportes técnicos de IEA, OCDE y Unesco.

Puede constatarse el gran entremezclamiento de las relaciones. Las conexiones con centros técnicos no son exclusivas. Statistics Canadá, ETS, WESTAT y otros han colaborado con los tres organismos, en diversas pruebas. La misma IEA ha cooperado con la OCDE y con la Unesco. Asimismo, la movilidad de los expertos entre uno y otro centro es intensa. Participantes del IEA, por ejemplo, que tempranamente adquirieron experiencia en el área fueron reclutados por otros programas, generando una activa circulación de conocimientos y experiencias (Keeves, 2001, p. 33). Incluso el mismo Andreas Schleicher, figura central en el desarrollo de la prueba PISA, trabajó inicialmente, en los años 1990, en la IEA, en Holanda, llegando a ser *Director for Analysis* (IEA, 2018).

Un actor técnico destacado en esa red es el ETS (Educational Testing Service). Es otra organización privada sin fines de lucro, tal como el IEA. Fue fundada en 1947, con fines de evaluación e investigación educacional, contando con los aportes de recursos y personal técnico provistos por el American Council of Education, la Carnegie Foundation y el College Entrance Examination Board. Desde temprano el ETS explora las evaluaciones internacionales, teniendo un exclusivo mercado en las pruebas de admisión universitaria, tales como el GRE y el Toefl, aplicados internacionalmente. Además de esas pruebas, que son para fines de certificación individual, en 1988 el ETS aplicó una prueba de evaluación en matemáticas y ciencia a seis países y luego una segunda versión, en 1991, a 20 países. No obstante, no continuó (Wagemaker, 2014). Después de eso, siguió participando en las evaluaciones internacionales, pero sin pruebas propias. Así, ha prestado diversos apoyos técnicos, tanto a la IEA como a la Unesco-Orealc, y desde 2010 ha establecido una estrecha colaboración con la OCDE.

172

Otras entidades, aun sin participar directamente en la elaboración y análisis de las pruebas, son también relevantes para su operación. Entre ellas, el Banco Mundial aparece como un componente financiero de especial importancia al aportar fondos para la realización de pruebas como TIMSS, PIRLS y PISA en países de menor desarrollo. Una fuente de su interés en apoyar las evaluaciones está en la necesidad de sus investigadores (economistas y sociólogos) de contar con datos para evaluar los resultados de las inversiones realizadas por el Banco, el cual tiene una agenda de mejoramiento de los resultados de aprendizaje a través de ambiciosas reformas sistémicas (Mundy y Verger, 2015). Las evaluaciones educacionales de IEA, OCDE y Unesco generan los datos necesarios para monitorear los resultados de innovaciones y programas que el banco financia.⁷ Sobre esa base, el Banco incluye “componentes de evaluación” en sus proyectos.⁸ Complementariamente apoya financieramente a los países en desarrollo para que hagan evaluaciones nacionales y participen en las evaluaciones internacionales (Lockheed, 2013, p. 166). Explícita o implícitamente, las pruebas internacionales se convierten en requerimientos para obtener los financiamientos del Banco Mundial.

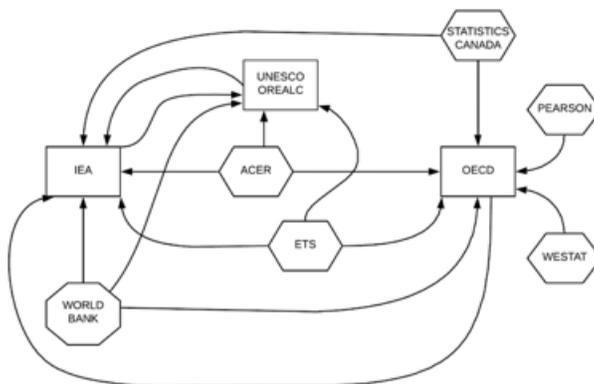
7. Estos criterios aparecen registrados en los documentos de los años 1990 del Banco Mundial y están reafirmados, por ejemplo, en su texto *Learning for All. Investing in People's Knowledge and Skills to Promote Development*, de 2011.

8. A principios de los 1990, el 50% de los proyectos educacionales del Banco contenían un “*testing component*”, proporción que después ha aumentado (Lockheed, 2011, p. 715).

Una organización del mundo de los negocios que ha adquirido creciente prominencia en las evaluaciones internacionales es Pearson. Ha ganado importantes licitaciones para participar en PISA 2015 y 2018, encargándose de su marco de referencia el 2018. Pearson Inc. es la compañía multinacional más grande en educación. Provee servicios de diseño curricular y de elaboración y procesamiento de pruebas. Tiene su casa matriz en Londres y opera en 70 países a través del mundo, con más de US\$ 1 billón en ventas durante 2011. Constituye un significativo exponente del *edu-business* que se expande globalmente (Bhanji, 2016; Sjoberg, 2019).

Como puede apreciarse, hay un denso tejido de relaciones entre los organismos internacionales. El gráfico siguiente muestra algunas de las principales relaciones de apoyo o colaboración internacional que sostienen la producción y mantención de las EIGE.

Gráfico 1. La red internacional



Fuente: elaboración propia a partir de datos de los organismos. Los cuadrados corresponden a los centros de producción y cálculo; los hexágonos, a las entidades de apoyo técnico y el octógono, a un organismo proveedor principalmente de financiamiento.

2.1.5. Conexiones nacionales

A esta red se agregan las conexiones con cada país. Los organismos a cargo de las EIGE —los nodos centrales de la red internacional— han llegado a establecer un modo común, muy similar entre ellas, de conexión con los países. El vínculo es, fundamentalmente, con los gobiernos y, más específicamente, con los ministerios de educación. En cada país hay un encargado nacional que es el interlocutor directo y quien debe viajar a encuentros internacionales para la planificación y diseño de las pruebas. Las propuestas de preguntas viajan entre los países y el organismo central, en ambas direcciones, para sucesivas revisiones y pruebas de campo. Periódicamente se realizan talleres y seminarios que congregan a expertos del organismo central y de los países, y a los coordinadores nacionales. A su vez, los países reciben la visita de expertos internacionales que los asesoran en sus actividades. Resulta así una estrecha conexión e intensa circulación de conocimientos, en formato interactivo.

Esta estructura operativa, consolidada a través de los años, basada en redes internacionales que integran una enorme cantidad de recursos cognitivos, humanos, financieros y tecnológicos, en estrecha conexión con los gobiernos nacionales, constituye un primer factor explicativo de la gran capacidad productora y reguladora de las EIGE.

2.2. EIGE: una expresión de *big science* en el ámbito de las ciencias sociales

El proceso de elaboración y análisis de estas pruebas es un trabajo científico, siguiendo los cánones positivistas, que busca validez, confiabilidad y capacidad de generalización de los resultados, empleando numerosos protocolos para asegurar su logro. El trabajo es riguroso y sistemático, llevado a cabo por una multitud de especialistas, proyectando, a través de su producción y difusión de números, la imagen de objetividad y neutralidad proclamada por este paradigma científico (Porter, 1995).

La investigación involucrada incluye una compleja elaboración conceptual de los dominios a ser medidos y una enorme operación de “circulación de la referencia” (Latour, 2001), transportando elementos experienciales provenientes de cientos de miles de estudiantes de todo el mundo.⁹ La escala de esta operación y la magnitud y calidad de las bases de datos generadas constituyen una empresa de *big science* que es completamente inusual para las ciencias sociales.

Son innumerables los aspectos en que se especifica la complejidad de esta operación.¹⁰ Mencionaremos solo algunos.

174

2.2.1. *Elaboración conceptual*

En cada ámbito temático de evaluación (ciencia, lenguaje, matemática, educación cívica, trabajo colaborativo, etc.) es necesario primero identificar conceptualmente el dominio a ser evaluado y las dimensiones involucradas, con las especificaciones que las precisen, a fin de llegar finalmente a las competencias, habilidades, conocimientos que se espera adquieran los estudiantes y, en algunas pruebas, los contextos o las situaciones de aplicación de los conocimientos. El punto de partida, en algunas pruebas, tales como TIMSS y ERCE, son los currículos nacionales; en otras, como PISA y CIVED, lo son nuevas elaboraciones más allá de las definiciones curriculares. Todo esto se alimenta de las diferentes disciplinas que constituyen lo que se ha llamado las ciencias de la educación: psicopedagogía, psicología educacional, psicología evolutiva, psicología cognitiva, sociología educacional, etc.

2.2.2. *Instrumentos*

En cuanto a la elaboración de los instrumentos, ellos son producto de un largo proceso a lo largo de al menos un par de años, con sucesivas iteraciones de propuestas de ítems y su revisión, de las que participan los integrantes de los ministerios de educación de cada país, o las personas designadas por ellos, y los expertos de los

9. Por “circulación de la referencia”, Latour entiende el proceso a través del cual elementos del mundo son seleccionados y reconfigurados a través de los procedimientos conceptuales y metodológicos de la investigación científica, en pasos sucesivos, hasta llegar a los textos científicos como “evidencia”.

10. Véase Rutkowski *et al.* (2014). Para una visión crítica, véanse: Lietz *et al.* (2017) y Hoptmann *et al.* (2007).

organismos centrales. Se realiza un intenso trabajo de estandarización, procurando que los ítems sean manifestaciones culturalmente equivalentes, sin sesgos, entre los diferentes países y las traducciones son sometidas a varios filtros de chequeo, en los organismos centrales y en los países. A esto le siguen pruebas piloto y nuevas revisiones.¹¹

Junto a las pruebas mismas, se acostumbra agregar cuestionarios que buscan información sobre factores de la experiencia en la escuela o en el hogar que puedan estar incidiendo en los resultados. Estos cuestionarios se aplican, en combinaciones cambiantes según las pruebas y en muchos casos opcionalmente, a estudiantes, profesores, directivos y apoderados. Se genera por esta vía una gran cantidad de variables que permiten reconstruir, a la manera positivista, aspectos del “contexto” escolar y familiar que son hipotetizados como potencialmente relevantes. Las sucesivas aplicaciones han permitido ir refinando y enriqueciendo estos cuestionarios y los análisis subsecuentes.

2.2.3. Tecnología computacional

Un factor crítico que ha posibilitado toda esta operación global, que hace posible la gran escala de los procesos de revisión de los instrumentos y de circulación e integración de los datos, y que ha permitido reducir los ciclos productivos hasta los escasos tres años de PISA, es la enorme multiplicación del poder computacional. Mientras en el caso del *Six Subject Study*, en los 1960, los datos debían almacenarse en discos magnéticos y ser llevados a Estados Unidos para su procesamiento en los pocos computadores disponibles que tenían los programas computacionales apropiados, y algunos análisis, como el análisis multinivel, debían desecharse por razones prácticas de dificultad técnica, hoy en día no parece haber límites técnicos para el procesamiento y análisis. La tecnología digital de comunicaciones, además, permite el diálogo continuo entre todos los participantes a través del mundo.

175

2.2.4. Análisis

El análisis de los datos incluye una gran cantidad de trabajo con estadísticas descriptivas, que lleva a los tan conocidos cuadros comparativos entre países, a los niveles de logro alcanzado por las poblaciones de estudiantes de cada país y a los famosos rankings mundiales. Junto a eso, los expertos hacen un intenso trabajo de análisis multivariado (regresión jerárquica, regresión logística, modelos de ecuaciones estructurales, etc.) rastreando factores explicativos del desempeño escolar, tanto a nivel agregado como individual: características de escuelas, actividad pedagógica, profesores, familias y de los estudiantes mismos que pueden tener incidencia en los resultados. Al respecto, se integran destacadamente los conocimientos acumulados de la sociología y psicología educacional, y estas pruebas han aportado abundante información para la realización de investigaciones en la materia, tanto nacionales como de comparación internacional. De hecho, los organismos centrales ponen sus bases de datos a disposición de los investigadores interesados y han nutrido una gran masa de estudios y publicaciones a través del mundo.

11. Véanse detalles de los procedimientos en Litz *et al.* (2017) y Rutkowski *et al.* (2014).

Estos análisis son, además, los que les sirven, a los organismos internacionales a cargo de las pruebas, de base para sus propuestas y recomendaciones de política. No obstante, debe anotarse que, pese a la sofisticación estadística, los análisis causales que realizan los centros de cálculo están lejos, por lo general, de poder ser categóricos, como para afirmar con certeza que una medida o política es más efectiva que otra. Tienden, así, a primar las preferencias explicativas de cada institución. En la Unesco, por ejemplo, con su tradición de preocupación por la paz, el clima escolar y ausencia de conflicto ha sido un factor explicativo privilegiado. En la OCDE lo son la autonomía escolar, la *accountability* y la equidad.¹² De ello derivan recomendaciones con parecida especificidad o dependencia institucional. En alguna medida cada agencia pone a disposición de sus usuarios “soluciones en busca de problemas”. En los contextos de recepción, los gobiernos las hacen calzar, contingentemente, con particulares problemas, y usan los resultados y las propuestas para legitimar sus agendas (Baker y LeTendre, 2005; Fischman *et al.*, 2018).

En resumen, lo que se tiene con las EIGE —en cada una de ellas— es una empresa global de investigación en que participan cientos de investigadores y expertos a través del mundo, con un gran soporte institucional y tecnológico, generando una gran masa de datos, provenientes de la experiencia de cientos de miles de estudiantes y de otros actores educacionales de todo el mundo. Tanto para generar estos datos como para su análisis se emplean complejos modelos y se realizan numerosos análisis buscando caracterizar el desempeño comparativo de los estudiantes entre países y explicar las diferencias entre unos y otros. Son muy pocos los casos en que la ciencia social opera en estas condiciones de *big science*, la cual es habitual en campos como el de la física de partículas, astronomía o biología genética. Este es un segundo factor explicativo de la fuerza de las EIGE.

176

2.3. Formas de presentación del producto que contribuyen a la potencialidad de las EIGE

Si bien el proceso seguido en la elaboración de las EIGE y en el análisis de sus resultados se rige por los protocolos científicos positivistas, el producto ha sido objeto, de manera creciente desde los años 2000, de un procesamiento y empaquetamiento que considera fundamentalmente su uso por los gobiernos y atiende a la necesidad, tanto para ellos como para la promoción de las propias EIGE, de la difusión pública en los medios de comunicación masiva. Este es un tercer factor o dimensión que explica la fuerza adquirida por estas evaluaciones.

La OCDE, en particular, concibe sus EIGE como elemento central de generación y propagación de conocimiento que ayuda a crear un “campo global de diseño de política educacional”, haciendo del mundo un “espacio de medición conmesurable” (Lingard y Sellar, 2016, p. 362), construyendo activamente sus propios públicos. Con sus estrechos vínculos con los gobiernos, con su entrega de informes diseñados atendiendo a su legibilidad política y con sus recomendaciones prácticas, la OCDE

12. Posteriormente, en PISA también se le ha dado importancia al factor clima escolar (OCDE, 2019).

se ha ido convirtiendo en un “nodo central de la estructuración del campo global de la política educacional” (Grek, 2018). Opera una producción simultánea de conocimiento técnico, propuestas de política educacional y construcción de una representación medial de la calidad educacional de los países.

En la circulación mundial de los resultados de las EIGE es de suma relevancia la forma en que ellos son presentados. Lo que encuentra una difusión más generalizada son los rankings, de inmediata claridad y síntesis informativa. Si la exhibición de los rankings es común a todas las EIGE, en otras materias de la presentación de resultados hay diferencias importantes. Mientras la Unesco es escueta, la OCDE es prolífica en su uso de imágenes, boletines y sitios web interactivos. En su manejo público de los resultados hay una preocupación por la dimensión estética, como una manera de influir, afectando emociones, esperanzas y temores (Lundahl, 2018). Además, la difusión pública de los resultados de la OCDE es una operación cuidadosamente orquestada a nivel internacional, que genera una ola noticiosa que impacta a través del mundo.

La efectiva fórmula de la OCDE ha sido parcialmente imitada por la Unesco y la IEA, al menos en cuanto a prestar una mayor atención a las orientaciones prácticas, pero sin que cuenten con una estructura organizativa, tipo de recursos humanos y forma de funcionamiento que permita equipararse a ella. El estatus de la OCDE en la Unión Europea, su vínculo con los gobiernos y la legitimidad lograda, le permite contar con mejores recursos. Nada más para PISA 2009 dispuso de 17 millones de dólares, incluyendo gastos para diseminación de resultados. Esto es más de tres veces el presupuesto de la IEA en TIMSS y PIRLS ese año (Lockheed, 2013). Además, la IEA y la Unesco están más centradas en la mejora de las prácticas pedagógicas, de manera más acotada, mientras que la OCDE pone su foco en la agenda de política pública en su conjunto (organización, financiamiento, etc.) (Ydessen y Andersen, 2019). El discurso de la OCDE sobre sus EIGE destaca, además, de modo directo e incluso cuantificado, la relación con el mejoramiento económico de los países, lo que refuerza su atractivo político.

177

3. ¿Qué es lo que las EIGE producen?

Las EIGE responden a objetivos prácticos de entregar información que permita, en los distintos países, tomar decisiones de mejora y, en última instancia, contribuir a una mejor formación; pero, en el camino hacia eso, tienen repercusiones diversas, algunas de las cuales son de carácter performativo: moldean la realidad, como efecto primario de la propia evaluación, de lo cual, a su vez, se derivan variados otros efectos prácticos. Analizaremos los resultados de las EIGE atendiendo a cuatro niveles:

(1) Los resultados directos: datos, informes, gráficos, *rankings*, producto de gran operación de transporte de la experiencia educacional del mundo

(2) La generación de estándares normativos

(3) La configuración y el despliegue de una realidad educacional global y de una noción de calidad

(4) Los efectos que esto tiene en la gestión educacional de los sistemas educativos a través del mundo

A continuación, se abordan los resultados o efectos en cada uno de estos niveles.

3.1. La gran operación de transporte de la experiencia educacional del mundo y sus productos

En cada prueba PISA o TIMSS participan más de 500.000 estudiantes de unos 60 o 70 países dispersos en todo el mundo, a quienes se les aplica un instrumento que permite aprehender una parte de su desempeño escolar y de su experiencia respectiva y transportarla hasta centros de cálculo que la procesan e integran analíticamente. Esta es una macrooperación diseñada y realizada con los criterios de la ciencia. En el paradigma científico positivista, este transporte del mundo se lleva a cabo luego de una extensa operación de recorte de la enorme complejidad del mundo escolar. La experiencia escolar, en particular el desempeño de los estudiantes, es especificada, formateada y “recogida” por la vía de las preguntas estandarizadas (no necesariamente cerradas), que son “indicadores” de ese desempeño, definido con respecto a un dominio cognitivo de interés.

178

Así, a través de operaciones de recorte, inscripción, traducción, transporte y rearmado, se produce la “circulación de la referencia”, como dice Latour (2001), desde el mundo escolar a los centros de cálculo. Es un proceso de estandarización y simplificación que tiene la gran ventaja de permitir armar un modelamiento estadístico de la realidad educacional del mundo, de docenas de países, la cual finalmente aparece aplanada y comprimida en unas cuantas páginas y en un set de gráficos, haciendo posible comparar países, tipos de escuelas y prácticas pedagógicas. Los sistemas educacionales, en este transporte y aplanamiento, se convierten en inscripciones en el papel y puntos en un gráfico, que son asumidos como una representación fáctica confiable lograda por la vía de la ciencia.¹³

La consecuencia de esa operación de recorte analítico es que lo que llega a los centros de cálculo, a las oficinas de la IEA, OCDE y Unesco, ineludiblemente no es toda la riqueza sociocultural de cada escuela, con su particular trayectoria, densidad experiencial e imbricación con su entorno. No comparece ese tejido de interacciones socioafectivas y materialidades que es cada escuela. Este estrechamiento de la mirada ha sido críticamente resaltado por autores como Gorur (2016a, 2016b) y Sjoberg (2019), quienes señalan los peligros que pueden derivarse de esta visión centralizada y uniformizadora. Parte de esa operación reductiva es la neutralización cultural, derivada intrínsecamente de los requerimientos comparativos del proceso

13. Esto es lo propio de la operación científica que, junto a la simplificación, gana en capacidad analítica y comparativa (Latour, 2001).

(Mendelovits, 2017, p. 85). La búsqueda solo de lo que sea comparable elimina lo propio de cada cultura y la contextualización que se lleva a cabo en el análisis es fundamentalmente cuantitativa, usando las variables sobre las que se tiene información. Para la ciencia, tal recorte es evidente y es inevitable: es la única forma de trabajar con este “laboratorio educacional mundial”. Esto es base de su fuerza, pero también es fuente de debilidades. Cuando se convierte en criterio de intervención, por parte de los gobiernos u otras entidades nacionales, corre el riesgo de tener consecuencias perjudiciales no anticipadas, ya que ni las escuelas ni sus integrantes corresponden a esa versión recortada analíticamente y desconectada que se manifiesta en los “datos”.

3.2. Generación de estándares normativos

Así como la estandarización es esencial para la mecánica de las pruebas, para el logro de comparabilidad, como producto de ellas se genera otro tipo de estándares, estándares de logro educacional que pasan a ser considerados por los países a través del mundo. Cada prueba genera sus niveles de desempeño; seis inicialmente en el caso de PISA, con un nivel inferior que después se dividió en dos (1a y 1b). Tal como en una escala de notas, hay niveles de desempeño destacables, positiva o negativamente. Los estudiantes que se encuentran bajo el nivel 2 de PISA no alcanzan las competencias mínimas que se requieren para participar de forma plena en una sociedad moderna; los niveles 5 y 6 reflejan un alto grado de logro. En esos niveles se clasifica a los estudiantes de cada país. En PISA 2015, por ejemplo, 13 millones de jóvenes de 15 años, en 64 países, eran de bajo rendimiento en al menos un área, concitando preocupación en los países e impulsando acciones de los gobiernos.

179

De manera análoga, a los países cuyos sistemas educacionales son evaluados se los ubica en rankings de desempeño, los cuales constituyen otro tipo de estándar (Busch, 2011). Estos ordenan a los países, permitiendo identificar, en lógica de *league tables*, a los distinguidos tres primeros y a los *top ten* y, por otra parte, a los que están por debajo del promedio y, por tanto, en zona de preocupación. El promedio se convierte en el estándar de suficiencia. Para los gobiernos, la meta básica es estar por encima de esa vara. De tal modo, los organismos internacionales, con sus EIGE, han contribuido con una nueva familia de estándares educativos.

Tales estándares son normas generadas por el procedimiento científico estadístico y las operaciones clasificatorias de los expertos. No existían antes de la operación de estas pruebas. Así como las normas morales provienen de una larga tradición cultural y las normas legales típicamente proceden de un proceso deliberativo que ocurre en el parlamento, estas otras normas tienen una fuente que no pasa por la sedimentación cultural ni por una deliberación representativa políticamente. La generación de estos estándares se sustenta, primariamente, en el trabajo de expertos, que definen los niveles de progresión en el aprendizaje a nivel teórico y luego los cortes numéricos entre niveles. Esto permite traducir los datos estadísticos y hacerlos legibles y usables para los *policy makers* y los legos en general, incorporándose estas normas en las políticas educacionales de los países, siendo empleadas para definir targets y metas. Poseen una legitimidad científico-técnica. En la medida, a su vez, que son utilizados para el gobierno de las conductas adquieren fuerza política y de mera norma técnica se transmutan en normas sociales y morales, aceptadas como tales. Atestiguan el

dicho de Latour de que “la ciencia es política por otros medios” (Latour, 1988, p. 229). En algunas de estas pruebas, por ejemplo, en el caso de PISA, tales estándares son definidos por miembros de países de industrialización avanzada, siendo aplicados en un proceso asimétrico al resto del mundo.

La manera en que los estándares derivados de las pruebas internacionales son asumidos y “traducidos” en cada país, y la forma como se articulan con los estándares preexistentes, varía de país a país. En el punto subsiguiente se mencionan algunos de los países que explícitamente han considerado los estándares de las EIGE y los han usado para definir sus propias metas, estándares y políticas.

3.3. Realidad educacional que se configura y se despliega globalmente

Estos grandes procesos de investigación y medición de la realidad, que operan entrelazados con grandes instituciones internacionales y nacionales, tienen como resultado de su acción no solo hechos científicos, sino que simultáneamente moldean lo percibido como calidad educacional en cuanto hecho social. Esto es una manifestación de la capacidad performativa de la ciencia que autores como Callon (1998), MacKenzie (2006) y Muniesa (2014) han estudiado con respecto a la ciencia económica y que otros han extendido a las ciencias sociales (Osborne y Rose, 1999; Ramos Zincke, 2012, 2014). Las evaluaciones internacionales operan como dispositivo sociotécnico con capacidad para enactar una nueva configuración de realidad educacional y de lo que es asumido como calidad. Esto repite a nivel global lo que ha ocurrido con las evaluaciones nacionales, articulándose ahora ambos procesos.¹⁴ Respecto a la calidad educacional, desde principios del siglo XX, los conocimientos de las ciencias sociales han ayudado a configurar lo que ella es, pero nunca hasta ahora de un modo tan articulado a nivel internacional, con fluidos circuitos de comunicación e interacción, ni tan sistemáticamente operacionalizado a través del mundo.

Esta es una operación performativa. Tal valor, tal calidad educacional del país definida a partir del desempeño de sus estudiantes en un registro comparativo internacional, no existía unas décadas atrás, tal como aparecen definidos por estos procedimientos, antes de que ellos tomaran lugar y circularan los resultados. Antes de esta macrooperación performativa, los alemanes no tenían la situación de crisis educacional —el “hecho” de la crisis— que confrontaron en 2000 luego de los resultados de la primera prueba PISA, ni los finlandeses se concebían como el país campeón mundial de la calidad educacional. Con la emergencia de esta nueva realidad, sus consecuencias ratifican y refuerzan su existencia. Los debates acalorados en la esfera pública alemana y de otros países, la definición de metas, la decisión de llevar a cabo diferentes medidas sustanciales, dan cuenta del peso de realidad de esa calidad educacional del país que fue configurada y comunicada por la OCDE y que luego tiene vida propia.

14. Algunas de las experiencias tempranas, a nivel nacional, son las de la NAEP, en USA, desde 1969 y la del PER, entre 1982-1984, seguida luego del SIMCE, desde 1988, en Chile.

De hecho, hoy día, de manera muy generalizada, los funcionarios de los gobiernos, expertos educacionales y una gran masa de profesionales y legos son capaces de identificar sistemas efectivos y no efectivos y, en particular, calificar al propio. Además, con respecto a la realidad propia, son capaces de hacer afirmaciones comparativas con otros países, con los de la región y sobre la tendencia a lo largo del eje temporal: “estamos mejorando”, “estamos estancados”. Es una percepción colectiva, una apreciación fáctico-normativa que unos 30 años atrás no tomaba forma.

Esta nueva configuración de elementos de la realidad educacional repercute en las definiciones, los criterios normativos y procedimientos de evaluación interna de los países. Los gobiernos ajustan su aparataje perceptual. Introducen cambios en sus definiciones de calidad y en sus currículos, e incorporan criterios y procedimientos evaluativos de las pruebas internacionales en sus propios sistemas de evaluación nacional.

Por cierto que, pese a la solidez fáctica y a la claridad del juicio normativo con que aparecen tales realidades, su facticidad y normatividad son dependientes y están condicionadas por el aparato productivo que las hizo emerger. Son *factiches*, como dice Latour (2010), contruidos pero reales. Tal aparato productivo tiene múltiples y variados elementos: particulares definiciones conceptuales de dominios, dimensiones de tales dominios, competencias en cada dimensión, preguntas específicas, formato de las preguntas, procedimiento de aplicación, tiempos permitidos de respuesta, muestras de estudiantes, procedimiento para establecer niveles de desempeño, forma en que se considera o no el contexto institucional, etc. Una variación de importancia en cualquiera de esos aspectos puede hacer que determinados países suban o bajen en los rankings. Así, por ejemplo, la aplicación del formato digital en PISA 2015 hizo que países como Chile y Uruguay bajaran sus puntajes y otros subieran (Rivas y Scasso, 2017). Los resultados pueden ser afectados por problemas en la representatividad muestral (Carnoy, 2015) y por el diferente grado de motivación y esfuerzo de los estudiantes (Eklof y Hopfenbeck, 2019).

181

3.4. Efectos en la gestión de los sistemas educacionales

Los efectos directos del dispositivo de reconfiguración fáctico-normativa de la realidad educacional tienen derivaciones subsecuentes. Si bien en cada país, dadas sus peculiaridades institucionales, su contingencia política y su trayectoria educativa, hay variaciones, es posible reconocer también ciertas recurrencias, según muestran los ya numerosos estudios en la materia. Atendiendo a los resultados de algunos de los principales de éstos (Breakspear, 2012; Baird *et al.*, 2011; Volante, 2016; Lockheed, 2013; Fischman *et al.*, 2018), se indican a continuación los principales efectos recurrentes, anotando una muestra de países en que ellos se presentan.

La manifestación más extrema está expresada en los *shocks* derivados de las evaluaciones, es decir, la construcción colectiva de un discurso cargado emocionalmente que exhibe una seria falla en el sistema educacional. Esto ha ocurrido en Alemania, Noruega, Dinamarca, Japón y Nueva Zelanda. Esta sensación colectiva de crisis ha abierto “ventanas de oportunidad” para reformas y cambios. La “crisis educacional”, que es una construcción narrativa en torno a los resultados del test, es

empleada para generar debate público y como gatilladora de cambios. Ha contribuido a destrabar bloqueos y resistencias. Por otra parte, aun sin *shock*, los resultados han servido para legitimar medidas, políticas y reformas, como ha sido el caso de Inglaterra, en que los buenos resultados sirvieron de respaldo a reformas en curso, y de México, Corea, China, Francia, Noruega, Rumania, Kuwait, Sud África, Singapur, Suiza y Nueva Zelanda.

Los resultados de las EIGE son procesados en cada país de modo particular por los actores colectivos concernidos. Pasan por diversas traducciones y son puestos al servicio de los diversos intereses de actores estatales, políticos, sociales y otros (Benveniste, 2002). Son, además, condicionados o filtrados por la trayectoria político institucional previa del país (Verger, Fontdevila y Parcerisa, 2019). No son las EIGE las que necesariamente imprimen la dirección de los cambios, sino que dependen de las decisiones de gobiernos y autoridades educacionales, presionados por los debates políticos. Los efectos resultantes son, por tanto, muy indirectos y los resultados de las evaluaciones son usados para apoyar políticas y medidas diversas, de maneras en general muy *ad hoc*. Entre ellas:

- Realizar revisiones y modificaciones curriculares, atendiendo a contenidos y criterios de las pruebas, poniendo mayor foco en problemas de la vida real (Polonia, Portugal, Rumania, Macedonia, España, Alemania, Malasia, Suecia, Japón, Estados Árabes, Latvia, Japón, Chile y Shanghái).¹⁵
- Fijación de estándares para evaluar a estudiantes y escuelas del país (Rusia, Alemania, Chile, Portugal, Suiza, Irlanda, Grecia y Noruega).
- Determinación de metas y targets para el mejoramiento del sistema educacional, en términos, por ejemplo, de posicionamiento en los rankings internacionales o de porcentaje de estudiantes nacionales en los niveles altos o bajos (Australia, Bélgica, Canadá, Dinamarca, Finlandia, Francia, Holanda, Irlanda, Japón, Israel, México, Tailandia, Polonia, Eslovenia, Turquía).
- Implementación de nuevas políticas para promover la equidad, sobre todo socioeconómica, pero también de género (Alemania, Austria, Dinamarca, Checoslovaquia, Chile, Hungría, Bélgica, Corea y Nueva Zelanda).
- Cambios en las formas de enseñanza (España, Australia, Japón, Canadá y Hungría). También incorporación de tests estandarizados, en países que no atendían a ellos (Rusia y Alemania).
- Cambios en la formación de profesores y en políticas respecto a su calidad (España, Cuba, Brasil, Colombia, Japón, Portugal, Israel, Singapur, Canadá, Finlandia e Irlanda).
- Contribución a la creación de agencias o de sistemas de evaluación educacional nacional (Rumania, Suecia, Malasia, Noruega, Israel, Japón, Luxemburgo, Irlanda y Polonia). De hecho, en los años 2000 se más que duplica la cantidad de evaluaciones nacionales respecto a los 1990 (Ramírez, Schofer y Meyer, 2018).

15. Ramírez *et al.* (2018, p. 355) muestran la positiva correlación de las pruebas con amplitud y cobertura curricular.

Se producen o refuerzan, al mismo tiempo, ciertas tendencias más generales, tales como:

- Un sustancial traslado de la atención de gobierno desde los *inputs* hacia los resultados educacionales.
- El afianzamiento de una creciente mirada internacional a la educación, la cual tradicionalmente había sido concebida como dominio doméstico. Ahora se le aplican criterios y normas internacionales.
- Gestación de un discurso globalizado para discutir sobre la “buena educación”. Ramírez *et al.* (2016, p. 58) hablan incluso de un “modelo de educación global” que estaría en desarrollo.

También ocurre un efecto reactivo, de rechazo, eventualmente contraperformativo. La amplia acogida de las EIGE, genera suspicacia, particularmente hacia las pruebas gestionadas por la OCDE, que tienen mayor impacto. Ha tomado forma, así, un bien nutrido discurso crítico, alimentado básicamente por académicos de países desarrollados. Algunos cuestionan materias técnicas (instrumentos, comparabilidad, muestras, etc.). Otros critican aspectos más sustanciales: el primado en las evaluaciones de una racionalidad instrumental por sobre una comunicativa, que atienda al mundo intersubjetivo; el recorte extremo de la realidad que involucran las pruebas estandarizadas con su empobrecimiento de la realidad vivida; el foco en lo cuantificable (la trampa de lo contabilizable); la atención privilegiada a una educación orientada hacia las necesidades de la economía; el crecimiento de la burocracia del aseguramiento de la calidad; su énfasis en la competencia, en sintonía con el espíritu neoliberal, entre otras materias. Los críticos han sido prolíficos y se han convertido en un interlocutor relevante para el desarrollo actual y futuro de las EIGE.¹⁶

183

Conclusiones: las EIGE como un dispositivo de gubernamentalidad global

Frente a los desafíos que enfrenta la educación, en un mundo en veloz cambio, con las presiones de la globalización, tecnificación y digitalización, a las respuestas nacionales se ha sumado y articulado un dispositivo internacional de gobierno cuyos elementos hemos ido presentando:

- (1) Un complejo aparato de producción de datos, información y conocimiento sobre la realidad educacional mundial, particularmente sobre el desempeño educacional y sus condicionantes, que opera a gran escala, empleando procedimientos y conocimientos de la ciencia social y estadísticas.

16. Véanse algunas de las críticas más consistentes y destacadas en Sahlberg (2016), Hopmann *et al.* (2007), Gorur (2016a, 2016b) y Sjöberg (2019).

- (2) Ese aparato de producción se despliega internacionalmente, conectando expertos diseminados por el mundo, institutos especializados, públicos y privados, que se vinculan con centros de cálculo y de diseño de propuestas.
- (3) Un proceso de creación de estándares normativos. Las orientaciones valorativas en circulación, promovidas por organismos de Naciones Unidas, respecto a una educación para todos, de calidad y con equidad, se especifican en estándares con respecto niveles de logro, grados de equidad y rankings internacionales, definiendo lo deseado y lo indeseado, y sirven de base para definir metas y *targets*.
- (4) Un proceso de difusión a escala global de resultados fácticos acompañados de estándares que los cualifican normativamente, con múltiples “traducciones” a diferentes formatos —informes, comunicados de prensa, presentaciones digitales, videos, etc.—, en una persistente iteratividad, con extensas cadenas citacionales, que incide en el encuadramiento y moldeamiento de la percepción pública en los respectivos países sobre su realidad educacional en comparación con la del resto del mundo. Esta configuración analítica se “imprime” sobre lo que se asume como realidad, enardece las preocupaciones y crea la necesidad de tomar decisiones, abriendo “ventanas de oportunidad” para realizar cambios.
- (5) Acompañando la distribución de material fáctico con sello normativo, y motivada por su contenido, se produce complementariamente una amplia circulación de propuestas de medidas de política educacional, que se manifiesta en parte en los numerosos boletines e informes de la OCDE, IEA, ETS, Unesco, Pearson, y se realizan procesos de *benchmarking* internacional. Se alimenta, así, un campo global de diseño de políticas educacionales (Ozga y Lingard, 2006).

184

Todo lo dicho resalta la importancia de las EIGE y sus peculiaridades como instrumento. Son, sin duda, un gran dispositivo de gobierno. Tal dispositivo es más que un instrumento de gobernanza internacional de la educación. Es un dispositivo que, por la potencialidad técnica de su construcción científica, por las conexiones de poder de las instituciones y redes que conectan con los gobiernos nacionales, por los intensos procesos de circulación de conocimientos y por su forma de operación en el espacio público e institucional, tiene una destacada capacidad performativa. Este dispositivo, con toda su batería de pruebas internacionales, ha estado creando, en su operación conjunta, una nueva realidad, una nueva configuración de la calidad educacional de los países, de lo que se percibe como tal. En una misma macrooperación le ha impreso un sello de calidad educacional particular a cada país y ha creado la norma para determinar el nivel de calidad asignado. Ni una ni otra, ni la calidad educacional del país ni el estándar normativo, existían previamente a las EIGE. Esa específica facticidad y esa evaluación normativa son un efecto performativo de las EIGE. Otras selecciones y combinaciones de elementos extraídos de las experiencias educacionales de los países y otros procesos de juicio normativo llevarían a otros resultados, tal como, por lo demás, ocurre entre las mismas pruebas internacionales, cuyos resultados exhiben diferencias importantes, asociadas a sus diferentes procesos constructivos. Esto no quiere decir que la operatoria de las EIGE sea arbitraria. Nada de eso. Es un proceso sistemático y riguroso. Pero, aun así, es posible imaginar otros procesos sistemáticos y rigurosos alternativos, con resultados diferentes, siendo en cada caso los resultados contingentes a cada particular operatoria productiva y con una fuerza de realidad concordante con el poder que posea la red de producción que la sustenta.

Esta capacidad productora de realidad, la cual regula la acción de los sujetos, cambiando sus condiciones de operación, favoreciendo que se comporten voluntariamente de modos diferentes, acorde a las orientaciones que provienen del dispositivo, queda mejor aprehendida en el concepto de gubernamentalidad de Foucault. Para este autor, la gubernamentalidad es una articulación de elementos —instituciones, procedimientos de análisis, reflexiones, cálculos y técnicas— que permiten el ejercicio de una forma de poder que tiene como su objeto a la población y que actúa sobre ella a distancia, creando condiciones que hacen más probables determinados comportamientos (Foucault, 2000a, pp. 219-220). La gubernamentalidad moderna es una forma de conducción de las conductas que se caracteriza por enlazarse el poder con saberes científicos, con regímenes de verdad (Foucault, 2000b, p. 225). Como componente de los procesos de globalización y de rearticulación de las formas de regulación social a través del mundo, las pruebas internacionales a gran escala constituyen un dispositivo de gubernamentalidad global, cuyas raíces vienen desde mediados del siglo XX, pero cuya plena consolidación ha ocurrido en el siglo XXI. Su operación regulatoria está redefiniendo el funcionamiento de los sistemas de educación a través del mundo. Estas evaluaciones están llevando a cambiar las nociones mismas de calidad, atendiendo a la adaptación a una sociedad global, donde reinan la ciencia y la tecnología, y a definir nuevos estándares normativos con los cuales juzgar los resultados educativos. Los nuevos criterios y las normas no solo van más allá de los currículos nacionales, sino que estos mismos han sido objeto de reflexividad y ajuste. Los currículos educacionales de los países, que históricamente han mantenido un fuerte carácter nacional, vinculados al propio desarrollo de la nación, pierden parte de tal carácter, con una referencia educativa dirigida ahora a una sociedad globalizada. El efecto de este dispositivo se extiende, a su vez, a la constitución subjetiva, afectando autopercepciones, expectativas e identidades, contribuyendo a descartar otras formas anteriormente prevalecientes.

185

La operación del dispositivo genera cadenas performativas que conectan a agencias internacionales, autoridades ministeriales, políticas educacionales, criterios de directivos de establecimientos, conducta de profesores y reacciones de estudiantes. Los cambios en el comportamiento son, no obstante, muy variados y responden al efecto conjunto de múltiples otros factores propios de cada país. Además, aunque se generaliza una pauta de regulación, su operación es en un marco de flexibilidad que abre mucho espacio para que se genere variedad en las respuestas.

Este macrodispositivo, como aquí se ha enfatizado, se sustenta en una estructura operativa basada en redes internacionales, de las que forman parte agencias internacionales, centros de investigación, expertos, corporaciones y organizaciones diversas. En esta red se lleva a cabo una labor investigativa a gran escala, a nivel de *big science*, manteniendo una directa conexión con la preocupación práctica por el diseño de políticas educacionales. En ello, el proceso comunicacional juega un rol fundamental en la proyección y potenciación de los resultados, y es objeto de una creciente atención por parte de las agencias internacionales. Estas redes son móviles y cambiantes. Continuamente hay nodos que se conectan y desconectan. Hay también corporaciones privadas que se integran, como es el caso de Pearson, lo que genera inquietudes por los intereses de negocios que involucran (Ball y Junemann, 2012; Sjoberg, 2019).

Se trata de un dispositivo que combina muy efectivamente poder y régimen de verdad, generación de facticidad y de normatividad, teniendo así una especial capacidad de reconfiguración de la realidad, de lo que se asume y percibe como realidad educativa, para todos los fines prácticos. Es una gran herramienta técnica, normativa y política, que moldea la realidad educacional a escala global.

Financiamiento

Esta investigación contó con el financiamiento del Consejo Nacional de Ciencia y Tecnología de Chile: Proyecto Fondecyt N° 1.170.477.

Agradecimientos

El autor agradece la colaboración de Matías Valderrama y los comentarios de Alejandra Falabella.

Bibliografía

Baird, J., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T. y Daugherty, R. (2011). *Policy Effects of PISA*. Report Commissioned by Pearson UK. Oxford: Oxford University Centre for Educational Assessment.

Baker, D. y Letendre, G. (2005). *National Differences, Global Similarities. World Culture and the Future of Schooling*. Stanford: Stanford University Press.

Ball, S. y Junemann, C. (2012). *Networks, New Governance and Education*. Bristol: Policy Press.

Bhanji, Z. (2016). *The Business Case for Transnational Corporate Participation, Profits, and Policy Making in Education*. En K. Mundy, A. Green, B. Lingard y A. Verger (Eds.), *The Handbook of Global Education Policy* (419-432). Nueva York: John Wiley & Sons.

Benveniste, L. (2002). *The Political Structuration of Assessment: Negotiating State Power and Legitimacy*. *Comparative Education Review*, 46(1), 89-118.

Breakspear, S. (2012). *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance*. OECD Education Working Papers, 71. París: OECD Publishing.

Burgi, R. y Trohler, D. (2018). *Producing the "Right Kind of People"*. En S. Lindblad, D. Pettersson y T. Popkewitz (Eds.), *Education by the Numbers and the Making of*

Society. *The Expertise of International Assessment (75-91)*. Nueva York: Routledge.

Busch, L. (2011). *Standards. Recipes for Reality*. Cambridge, Massachusetts: MIT Press.

Callon, M. (1998). *The Law of the Markets*. Oxford: Blackwell Publishers.

Carvalho, L. (2018). *International Assessments and Its Expertise Fabricating Expert Knowledge for Policy*. En S. Lindblad, D. Pettersson y T. Popkewitz (Eds.), *Education by the Numbers and the Making of Society. The Expertise of International Assessment (110-126)*. Nueva York: Routledge.

Carnoy, M. (2015). *International Test Score Comparison and Education Policy: A Review of the Critiques*. Boulder: National Educational Policy Centre.

Eklof, H. y Hopfenbeck, T. (2019). *Self-Reported Effort and Motivation in the PISA Test*. En B. Maddox (Ed.), *International Large-Scale Assessments in Education (121-136)*. Londres: Bloomsbury Academic.

Fischman, G., Topper, A., Silova, I., Goebel, J. y Holloway, J. (2018). *Examining the Influence of International Large-scale Assessments on National Education Policies*. *Journal of Education Policy*, 34(4), 470-499.

Foucault, M. (2000a). *Essential Works, 1954-1984. Volume 3: Power*. Nueva York: The New Press.

Foucault, M. (2000b). *Defender la Sociedad. Curso en el College de France (1975-1976)*. México DF: Fondo de Cultura Económica.

Fulge, T., Bieber, T. y Martens, K. (2016). *Rational Intentions and Unintended Consequences: On the Interplay Between International and National Actors in Education Policy*. En K. Mundy, A. Green, B. Lingard y A. Verger (Eds.), *The Handbook of Global Education Policy (453-469)*. Nueva York: John Wiley & Sons.

Gorur, R. (2011). *ANT on the PISA Trail: Following the Statistical Pursuit of Certainty*. *Educational Philosophy and Theory*, 43(1), 76-93.

Gorur, R. (2016a). *Seeing like PISA: A Cautionary Tale about the Performativity of International Assessments*. *European Educational Research Journal*, 15(5), 598-616.

Gorur, R. (2016b). *The "Thin Descriptions" of the Secondary Analyses of PISA*. *Educação & Sociedade*, 37(136), 647-668.

Gorur, R. (2018). *Standards: Normative, Interpretative, and Performative*. En S. Lindblad, D. Pettersson y T. Popkewitz (Eds.), *Education by the Numbers and the Making of Society. The Expertise of International Assessment (92-109)*. Nueva York: Routledge.

Grek, S. (2009). Governing by Numbers: the PISA “Effect” in Europe. *Journal of Education Policy*, 24(1), 23-37.

Grek, S. (2018). OECD as a Site of Coproduction: The European Education Governance and the New Politics of “Policy Mobilization”. En S. Lindblad, D. Pettersson y T. Popkewitz (Eds.), *Education by the Numbers and the Making of Society. The Expertise of International Assessment (185-200)*. Nueva York: Routledge.

Hoptmann, S., Brinek, G. y Retzl, M. (2007). *PISA According to PISA*. Berlín: LIT Verlag.

Husen, T. (1975). Multi-national evaluation of school systems. II?? *Occasional Papers*, 37. París: Unesco, International Institute for Educational Planning.

IEA (2018). *60 Years of IEA (1958-2018)*. Amsterdam: IEA.

Keeves, J. (2011). IEA – From the Beginning in 1958 to 1990. En C. Papanastasiou, T. Plomp y E. Papanastasiou (Eds.), *IEA 1958-2008: 50 Years of Experiences and Memories (3-40)*. Amsterdam: IEA.

Latour, B. (1988). *The Pasteurization of France*. Cambridge, Massachusetts: Harvard University Press.

Latour, B. (2010). *On the Modern Cult of the Factish Gods*. Durham: Duke University Press.

Latour, B. (2001). *La esperanza de Pandora. Ensayos sobre la realidad de los estudios de la ciencia*. Barcelona: Gedisa Editorial.

Lawn, M. (2008). *An Atlantic Crossing. The Work of the International Examination Inquiry, its Researchers, Methods and Influence*. Oxford: Symposium Books.

Lingard, B. y Sellar, S. (2016). The Changing Organizational and Global Significance of the OECD’s Education Work. En K. Mundy, A. Green, B. Lingard y A. Verger (Eds.), *The Handbook of Global Education Policy (357-373)*. Nueva York: John Wiley & Sons.

Lietz, P., Cresswell, J., Rust, K. y Adams, R. (2017). *Implementation of Large-Scale Education Assessments*. Chichester: John Wiley & Sons.

Lockheed, M. (2011). Reflections on IEA from the Perspective of a World Bank Official. En C. Papanastasiou, T. Plomp y E. Papanastasiou (Eds.), *IEA 1958-2008: 50 Years of Experiences and Memories (711-720)*. Amsterdam: IEA.

Lockheed, M. (2013). Causes and Consequences of International Assessments in Developing Countries. En H. Meyer y A. Benavot (Eds.), *PISA, Power and Policy. The Emergence of Global Educational Governance (163-184)*. Oxford: Symposium Books.

Lundahl, C. (2018). The “Beauty” of PISA: The Politics of How PISA Scores Are Used to Represent Public Education. Ponencia presentada en AERA 2018 Annual Meeting. Abril 13-17, Nueva York.

Mackenzie, D. (2006). *An Engine not a Camera. How Financial Models Shape Markets*. Cambridge: MIT Press.

Mendelovits, J. (2017). Test Development. En P. Lietz, J. Cresswell, K. Rust y R. Adams (Eds.), *Implementation of Large-Scale Education Assessments (63-91)*. Chichester: John Wiley & Sons.

Meyer, H. y Benavot, A. (2013). *PISA, Power and Policy. The Emergence of Global Educational Governance*. Oxford: Symposium Books.

Mundy, K. y Verger, A. (2015). The World Bank and the Global Governance of Education in a Changing World Order. *International Journal of Education Development*, 40, 9-18.

Muniesa, F. (2014). *The Provoked Economy. Economic Reality and the Performative Turn*. Londres: Routledge.

Novoa, A. y Yariv-Mashal, T. (2003). Comparative Research in Education: A Mode of Governance or a Historical Journey? *Comparative Education*, 39(4), 423-438.

OCDE (2014). *PISA 2012. Technical Report*. París: OECD Publishing.

189

OCDE (2019). *PISA 2018 Results. What School Life Means for Students' Lives. Volume III*. París: OECD Publishing.

Osborne, T. y Rose, N. (1999). Do the Social Sciences Create Phenomena? The Example of Public Opinion Research. *British Journal of Sociology*, 50(3), 367-396.

Ozga, J. y Lingard, B. (2006). Globalization, Education Policy and Politics. En B. Lingard y J. Ozga (eds.), *The RoutledgeFalmer Reader in Education Policy and Politics*. Londres: Routledge Falmer.

Porter, T. (1995). *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Ramos Zincke, C. (2012). *El ensamblaje de ciencia social y sociedad. Conocimiento científico, gobierno de las conductas y producción de lo social*. Santiago: Ediciones Universidad Alberto Hurtado.

Ramos Zincke, C. (2014). Datos y relatos de la ciencia social como componentes de la producción de realidad social. *Convergencia. Revista de Ciencias Sociales*, 21(66), 151-177.

Ramírez, F., Schofer, E. y Meyer, J. (2018). International Tests, National Assessments, and Educational Development (1970-2012). *Comparative Education Review*, 62(3), 344-364.

Ramírez, F., Meyer, J. y Lerch, J. (2016). World Society and the Globalization of Educational Policy. En K. Mundy, A. Green, B. Lingard y A. Verger (Eds.), *The Handbook of Global Education Policy* (43-63). Nueva York: John Wiley & Sons.

Rivas, A. y Scasso, M. (2017). ¿Qué países mejoraron la calidad educativa? América Latina en las evaluaciones de aprendizaje. Documento de Trabajo, 16. Buenos Aires: CIPPEC.

Rizvi, F. y Lingard, B. (2010): *Globalizing Education Policy*. Londres: Routledge.

Rutkowski, L., Von Davier, M. y Rutkowski, D. (2014). *Handbook of International Large-Scale Assessment. Background, Technical Issues, and Methods of Data Analysis*. Boca Raton: Taylor & Francis Group.

Sellar, S. y Lingard, B. (2013). PISA and the Expanding Role of the OECD in Global Educational Governance. En H. Meyer y A. Benavot (Eds.), *PISA, Power and Policy. The Emergence of Global Educational Governance* (185-206). Oxford: Symposium Books.

Sahlberg, P. (2016). The Global Educational Reform Movement and Its Impact on Schooling. En K. Mundy, A. Green, B. Lingard y A. Verger (Eds.), *The Handbook of Global Education Policy* (128-144). Nueva York: John Wiley & Sons.

Schneider, S. (2013). The International Standard Classification of Education 2011. *Comparative Social Research*, 30, 365-379.

Sjoberg, S. (2019). The PISA Syndrome. How the OECD has Hijacked the Way we Perceive pupils, Schools and Education. *Confero*, 7(1), 12-65.

UNESCO-OREALC (1998). *Primer Estudio Internacional Comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado*. Santiago: Unesco.

UNESCO-OREALC (2008). *SERCE. Los aprendizajes de los estudiantes de América Latina y el Caribe. Primer reporte*. Santiago: Unesco.

UNESCO-OREALC (2014). *Comparación de resultados del segundo y tercer estudio regional comparativo y explicativo. SERCE y TERCE 2006-2013*. Santiago: Unesco.

UNESCO-OREALC (2015). *TERCE. Informe nacional de resultados*. Santiago: Agencia de Calidad de la Educación y Unesco.

Verger, A., Fontdevila, C. y Parcerisa, L. (2019). Reforming Governance through Policy Instruments: How and to What Extent Standards, Tests and Accountability in Education Spread Worldwide. *Discourse Studies in the Cultural Politics of Education*, 40(2), 248-270.

Volante, L. (2016). *The Intersection of International Achievement Testing and Educational Policy. Global Perspectives on Large-Scale Reform*. Londres: Routledge.

Wagemaker, H. (2014). *International Large-Scale Assessments: From Research to Policy*. En L. Rutkowski, M. von Davier y D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessment. Background, Technical Issues, and Methods of Data Analysis* (11-36). Boca Raton: Taylor & Francis Group.

Ydesen, C. y Andreasen, K. (2019). Los antecedentes históricos de la cultura evaluativa global en el ámbito de la educación. *Foro de Educación*, 17(26), 1-24.

Cómo citar este artículo

Ramos Zincke, C. (2021). Las evaluaciones internacionales a gran escala y la regulación global de los sistemas educacionales: un análisis integrativo. *Revista Iberoamericana de Ciencia, Tecnología y Sociedad —CTS*, 16(48), 159-191.