

# r a s e

Revista de la Asociación de Sociología de la Educación

Volumen 10, n.º 2, mayo 2017

<http://rase.ase.es>

## Artículos

NELLY P. STROMQUIST. Testing and the Promise of Academic Progress.

ALEXANDER OSIPOV. Institutional Functions as a Theoretical Nexus and Measurement Target in Sociology of Education.

DIETMAR K. PFEIFFER. As medidas da educação. Medição e gestão mediante indicadores.

ALMERINDO JANELA AFONSO. Para discutir la hegemonía epistémica evaluativa eurocéntrica: un enfoque exploratorio.

RUBIA SALHEB FONSECA, JOAQUIM ESCOLA, ARMANDO LOUREIRO Y AMÂNCIO CARVALHO. Competências de Estudo dos Estudantes Universitários: Estudo comparativo entre uma universidade portuguesa e brasileira.

SANDRA I. GONÇALVES SAÚDE, SANDRA C. SANTANA LOPES, CARLOS M. LOPES BORRALHO E ISIDRO L. RODRIGUES GÓIS FÉRIA. The Performance Evaluation of Polytechnic Higher Education Institutions: Tensions and Reconfigurations Resulting from the Adoption of Specific metrics.

IAN HARDY. Measuring, Monitoring, and Managing for Productive Learning? Australian Insights into the Enumeration of Education.

UYGUACIARA VELOSO CASTELO BRANCO, PAULO HIDEO NAKAMURA Y EDINEIDE JEZINE. Permanência na educação superior no Brasil: construção de uma escala de medida.

ARMANDO ALCÁNTARA SANTUARIO Y ALEJANDRO MÁRQUEZ JIMÉNEZ. La medida de la investigación en educación y su impacto social: las revistas de educación de Iberoamérica en los índices bibliométricos internacionales.

DANIEL A. WAGNER Y NATHAN M. CASTILLO. El aprendizaje en la base de la pirámide: Restricciones, comparabilidad y política en países en vías de desarrollo.

REGINA CÉLIA LINHARES HOSTINS. A tradução da política de avaliação pelos programas de pós-graduação em educação no Brasil.

CARMEN CARMONA RODRÍGUEZ. Tested: Una reflexión sobre medir en educación.

ALEJANDRA MONTANÉ, JOSÉ BELTRÁN Y ANTONIO TEODORO. La medida de la calidad educativa: acerca de los rankings universitarios.

## La medida de la educación

**Coordinadores** José Beltrán  
y António Teodoro

# r a s e

Revista de la Asociación de Sociología de la Educación

Volumen 10, n.º 2, mayo 2017

<http://rase.ase.es>

## Consejo Editorial Asesor

PEDRO ABRANTES, *Instituto Universitário Lisboa*, Portugal.

ARI ANTIKAINEN, *University of Eastern Finland*, Finlandia.

MICHAEL APPLE, *University of Wisconsin*, EE. UU.

STEPHEN BALL, *University of Oxford*, Gran Bretaña.

ANA BENAVENTE, *Universidade Lusófona de Humanidades e Tecnologias*, Portugal.

ANTONIO BOLIVAR, *Universidad de Granada*.

LINDOMAR WESSLER BONETTI, *Pontifícia Universidade Católica do Paraná*, Brasil.

JULIO CARABAÑA\*, *Universidad Complutense de Madrid*.

BETTINA DAUSIEN, *Universität Wien*, Austria.

MARIA JOSE CASA NOVA, *Universidade do Minho*, Portugal.

JEAN-LOUIS DEROUET, *Institut N. de Recherche Pédagogique*, Francia.

FRANÇOIS DUBET, *Université Bordeaux II*, Francia.

INÉS DUSSEL, *FLACSO*, Argentina.

ANTHONY GARY DWORKIN, *University of Houston*, EE UU.

JUAN MANUEL ESCUDERO MUÑOZ, *Universidad de Murcia*.

MARIANO FERNÁNDEZ ENGUITA\*, *Universidad Complutense de Madrid*.

FRANCISCO FERNÁNDEZ PALOMARES\*, *Universidad de Granada*.

JUAN SEBASTIÁN FERNÁNDEZ PRADOS\*, *Universidad de Almería*.

GUSTAVO FISCHMAN, *Arizona State University*, EE UU.

GAUDÊNCIO FRIGOTTO, *Universidade do Estado do Rio de Janeiro*, Brasil.

PABLO GENTILI, *Universidade do Estado do Rio de Janeiro*, Brasil.

AMADOS GUARRO PALLÁS, *Universidad de La Laguna*.

BEATE KRAIS, *Tech. Universität Darmstadt*, Alemania.

ADRIANA MARRERO, *Universidad de La República*, Uruguay.

MARIA ALICE NOGUEIRA, *Universidade Federal de Minas Gerais*, Brasil.

EMILIA PRESTES, *Universidade Federal da Paraíba*, Brasil.

BETANIA RAMALHO, *Universidade Federal do Rio Grande do Norte*, Brasil.

JUAN CARLOS TEDESCO, UNESCO, *Instituto Internacional de Planeamiento de la Educación*, Argentina.

EMILIO TENTI FANFANI, UNESCO, IPE, *Universidad de Buenos Aires*, Argentina.

ANTÓNIO TEODORO, *Universidade Lusófona de Humanidades e Tecnologias*, Portugal.

DENISE VAILLANT, *Universidad ORT*, Uruguay.

AGNÈS VAN ZANTEN, CNRS, *Institut d'Etudes Politiques de Paris*, Francia.

JULIA VARELA, *Universidad Complutense de Madrid*.

MARIOS VRYONIDES, *European University*, Chipre.

MICHAEL YOUNG, *University of London*, Gran Bretaña.

\* Miembros de la Asociación de Sociología de la Educación

## Comité Editorial Ejecutivo

ALICIA VILLAR AGUILÉS, *Universitat de València* (Directora)

MAR VENEGAS MEDINA, *Universidad de Granada* (Vicedirectora)

FRANCESC J. HERNÁNDEZ DOBON, *Universitat de València*

BEGOÑA ASUA BATARRITA, *Universidad del País Vasco*.

LEOPOLDO CABRERA RODRÍGUEZ, *Universidad de La Laguna*.

MARÍA FERNÁNDEZ MELLIZO-SOTO, *Universidad Complutense de Madrid*.

ENRIQUE MARTÍN CRIADO, *Universidad Pablo Olavide, Sevilla*.

XAVIER MARTÍNEZ CELORRIO, *Universitat de Barcelona*.

JUAN CARLOS RODRÍGUEZ PÉREZ, *Universidad Complutense de Madrid*.

MARINA SUBIRATS MARTORI, *Universitat Autònoma de Barcelona*.

JUNA SEBASTIÁN FERNÁNDEZ PRADOS, *Universidad de Almería*.

## Revista electrónica de la Asociación de Sociología de la Educación (ASE)

<http://rase.ase.es>

Email: [raserevista@gmail.com](mailto:raserevista@gmail.com)

ISSN 1988-7302

## Tabla de contenidos

### Presentación

JOSÉ BELTRÁN Y ANTONIO TEODORO: Medir la educación: perspectivas desde la crítica sociológica .....	110
--	-----

### Artículos

NELLY P. STROMQUIST: Testing and the Promise of Academic Progress .....	115
--	-----

ALEXANDER OSIPOV: Institutional Functions as a Theoretical Nexus and Measurement Target in Sociology of Education .....	128
--	-----

DIETMAR K. PFEIFFER: As medidas da educação. Medição e gestão mediante indicadores .....	143
---	-----

ALMERINDO JANELA AFONSO: Para discutir la hegemonía epistémica evaluativa eurocéntrica: un enfoque exploratorio .....	156
--	-----

RUBIA SALHEB FONSECA, JOAQUIM ESCOLA, ARMANDO LOUREIRO Y AMÂNCIO CARVALHO: Competências de Estudo dos Estudantes Universitários: Estudo comparativo entre uma universidade portuguesa e brasileira .....	167
--	-----

SANDRA I. GONÇALVES SAÚDE, SANDRA C. SANTANA LOPES, CARLOS M. LOPES BORRALHO E ISIDRO L. RODRIGUES GÓIS FÉRIA: The Performance Evaluation of Polytechnic Higher Education Institutions: Tensions and Reconfigurations Resulting from the Adoption of Specific Metrics .....	179
--	-----

IAN HARDY: Measuring, Monitoring, and Managing for Productive Learning? Australian Insights into the Enumeration of Education .....	192
---	-----

UYGUACIARA VELOSO CASTELO BRANCO, PAULO HIDEO NAKAMURA Y EDINEIDE JEZINE: Permanência na educação superior no Brasil: construção de uma escala de medida .....	209
---	-----

ARMANDO ALCÁNTARA SANTUARIO Y ALEJANDRO MÁRQUEZ JIMÉNEZ: La medida de la investigación en educación y su impacto social: las revistas de educación de Iberoamérica en los índices bibliométricos internacionales .....	225
--	-----

DANIEL A. WAGNER Y NATHAN M. CASTILLO: El aprendizaje en la base de la pirámide: Restricciones, comparabilidad y política en países en vías de desarrollo .....	240
---	-----

REGINA CÉLIA LINHARES HOSTINS: A tradução da política de avaliação pelos programas de pós-graduação em educação no Brasil .....	254
--	-----

CARMEN CARMONA RODRÍGUEZ: Tested: Una reflexión sobre medir en educación .....	268
---	-----

ALEJANDRA MONTANÉ, JOSÉ BELTRÁN Y ANTONIO TEODORO: La medida de la calidad educativa: acerca de los rankings universitarios .....	283
--	-----

### Reseñas

JUAN PECOURT GRACIA: Michael James Grenfell (2014) [2007]. <i>Pierre Bourdieu</i> . Londres: Bloomsbury .....	301
--	-----

ADRIANA MARRERO: Tim Rudd and Ivor Goodson (ed.) (2017). <i>Negotiating Neoliberalism. Developing Alternative Educational Visions</i> . Rotterdam: Sense Publishers .....	303
---	-----

<b>Anuncios y próximos números</b> .....	308
--	-----

## Presentación

# Medir la educación: perspectivas desde la crítica sociológica

José Beltrán y António Teodoro<sup>1</sup>

**E**l interés y la preocupación por una política educativa gobernada por los números se reflejan claramente en una literatura académica creciente desde disciplinas que en muchas ocasiones atraviesan fronteras epistemológicas y dialogan entre sí a nivel internacional. Si bien el problema acerca de la medida de la educación no es nuevo, es cierto que la generalización y proliferación de estudios, informes y pruebas que tratan de capturar resultados cuantitativos en la esfera educativa, estrechamente asociados a principios y criterios evaluativos, despierta tanto interés como inquietud. Recientemente se han desarrollado foros y debates que han planteado la necesidad de ofrecer alternativas a una corriente neopositivista que parece estar imponiéndose en las ciencias sociales, con efectos e impactos en las políticas y en las prácticas educativas. Así, en 2015, la revista *Journal of Education Policy* había prestado atención, en un número especial (vol. 30, n.º 3), a «la educación, la gobernanza y la tiranía de los números», con un editorial firmado por Stephen J. Ball (2015:299-301), en el que advierte del papel singular que en la actualidad juegan las maneras de medir la educación dentro de la relación entre verdad y poder. Por su parte, la Sección de Educación Comparada de la Sociedad Portuguesa de Ciencias de la Educación (SPCE-SEC) celebró en Lisboa, en enero de 2016, una Conferencia Internacional en torno al tema «La Educación Comparada más allá de los números: contextos locales, realidades nacionales y procesos transnacionales», con el propósito de reflexionar sobre los modos de regulación de las políticas a través de grandes estudios internacionales. También a finales de 2016, *Profesorado. Revista de curriculum y formación del profesorado* editaba un monográfico (vol. 20, n.º 3) con el título «políticas de evaluación y *accountability* en educación», coordinado por Almerindo Janela Afonso en cuya presentación reclamaba «una reflexión más densa» (2016:1-12). En enero de 2017 la revista *Comparative Education*, plantea un monográfico (vol. 53, n.º 1) cuyo primer artículo, a cargo de Elaine Unterhalter (2017:1-16) cuestiona la pretensión de «medir lo inmensurable en educación». Estos ejemplos no pretenden agotar el inventario, pero sin duda constituyen un síntoma claro del abanico de problemas al que da paso la pregunta acerca de «la medida de la educación». No es casual, entonces, que la sociología de la educación, a través de la RASE, se muestre sensible a un debate que afecta, directa o indirectamente, a cuantos formamos parte de la comunidad educativa.

La discusión sobre los modos de medir la educación no es trivial. El debate se dirime en una tensión –sí no en una clara oposición– entre la lógica y la dialéctica. Ambas son herramientas auxiliares legítimas

<sup>1</sup> José Beltrán es profesor del Departament de Sociologia i Antropologia Social de la Universitat de València y miembro de RIAIPE. António Teodoro es Director del Centro de Estudos Interdisciplinares em Educação e Desenvolvimento (CeIED), de la Universidade Lusófona. Coordinador de la Red Iberoamericana de Investigación en Políticas Educativas (RIAIPE).

y complementarias para comprender la realidad, pero no debería olvidarse que la validez científica no puede reducirse al valor numérico (ni tampoco al valor discursivo), y que hay aspectos de la realidad social que no son calculables: «¿cuánto vale una vida?» (García Ruíz, 2016:88), «¿cómo se mide una experiencia?», se preguntaba hace cien años John Dewey (1998 [1916]:287). No podemos contabilizar la intensidad de una discusión científica o de una argumentación racional, de la misma manera que no podemos discutir dialécticamente los resultados numéricos de una operación aritmética o una regla de la lógica formal. Tanto las teorías que elaboramos, como las hipótesis que sometemos a prueba y las creencias que tenemos están condicionadas por nuestra manera de medir, un término que está emparentado con la voz griega *metron*, y que significa trazar correctamente los propios límites, tener medida (Grün, 2016:9). Por eso, desde el ámbito de la economía, de la política y de la sociología se están replanteando los sistemas de medición, en el sentido de cuestionar sus límites (Meyer, 2017:17-34), sus sesgos (Ozga, 2017), y sus posibilidades más allá del «espíritu de Arquímedes» (Sivesind, 2014:57-77).

Buena parte de las cuestiones que pivotan en torno a la medida de la educación están asociadas a la producción de indicadores, que está basada en la presunción de una suerte de superioridad epistemológica. En el caso de los indicadores de educación, está produciéndose un auténtico ascenso de «la métrica», favoreciendo la construcción de la figura de «expertos», que en ocasiones resultan ser «débiles» (Lim, 2017). Algunos de ellos prestan una atención cada vez mayor al impacto de la producción científica o académica en instituciones educativas, asociando de manera descontextualizada y desproporcionada calidad a cantidad. La proliferación en los últimos años de una serie de informes internacionales que se centran en la esfera educativa (PISA, PIRLS, TALIS...) refleja nuevos modos de regulación transnacional, sustentada en una evaluación de los resultados y de su comparación permanente a través de grandes encuestas e informes estadísticos. (Teodoro, 2010). Y esa industria contable contribuye a configurar a su vez, en el marco de la globalización hegemónica, un nuevo sentido común, una gobernanza epistémica. Todo ello nos devuelve, de nuevo, a la cuestión central, en la que se entrecruzan interés y conocimiento: ¿cómo se mide la educación?

Esta cuestión, a su vez, da paso a una serie de interrogantes: ¿Con qué propósitos (científicos, sociales, económicos...) medimos la educación? ¿Qué instrumentos se utilizan para ofrecer información sobre las diferentes variables del campo educativo? ¿Qué valor tiene la experiencia educativa? ¿Qué efectos tienen los informes que se emiten periódicamente desde las agendas educativas a nivel nacional e internacional? ¿Qué relación existe entre las mediciones educativas y las reformas educativas? ¿Qué marcos culturales e interpretativos subyacen a nuestros modos de medir la realidad educativa? ¿El papel de los sistemas de indicadores es informativo o también tiene consecuencias normativas? ¿Es necesaria o posible la construcción de indicadores alternativos a los que ahora mismo son dominantes?

Se trata de cuestiones que ya hemos planteado en anteriores trabajos cuando nos preguntábamos si los indicadores acaban poniendo más el foco en los productos que en los procesos (Beltrán y Villar, 2010). En una reflexión previa sobre evaluaciones internacionales (Teodoro y Montané, 2009), se advertía que los indicadores educativos no capturan hechos puros sino datos que ya se nos ofrecen filtrados por el propio artefacto de medición. Nada que objetar, siempre que no perdamos de vista esta diferencia. Desde el siglo pasado, como respuesta a las exigencias del empirismo lógico, se fundamentó la necesidad de contextualizar las representaciones de la realidad, evitando confundir el «mapa» con el «territorio». El autor que formuló esta distinción metafórica, Wittgenstein, comparó la aritmética con la institución bancaria, sosteniendo que podría desmoronarse en cuanto la gente desconfiara de ella y se apresurara a sacar de

allí su dinero (Beltrán, 2009:11). No se pretende aquí cuestionar el valor de los números, sino más bien mostrar cautelas epistemológicas ante la suplantación de los significados por significantes, y de mensajes por los medios que les sirven de transporte. Ningún lenguaje es neutral en su uso ni es inocente en su aplicación. Desde un punto de vista social, no es posible ni deseable una descripción lógica aislada de una comprensión sociológica. La medida de la educación, al servicio de la lógica de la investigación científica, ha de explicar y explicarse en relación con prácticas sociales, con formas de vida. La superioridad epistemológica que se atribuye a la numérico (lo mensurable con números) se debe a su importante e innegable capacidad de abstracción, de separar elementos para su análisis, pero para la esfera de la educación no debemos olvidar la importancia de integrar estos elementos en un marco de análisis, crítico e interpretativo, más amplio, en una cultura común (superando de paso el abismo entre las dos culturas que ya denunciara en su momento C. P. Snow). Como toda tecnología al servicio de un fin determinado, la medida (de la educación) se puede considerar también un artefacto, un arma. Pero un arma tiene un significado ambivalente: puede convertirse en una herramienta constructiva, o puede convertirse en una amenaza, con efectos no previstos o no deseados. Todo depende del uso que decidamos darle. La dicotomía entre medición (lógica) y argumentación (dialéctica), puede superarse mediante una crítica sociológica que procure una «alianza tensa» con fines emancipadores. En este sentido, los artículos de esta revista ofrecen una buena muestra de los límites y posibilidades, de los usos y abusos, de las políticas y prácticas, así como de la creatividad e imaginación sociológica, a la hora de pensar y operativizar la medida de la educación.

Este monográfico reúne contribuciones que abordan el problema de la medida de la educación desde la sociología de la educación, con el concurso de otras disciplinas cercanas y convergentes, a través de una serie de miradas críticas y de enfoques plurales. Sin duda, las contribuciones que aquí se dan cita pueden estimular y enriquecer la discusión científica y el debate académico en torno a las implicaciones sociales de nuestros modos de medir el campo educativo en sus múltiples facetas.

En este número participan, con trece artículos y dos reseñas, veintiséis investigadoras e investigadores procedentes de nueve países de diferentes regiones del mundo: Brasil, México, Uruguay, Estados Unidos, Australia, Rusia, Alemania, Portugal y España. Las aportaciones para este número monográfico de la RASE exploran un conjunto de dimensiones muy variadas en torno al tema propuesto de la medida de la educación. Algunas adoptan una perspectiva principalmente teórica, abordando cuestiones acerca de la validez científica, la fundamentación epistemológica y las implicaciones sociales y políticas en la concepción y en el modo de medir la educación; otras tienen un carácter más empírico, y presentan resultados a partir de investigaciones ya realizadas o en curso que se han servido de herramientas de medición educativa; y también, en buena parte de los textos se combinan las perspectivas teóricas y empíricas al servicio de análisis que ilustran diferentes metodologías, aplicaciones y efectos, en ningún caso neutrales, de la medida de la educación. Estos tres tipos de enfoques diferenciados, que sirven a modo de clasificación, es el que nos ha facilitado el criterio de presentación y ordenación de los artículos.

En un primer bloque, de orientación principalmente teórica, se reúnen cuatro artículos. En el primero de ellos, Nelly P. Stromquist, partiendo de la realidad estadounidense pero más allá de la misma, lleva a cabo una crítica fundamentada sobre los efectos no previstos de la aplicación de los test estandarizados y cuestiona la supuesta relación entre la calidad educativa y su medición. Por su parte, Alexander Osipov propone un giro cualitativo, argumentando en su análisis la necesidad de un enfoque sobre las funciones de la educación considerada como institución social, en tanto que nexo teórico para dimensionarla y operativizarla conceptualmente. Dietmar Pfeiffer, en su texto, subraya el carácter ambiguo de los indicadores

más extendidos, la descontextualización en su aplicación, al tiempo que analiza la confusión común entre medir (un acto descriptivo) y evaluar (un acto normativo) y observa la configuración de una «cultura de los indicadores» que actúa como mecanismo de control. Desde un punto de vista poscolonial, Almerindo Janela Afonso cuestiona y denuncia la hegemonía epistémica eurocéntrica que reflejan los programas de evaluación comparativa internacional, subordinados a una agenda global asimétrica en sus intereses y en sus relaciones de poder.

En un segundo apartado, se dan cita una serie de cinco textos en los que prevalece la orientación empírica. Así, Rubia Fonseca, Joaquim Escola, Amâncio Carvalho y Armando Loureiro presentan un estudio comparativo en el que se analizan, sirviéndose de diferentes escalas de medición, las competencias de estudio de los estudiantes universitarios de universidades de Brasil y Portugal, con resultados sobre sus condiciones, actitudes y estrategias. El trabajo en equipo de Sandra Saúde, Sandra Lopes, Carlos Borralho e Isidro Féria, contextualizado en Portugal y en el caso del sistema politécnico de educación superior, se centra en el valor y las tensiones que pueden suponer la introducción de nuevas formas de medir el desempeño profesional en la dinámica educativa y organizativa, a partir de un proyecto en el que se ensayan la aplicación de una serie de indicadores. Ian Hardy reflexiona sobre cómo la formación continua de maestras y maestros está condicionada por una creciente medición de los resultados de aprendizaje de los estudiantes. Su estudio se centra en una escuela de Australia y en el análisis de una serie de reuniones en las que los docentes expresan cómo la manera de medir el aprendizaje afecta y altera su propia formación docente. A continuación, Uyguciara Veloso Castelo Branco, Paulo Hideo Nakamura y Edineide Jezine presentan la elaboración de una escala de permanencia discente, sirviéndose de la Lógica Paraconsistente, en el marco de un proyecto sobre la permanencia en la Educación Superior en Brasil. Desplazando el punto de enfoque, Armando Alcántara presta atención al modo de medir la investigación a través del estudio de las revistas de educación iberoamericanas. El artículo presenta los primeros resultados de un proyecto que analiza los efectos de la información bibliométrica en las revistas académicas.

Finalmente, cuatro nuevas entregas completan el tercer bloque de esta revista, con artículos elaborados desde aproximaciones teórico-empíricas. Daniel A. Wagner y Nathan M. Castillo ofrecen un análisis de las tensiones científicas en la comprensión del aprendizaje en poblaciones pobres y marginadas, que se sitúan en la base de la pirámide. Los autores argumentan la necesidad de que las Naciones Unidas prioricen el aprendizaje de la población empobrecida con el fin de que los objetivos para el desarrollo post-2015 aborden de manera crítica y comprometida las inequidades sociales y económicas. Desde una óptica distinta, Regina Célia Linhares Hostins aborda las políticas educativas de evaluación de postgrado implementadas en la reforma educativa neoliberal de Brasil en los años noventa. Y lo hace en una investigación de análisis documental inspirada en la teoría de la interpretación-traducción de la política propuesta, entre otros, por Stephen Ball. Cambiando de nuevo la mirada, Carmen Carmona Rodríguez se sirve del documental *Tested* (2015) –centrado en la aplicación en la ciudad de Nueva York del polémico *Specialized High School Admission Test* (SHSAT) dirigido a estudiantes de entre 12-13 años– para reflexionar sobre las consecuencias en términos de distribución de oportunidades que tiene la realización de evaluaciones estandarizadas al alumnado. Por último, Alejandra Montané, José Beltrán y António Teodoro revisan la construcción de sistemas transnacionales de evaluación de la calidad, a través de los principales *rankings* internacionales, desde una perspectiva crítica. Los autores sostienen que bajo el modelo de la «calidad educativa» se ha legitimado el dominio de una poderosa industria contable, alejada del horizonte de equidad y de justicia social, cognitiva y curricular.

El panorama de perspectivas que ofrecen todas estas contribuciones desde la sociología de la educación supone una muestra de la necesidad de llevar a cabo reflexiones que se puedan nutrir al mismo tiempo de la lógica de la investigación científica y de la dialéctica social. Inspirándose en ambas, la crítica sociológica invita, además, a dar el paso desde el compromiso académico a la acción social fundamentada.

La escritura compartida en el espacio que brinda una revista también es una forma de conversación y de acción. Los editores quieren agradecer a todos y cada uno de quienes han hecho posible este número, con sus valiosos trabajos, de manera desinteresada, y más allá de sus obligaciones profesionales.

## Referencias bibliográficas

- Afonso, Almerindo Janela (2016): “El campo de las políticas de accountability en educación. Para una reflexión más densa”. *Profesorado. Revista de curriculum y formación del profesorado*, 20 (3), 1-12.
- Ball, Stephen J. (2015): “Education, governance and the tyranny of numbers”. *Journal of Education Policy*, 30 (3), 299-301.
- Beltrán, José (2009): “Presentación” en António Teodoro y Alejandra Montané (coord.): *Espejo y reflejo: Políticas curriculares y evaluaciones internacionales*. Alzira, Germania.
- Beltrán Llavador, José y Villar Aguilés, Alicia (2010): “La medida de la educación: algunas consideraciones a propósito de los indicadores”. *Educação & Linguagem*, 13, (21), 134-149.
- Dewey, John ([1916]1998). *Democracia y educación*. Madrid, Morata.
- García Ruíz, Alicia (2016). *Impedir que el mundo se deshaga. Para una emancipación ilustrada*. Barcelona, Catarata.
- Grün, Anselm (2016). *El arte de la justa medida*. Madrid, Trotta.
- Lim, Miguel Antonio (2017): “The building of weak expertise: the work of global university rankers”. *High Education*. 13<sup>th</sup> of April.
- Meyer, Heinz-Dieter (2017): “The limits of measurement: misplaced precision, phronesis, and other Aristotelian cautions for the makers of PISA, APPR, etc.”. *Comparative Education*, 53 (1), 17-34.
- Ozga, Jenny (2017): “Education policy should not be driven by performance data”. *Nature Human Behaviour*, 1, 0014.
- Sivesind, Kirsten (2014): “Education in the Spirit of Archimedes: pertaining to the buoyancy of PISA” en Andreas Nordin & Daniel Sundberg (ed.): *Transnational Policy Flows in European Education: the making of knowledge in the education policy field*. Oxford: Symposium Books.
- Teodoro, António y Montané, Alejandra (coord.) (2009). *Espejo y reflejo: Políticas curriculares y evaluaciones internacionales*. Alzira: Germania.
- Teodoro, António (2010). *Educação, Globalização e Neoliberalismo. Os novos modos de regulação transnacional das políticas de educação*. Lisboa: Edições Universitárias Lusófonas.
- Unterhalter, Elaine (2017): “Negative capability? Measuring the unmeasurable in education”. *Comparative Education*, 53 (1), 1-16.



# Testing and the promise of academic progress

## Las pruebas estandarizadas y la promesa del progreso académico

Nelly P. Stromquist<sup>1</sup>

### Abstract

In light of the constantly expanding use of standardized tests (STs) not only in industrialized countries but also in developing nations, this article seeks to offer some reflections on the assessment path that educational systems across the world are taking. In unpacking the STs we must not take for granted a number of aspects of standardized tests—from their construction and application to their unintended consequences. The article proceeds in four distinct but interrelated parts: (1) understanding the assumptions and process in the development of STs, (2) probing the assumptions underlying the use of STs, (3) recognizing that STs are limited in what they can tell teachers about how to improve their instructional practices, and (4) forming a broader philosophical critique about how knowledge transmission should be measured and by whom. There are multiple procedures and features that render STs an easy-to-use assessment tool for student learning. However, assumptions used test design create considerable distance between their construction and any measurement of ability to teach. This distance impedes the use of STs as practical tools to improve instruction. Nonetheless, the testing industry in the US is solid and there are global efforts underway not only to expand the application from pre-Kindergarten to secondary schools but also to the university level. The article concludes with a few challenging ideas for further debate—ideas such as the need to rely more on teachers' evaluation of their students and to prepare teachers to fulfill better their instructional and assessment functions.

### Key words

Standardized testing, normal distribution curve, teacher performance, instructional practice, academic achievement, testing industry.

### Resumen

A la luz del uso en constante expansión de las pruebas estandarizadas (STs) no solo en los países industrializados, sino también en los países en desarrollo, este artículo pretende ofrecer algunas reflexiones sobre la trayectoria de evaluación que los sistemas educativos de todo el mundo están tomando. La deconstrucción de las ST requiere que no se den por sentado varios aspectos de las pruebas estandarizadas, desde su construcción y aplicación hasta las consecuencias imprevistas de tales pruebas. El artículo se desarrolla en cuatro partes distintas pero interrelacionadas: (1) comprender las suposiciones y el proceso en el desarrollo de las STs, (2) sondear las suposiciones subyacentes al uso de STs, (3) reconocer lo que las STs pueden y no pueden decir a los maestros en cuanto a cómo mejorar sus prácticas de instrucción y (4) formar una crítica filosófica más amplia sobre cómo debe medirse la transmisión del conocimiento y por quién debe ser hecha. Existen múltiples procedimientos y características que hacen que las STs sean fáciles de usar como herramientas de evaluación del aprendizaje de los estudiantes. Sin embargo, los supuestos del diseño de la prueba estandarizada crean una distancia considerable entre la construcción de la prueba y la medición precisa de la capacidad del docente para enseñar. Esta distancia impide el uso de las STs como herramientas prácticas para mejorar la instrucción. Sin embargo, la industria de las pruebas en los Estados Unidos está fuertemente arraigada y hay esfuerzos globales en curso no solo para ampliar la aplicación las mismas desde los niveles de pre-jardín de infancia hasta la escuela secundaria, sino también a nivel universitario. El artículo concluye con algunas ideas desafiantes para el debate posterior: ideas como la necesidad de confiar más en la evaluación realizada por los profesores de sus estudiantes y la preparación a los profesores para que desempeñen mejor sus funciones.

### Palabras clave

Pruebas estandarizadas, curva de la distribución normal, desempeño docente, prácticas de instrucción áulica, logro académico, industria ligada a la construcción de pruebas estandarizadas.

Recibido: 30-03-2017  
Aceptado: 02-05-2017

<sup>1</sup> University of Maryland, [stromqui@umd.edu](mailto:stromqui@umd.edu)

## 1. Introduction

In the era of the «knowledge society» a growing and global interest among policymakers is to monitor the performance of public schools so that the contribution of public investment may be justified. With advances in test design, standardized tests (STs) have become ubiquitous. They range from those used merely for national purposes to those deployed for international comparisons. They also vary greatly in purpose, from whether they are used to assessing school performance to serving high-stakes decision-making, in which case student performance is used to determine teacher competence and, thus, to reward or punish individual teachers.

Reliance on STs is significant across the world, but perhaps greatest in the US, where high-stakes tests have been mandated by successive school reforms, with the greatest growth in usage occurring in the mid-1980s (Stecher, 2002). Today, STs are in their third generation. They began as part of the legislative mandates for the Elementary and Secondary Education Act (ESEA) of 1965, then continued as part of high-stakes testing with the No Child Left Behind (NCLB) Act of 2001 and its reauthorization of ESEA (Hamilton, Stecher & Klein, 2002), and now with the Every Student Succeeds Act (ESSA) of 2015. The NCLB of 2001 mandated universal testing in reading and math in grades 3 through 9.

The Council of the Great City Schools, which represents large urban districts in the US, found that students take an average of 113 standardized tests from pre-K (before Kindergarten) to the 12<sup>th</sup> grade (the last year of secondary school) (Hefling, 2015). Notwithstanding the reliance on test scores as instruments of educational policy reform, major professional associations dealing with education in some form, have expressed concern with the use of testing to reward or punish individual teachers, arguing that «No single test score can be a perfectly dependable indicator of student performance and high-stakes decisions about individuals should be based on factors other than the score on a single test» (AERA, American Psychological Association, and Council on Measurement in Education, 1999).

By 2012, the K-12 testing had become a \$1.7 billion industry, an amount that will increase as tests are mandated by the Common Core Standards of ESSA, which represents a continuous strategy by the US federal government to use testing as a major policy reform mechanism (Resmovits, 2012). In the US, to comply with NCLB alone, some 68 million students were estimated to take STs in 2007 (Scherer, 2005). States have the flexibility to develop their own testing accountability system as long as they have «an appropriate feedback mechanism» (Hamilton, Stecher & Klein, 2002: xiii). The fact is that many education systems rely on STs developed by only a handful of test developers, all major national firms.

STs are unquestionably an American model of learning assessment that is attaining widespread diffusion across the world. We have today a growing array of cross-national tests: Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), the South African Consortium for Monitoring Essential Quality (SACMEQ), Progress in International Reading Literacy Study (PIRLS), Programme for the International Assessment of Adult Competencies (PIACC), Programme for the Analysis of Education Systems of the CONFEMEN (PASEC) and, more recently, efforts by the Organisation for Economic Co-operation and Development (OECD) are in place to measure student learning at the higher education level in the future through the Assessment of Higher Education Learning Outcomes (AHELO).

The growing diffusion of international large-scale assessments through STs has been compared to old colonial practices of exporting their goods as the best possible to the colonies. One observation catches this well:

*«Thus, while in previous centuries Christian missionaries followed in the wake of the military to promote their brand of education across the European colonial empires, we now face the prospect of the West exporting its vision of schooling around the world through the auspices of cross-national tests supported by the modern missionaries and camp followers of our time: the think tanks and multinational companies who specialize in identifying and delivering “what works”» (Morris, 2016:2).*

An international institution that is heavily invested in the promotion of STs is OECD. This organization is engaged in two initiatives. The PISA for Development Initiative seeks to support the Education 2030 agenda of the Sustainable Development Goals (SDGs) and thus to increase the number of participant countries engaged in STs beyond the over 70 that participate in PISA today. The second initiative refers to OECD efforts underway to develop STs to assess student learning and institutional performance across the world at the higher education level. Its Assessment of Higher Education Learning Outcomes (AHELO) is now undergoing test development in 17 countries participating in a feasibility study to develop standardized tests in three areas: «generic skills,» economics, and engineering. OECD hopes to have the project initiated sometime in 2017. The American Council on Education and Universities Canada have expressed opposition to AHELO, stating that «the AHELO approach fundamentally misconstrues the purpose of learning outcomes, which should be to allow institutions to determine and define what they expect students will achieve and to measure whether they have been successful in doing so» (Redden, 2015).

In light of the growing presence of STs, this article proposes to engage in reflection about the value of such tests, the appropriateness of their use for different purposes, and their major consequences for the production of knowledge. To do so, the article proceeds in four parts: (1) understanding the assumptions and processes in the development of STs, (2) probing the assumptions underlying the use of STs, (3) coming to terms with what STs can and cannot tell us about how teachers might improve their practice, and (4) forming a broader philosophical critique about how knowledge acquisition should be measured and by whom. This article then concludes with a few challenging ideas for further debate. It should be noted that I am a sociologist of education, not an expert in STs. This article is based primarily on a review of English language sources, comprising books, articles, and several reports and presentations from reputable education institutions and research organizations. It is in the English-speaking world that most of the pertinent literature has been produced. I make an effort to synthesize the pertinent literature on STs and to offer some ideas of my own based on the argument that, in the same way war is too important to be left to the generals, knowledge is too important to be left to test developers and politicians.

## 2. Assumptions and processes in the development of STs

The design of a norm-referenced achievement test starts with the consideration of content: an examination of the curriculum and textbooks (Hamilton & Koretz, 2002). But soon other factors enter into play producing particular effects and consequences. Decisions on test design (test length, item format, content coverage) require trade-offs with respect to reliability, validity, and costs. Some of these considerations tend to take us far away from classroom realities.

There are many ways to assess the acquisition of learning: multiple choice, true/false, student portfolios, self-reporting inventories, among others. Further, the assessment can be formal (grades) or informal (Popham, 2009). STs derive their popularity from the relatively easy procedures they create: uniform administration, uniform scoring, convenience in reporting, and – with increasingly sophisticated digital computation – results highly amenable to quantitative analysis. STs are timed by design: usually, test administration is limited to one or two hours, to reveal what has been learned over a considerably longer period. Most STs are based on multiple choice (an American invention par excellence), although some include true/false responses.

Some reflections on the nature of knowledge are pertinent here. Multiple choice questions require precise answers (e.g., choosing response a, b, c, d, or e). Life, however, is not black and white and frequently engages us in situations where recognizing degrees of grayness is crucial, where a, b, c, d, or e choices are not so neatly laid out. Certain skills may not be measured with multiple-choice items (Le & Klein, 2002). I would add: certain skills *should not* be measured with multiple-choice items. It is paradoxical that STs are so prevalent in a country (the US) that places a premium – at least, discursively – on student-centered learning. Such learning calls for exploration and discovery, which means that the knowledge being developed does comprise degrees of ambiguity and contradiction. Student-centered learning implies that children are encouraged to pay attention to a variety of experiences; student-centered learning also implies that the student is a key agent in the knowledge acquisition process. So, it is the process of self-determination that should be functioning, not the forced selection of items in a multiple-choice format.

From a construction perspective, STs are based on two samples: a sample of the knowledge being tested and a sample of the population upon which the test will be normed. Regarding the first sample, a test contains only a fraction of all questions that could be asked on a given subject area (Le & Klein, 2002). In the second sample, norm-referenced means comparing the standing of the students being tested to «a larger group,» which really means a sample of students of similar age and levels of education (Hamilton, Stecher & Klein, 2002).

The logic of ST construction gives priority to the test's ability to discriminate among the test takers. It does so by employing the normal curve as the referent for the distribution of test scores. The normal distribution curve – whose precise name is normal distribution *probability* curve – assumes that the majority of people (64% to be exact) will have a satisfactory performance and that a few will have an excellent performance while another more or less equal few will have a poor performance. Basing ST scores on the normal distribution necessarily places half the students above the median score and the other half below it (McMillan, 2001). Examining this logic more closely, what the distribution of scores according to the normal curve implies is that satisfactory/complete knowledge on any discipline/area can never be acquired by all. This assumption is plausible, but an additional assumption of the normal curve is that while excellence is reserved for a few, failure is also created and assigned to one unfortunate group of students. This group is, of course, small, but nonetheless placed in the category of failure. So STs create both failure and excellence – warranted or not – by design. Further adding to the arbitrariness of test construction is the concomitant use of cut points to create categories. Hamilton and Koretz (2002:27) note, «The use of cut points is fundamentally judgmental and is often called into question.» Often, a difference of one point in ST performance can mean student placement in a different performance level. To be sure, teachers' judgments of student performance also have an irreducible element of arbitrariness, but it is the STs that claim full objectivity, while denying the considerable element of subjectivity behind the numbers.

From the perspective of test developers, these tests must be created to produce well-spread score distributions to attain high reliability indices. Popham (2001:48), a well-known US expert in instruction, evaluation, and measurement has observed that, «The more important the content, the more likely teachers are to stress it. The more that teachers stress important content, the better students will do on an item measuring that content. But the better students do on an item, the more likely it is than an item will disappear from the test.»

One may assign to a test all the good purposes one wishes but if the test has not been designed to meet those objectives, it becomes a victim of wishful thinking, and even dangerous. One reason tests cannot be used to measure teacher effectiveness or education quality is the frequent mismatch between local curriculum content and standardized test content. Popham (2001:43) remarks, based on his long experience with test development, «If you look carefully at what the items in a standardized test are actually measuring, you will often find that half or more of what's tested wasn't even supposed to be taught in a particular district or state.»

The logic of ST construction – actually, any form of testing – reduces knowledge to a few competencies. This is to be expected and does not exclude standardized testing from the array of valid assessments of knowledge. What this logic should do is to invalidate any claim made by educational policy makers that standardized testing is the sole measure of student knowledge and that teachers are the sole agent responsible for student performance in tests. The dominance of the testing regime is characterized by intellectual contradictions. Despite the strong support by policymakers on the use of STs to monitor schools and learning, it has been observed that, «Many educators and educational policymakers are largely untrained in test design and validation.» (Hamilton, Stecher & Klein, 2002: xiv).

### 3. Assumptions about the use of STs

The ultimate purpose of testing students is to improve educational practice: Teachers will adopt more affective practices and students will be motivated to work harder (Stecher, 2002). This objective makes three assumptions of how learning occurs: (1) learning is based on what the teacher teaches, (2) all students have the same possibility of learning regardless of differences in personal context, (3) the distribution of test scores along certain subjects (e.g., math, reading) or, even better, certain specific test items, enables the teacher to figure out how to improve her instructional strategies or her content knowledge about the subject.

While by no means arguing that teachers do not make a difference in learning, we have accumulated massive research findings that indicate that family influences on learning are powerful and surpass school/teacher influences. Some observe that family factors can explain up to 70% of the observed variance in test scores. A key assumption in the use of STs is that the determinants of student performance lie primarily with the education system and that the determinants of superior performance can be isolated and policies based on these causes can be transplanted into different contexts (Morris, 2016).

For teachers to use test scores to improve their practice with some disadvantaged students, scores would have to be given at both the item level and the specific student level. Test score distribution at such a molecular level seldom happens, even though it is possible with current digital analysis capabilities.

Hamilton *et al.*, (2002) ask a critical question: how do teachers use test data to change instructional practices? In other words, how would a teacher move into corrective action given a student's poor per-

formance in a given subject? Let us assume for a moment that teachers have access to the most detailed of test performance data: If I were the teacher and saw that Jim, my fifth grade math student, did very poorly on item three of the ST for his grade, how can I tell what went wrong? Did I not explain it well to him? Did he need more time to understand it? Was my instructional approach not effective with him? And if the majority of students in the class did better than Jim on that item, should I change my practice? In other words, what information does the ST item give me to personalize my practice vis-à-vis Jim? I would argue that the best source of data to improve my teaching skills would have to come from my daily observations and my knowledge of specific students. ST data – given their relative distance from the curriculum covered in class, the time elapsed between my teaching of the relevant module and the time I got the test results, and the unspecified connection between my teaching and the test results – would make the whole exercise a difficult piece of detective work for which I, as a teacher, would have few clues or the time to piece together those classroom memories into an answer. I would join other teachers in a common response: smile, accept the test results, and shelve them.

How many teachers receive feedback on their students' performance – a meaningful feedback that enables them to trace which aspects of a specific discipline might have been better explained or treated in the classroom? Hamilton and Koretz (2002:25) remark, «If performance is what the student knows and can do, neither the difficulty of test items nor the items' ability to differentiate between high and low performing students is necessarily important.» The same authors add, «If all students show maturity, then that is what the test should reveal.» (2002:25).

McMillan (2001:132) holds that one of the most serious misuses of STs is to use them to evaluate teachers. He states:

*«This is a misuse of the results for several reasons: (a) Standardized tests are not designed to evaluate teaching or teachers; (b) the content and skills tested will not have a perfect match with local curriculum or what individual teachers stress in the classroom; (c) each year brings a unique group of students to a teacher, with knowledge, skills, motivation, and group chemistry that may be different from other years; (d) it is difficult, if not impossible, to isolate the influence of differences between teachers (most of what students experience is common); and (e) a standardized test provides only one indication of student performance.»*

Porter (1995) and Apple (1996) were the first to observe a disjuncture between test developers/proponents and test users. Porter, a philosopher of science, has noted that in the production of STs there are two dissimilar expert groups: those who are experts in the development of tests and their quantification and those who are the experts in the content being measured. Morris makes a similar comment when he remarks that there is an «increasing detachment of policymaking from academia and the public sector.» In this view, «Those advocating reform are now directly and heavily involved in the policy-making process, primarily as the providers of independent “evidence-based research,” which serves to define both the nature of policy problems and their solution.» (Morris, 2016:8). Describing this state of affairs, the Graduate Institute of International and Development Studies (2016) puts it this way: «The epistemic community of education statistics and measurement is distant from the objects of measurement.» In this view, these communities are so distant that the experts in content seldom influence the experts in test development. It is to be noted that the UNESCO Institute for Statistics (UIS) has launched the Global Alliance to Monitor Learning (GAML) initiative, which seeks to ensure «technically sound and reliable

approaches to measure learning» (Graduate Institute of International and Development Studies, 2016). It does not seem clear to what extent this initiative will bring the two epistemic communities together.

Positive teacher responses to test results include: providing more instructional time, working harder to cover more material, working more effectively (Stecher, 2002). Despite the good intentions of teachers and school administrators, it is not clear how tests can lead to more equal opportunities or outcomes for students from different backgrounds. The distribution of scores may tell us that certain minorities are doing less well than the dominant ethnic group, but to move from there to improved practices would require institutional responses – not teacher responses alone. Unfortunately, there often emerges less an institutional than an individual response. In some US states, teachers are spending large portions of class time on test preparation, particularly in urban, low-income, and high minority school districts (Stecher, 2002)<sup>2</sup>.

Psychometricians develop tests but soon afterwards statisticians take over. The role of statisticians is to analyze the distribution of scores across national and various subnational levels and, often, to examine the correlates of test scores with various factors, such as student location (urban, rural), sex, age, and – when available – education and occupation of parents. What is constantly found is that family background significantly affects student performance and, thus, that economically and socially disadvantaged minority students persistently perform less well than wealthier students. Here, statisticians play a confirmatory, reproductive role.

At global levels today, the role of the statistician is growing in complexity and authority. Statistical experts from UN international agencies have been assigned to work on the development of indicators to measure progress in the implementation of the Sustainable Development Goals (SDGs), a global policy that will influence education in all countries between 2015 and 2030. Consequently, statisticians are spending considerable amounts of time searching for and defining indicators of «learning.» «Expert groups» have been appointed by governments and international agencies and are working hard on the identification of global indicators. Since the SDGs include education – i.e., goal number 4: «Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all» – efforts have been underway to design STs to measure learning across the world. Needless to say, the knowledge to be measured by the tests will be greatly decontextualized and subjected to standard features of ST construction, which include the identification of excellence – but also that of failure. What will come out of this? Both outcomes may be expected, but both are greatly influenced by family, community, and national circumstances that lie beyond the scope of any standardized educational testing and, therefore, beyond the capacity of such tests to judge teachers.

#### 4. What STs can and cannot tell teachers that will improve their practice

Popham (2009) makes a distinction between assessment *of* learning and assessment *for* learning. He is much more in favor of assessment for learning. According to him, there is strong evidence that instructionally oriented assessment, *if effectively implemented*, will improve student learning. Since STs are designed to discriminate among students, to assess student performance more completely (showing more of what the student knows), other forms of learning evidence should also be allowed.

<sup>2</sup> To see a full discussion of unintended consequences of STs on learning, see the study by Barrenechea (2010), which makes six additional critiques to standardized tests, i.e., the tension between multiple intelligence levels and the standardized form of evaluation, lack of attention to the real curriculum, risks to teaching to the test, incentives to cheat in the production of results, failure to consider student socioeconomic differences, and test limitations in predicting entrance to the labor market. Also relevant is an earlier work by Hamilton, Stecher, & Klein, 2002.

Several educators argue that STs cannot serve teachers because these tests do not have an instructional component, i.e., they are not designed to improve instruction; they are designed to determine system performance in general (accountability) (DePascale, c. 2004). Popham (2006:82) has remarked that most of the key state standardized tests are «instructionally insensitive – that is, they are unable to detect even striking instructional improvements when such improvement occurs.» This shortcoming arises in part because scores are forced to fit the normal curve (see above) and in part because the tests are so strongly linked to socioeconomic status that they measure what students bring to school rather than what they are taught there. If a student did not respond well to an item, it is not clear that this happened because the knowledge measured by that item was never taught by the teacher, badly taught, or the student learned it but subsequently forgot. Looking at aggregate data (whether national or even classroom) would not tell us this. Teachers would need to go item by item to identify response patterns.

Teachers are expected to be beneficiaries of testing, not by their own request but rather as the intention of policymakers. By having access to student standardized test results teachers are supposed to learn how to improve their teaching in terms of covering the content better, improving their instructional strategies, or both. The core assumption underlying testing is that by enabling teachers to confront the actual distribution of scores among their students, this constant monitoring will lead to constant professional development by education authorities and through the individual initiatives of the teachers themselves.

Let's examine the role of teachers in greater depth. To create a comprehensive conceptual framework to understand learning, teachers would have to be assigned a key function. To expect teacher performance that is satisfactory or higher, teachers would have to be:

- credentialed (initially trained in the details of pedagogical skills and in the subject to be taught),
- provided constant professional development, to hone their skills in the face of diverse and new socioeconomic contexts and challenges,
- well remunerated and recognized as professionals, to ensure high levels of professionalism and permanence in the educational system, and
- allowed access to an appropriate classroom environment, in terms of student class size, amount and quality of education materials, amount and quality of furniture and equipment.

In developed countries, teachers' conditions are generally satisfactory, although their training regarding «diverse and new socioeconomic contexts and challenges» could be much better. In many developing countries, the four conditions above are a remote fantasy. If many teachers operate under poor working conditions, how realistic is it to expect that learning should improve by confronting them with ST results? An analogy that comes to mind is to expect people to gain weight by weighing them regularly but without altering their diet. As noted earlier, one of the strongest findings we have gained from research on schools is the great influence of family wealth (or lack of wealth) on the academic progress of students. To neutralize this influence requires special measures and not just constant testing that ranks students and schools in highly predictable ways.

A relatively recent review of the use of STs by a prestigious commission appointed by the National Research Council in the US (Hout and Elliott, 2011) summarized contributions from economics, psychology, and educational measurement on the use of STs for educational improvement. It concluded that not only had these tests not functioned as incentives to improve student achievement but that they had actually decreased the rate of high school graduation. The study noted somberly that «despite using test-



ing for several decades, policy makers and educators do not yet know how to use test-based incentives to consistently generate positive effects of achievement and improve education.» (2011:92).

The global application of STs will bring additional difficulties. Such tests compare very different populations: some education systems include children with special needs or children of migrant workers, sometimes cities (e.g. Shanghai) are compared with countries (Morris, 2016). Moreover, rankings across countries introduce an element of unreality. Under what criteria can we expect students in developed countries that spend on average \$212 annually per primary school student and about \$368 annually per lower secondary student to register higher ST results than students in OECD countries, where investment per student in combined primary and lower secondary is about \$9000, not to mention \$11,000 in the US? (International Commission on Financing Global Education Opportunity, 2016). We should also not forget that STs are an acquired practice, which means that some students are more acquainted with such tests than others. For instance, STs are less frequent in developing countries than industrialized nations. Given the multiplicity of factors, Morris (2016:19) concludes, «The reasons underlying different levels of pupil achievement are inherently complex and explanations are conditional.» Further adding to the imprecise nature of STs are pressing logistical conditions. It has been noted that the Programme for the Analysis of Education Systems of the CONFEMEN (PASEC), which comprises 13 countries and aims to develop a common qualification framework, has a lag of two years between assessments and their publication. A similar time lag has been detected in the case of the South African Consortium for Monitoring Essential Quality (SACMEQ) (Graduate Institute of International Development Studies, 2016). Such delays likely remove test results as practical information that teachers might consider to improve their practice and situate the tests as general barometers of national education performance.

## 5. How knowledge should be measured and by whom

Through the widespread use of STs, we have seen a drastic reduction of knowledge transmission at the classroom level across the world as STs focus mostly on two disciplines. It is estimated that 99% of national assessments today measure math and reading, while topics linked to culture and art are covered in 36% of assessments, socio-emotional issues are treated in 18%, and physical wellbeing and learning approaches/cognition are covered is less than 1% (Anderson & Winthrop, 2016). In other words, important topics dealing with the joy and meaning of life as well as knowledge of how attain and negotiate relationships and conditions are not assessed. It is well understood that STs are not capable of measuring skills that have been learned but are not susceptible to easy quantification (Beck, 1976). It is encouraging that some knowledge and skills are *not* tested, yet it must be acknowledged that in heavy testing environments, what is not tested is simply taken out of the curriculum. In the US, a well-known critic of educational reforms laments that, despite the importance of fields such as history, science, literature, the arts, and politics, attention to them has diminished considerably (Ravitch, 2016). Echoing this concern, Gordon and Rajagopalan (2016) fear that the simultaneous measurement of decontextualized knowledge, fully deployed in STs, and the strong claim to precise quantification by measurement science may actually impede students' development of their own capacity to analyze, document, and appraise.

We live in a world of contradictions. On the one hand, we recognize that «education... should be directed to the all-round development of the human personality and to the spiritual, moral, social, culture and economic progress of the community, as well as to the inculcation of deep respect for human rights and fundamental freedoms.» (ILO/UNESCO, 1966). On the other hand, we have embarked on a

course of testing regimes that promote knowledge and skills that focus on labor market productivity, to the exclusion of more humanistic endeavors. Another significant contradiction is that, again on the one hand, there is growing evidence that improving the education quality in developing countries has not correlated with the increasing use of ST-based assessments, which has led several researchers to assert that large-scale assessments are weak tools for changing the classroom experience (Graduate Institute of International and Development Studies, 2016); yet on the other hand industrialized nations, led by the US and those in OECD, are resolutely committed to the use of STs and global comparisons.

No one disputes the authority of a university professor in grading her students. After watching student performance over a semester, we consider the professor as qualified to assess student learning. Why should we not give the same trust to teachers working at primary and secondary education levels, who spend not a semester but a year with their students? This lack of trust in teachers would be justified if the teachers are not credentialed or not given constant professional development. If this is the case, would not teacher training be the point at which to begin academic improvement efforts?

## 6. Conclusions

Standardized testing and the accompanying test-based accountability have taken solid root in the US since 2001 and are now integral to the evaluation of teacher performance. And, as noted, under the influence of organizations such as OECD, the march of STs across the world seems inexorable. It would be difficult to remove standardized student testing from the policy arena, but what can be done is to question their properties and thus erode their pervasive influence in determining the fate of teachers and dictating school behaviors (Hamilton, Stecher & Klein, 2002).

Administering STs is not inexpensive. In 1996, testing cost \$165 million and was projected to reach \$300 million in 2000 in the US (Hamilton, Stecher & Klein, 2002). By 2015, these projections had been notably exceeded when the testing industry was reported to be generating about \$2 billion annually (Strauss, 2015). There are additional costs linked to test administration and related activities. The General Accounting Office (1993) estimated the value of class time required for test administration to be seven hours per student or \$516 million during the academic year 1990-91 (Hamilton, Stecher, & Klein, 2002). So, why is standardized testing so prevailing and increasingly endorsed by policymakers? Several plausible reasons emerge, not necessarily mutually exclusive. One is that «Tests are cheaper than changing classroom practice though direct intervention.» (Linn, 2000:8). Asserting that tests will produce the information needed to improve the system tends to increase the government's (i.e., the backer's) legitimacy. Another possibility is that the current test regime involves so many firms and professionals that it would be difficult to erase their influence on policymakers. In 2015, the large test developers spent \$20 million in lobbying activities in the US (Strauss, 2015). A third explanation is that STs have generated their own research constituency: Some advocate the use of international tests for the opportunity they open to create data that enable all sorts of related quantitative analyses. Multiple regressions that use test scores as the dependent variable dominate studies of school effectiveness and these regression models generally look for causality, which is a noble endeavor in most sociological research. Many of these analyses, unfortunately, continue to be devoid of clear conceptual frameworks to explore such causality. Yet a fourth explanation would argue that as we move into the «knowledge science,» we enter the delusion that all things ought to be – and *can* be – quantified.

The logic of STs, as discussed above, incorporates several assumptions that intervene between teacher practices and a *true* assessment of student learning outcomes. It makes sense to want to know how teachers and schools are doing, but we have to remember that there is no direct link between this and ST test scores. Teachers observe ongoing student performance from the beginning to the end of the class or academic year. Because of this, teachers know their students' strengths and weaknesses more than any ST could discover, but the trend today is not to trust teachers' judgments. Conversely, teachers do not think much of STs. A teacher survey conducted by the Teachers Network (2007), based on a nationwide sample of teachers covering all levels of schooling, probed their opinions about NCLB testing. It found that, while 37% of the teachers saw such testing as «somewhat helpful,» a larger proportion (42%) found it «not at all helpful.» About 40% of the participating teachers also stated that those tests encouraged rote drilling and eliminated curriculum materials not tested. Most damaging of all, was the teachers' assertion, with 69% strongly agreeing, that the Adequate Yearly Progress goals (a provision of NCLB and that was based on ST performance) had «contributed to teacher burnout.» ST objectives, therefore, may end up functioning as highly oppressive requirements.

Educators and policymakers alike must continue to address three crucial questions: How to encourage learning? How to encourage teacher professionalism? And, what is the relationship between measurement and quality of education? Standardized testing is incapable of measuring with any precision levels of critical thinking and high-order reasoning. Further, educators fear that STs remove, if not erase, teachers' judgments, all the more so since tests are designed by experts distant from the classroom environment (Apple, 1996).

STs can play a useful role in understanding basic trends in knowledge acquisition by successive cohorts. They can also play a role in overall monitoring of the educational system and *societal* performance. But two points should be made clear: (1) the assumption that ST results can be readily used to improve teaching practice is simply erroneous, and (2) it is possible to monitor national performance without necessarily having a regional or global comparison. We should make sure that tests cover all important elements of the domain; for that reason, it is necessary to rely on content experts, not test developers alone (Le & Klein, 2002).

Poor performance of a country's education system gives rise to expressions of alarm and, in some cases, to serious concern over teaching quality. The unintended consequences of STs, and particularly of their misuse, have been «a decline of trust in public education and increased pressure from both users and business firms to turn to the private sector partially or wholly for solutions.» (Graduate Institute of International Development Studies, 2016). In developing countries, teacher training, whether pre-service or in-service, continues to be in need of reform and improvement. Enabling teachers to become better professionals rather than expanding the use of STs – this is where our priorities should be directed.

## References

American Education Research Association, American Psychological Association, & Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington: American Education Research Association, American Psychological Association, & Council on Measurement in Education.

- Anderson, Kate & Winthrop, Rebecca (2016): "Building Global Consensus on Measuring Learning" in Simon McGrath and Qing Gu (eds.): *Routledge Handbook of International Education and Development*. New York: Routledge.
- Apple, Michael (1996). *Cultural Politics and Education*. Buckingham: Open University Press.
- Beck, Robert (1976): "Equal Time: Standardized Test Results: Answering Our Critics and Ourselves". *The English Journal*, 65 (3), 65-71.
- Barrenechea, Ignacio (2010): "Evaluaciones Estandarizadas: Seis Reflexiones Críticas". *Archivos Analíticos de Políticas Educativas*, 18 (8), 1-27.
- DePascale, Charles. (2003). *The Ideal Role of Large-Scale Testing in a Comprehensive Assessment System*. National Center for the Improvement of Educational Assessment.
- Gordon, Edmund & Rajagopalan, Kavitha (2016). *Testing and the Learning Revolution: The Future of Assessment in Education*. New York: Palgrave Macmillan.
- Graduate Institute of International and Development Studies. (June 2016). *Learning from Learning Assessments: The Politics and Policies of Attaining Quality. Roundtable Report*. Geneva: Graduate Institute of International and Development Studies.
- Hamilton, Laura; Stecher, Brian & Klein, Stephen (eds.) (2002). *Making Sense of Test-Based Accountability in Education*. Santa Monica: Rand Corporation.
- Hamilton, Laura & Koretz, Daniel (2002): "Tests and their Use in Test-Based Accountability Systems" in Laura Hamilton, Brian Stecher & Stephen Klein (eds.): *Making Sense of Test-Based Accountability in Education*. Santa Monica: Rand Corporation.
- Hefling, K. (17 January 2015): "Do students take too many tests?" Congress to weigh the question. The Rundown (a blog of news and insights).
- Hout, Michael & Elliott, Stuart (eds.) (2011). *Incentives and Test-Based Accountability in Education*. Washington: The National Academies Press.
- ILO/UNESCO (1966). The ILO/UNESCO Recommendations concerning the Status of Teachers. <[http://www.ilo.org/sector/Resources/sectoral-standards/WCMS\\_162034/lang--en/index.htm?ssSourceSiteId=global](http://www.ilo.org/sector/Resources/sectoral-standards/WCMS_162034/lang--en/index.htm?ssSourceSiteId=global)>
- International Commission on Financing Global Education Opportunity. (2016). *The Learning Generation: Investing in Education for a Changing World*. <<http://report.educationcommission.org/report/>>
- Le, Vi-Nhuan & Klein, Stephen (2002): "Technical Criteria for Evaluating Tests" In Laura Hamilton, Brian Stecher, & Stephen Klein (eds.): *Making Sense of Test-Based Accountability in Education*. Santa Monica: Rand Corporation.
- Linn, Robert (2000): "Assessments and accountability". *Educational Researcher*, 29 (2), 4-16.
- McMillan, James H (2001). *Essential Assessment Concepts for Teachers and Administrators*. Thousand Oaks: Corwin Press.

- Morris, Paul (2016). *Education policy, cross-national tests of pupil achievement, and the pursuit of world-class schooling*. [Based on an Inaugural Lecture delivered December 2015 at the UCI Institute of Education]. London: UC Institute of Education.
- Popham, W. James (2009). *Instruction that Measures Up: Successful teaching in the age of accountability*. Alexandria: Association for Supervision and Curriculum Development.
- Popham, W. James (2006): "All About Accountability/Assessment for Learning: An Endangered Species?". *Educational Leadership*, 63 (5), 82-83.
- Popham, W. James (2001). *The Truth about Testing: An Educator's Call to Action*. Alexandria: Association for Supervision and Curriculum Development.
- Porter, Theodore (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Ravitch, Diane (2016). *The Death and Life of the Great American School System: How Testing and Choice are Eroding Education*. New York: Basic Books.
- Redden, Elizabeth. (5 June, 2015). "Objections to the OCED's AHELO", Inside Higher Ed. <<https://www.insidehighered.com/quicktakes/2015/06/05/objections-oecd-ahelo>>
- Resmovits, Joy (29 November 2012): "School Testing in US Costs \$1.7 Billion But That May Not Be Enough" [Report]. *The Huffington Post*.
- Scherer, Marge (2005): "Reclaiming Testing". *Educational Leadership*, 63 (3), 9.
- Stecher, Brian (2002): "Consequences of Large-Scale High-Stakes Testing on School and Classroom Practice" in In Laura Hamilton, Brian Stecher, & Stephen Klein (eds.): *Making Sense of Test-Based Accountability in Education*. Santa Monica, California: Rand Corporation.
- Strauss, Valerie (March 2015): "Big education firms spend \$20 million lobbying for pro-testing policies" [report]. *The Washington Post*.
- Teachers Network (2007): Teachers Network Learning Institute: Survey: No Child Left Behind: What people are saying. Teachers Network. <[teachersnetwork.org/tnli/survey\\_highlights.htm](http://teachersnetwork.org/tnli/survey_highlights.htm)>

## Nota biográfica

Nelly P. Stromquist es profesora de políticas educativas internacionales de la Facultad de Educación de la Universidad de Maryland. Especializada en temas relacionados con el género, las organizaciones dirigidas por mujeres, la educación popular y el impacto de la globalización en el profesorado, que examina desde una perspectiva crítico-política. Algunos de sus libros más recientes son: *Critiques and alternatives* (coeditado con S. Klee y J. Samoff) y *Feminist Organizations and Social Transformation in Latin America*. Fue premiada con el galardón Krestin Hesselgren al profesor invitado 2012, otorgado por el Consejo de Investigación de Suecia. Su último libro, publicado en 2017, es: *Women Teachers in Africa. Challenges and possibilities* (coeditado con S. Klee y J. Lin).

# Institutional Functions as a Theoretical Nexus and Measurement Target in Sociology of Education

Funciones institucionales como nexo teórico y objetivo de medición en sociología de la educación

Alexander Osipov<sup>1</sup>

## Abstract

The paper is devoted to methodological aspects of the study of institutional functions of education as one of the basic theoretical constructs of contemporary sociology of education. The author argues that the elaboration into the function problem also provides a direction for social-practical outputs of this branch of sociology, conceptual ground for its own theoretical, revitalizing its connection with general sociology. Without such elaboration, the author assumes, most of theoretical and empirical studies and measurements conducted by sociologists are doomed to thematic fragmentation, have slim chances to contribute to a shared sociological vision of education, delineate the area of institutional responsibility of education in society, improve educational politics, resist the neoliberal «Trojan horses» in education. The author brings a set of criteria to define the institutional functions of education, including the empirical measurability of each function, and a hypothetical model of functions detailed by four major areas of society: economic, social, cultural and political.

## Key words

Education, institutional functions, criteria of definition, sociology.

## Resumen

El presente texto aborda aspectos metodológicos del estudio de las funciones institucionales de la educación como uno de los constructos teóricos fundamentales de la sociología contemporánea de la educación. El autor sostiene que la reflexión sobre el problema de la función proporciona además una dirección para resultados sociales prácticos de este campo de la sociología, revitalizando su conexión con la sociología general. Sin tal elaboración, a juicio del autor, buena parte de los estudios teóricos y empíricos llevados a cabo por los sociólogos están condenados a la fragmentación temática, y tienen pocas ocasiones de contribuir a una visión sociológica compartida de la sociología de la educación, a una definición del área de la responsabilidad institucional de la educación en la sociedad, a una mejora de la política educativa, a una resistencia de «los caballos de Troya» institucionales en educación. El autor presenta un conjunto de criterios para definir las funciones institucionales de la educación, incluyendo la mensurabilidad empírica de cada función, y un modelo hipotético de funciones detalladas por cuatro áreas principales en la sociedad: económica, social, cultural y política.

## Palabras clave

Educación, funciones institucionales, criterios de definición, sociología.

Recibido: 23-03-2017  
Aceptado: 23-04-2017

<sup>1</sup> Yaroslav-the-Wise Novgorod State University (NovSu), Moscú, [osipoval58@gmail.com](mailto:osipoval58@gmail.com)

## 1. Concepts of functions-nexus of the sociology of education

In the early 21<sup>st</sup> century, the sociology of education became the most popular area of sociology in the world. But the area's leading experts are now speaking openly about the methodological, conceptual and political factors slowing and even halting its development in their respective countries (Osipov *et al.*, 2013). Russian scholars are also writing about «working on the mistakes» in the field and advancing a list of its priorities upon whose achievement, they believe, it can become helpful in formulating a strategy and concrete solutions in educational policy.

In our opinion, the weakness of today's sociology of education in Russia and around the world is largely associated with the unpreparedness of scholars to solve the *threshold* problems which, firstly, form the theoretical carcass of this scientific field, and secondly, ensure its greatest-possible socially-practical return. Indeed, this science has thus far failed to prove that it is essential to society, to the administration of the education system.

## 2. Concept of functions – nexus of the sociology of education

The main and most difficult *threshold* was outlined by the classics of sociology and focused on the problem of the functions of education as a social institution. The centerpiece of this problem in the subject area of the sociology of education is not as clear for many scholars. In general sociology and the different branches of sociological theory, the problem of institutional functions has been posited with less thoroughness than other, more peripheral and emergent concerns (education accessibility, satisfaction with the education one receives, values-based attitude towards education, etc.). This problem is all but ignored in debates on the efficacy of education policy or the prospects of particular education systems (Antikainen *et al.*, 2011). It is often perceived as an attribute of the long-since repudiated functionalism. But is it really possible to turn away from this problem if sociology itself endeavors to identify sustainable links between major phenomena and organizations in society? And if sociology withdraws into the shifting structures of the signs, metaphors, identities, networks and practices of everyday life, will it not run itself ragged in an endless circle without making any discernable progress for the theory and system of education? Thus, insofar as it is impossible to expel the problem of institutional functions from the framework of sociology, we shall briefly endeavor to return to its sources.

Durkheim wrote that education, as the organized («methodical») socialization of individuals, ensures the homogenization of society. He ascribed to functions the sense of sustained activity aiming the long-term impact of this institution towards the greater benefit of the entire social system. But he has not marked any theoretical boundaries of the category of «function.» While Durkheim's arguments on the functions of education remain unproven, they retain the hope, moving forward, of the application of sociological knowledge about the functions for the purposes of educational governance (Durkheim, 1922:51). Later, T. Parsons hypothetically identified four functions of education: academic, distributive, economic and political (Parsons, 1959). But this set of functions can be found attributed to almost every social institution – it does not reflect the specifics of education. Concordant with the hope referenced by Durkheim was the function problem advanced by R. Merton. As Merton stated in his more general theorem, «the social functions of an organization help determine the structure... just as the structure helps determine the effectiveness with which the functions are fulfilled» (Merton, 1968:136). In other words, the organizational structure of a system must be adequate to the given functions, while knowledge of the functions is consequently applicable in organizational and social management.

The classicists of the sociology of education did not describe the method to identify functions. They used to postulate certain functions without regard to the criteria or methods to be utilised in the search of functions, and this lack of substantiation persists to this day. Well-known sets of functions (those given by Durkheim, Parsons, Merton or other contemporary authors) are cited speculatively without any theoretical justification, and only sporadically accompanied with some empirical illustrations.

The function problem is, in fact, the theoretical backbone of the sociology of education. This certainly does not exclude other planes for the analysis and theoretical constructs of phenomena occurring in the educational arena, but it is precisely the concept of the institutional functions of education which is capable of encompassing the multiplicity of other constructs (including the microsocial) within the overarching architecture of sociological knowledge, giving it both compatibility and operability. The stream of education functions (both manifest and latent ones) encompasses all social interactions among actors in the sphere of education. It would be expedient to conduct the sociological study of such interaction in the context of the functions of education, which would make it possible to create a more holistic picture of the life of educational commonalities from the standpoint of the efficacy of their particular education programs or strategies. Proceeding from the concept of functions, researchers can move beyond the interaction between teachers and students traditionally investigated by pedagogy and psychology and move into the other social spheres with which education is connected by virtue of its functions, including, *inter alia*, economics, social inequality, politics and culture.

The concept of institutional functions plays a methodological, navigational role in the sociology of education, without replacing or substituting for other problems and levels of their investigation. This concept is not limited to the structural level, but reflects the totality of social interactions in education, including those at the micro-level. It can provide impetus to the internal branches of the sociology of education and serve as the basis for the integration of their results into the sociological theory of education.

A detailed version of the functions of education was published about twenty years ago in Russia, without critical response nationally (Osipov, 1998). Hypothetical in nature, it searched for the universal socio-historical features of education. It would be worthwhile to continue the discussion on theoretical backgrounds of the study of the institutional functions of education, the scope of the search for the functions, the boundaries of education's responsibility in society, the criteria of functions and the methods in justifying the functions. It is also important to specify the practical potential of the concept of the functions of education.

The concept of the institutional functions and understanding of the system structure of education are complementing each other (Osipov, 2006:171-211), thus delineating the problem field of the sociology of education and creating a theoretical foundation for its integration and interactions with adjacent research areas (those outside of the domain of education).

### 3. The search for functions in the context of social practice

The reason of our addressing the functions of education lies in determining the scope of the institutional responsibility of education, as well as in arranging a rational organizational structure of education at all of its functional levels – from administering in individual school settings to the state educational governance.

We usually accept that one of the education's institutional functions is its involvement in the reproduction of the occupational and qualification composition of the population. But then one can view as a fundamental misunderstanding the fact that the administrative structures of vocational education (from



individual colleges to regional committees and the federal ministry) has no internal divisions (departments, units) responsible for this particular function. Such divisions should have systematically keep up studying demands in the employment realm. Russian colleges and universities used to set their admission plans basing on their own corporate interests (which is viewed as their *corporacentrism* – organisational ignoring the actual needs they are thought to serve to). This regularly led to imbalances in the structure of graduates, whereby half of them now are working at jobs irrelevant to their specialization and one in six is unemployed. The situation was that wrong for decades, while primary and secondary vocational education badly suffered from an acute shortage of applicants.

Heeding this example, the organizational structures of the education system and its local establishments should be built in accordance with the appropriate functions. From here, elaborating the problem of education's functions can call into question certain approaches and criteria of its administration – might expose the *corporacentrism* in the system and its establishments (Osipov, Ivanov, 2004). It could help find a balance between responsibility for fulfilling the functions of education and the countervailing array of financial-economic and legal norms at the state level.

In contemporary sociology, most of interpretations of the functions of education have «frozen» at the level of Durkheim, Parsons and Merton. Mention is usually made of some common functions: socialization, cultural transmission, social control, selection and the attainment of social status (Ballantine, Hammack, 2009). They usually refer in a very general way to the societal mission of schooling, but are limited to the social sphere. In this form, the knowledge of functions is not transmittable into education practice and policy. In Russian sociology, the interest towards institutional functions of education has been lost. Yet it does not entail the practical application of the function knowledge.

Thus, the idea of the *functionality* of social institutions remains, while the study of the functions of the same institutions continues to linger in the shadows. Yet, there can be no doubt that two phenomena and concepts – *social institution and institutional function* – share a common fate. These concepts will either be conceptually formed together or fade from scholarly horizons, at which point the *social institution* as a sociology's fundamental notion will come into question (or just die). If sociology refrains from elaborating the problem of functions as it applies to those social institutions which are administered by the state (education, healthcare, social security, etc.), then these same institutions will be doomed to remaining under bureaucratic control. In Russia, the literature on educational management totally ignores the problem of the functions of education, which helps explain the administrative leapfrogs and frequent revisions in the rules, standards, competencies and reporting (accounting) forms.

And what is the range of phenomena that encompasses the problem of functions? Sociologists usually define education as organized and *purposeful* (methodical) socialization (Durkheim, 1922). While the criterion of this purposefulness requires its sociological interpretation, it does distinguish *education* from the multitude of *spontaneous socialization* processes. Education can be logically divided by the criterion of formal organization into two spheres: manifest (formal) and latent (informal). *Formal education* is represented with schooling (and various programs) that concludes with the issuance of the respective certificates. *Informal education* encompasses the socializing influences exerted by family, religion, media, advertising, cinematography, etc. which are subject to sociological study.

The actors, who are methodically included in the processes of socialization, implement certain education programs that meet their own social strategies. A program assumes the objectives and content of

socialization in a particular area, as well as the respective methods and providers. Methods are frequently highly sophisticated, with an emphasis on special target-audience research and effectiveness metrics, etc. An education program is a systemic substantiation of any purposeful socializing influences – from those emanating from the family or the mass media, which are seemingly routine and spontaneous, to those more long-term and organizationally regimented, as practiced in the instruction of the professions. While the agencies of formal education implement a quasi-synthesis of manifest and latent programs, the programs of informal education are frequently latent, advancing the legitimization of particular values, norms, styles and practices.

Latent education programs promote certain cultural types that are at odds with the aims of manifest programs. Thus, countervailing the officially-declared array of social values and norms (life, health, work, justice, accountability, law-abiding behavior, patriotism and healthy lifestyle, etc.), latent programs, acting according to the principle of the freedom of choice in entertainment and the idea that «whatever is not prohibited is allowed,» advance the opposite values (force, risk, chance, independence, individualism). Unfortunately, this often goes at the expense of the citizenry – the taxpayers. Manifest programs are often designed to overcome the effects of latent ones, thus reflecting hidden institutional conflicts in the society.

The criterion for assigning latent programs to the sphere of education is the fulfillment of the general functions intrinsic to education as social institution. Once these general functions have been identified, it becomes possible to search for the specific functions that distinguish one sector or level of education from the other.

#### 4. Institutional function in application to education

The dissonance in defining the functions of education is explained by the fact that the education's social output is often delayed across time and manifested in sweeping results that are resistant to accounting, as well as by the crossover interference (juxtaposition) of a number of discrete functions. The same may be said of the return on *manifest* and *latent* education programs (*formal* and *hidden* curricular). Theorists would tend to keep off the function problem. Sociologists have made little (if any at all) effort to elaborate approaches to the measurement of functions, restricting themselves to philosophical interpretation in such vague terms like «mission», «social role of schooling», etc. But the very concept of «function» as it applies to the social institution is deserving of elaboration across a whole range of issues: on the correlation between functions and other adjacent phenomena and categories, on the methods for identifying functions, on the stability or variability of functions, and on their specificity as intrinsic to a particular institution.

Some functions of education are not recognized for long by public perception and science at certain intervals of history, at which point those functions fade from the ranks of education system's public tasks. On the other hand, dominant groups and organizations may emphasize certain function in an effort to elevate it to the status of the manifest tasks of education. Thus, the fulfillment of functions is conditioned by the social structure, social ideology and society's dominant groups in economics, politics, social development and culture.

The vocabulary of contemporary sociology insufficiently delineates the adjacent phenomena in the problem under consideration. Clarification of these phenomena and concepts has been partially provided in descriptions of the interrelatedness and disparity of *functions*, *public tasks*, *social expectations* and *social requests* in the area of education. Institutional function is viewed as a socio-historical universal feature of this institution (Osipov, 2006:212-227, 462-478). Thus, the identification of functions may feature the

application of historical-comparative methods making it possible to demonstrate the social-historical universality of particular characteristics of education as social institution.

Publications touching on the problem of the education's *functions* have identified immediate or prospective *tasks* of education, formulated from the position of a particular sociological or socio-political paradigm, but insufficiently addressed the institutional specifics of education. We shall endeavor to set out those aspects of the *function* as a category that would be helpful in approaching a clearer theoretical understanding of the phenomenon.

Function is a stable sort of influence that education is exerting on the main spheres of society. Within those spheres, the influence is exerted on their structures, connections and actors. Institutional functions are those socio-historical universal features of education that are resistant to rapid or frequent change. The method for the theoretical delineation of the functions lies in the critical filtration of education's manifestations from ideological stamps, situational and transient characteristics in the sphere of education. The search for functions rests on comparative and socio-historical analysis, and this method – in combination with the criterion of empirical operability – has made it possible to develop a fuller, albeit not exhaustive, version of the institutional functions of education. The weakness of this proposed concept remains the dearth of experience in its empirical and socio-practical approbation.

## 5. Criteria of *function*

In defining the functions of education as a social institution and of individual organizations as its discrete elements, it would be prudent to set the appropriate boundaries so as to avoid arriving at theoretically and practically unproductive results, confusing adjacent social phenomena in education. This is particularly important at the initial stage of formulating the problem of institutional functions, when a preliminary, theoretical (logical, hypothetical) search is conducted and fragmentary argumentation is accumulated. This search and argumentation have thus far not been elevated to the sort of comprehensive investigations that would make it possible to measure functional processes and evaluate the substantiation of theoretical discourse in this problem area.

Firstly, all of the elements of the function model of education as an institution must be compatible with the categorical apparatus of sociology, and moreover – serve as a constructive basis for empirically-measurable indicators of the function of education encompassing in their totality all of the realities of this particular social institution. Important in this respect is the general-sociological definition (conceptual operationalization) of education as an institution (Osipov, 2006:311-313). This general definition in no way excludes other, more specialized definitions of education which may be necessary in revealing, by way of intensive operationalization and empirical interpretation, the specific properties of this social institution – for example, from the standpoint of the theory of cultural-linguistic codes or the labeling theory. However, the conceptual terminological compatibility as a methodological requirement for the search of functions within the scope of the social sciences remains.

Secondly, the content of formulations of the institutional functions of education must reflect its most stable characteristics, associated with its socio-historical universals. This makes it possible to distinguish functions ensuring the integrity of society and its reproduction from fleeting and short-term *social requests* and, consequently, imbalances in sociological constructions of institutions.

It's important to distinguish logically the two sides of the problem from one another. *Function*, as a logically – and empirically – definable sort of stable connection ensuring the integrity of society and characterized with socio-historical universality. Function shows up always and everywhere in the sphere of education. Function is an objective parameter of education allowing for an evaluation of the conditions of education as a social institution, its dynamics, scope and impact on society as a social, socio-cultural system. *Functioning* is a process unfolding within the bounds of a particular kind of stable connection. The content of this process may change in the same sense as the content of socialization processes, in which a wide range of values, norms and behavioral patterns are reproduced.

Proceeding from this distinction, one of institutional functions of education is the organized reproduction, whether at the group, community or societal level, of certain types of culture. But in different socio-historical situations, different types of culture are subject to reproduction – national, peasant, religious, proletarian, bourgeois, professional, etc. This in no way leads to the *changeability* of functions, since what changes is the *content* of functioning – but not the *sort* of stable influence (that is, function).

Using a well-known analogy, we note that under normal conditions, an individual bodily organ (digestive tract, eye, nervous system, etc.) as a system fulfills its own (the same stable) functions, irrespective of the content of the functioning process (food ingredients, lighting conditions, nature of external or internal irritants), thus supporting the wholeness of body. In physiology, this distinction makes it possible to identify the *functions* of each of the organs by distinguishing them from the various situational manifestations of the *content of functioning* (the digested food, conditions, influences). That allows medicine to monitor the organ's status according to its stable parameters (functions), cure it (that is, overcome dysfunction), and thereby maintain conditions that are favorable for the body as a whole.

In our view, it is essential to clarify the categorical apparatus used in studying the functions of institutions, including education. Otherwise, the classical terms and concepts of sociology with which the theoretical constructs of social institutions and sociology of education are associated (just as other mid-level theories) will remain shaky elements of social theories.

Thirdly, formulations of institutional functions must serve as the specific characteristic of this institution, and be minimally superimposed on the characteristics of other institutions. Otherwise, the concepts of other institutions will begin merging and converging – first and foremost through the convergence of the characteristics of their functions. This criterion is rooted in a sociological concept that defines the social institution as a relatively stable assemblage of norms and organizations regulating the satisfaction of a particular sphere of societal and human needs. This criterion is associated with the development not only of the sociology of education, but also of other mid-level sociological theories.

This criterion upsets the usual notions of the institution of education, but meeting this criterion is still possible. Thus, the aspect of the institutional functions of education from the standpoint of social reproduction (of human, of culture) could result in the recognition that reproduction occurs through the efforts of all of the main social institutions. The reproduction paradigm states that all institutions, in one way or another, fulfill certain, the same range of functions: they all regulate the needs of individuals and groups, conduct socialization, exercise social control, differentiate communities and reproduce certain elements of culture, thereby ensuring the integration of society and maintaining its wholeness as a system. Admitting this vein of functions could lead to the conclusion that in the functional sense, all social institutions sometimes duplicate one another in fulfilling their respective functions in the *socium*.

In that case, the assertion that one of the institutional functions of education is reproduction of culture fails to express the specifics of this institution, since the *reproduction of culture* in society is supported regularly by the state, the economy with its organization and division of labor, family and religion. But sociology would need to distinguish stable parameters expressing the education's institutional specifics. This makes it possible to develop the concept of social institution, as well as its specific efficacy criterion and the optimal model of its organisational structure, in view of its own functions for the purposes of educational management.

From here, formulations of institutional functions of education must minimally superimpose (overlap) on the functional characteristics of other institutions. However, to avoid mutual interference, overlapping of formulations functions of different institutions completely will hardly be possible. This is due, firstly, to the fact that social life exhibits the complementarity of role structures and institutions as normative complexes regulating interactions of people. Secondly, the theoretical study of institutional functions has thus far failed to achieve a clear distinction between structures and processes. Thus, the socialization processes reveal partial interference or conflict of functions on the part of the family, religion and school. From here, the absolute delineation of institutional functions is unachievable, but this does not discard the task of formulating a specific concept of the functions of a particular institution, including education.

Fourthly, the concept of the education's functions should not suffer from excessive conciseness. Formulations of specific functions must be clearly operational within the system of sociological terminology, but most importantly – be open to empirical interpretation. This would make it possible for the theoretical constructs developed in the sociology of education to approach empirical measurements, applied research, and practical solutions based on the concept of the institutional functions of education.

These are the general theoretical concerns to take into consideration to approach a clearer sociological vision of the place and role of the institution of education in society. We believe that sociologists can accept the importance of such a consideration when they design their theoretical and/or empirical research on education, when they feel concern about how their future research results would contribute to developing a sociological theory of education, improving educational policy. We also dare hope that the proposed theoretical model of education functions can be corrected and developed in the future. Then, the sociology of education would begin approaching the level of knowledge at which, according to R. Merton, the interpretation of functions serves to determine the optimal structure of its organization and efficacy, and the sociology of education ceases to play only a role of delayed/late criticism of education policy. For now, the general impression is that new areas of thematic and conceptual advancements are emerging within the sociology of education, while its basic constructs and concepts remain foundationally unsubstantiated. Given the obscurity of these basic structures, overcoming the theoretical crisis in contemporary sociology, including in its mid-level theoretical areas, would seem elusive.

Interaction between the education system and other institutions is bilateral in nature. The system itself reacts to the dynamics of other institutions, altering its internal structure or content according to changing conditions. This reaction most often is lagging behind, as the classics of sociology noted back in their day. Now let us discuss a version of the institutional functions of education in application to the main spheres of society's life: economic, social, cultural and political.

## 6. Functions of education in the economic sphere

- a) Formation of the structure of workforce. This function appears locally, nationally and globally in both its quantitative and qualitative aspects. In the quantitative aspect, the education system is responsible for the formation of occupation-qualification proportions in the workforce. In practice, however, this influence is almost never optimal. Due to its institutional and organizational inertia or administrative errors, the education system leads to the overproduction or underproduction of individual professional corps. This results in a situation where certain occupational groups experience an influx of people without the appropriate preparation, and workplace training becomes widespread (without the appropriate theoretical knowledge and creative skills). Both, overproduction and underproduction has a negative impact on the vocational structure of the population, leading to an *erosion of professional cultures*.

The qualitative side of this function ensures the formation of the productive qualities of workers. It is primarily conducted by vocational-education institutions, but also in the framework of labor activity and general education. According to statistics, a worker's labor productivity and innovation activity increases as his educational attainment level expands.

- b) Formation of the public's consumer standards. Education's influence on the economy concerns not just the production, but also the consumption of goods (information, cultural values, financial and natural resources, etc.). This function has always been inherent to education – suffice it to recall the biblical commandments about the moderation of consumption, and this function also largely determines the content of informal education in the family and neighborhood community.

The education system can instill rational standards to social, material and spiritual needs. In this way, it contributes to the emergence of a resource-saving economy and a human-friendly environment. But in the market economy this function of education is opposed to the interests of business, as embodied in commercial advertising with its hidden message to «consume more!» This advertising intensifies *consumerism* – a strain of the mass culture in which consumption serves as a criterion of progress. An example of the formation of consumer standards is the introduction of academic courses on environmental protection, social ecology, healthy eating, personal hygiene, public health and safety, civil and consumer rights, etc.

- c) Attracting economic resources for the purposes of the education system. The education system is a major consumer of public resources. It attracts and expends the economic (material, financial, etc.) resources without which its functioning is just impossible. The sources of these resources vary largely, ranging from the state budget to private investments and paid services, but these sources point to the *dependence* of education on other institutions and organizations.

The function of attracting economic resources explains a lot in the content and forms of education. Reliance on the state budget intensifies unification, while the orientation towards business circles strengthens the autonomy of education in terms of its content and organization. As schools shift from federal state to local (municipal) budgets, the development of the regional and local component in academic content has been activated.

*Money-making* by educational settings has become a special feature of this function. This feature reflects a liberal-market orientation towards education as a service which, together with the expansion of *the right to choose* leads to the segmentation (organizational stratification) within the system of education, to the weakening of its mission of forming the moral character of the next generations and the social

integration of society. The commercialization of education is generally frowned upon by sociologists and educators, but remains among the performance indicators of its organization, as applied by the top management of the education system.

- d) Internal distribution of economic resources. As represented by its various management bodies at all levels (federal ministry, regional departments, school administrations, etc.), the education system distributes the funds it attracts among its many regions, internal subsystems, establishments, organizational subdivisions and jobs. As a result, resources are allocated for different sectors – including those outside the educational process itself (support for commercial structures, research and ranking agencies, security services, social and health protection services for students and teachers, advertising, etc.).

This distribution is capable of intensifying social inequality when some subsystems of education can be lacking the required resources. Thus, the transmission of kindergartens in Russia to the municipal budget may reduce their ability to maintain a high level of preschool education. Children deprived of preschool training are frequently incapable of subsequently coping with primary school programs. Social consequences of such practices of the resource distribution turn out to be of tremendous importance to society as a whole.

The education system is also capable of modifying the internal criteria for stimulating the performance of students and staff, thereby changing their socio-economic statuses.

The distribution of economic resources is predicated on the nature of the social organization of educational establishments or their networks. The social predicate of this distribution turns out to be even more substantial than the economic predicate – after all, economic criteria often lack direct application in the sphere of education. Pushed to the forefront are those criteria that reflect the normative outcome of consensus among certain professional groups inside education or around it. At times, the education system undertakes this distribution in defiance of all common sense. For example, the Soviet system of teachers training overproduced teachers, which in turn stunted growth in teacher incomes, the improving teachers' housing conditions, and the technical equipping of schools. Another example of distribution contrary to the interests of the education system is the annual generous allocation by the RF Government of billions in budgetary subsidies for a select few of the leading universities with the aim of elevating their positions in the global rankings and foreign citation indices. This kind of «selective» support for rich universities does nothing to address the financial plight of the majority of universities, but hints at the lobbying of foreign business interests (ranking agencies, registration-publication systems, etc.).

From the standpoint of sociology, this raises the question of the social mechanisms behind the distribution of economic resources in the education system. What status groups within the education system – or beyond it – have a decisive influence on these mechanisms? Such questions must be focal to the sociological understanding of emerging trends in the development of the education system.

## 7. Functions of education in the social sphere

- a) Homogenization of society through the organized socialization of individuals. Durkheim was the first to mention the function of homogenization. In a society undergoing social and cultural differentiation, this function ensures the reproduction of a common system of social norms and substantive knowledge with the organizational support of individual educational establishments and educational communities.

- b) Formation and reproduction of educational communities. The sphere of education is regularly forming and reproducing educational communities, within which individuals are connected with each other by their mutual participation in educational (learning, instructional, managerial) processes and their particular value-conscious attitude towards education and knowledge. These communities spring up inside and around the domain of education. They are exemplified by various professional, parental, student and other groups (including mixed ones) participating directly or indirectly in the performance, maintenance or management of educational activity. These groups and communities occupy their own place in the social structure of society, exerting a real influence on the functioning of education through the various channels of social mandates and support for particular education strategies (trajectories).
- c) Activation of social mobility – a perceptible function of education. It manifests itself to the extent that in a system of inequality, achievable statuses depend largely on formal educational attainments of individuals.

Education has evolved to become one of the mass *social channels* and *social elevators (lifts)* of social mobility, leading mostly to more sophisticated types of labor, higher incomes and greater prestige. But the actual start of this process was marked by the introduction of informal (as verbal recommendations) and formal certification in different types of jobs and the emergence of the corresponding types of vocational training – from a child doing an apprenticeship with a master tradesman to cutting-edge forms of education, including contemporary IT-based distance education.

Social movement through the channels and elevators of education is the main prerequisite for a more open social-class structure, for society's movement towards more egalitarian values and norms, towards the alleviation or even erasure of painful disparities in the social conditions of different strata and groups. That was the real impact of education in the USSR, when two-thirds of the augmentation of intellectual-labor groups were achieved due to upward social mobility (from the bottom of the social pyramid, so from the peasantry and working class).

Yet, it is important not to overemphasize the upward mobility associated with education. In today's Russia, a university graduate working as a provincial teacher or physician earns a salary on par with the subsistence minimum, which does not allow for economic independence. A PhD qualification and employment as a university professor signifies nothing more than a very heavy burden of teaching load and approximation with the average income level in the regional economy. This is associated with the weakness of state control over the economic disparities and stratification of the population. RAS academician M.K. Gorshkov, speaking at the Fifth All-Russian Sociology Congress (October 2016), reported that while the decimal coefficient of economic inequality in Russia officially stands at 13, it actually exceeds 20 and approaches 40 in metropolitan areas.

Education also determines the mobility of individuals within the demographic structure of the population (settlement, family, marital structures).

- d) Social selection. In the education system, students are formally and informally divided by indicators deemed significant to the system, and these kinds of divisions are often combined (*streaming, tracking*). Informal division may show up in localization of students inside classroom, in uneven amount of attention and encouragement provided on the part of the teacher, etc. Selection becomes obvious according to empirical research.



The formal basis for *tracking* is provided by the nature and level of the intellectual capabilities exhibited by the child in the educational process. These capabilities, in turn, are identified through the tests, but assignment to particular *streams* may also take into consideration the personal recommendations by educators, relatives or sponsors. Academic tests always contain certain cultural context, and the student's ability to cope with the test depends on the distance between the school culture (professional cognitive culture) and the cultural traits of student's primary socialization environment. The most advantageous roles, the longest attention and the greatest encouragement usually go to those students who satisfy the requirements of the dominant (school) culture. Cultural differences often have an impact on how quickly and fully the abilities of students are developed. In the more prestigious forms of education, payment for education or accommodation is becoming an important condition.

- e) Reproduction of social groups and strata, affiliation with which is conditioned by educational attainments. The system of formal education supports the development on the part of individuals of skills and aptitudes corresponding to particular social and professional functions and roles. This function was demonstrated in previous epochs by informal evidences, and now – by credentials as a condition of occupying certain positions in the economy and social stratification, of ascending to certain social groups and strata, as proved by R. Collins in his *The Credential Society*.
- f) Substituting for parents, providing social support for students, most of whom are children or adolescents, has historically been and continues to be a function of education. For centuries, schooling represented the continuation of the traditional care provided by the older for the younger, but is now associated with society's orientation towards the creation of early access to education and a comfortable learning environment. This support is provided over the extended period of time spent within the walls of an educational setting, where specialized organizational and role structures reminiscent of the home environment are recreated – meals, communications and recreation, bedding, medical care, play areas and safe outdoor grounds. In this function, education reproduces the role differentiation of the sexes and genders typical of the family.

## 8. Functions of education in the sphere of culture

Speaking of the functions of education with respect to culture, many philosophers and sociologists habitually refer to the *transmission of culture* to the younger generation, or of the *reproduction of culture*. Yet, this assertion is only valid in the global context. Even in the best-case scenario, an individual gains from the education system only a fraction of the national culture and minimal insight into the world culture. Schooling gives an individual the basic elements of a society's core culture, and university – the basic elements of professional culture. The social conditions of the transmission of culture constitute the central problematics of the sociology of education. But education also reproduces the entire specter of cultural differences between different social groups and strata.

- a) The reproduction of particular social types of culture (ethnic, historical, local, regional, group, etc.) and their components – systems of values, norms and practices – is the leading function of education. Yet education reproduces far not all types of culture, but only those supported by the leading social request. The remaining types of culture, including minority cultures, are not included in reproduced cultural content.

School increasingly transmutes intellectual and professional experience into formalized constructs, allowing it to be systematized, configured, scientifically analyzed, accumulated and transmitted. The edu-

education system reduces the amount of time spent on mastering certain types of culture – it encourages a more open social structure, a broader exchange of culture locally, nationally and internationally.

One of the manifestations of this function is the reproduction of school subculture (college, academic, etc.).

Educational communities, unlike others, are mostly mono-aged (each consisting of individuals of similar age). The cultural conflict between the generations is likely associated with this particularity. School, encompassing youth, is the breeding ground of a subculture in which the core value has become the continuous process of updating knowledge, renewing intellect and skills. This renewal is less typical in other cultural environments. The daily activity of the student, his/her status and formal evaluation in the microenvironment of the school depend completely on his/her success in renewing intellect and skills. This renewal (increasing complexity/qualification of educational exercises from grade to grade with each passing academic year) becomes the overarching principle of the school culture. The value of this renewal is not typical of most older people, and this difference is likely the latent basis for the intergenerational cultural contradictions observed in society.

- b) Innovations in the sphere of culture. Sociologists from the functionalist paradigm were the first to point to the fact that education drives innovations in the sphere of culture. This function complements the function of the reproduction of socio-cultural types. In the broader socio-historical sense, education supports cultural innovations, but in specific historical situations, education also counteracts certain cultural innovations.

This function has two interconnected prerequisites – cultural and social. From the standpoint of the former, the public education system relays but a fraction of the innovations achieved in sciences, arts, law, religions and other institutions. To this end, it sorts and wraps these selected innovations (norms, information, technologies, values, etc.) in the form of educational (cognitive) materials. The selectivity with which education sorts out innovations points to the social nature of the organization of education and its interconnectedness with other institutions and social groups. As a rule, the innovations that receive support are those lying in the mainstream of the *core culture* and not posing a danger to the overall integrity of this social organization (or the stability of its management structure). We here, of course, mostly consider the public schooling programs. There can be quite opposite programs of *civil education*, *radical activists training*, which usually precede all sorts of modern revolutions (like the «Arab Spring», «the Rose Revolution», the «Dignity Revolution», etc.) and are sponsored from abroad, bringing to the destruction of national states and the split of societies.

- c) Formation and reproduction of public intellect (mindsets, disciplines and social technologies of intellectual activity). Since its inception, the public education system has been transforming into a multidisciplinary complex whose purpose is not merely the transfer of knowledge, but also the reproduction of public intellect.

Substantive knowledge appears not in the form of individual facts, but takes the shape of relatively independent cognitive sets (schooling subjects, scientific disciplines, professional standards, etc.). Precisely this function could play a key role in the pace of civilizational progress in the future, and has already transformed into one of the factors of international cooperation and competition. Confirmation of this fact are the attempts on the part of states and international organizations to establish conceptual control over education systems in different regions of the world. In this case, education and public intellect of a

recipient country (so, a receiver of educational technologies) fall into dependence of the donor country or donor organization (so, developer and distributor of educational technologies). This dependence is possibly even more influential than technological dependence in industry, insofar as it makes profound changes to the *elemental base* of the recipient's public intellect and guarantees the primacy of the donor. Martin Carnoy described it in his *Education as Cultural Imperialism*.

But even countries with longstanding traditions in their own education systems, like Russia, can find themselves in a position of intellectual dependence. This becomes more likely when the state and society lose conceptual control over the development of education, at the same time as a large portion of the political and intellectual elite ignore national interests.

## 9. Functions of education in the socio-political sphere

The formation of the civic identities of individuals comprises one of the priority interests of the state, society and communities. An essential component of *content* in any education system consists of the values and norms of legal and political conduct in society and the state. This component comprises law courses, Constitution lessons, but is most commonly associated with the political culture of dominant groups. Thus, the education system fulfills functions in the socio-political sphere.

- a) Cultivating acceptable legal and political values and norms, methods of participating in political life among educational communities. This function is typical of both formal and informal education. While the education system sometimes lends approval to certain alternative types of culture or cultural styles, there are virtually no examples of the institution of education condoning legal or socio-political deviations. Due to the importance of this function, education often is viewed an instrument of political supremacy (Gramsci, Freire, Apple, etc.). It is not by coincidence that each newly-formed political regime endeavors to establish control over the content of education or creates its own *new* school system. Recognition of this function leads to the ideologization of schooling. In this sense, school also fulfills another function with similar content.
- b) Encouragement of law-abiding legal and political behavior, reproduction of state (leading) ideology. Social groups that are carriers of alternative political and legal ideologies attempt either to establish an alternative school, or to add their own sets of legal and political values to existing education content. From here, the school becomes a field for social movements and political battles, as the legal consciousness of the masses is preconditioning the stability of political regime.

The education system is not sterile in the ideological sense. It is always subject to the ideological control exerted by dominant groups. In the USSR, this function appeared publicly with ubiquitous school – or college-based Communist Party committees. Today it operates in latent forms, in which most of social sciences and humanities in public education programs are following the values of liberalism, individualism and globalism.

This function aims at the legal and organizational integration of society, that is – at maintaining the integrity of the system as a whole.

- c) Encouragement of patriotism becomes a function of education in the context of each national state. Within the content of education, it is transmitted by the *citizen's code* and expresses the expectations of the state. This function serves as a precondition for the preservation of any national-state community in the unstable geopolitical environment. Moreover, national culture leaves a mark

on the content of education, highlighting within it the totality of values that underscore the specific interests of its ethnos. The encouragement of patriotism is a feature of most state systems of public education.

This version of the functions of education aims at application of the four criteria for defining institutional functions substantiated above. It certainly makes no claims of being complete, and is subject to both theoretical criticism and empirical approbation. This version intends to direct the attention of educational sociologists towards a problem, which, in our view, is creating clearer theoretical architecture for the sociology of education, and with it, opportunities for the integration therein of new thematic areas and empirical studies, as well as social practical outcomes in the interests of the education system, society and the state.

## References

- Antikainen, Ari *et al.*, (2011): "Contemporary Themes in the Sociology of Education". *International Journal of Contemporary Sociology*, 48 (1), 117-147.
- Ballantine, Jeanne H. and Hammack, Floyd M. (2009). *The sociology of education: A systematic analysis* (6<sup>th</sup> ed.). Upper Saddle River: Prentice Hall.
- Durkheim, Émile (1922). *Education et sociologie*. Paris: Presses Universitaires de France.
- Ivanov, S.V. and Osipov, Alexander M. (2004): "The university as regional corporation". *Sotsiologicheskie issledovaniya [Sociological studies]*, 11, 162-172.
- Merton Robert King (1968). *Social Theory and Social Structure. Enlarged edition*. New York: Free Press.
- Osipov, Alexander M. (1998). *Society and education. Lectures on the sociology of education*. Novgorod: NovSU.
- Osipov, Alexander M. (2006). *Sociology of education: Essays on theory*. Rostov-on-Don: Phoenix.
- Osipov, Alexander M.; Ivanova, Viktoria and Dobrenkov, Vladimir (2013): "Global Sociology of Education: General Perspectives. Monograph Supplement to 'Sociological Studies'". *Journal of the Russian Academy of Science. Veliky Novgorod*: Yaroslav-the-Wise Novgorod State University.
- Osipov, Alexander M. (ed.) (2014). *Sociology of Education Abroad*. Veliky Novgorod: NovSU.
- Parsons, Talcott (1959): "The school class as a social system". *Harvard Educational Review*, 29 (4), 297-318.

## Nota biográfica

**Alexander Osipov** es Doctor en Ciencias Sociales; Yaroslav-the-Wise Novgorod State University (NovSU). Profesor en el Departamento de Sociología y Educación Bilingüe. Miembro del consejo editorial de la revista *Education and Society*; co-director del comité de investigación (RC) de Sociología de la Educación. Entre sus líneas de investigación, destacan: sociología de la educación, juventud, salud pública. Entre sus recientes publicaciones, cabe mencionar: *Global Sociology of Education. General Perspectives* (coeditado con Viktoria Ivanova y Vladimir Dobrenkov), Veliy Novgorod: Moscú, 2013.

# As medidas da educação. Medição e gestão mediante indicadores

Measures of education.

Measurement and management through indicators

Dietmar K. Pfeiffer<sup>1</sup>

## Resumo

O processo da globalização exerce um impacto forte sobre os sistemas educacionais nacionais no mundo inteiro. Parte deste processo é a implementação de um sistema de indicadores para a mensuração da qualidade educacional. De importância especial neste contexto é a disseminação de avaliações de larga escala (PISA). Medir não é mais uma prática exclusiva para as ciências exatas, mas se tornou em um dos objetivos principais nas ciências educacionais, também. Porém, embora haja uma plentora de indicadores educacionais, uns deles fazendo parte de conhecidos índices (HDI, SESC, ICM), há apenas poucas iniciativas até agora para criar um índice composto para o desenvolvimento educacional.

O artigo presente tem como objetivo revisar de forma crítica os problemas de mensurações representativas nas ciências humanas apontando que a maioria dos indicadores compostos e índices não correspondem à axiomática da teoria representacional de medição. Medições mediante índices são consideradas como substitutos para medições representativas. Seus níveis escalares e o significado são muitas vezes desconhecidos de modo que precisam de ser aproximados através de critérios de fiabilidade e validade. A diversidade teórica é considerado como um dos motivos principais para a falta de medidas consistentes e relevantes no campo educacional. Além do mais, muitas variáveis na esfera educacional não são quantificáveis.

A preocupação principal no tocante a medição por indicadores se refere ao caráter ambivalente deles. Por um lado eles podem proporcionar informações relevantes ao nível descritivo e subsidiar decisões políticas, por outro lado, a «cultura de indicadores» assume de forma crescente vertentes de um sistema normativo de controle.

## Palavras chaves

Índice, indicador, medição, PISA, gestão de educação, globalização.

## Resumen

El proceso de globalización ejerce un fuerte impacto sobre los sistemas educativos nacionales de todo el mundo. Parte de este proceso es una implementación de un sistema de indicadores para la medida de la calidad educativa. En este contexto adquiere una especial importancia la difusión de evaluaciones de larga escala (PISA). Medir ya no solo no es una práctica exclusiva de las ciencias exactas, sino que también devino uno de los objetivos principales de las ciencias educativas. Pero, aunque hay un abanico de indicadores educativos, algunos de ellos formando parte de índices conocidos (HDI; SESC, ICM), hasta ahora apenas hay unas pocas iniciativas para crear un índice compuesto para el desarrollo educativo.

El presente artículo tiene como objetivo revisar de forma crítica los problemas de las mediciones representativas en las ciencias humanas señalando que la mayoría de los indicadores compuestos e índices no corresponden a la axiomática dela teoría representativa de la medición. Sus niveles de escala y sus significados son muchas veces desconocidos, de modo que necesitan ser abordados a través de criterios de fiabilidad y valide. Una diversidad teórica es considerada como uno de los motivos principales para la falta de medidas consistentes y relevantes en el campo educativo. Además, muchas variables en la esfera educativa no son cuantificables.

Una preocupación principal respecto a la medición a través de indicadores se refiere a su carácter ambivalente. Por un lado, pueden proporcionar informaciones relevantes a nivel descriptivo y avalar decisiones políticas, por otro lado, una «cultura de los indicadores» asume de forma creciente aspectos de un sistema normativo de control.

## Palabras clave

Índice, indicador, medición, PISA, gestión educativa, globalización.

<sup>1</sup> Universität Münster, Alemania, [dicape@uni-muenster.de](mailto:dicape@uni-muenster.de)

## Abstract

The process of globalization exerts an increasing impact on national educational systems worldwide. Part of this process is the attempt to establish an indicator system for educational quality measurement. Of outstanding importance in this context is the dissemination of large scale assessment studies (PISA). Measuring is not any longer a practice of natural sciences only, but has become a main objective in educational sciences as well. But although there are a plethora of educational indicators, some of them making part of well-known indexes (HDI, ESCS, GCI) there are up to now only few attempts to create a composite index for educational development.

The present article aims to review critically the problems of representative measurement in human sciences pointing out, that most of the indicators and indexes do not fit the axiomatic of the representational theory of measurement. Index measurement must be considered as a substitute for representative measurement. Its scale level and the meaning of the data remain often unknown, so that their relevance must be approximated by reliability and validity criteria. Theoretical diversity is considered as one of the main reasons for the lack of consistent and meaningful measures in the field of education. Furthermore many variables in the field of education are not quantifiable.

The main concern of measurement by indicators in education is its ambivalent character. On the one hand it can provide on the descriptive level relevant information and support for political decision-making, on the other hand the «culture of indicators» takes on increasingly features of a normative control system.

## Key words

Index, indicator, measurement, PISA, educational management, globalization.

Recibido: 31-03-2017  
Aceptado: 02-05-2017

## 1. Globalização e Indicadores Educacionais

Os sistemas educacionais até hoje são mais do que outros subsistemas sociais (economia, ciência) caracterizados pelas tradições culturais de cada nação inclusive as suas contradições que resultam do seu desenvolvimento histórico. Porém, apesar de todas as diferenças e particularidades existentes, o processo da globalização exerce impactos também sobre a realidade educacional. Organizações internacionais, tais como EU, OECD, UNESCO, Banco Mundial, estendem cada vez mais sua esfera de influência às políticas educacionais nacionais. Desse modo, a relação entre estado nacional e sistema educacional está em processo de mudança e um regime de educação internacional está emergindo (Parreira do Amaral, 2011).

Devido à importância crescente do denominado capital humano nos sistemas modernos de produção, está desenvolvendo-se um mercado global de serviços educacionais, cujo volume está chegando a vários bilhões de dólares segundo dados do Banco Mundial. A questão não é apenas se em certos setores a iniciativa privada pode ou até deve complementar a oferta pública, a questão é se a educação deve se converter em um serviço a ser vendido como qualquer outra mercadoria. Esta discussão se desencadeou a partir das regulações da OMC e do GATS e as consequências desta privatização e comercialização até agora são pouco conhecidas. Mas sem dúvida, no caso de uma abertura dos mercados nacionais abrangendo todos os setores de serviço se trataria de um novo nível de internacionalização que terá amplas repercussões nas condições nacionais da educação pública.

A crescente concorrência internacional na qual a educação representa de acordo com a declaração do Conselho Europeu Declaração de Lisboa (2000) um fator importante, exerce uma pressão enorme sobre os sistemas educacionais em direção a uma standardização de estruturas escolares e conteúdos curriculares. Globalização significa também na área educacional standardização, significa parâmetros estáveis de qualidade, indicadores quantitativos, normas e procedimentos, competências definidas consideradas como sendo relevante no contexto da concorrência internacional e finalmente *accountability*. A

manifestação mais óbvia deste desenvolvimento são os sistemas de indicadores educacionais como, p.ex., o sistema da OECD (2016a) que abrange 115 indicadores de diversas áreas (anos escolares, taxas de escolarização, estruturas e organização escolar, financiamento) para todos os membros da OECD e diversos países associados. Junto às avaliações dos conhecimentos e competências dos alunos ao nível internacional, se desenvolveu deste modo um regime poderoso através do qual a globalização está chegando nas escolas no mundo inteiro.

A mais conhecida destas avaliações é o «Programa Internacional de Avaliação dos Estudantes» (PISA). No primeiro estudo PISA no ano 2000 participaram 32 países, a grande maioria deles membros da OECD. No estudo atual (OECD, 2016b) já participaram 72 países sendo 37 fora da OECD entre eles nove da América Latina e dez do mundo islâmico. Isso comprova que esta comparação tem assumido uma dimensão global sendo um produto das novas relações de concorrência internacional e ao mesmo tempo uma manifestação da integração avançada da sociedade mundial. No final desta cadeia de internacionalização se encontram os alunos, os professores das escolas, os funcionários, todos que são envolvidos na base do sistema de ensino. Ninguém nunca os perguntou se eles querem participar neste tipo de avaliação, e quais são suas vantagens em participar.

PISA e outras avaliações internacionais, completadas por avaliações nacionais estão estreitamente vinculadas ao conceito da qualidade da educação escolar. Embora os agentes promotores de PISA evitem interpretar os resultados dos testes como indicador direto para a qualidade do ensino de um país ou mais ainda, de um sistema educacional em geral, na discussão internacional prevalece esta interpretação. No Brasil, p.ex., o Ministério da Educação (MEC) formulou de forma muito clara a relação entre qualidade de educação e indicadores a fim de medi-la.

*Ideb é o Índice de Desenvolvimento da Educação Básica, ... formulado para medir a qualidade do aprendizado nacional e estabelecer metas para a melhoria do ensino. O Ideb funciona como um indicador nacional que possibilita o monitoramento da qualidade da educação pela população por meio de dados concretos, com o qual a sociedade pode se mobilizar em busca de melhorias. Para tanto, o Ideb é calculado a partir de dois componentes: a taxa de rendimento escolar (aprovação) e as médias de desempenho nos exames aplicados pelo Inep. Os índices de aprovação são obtidos a partir do Censo Escolar, realizado anualmente.*

A reflexão sociológica crítica referente ao uso e abuso de indicadores está focalizando-se predominantemente ao aspecto filosófico, à noção dominante do conceito do indicador e sua função controladora, a elaboração de novos e melhores indicadores e finalmente, seu caráter normativo que afeta cada vez mais as políticas nacionais na esfera econômica, social e educacional (Beltrán e Villar, 2010). Entretanto, menos atenção recebe a questão fundamental, se construtos tais como qualidade educacional, progresso social, bem estar, ou exclusão/inclusão são realmente mensuráveis ou se os indicadores simplesmente geram artefatos matemáticos-estatísticos que sugerem certezas que não existem. Uma breve revisão dos conceitos de medição, de indicador e de índice pode servir para aproximar-se a uma resposta tentativa desta questão controvertida.

## 2. Medição e Indicadores

Historicamente e metodologicamente a medição é uma base do progresso da ciência. No entanto, antes de medir é mister de esclarecer a questão básica, se o fenômeno a ser medido é realmente mensurável. A solução deste problema é o objetivo central de cada teoria de medição que conecta, de uma forma ou a

outra, a realidade com os dados. Falando hoje em dia de teoria de medição se refere à teoria representacional de medição<sup>2</sup> que concebe medição como uma atribuição de números a objetos ou eventos de tal maneira que as estruturas do conjunto empírico (E) são representados por um conjunto numérico (Z) de forma homomórfica respectivamente isomórfica (f: E→Z). Trata-se não somente de um problema técnico, mas de um problema metodológico, mais ainda epistemológico em quanto se refere à relação entre sujeito e objeto. Sem uma reflexão da base de medição, a análise, a interpretação, e a relevância dos dados sempre estão incertos.

Ninguém duvida da necessidade e da utilidade de medições nas ciências exatas, na economia, na educação física, entre outras. E nosso dia-a-dia está completo de medições, muitas vezes sem serem percebidos conscientemente: horas de trabalho, temperatura de água, salário mensal, peso dos ingredientes na hora de preparar uma comida, ritmo cardíaco, etc. No entanto, nas ciências humanas e sociais (sociologia, psicologia, pedagogia, e outras) existe uma controvérsia veemente com respeito ao processo de quantificação, uma controvérsia que marca uma das divisórias metodológicas mais importantes entre a abordagem quantitativa e qualitativa. É necessário de sublinhar, que medição neste contexto implica a existência de uma propriedade métrica (escala intervalar, escala de razão, escala absoluta) porque medições ao nível nominal (categorias) ou ordinal (postos) se encontram também frequentemente em pesquisas qualitativas, seja de forma explícita ou implícita.

Não cabe ao presente artigo de retomar as polêmicas eternas, se as duas abordagens são complementares ou incompatíveis, mas, esboçar as possibilidades e os equívocos das práticas correntes de medição na esfera educacional. Recordar-se, que a mais de 50 anos atrás, Cicourel (1964) expôs convincentemente que a maioria das medições nas ciências sociais não correspondem às exigências axiomáticas<sup>3</sup> da teoria representacional de medição, mas são mais ou menos arbitrárias (*measurement by fiat*). Hoje em dia, onde se produzem permanentemente novos índices, indicadores, rankings, escalas, muitas vezes sem base conceitual sólida, as reflexões críticas de Cicourel parecem ser mais atuais e pertinentes de que nunca. Cabe esclarecer que ele nunca foi um adversário fundamental de mensurações per se, mas de mensurações desvinculadas da realidade social e de significado duvidoso. «Let me underscore the fact, that I am NOT opposed to quantification or formalization or modeling, but do not want to pursue quantitative methods that are not commensurate with the research phenomena addressed» (Cicourel, 2004).

Este distanciamento referente ao processo de quantificação tem sua origem metodológica, entre outros motivos, nas severas dificuldades que as ciências sociais encontram na hora de medir (Felbiger, 2010; Pasquali, 2009), nomeadamente:

- A falta de um sistema de unidades internacionalmente reconhecido como existem nas ciências exatas.
- Os conceitos usados são muitas vezes imprecisos, e suas definições não são aceitas de forma unânime.
- Os construtos são complexos e multidimensionais e muitas vezes não diretamente observáveis.
- As medições são vulneráveis a erros sistemáticos e, portanto não alcançam a precisão necessária e desejável.

2 As outras duas teoria de medição conhecidas são (1) a teoria operacional e (2) a teoria clássica (Hand, 1996).

3 As relações empíricas tem que salvar os axiomas básicos da identidade, ordem (transitividade e conectividade), e aditividade.



Perante aos problemas expostos, medições representativas *stricto sensu* dos construtos teóricos são inviáveis nas ciências sociais. Por motivos pragmáticos se observam em vez dos verdadeiros construtos (variáveis latentes), uma série de indicadores (variáveis manifestas) dos quais se supõe que estes estão em uma relação estreita com o construto que interessa (*latino: indicare*). Como Rost (2005:6) destacou, nas medições das ciências exatas, o instrumento de medição (p.ex. um termómetro) permite diretamente a leitura do escore da variável em questão. Nas ciências sociais, contudo, há uma diferença fundamental entre conceito medido (p.ex. satisfação com a vida) e indicadores observáveis (p.ex. respostas a uma ou mais perguntas). A maneira como chegar dos indicadores a um escore para o conceito é justamente o objeto dos modelos de medição.

Deste modo, p.ex., a variável «anos escolares» é considerada como indicador para o nível de «educação formal» de uma pessoa, a «taxa de conclusão» dos alunos como indicador para «produtividade do sistema universitário», e o BIP per capita, apesar de todas as críticas, como indicador para o bem-estar de uma população.

Geralmente se usa mais de que um indicador para representar determinados aspectos da realidade porque «os fenômenos estudados pelas ciências sociais são demasiadamente complexos para serem interpretados e analisados sob uma ótica unidimensional (Soligo, 2012:14). Muitas vezes, a mensuração por meio de indicadores é um processo *multi-level*, quer dizer, os próprios indicadores podem resultar da junção de um conjunto de diversos sub-indicadores. Mediante de uma agregação dos escores dos indicadores resultar finalmente um *composite indicador*, ou melhor: um índice. De acordo com Siche *et al.*, (2007:139), um índice é o valor agregado final de todo um procedimento de cálculo onde se utilizam, inclusive, indicadores como variáveis que o compõem». Exemplos internacionalmente conhecidos são o *Índice de Desenvolvimento Humano* (IDH) das Nações Unidas, o *Índice de Status Econômico, Social e Cultural* (ISEC) da OECD, o *Índice de Sustentabilidade Ambiental* (ESI), e o *Índice de Competividade Mundial* (ICM) do Fórum Econômico Mundial. Os escores obtidos nos índices, por sua parte, são usados muitas vezes para construir uma escala de ranking, um procedimento que se virou na atualidade em uma das modas mais questionáveis da qual nada e ninguém escapa. Há rankings de tudo: hospitais, universidades, cidades, países; falta somente, como Kaube (2014) comentou ironicamente, um ranking das melhores religiões do mundo<sup>4</sup>.

O problema da transformação de indicadores em escores de um índice consiste obviamente na seleção, na justificativa teórica e na ponderação<sup>5</sup> dos diversos indicadores. Em princípio o número de indicadores per se é infinito e portanto, a decisão por um indicador ou outro é inevitavelmente o resultado de um processo seletivo. O universo dos indicadores que refletem um construto sempre representa somente uma amostra deste universo (Pfeiffer e Püttmann, 2011:44). Indicadores são um tipo de filtro, eles selecionam certos aspectos da realidade e eliminam outros. Em consequência, indicadores sempre podem ser objetos de críticas e assim incentivar a busca de indicadores alternativos. Além do mais, eles não são independentes do tempo e do espaço. «Um indicador útil em determinado lugar e época pode não ser proveitoso para outra região, ou até mesmo para própria região em outro momento» (Soligo, 2012:17).

4 Por mais improvável que possa parecer, a origem desta moda se encontra no século dezoito em forma de pontos e *rankings* para pintores e escritores famosos da época (Spoerhase, 2014).

5 A ponderação dos indicadores é um problema sumamente importante que tem um impacto aos resultados. Diferentes pesos para os indicadores e diferentes modelos de agregação (aditivo, multiplicativo) dos indicadores levam a diferentes escores no índice e em consequência a diferentes posições no ranking.

Pressuposto que existem indicadores mensuráveis, segue no próximo passo o *problema de correspondência*, ou seja, a questão como se pode ter certeza que os indicadores realmente medem as propriedades dos objetos/sujeitos que eles pretendem a medir. Na literatura metodológica este problema é abrangido pelo conceito da *validade* que em termos gerais se refere ao grau da congruência entre indicadores e construto teórico. Conforme Pasquali (2009:994), a validade é tipicamente discutida nas ciências psicossociais. Mas também nas outras ciências humanas, economia, educação, letras, a validade é um problema permanente, enquanto nas ciências exatas normalmente não há necessidade de provar que um instrumento (p.ex. uma balança) mede a propriedade intencionada (a massa do objeto). Em contraste, se um índice de qualidade educacional mede realmente o que pretende, depende do conceito subjacente como será explicado mais a frente. Resumidamente: medições por meio de índices podem ter uma utilidade prática, caso que sejam cuidadosamente construídos e validados. Mas a qualidade informacional dos escores e o nível de escala são desconhecidos e portanto, a interpretação do significado dos escores sempre tem que ser cautelosa. Os diversos procedimentos de comprovar a validade que representam de certo modo nas ciências sociais um substituto para compensar a falta de uma representação estrutural. Entre os procedimentos estatísticos de testar a correspondência entre o conceito teórico e o conceito observável se destacam modelos causais que estabelecem relações entre construto e indicadores que são passíveis a testes empíricos.

A luz destas reflexões se evidencia de forma muito clara que indicadores não são simplesmente uma coleção qualquer de variáveis, mas supostas manifestações observáveis de construtos teóricos cujo objetivo consiste em elaborar um instrumento de medição. Portanto, a seleção das variáveis observáveis e medíveis precisa de uma justificativa teórica, assim como a ponderação dos indicadores na construção do índice. Devido a falta de uma base teórica sólida se usam preferencialmente pesos iguais o que implica que todos os indicadores são considerados mutuamente como substitutos, um pressuposto que, sobretudo no contexto da avaliação educacional é altamente duvidoso.

A diversidade de teorias, às vezes até contraditórias, é um dos motivos principais para a plethora de índices, escalas e ranking que invadem as ciências humanas e as políticas públicas. Por exemplo, Rohwer e Hülsewig (2011:73) relatam que uma comparação de três estudos relevantes a nível profissional sobre a competitividade de países mostrou resultados significativamente diferentes de acordo com o enfoque teórico. Na sociologia, psicologia e pedagogia, a confusão no campo teórico é maior ainda. Os indicadores usados muitas vezes são irrefletidos e os instrumentos construídos superficialmente o que leva a resultados sem significado.

### 3. Medição da qualidade educacional

Os problemas da medição mediante de indicadores se podem mostrar de forma exemplar através do construto da qualidade educacional<sup>6</sup>. O conceito «qualidade» se deriva da palavra latina «*qualis*» o que significa «de tal natureza». Neste sentido, qualidade designa de forma descritiva, certas características, propriedades ou traços de um objeto qualquer. O escalamento da qualidade, ou seja, a construção de um índice para medi-la exige, igual como no caso de outros construtos, indicadores observáveis e mensuráveis. Em caso de produtos e serviços no âmbito econômico, a qualidade se refere (1) aos requisitos (normas) técnicos ou processuais (NBR, ISO) estabelecidos por órgãos nacionais e/ou internacionais;

<sup>6</sup> As reflexões a seguir estão limitando-se à educação escolar. Este escola-centrismo é um restrição conceitual, (Beltrán e Aguilés, 2010: 13) mais não afetará os parâmetros básicos do nosso raciocínio.

(2) à capacidade de satisfazer determinadas exigências, ou necessidades estabelecidas pelo consumidor. Portanto, a qualidade no sentido descritivo não está necessariamente limitada ao resultado, senão pode incluir também aspectos processuais, ou seja, como o produto ou o serviço foi produzido (*quality management*). De qualquer modo, o consumidor precisa ter parâmetros, conhecimentos e informações, os quais o possibilitam medir e avaliar a qualidade do objeto ou serviço.

Este marco referencial corresponde amplamente aos princípios de uma economia de mercado. Na esfera educacional se tenta medir a qualidade educacional basicamente por meio de dois indicadores: anos escolares e rendimento dos beneficiados, seus conhecimentos, competências, hábitos, etc. (Beltrán e Villar, 2010:140). Enquanto o primeiro é fácil de medir, a medição dos conhecimentos e habilidades em aplica-los se efetua por testes mais ou menos sofisticados cujos resultados são expressados em escores. Aspectos processuais, ou contextuais não são levados em consideração. Esta operacionalização parece ser uma solução objetiva e teoricamente fundada, porque ninguém pode negar que o objetivo de uma educação escolar consiste em aprender e proporcionar aos alunos conhecimentos e competências necessárias para a vida. Mas a pergunta central, quais são em termos concretos as competências, conhecimentos e valores relevantes para a qualidade da educação ficam abertos. As empresas, as instituições públicas e privadas, os pais, os alunos, os professores, organizações internacionais e outros *stakeholder*, tem muitas vezes uma visão diferente de que uma educação de boa qualidade.

Como anteriormente mencionado, os indicadores (critérios) para medir a qualidade dependem do marco teórico subjacente e no caso da educação mais ainda da imagem do homem (*Menschenbild*) e da visão do mundo (*Weltbild*). Sendo que o universo dos indicadores que refletem um construto, neste caso o construto «qualidade do aprendizado» é infinito e o conjunto de indicadores usado sempre representa somente uma amostra deste universo. Para os empresários o critério relevante é a utilidade da pessoa no processo de produção, para a igreja a vivência com os valores cristãos, para o estado um cidadão integrado à sociedade, para intelectuais críticos a capacidade de resistência contra processos restritivos e o uso da razão crítica, e para os adeptos do ideal clássico de «Bildung» o cultivo de si mesmo, a evolução da interioridade do sujeito, a autopoiesis da consciência.

Diante a esta pluralidade, Psacharopoulos (1995), levantou a pergunta, quem define e implementa os indicadores e quais são os objetivos a serem alcançados através dos conhecimento que os indicadores proporcionam? De um certo modo é uma questão de poder o que se mede e quem avalia o que e através de que critérios. Leis são o resultado de discursos públicos e decisões políticas, preços são, pelo menos em mercados eficientes, o resultado de oferta e demanda. Entretanto, conteúdos curriculares e estruturas do sistema educacional são resultados de negociações complexas entre setores administrativos, atores políticos e econômicos, expertos acadêmicos e, como resultado da globalização cada vez mais organizações internacionais tais como a OECD ou o Banco Mundial. Eles implicam controle mediante das avaliações, mas sua legitimidade é duvidosa, porque uma discussão pública quase não existe e dificilmente a realizar porque o assunto é teoricamente complicado e pouco atraente para a mídia e o público em geral.

A situação quanto aos indicadores educacionais se complica sob uma perspectiva sistêmica. Os indicadores usados para medir a qualidade, ou melhor, dito, o *performance* do sistema são compostos via regra como agregados de indicadores individuais. A maioria deles se incorpora ao modelo internacionalmente reconhecido hoje em dia: *Context-Input-Process-Output-Outcome*. Os resultados dos mencionados testes nacionais e internacionais, baseados em complexos modelos de medição, como p.ex. o modelo de resposta

ao item (TRI), o nível de conclusão escolar, as taxas de conclusão nos diferentes níveis, os anos médios de escolaridade etc. são os mais conhecidos indicadores do *output*. Ultimamente, indicadores referentes à paridade dos gêneros e a equidade social estão ganhando mais consideração. Os números e as taxas de matrículas, os recursos financeiros investidos (professores, materiais, prédios, infraestrutura, etc.) são considerados como *input* e o *outcome* são os benefícios individuais e sociais fornecidos pela educação ao longo prazo.

Um indicador importante de processo é a taxa de aprovação durante o ciclo escolar, enquanto indicadores que se referem às metodologias de ensino e os estilos de aprendizagem são abordados raras vezes porque são difíceis de quantificar. Este modelo tem, sem dúvida, um certo valor heurístico, contudo ele não representa um conceito teórico (Wolter, 2008), senão parte simplesmente da definição de certas áreas temáticas ou problemas relevantes (custos, resultados, equidade) as quais em seguida são descritas de forma pragmática através de dados quantitativos disponíveis.

Muitos dos indicadores educacionais entram na construção de conhecidos índices (Índice do Desenvolvimento Humano, Índice de Competividade Internacional, Índice de Habitabilidade Urbana entre outros). Entretanto, um índice composto de diversos indicadores, que podemos denominar «Índice do Desenvolvimento de Educação» até hoje não existe. A falta de um conceito teórico consensual, capaz de guiar a seleção e ponderação de indicadores e o fato que muitos indicadores educacionais não são mensuráveis são apenas dois de diversos obstáculos<sup>7</sup>. Além do mais, os escores de muitos indicadores se devem às condições histórico-culturais do país ou da região, que nenhuma medição ou comparação pode mudar em curto prazo. «Para comprender la realidad educativa de un contexto se ha de conocer sus condiciones de posibilidad, de no ser así, los resultados se naturalizan» (Beltrán e Villar, 2010:140). Se as condições econômicas de um país limitam os gastos por aluno, se a formação dos professores é fraco e a estrutura administrativa ineficiente, é difícil de esperar um rendimento alto. Por isso já vinte anos atrás, Vedder (1997) colocou a pergunta se os resultados de avaliações a nível internacional realmente fornecem um aporte relevante aos países com problemas nos seus sistemas educacionais ou se eles servem somente para comprovar o que estes já sabem, ou seja que se encontram no último quintil da escala internacional *pot empo* indefinido.

Neste contexto é importante frisar que não existe uma relação linear entre gastos por aluno e rendimento escolar. Uma análise de Pfeiffer (2011:31) revela a presença de uma função logística entre as duas variáveis, o que significa que a partir de certo nível, um aumento dos gastos não leva mais a um aumento significativo do rendimento (curva de saturação). Além do mais, a relação não é determinística, quer dizer, que sempre há países que com gastos modestos alcançam resultados bons, assim como sempre há escolas pobres que por meio de práticas educativas inovadoras conseguem resultados acima da média.

#### 4. Indicadores para quê?

Igual às outras ciências, a implementação de indicadores a nível internacional e nacional tem uma face múltipla sendo ao mesmo tempo 1. uma ferramenta teórica indispensável para a pesquisa acadêmica, 2. uma fonte importante de informação para a sociedade civil, 3. um auxílio para desenhar políticas públicas e 4. um instrumento para o controle e o exercício de poder.

1. Para analisar se um sistema educacional garante a igualdade de chances e promove a equidade social no âmbito educacional, precisamos de indicadores mensuráveis que representam estes conceitos.

<sup>7</sup> Um problema específico consiste na possibilidade de influenciar os escores de certos indicadores por motivos políticos, p.ex. as taxas de aprovação.

Estudos acerca dos impactos do uso de modernas tecnologias de informação e comunicação sobre competências de letramento e satisfação dos alunos na aprendizagem precisam de instrumentos para medir estes construtos. Evidentemente, sobre definições e medições sempre há opiniões divergentes, mas isso é uma característica de cada processo científico. Por outro lado é importante de ter em consciência que nem tudo é mensurável e passível a ser representado por indicadores: os estilos de aprendizagem, os processos extracurriculares, o desenvolvimento da personalidade dos educandos.

2. No tocante à função dos indicadores para a sociedade, vale sublinhar que estes podem funcionar como base informacional para reivindicações justas. Sempre quando existe um financiamento público a consideração de critérios econômicos, entre outros, é indispensável. A sociedade civil tem o direito de saber quais são os destinos e os efeitos dos recursos que o contribuinte destinou ao Estado. Indicadores quantitativos podem contribuir para aumentar a transparência referente ao uso dos recursos, para detectar distorções e inequidades, para subsidiar melhoramentos nas práticas pedagógicas e também para sensibilizar as escolas referentes às suas responsabilidades. Neste sentido é importante e até necessário para um país, saber quais são as competências de seus jovens levando em consideração a magnitude dos recursos investidos e as exigências do mercado de trabalho.
3. Para o setor público em geral e os atores políticos no sistema educacional em especial, indicadores e índices constituem um tipo de informações condensadas, uma ferramenta eficiente para a redução de complexidade, para facilitar a comunicação e a tomada de decisões. Eles fornecem a base para um *monitoring* contínuo e sistemático das condições, processos, resultados e efeitos no sistema educacional, e facilitam o planejamento de processos futuros. Tudo isso é compreensível sob a perspectiva do sistema político, mas não sem risco: situações complexas são simplificadas, uma base teórica dos índices muitas vezes não existe ou fica irrefletido, um alto grau de agregação resulta em índices sem muito sentido, e a preocupação com o posicionamento em rankings se torna uma obsessão. Além do mais, medir e definir metas é uma coisa, achar os caminhos para alcançá-las é outra.
4. O perigo iminente da nova «cultura de indicadores» é a direção e o controle global dos sistemas através da transformação de um conceito descritivo a um conceito prescritivo. Índices e indicadores sobre o *status quo* tem a função de oferecem informações sobre resultados e problemas persistentes, mas na prática atual, eles se tornam cada vez mais em normas expressando uma aura de autoridade. Evidentemente, cada decisão educacional tem implicações normativas, mas de informações não se pode deduzir normas, nem objetivos. Embora, a OECD sempre enfatize oficialmente que os dados fornecidos não sirvam como mecanismo de controle, mas como um apoio para políticas educacionais dos países, os comentários e discursos dos atores responsáveis da OECD vão muito além de meras recomendações. Uma tríade de lições, críticas e louvores devem servir para exercer pressão aos países que não correspondem aos padrões desejados<sup>8</sup>.

Sem dúvida, há *benchmarks* consensuais para determinados indicadores educacionais em uma perspectiva global, mas se observa no discurso público que muitas vezes que os dados são compatíveis com objetivos políticos distintos e que os atores os interpretam de acordo com suas posições ideológicas.

As ambivalências da medição da qualidade de educação se tornam óbvias sempre quando se trata de medir a qualidade da educação mediante de testes como PISA ou PIAAC (Programme for the International Assessment of Adult Competencies). Diversos críticos suspeitam a evolução de um novo tipo de

<sup>8</sup> Assim, Andreas Schleicher, Diretor na OECD e Coordenador Geral de PISA criticou ano por ano em público a número de estudantes universitários na Alemanha como sendo insuficiente, desconsiderando por completo a estrutura específica do sistema educacional do país.

dominação cultural e um empobrecimento dos conteúdos curriculares e dos processos de aprendizagem mediante uma normalização (Meyer e Benavot, 2013) baseado em critérios de demanda dos mercados laborais. Outros esperam que eles podem contribuir para o desenvolvimento de uma base comum de competências internacionalmente reconhecidas, talvez até um espaço de pensamento no sentido de um universo reciprocamente compartilhado. Uma resposta definitiva até o momento não é possível.

## 5. Possibilidades e limites de gestão através de indicadores

Na economia e na administração empresarial, a gestão por meio de indicadores quantificáveis e/o índices tem uma longa tradição de acordo com o conhecido dogma: «*If you can't measure it, you can't manage it*». Em contraste, a tentativa de direcionar sistemas educacionais baseado em conceitos derivados dos princípios da administração empresarial é uma evolução mais recente que resulta da sensação dos atores políticos que nas estruturas e nos processos operativos do sistema educacional como tal e suas unidades (escolas, universidades) existem elementos irracionais e arbitrários, o que exige implementar *standards* de racionalidade e de gerência, já que não existem mercados para avaliar os produtos e serviços pelos consumidores. Neste contexto, indicadores tem a função de fornecer dados relevantes para avaliar a *performance* do sistema educacional. Avaliações do tipo PISA, mas também provas a nível nacional, são explicitamente ou implicitamente declarações de desconfiança com o objetivo de exercer controle e pressão para que as escolas se esforcem, proporcionando mais eficácia e eficiência, e finalmente melhores resultados de aprendizagem.

Certamente é um desejo legítimo da sociedade, do estado, dos pais, dos alunos, que as escolas façam um uso racional dos recursos destinados e que os professores assumam sua responsabilidade no processo educativo, o qual é a base para a autonomia e a realização social do ser humano. As grandes interrogações são se medições e avaliações através de sistemas de indicadores, estruturas de controle e *accountability* realmente produzem conhecimentos que proporcionam decisões racionais com a finalidade de alcançar os objetivos desejados, se as informações quantitativas proporcionadas pelos indicadores resultam em uma melhor qualidade das decisões dos atores na área da política educacional, ou se eles causam somente turbulências, incertezas e reações irracionais.

Para aproximar-se a este questão é necessário recorrer às análises fundamentais sobre as características específicas de *people processing organizations* em geral e de escolas em especiais. Já trinta anos atrás, Meyer e Rowan (1977, 1978) analisaram as características estruturais de organizações educacionais, identificando as formas específicas através das quais eles atuam e legitimam seus processos. O modelo input-processo-produto com o uso de uma tecnologia bem definida, característico para processos de produção material é inadequado para processos de aprendizagem em organizações educacionais, porque a relação entre meios e objetivos é mais complexo e menos determinado. Eles exigem o *médium* de comunicação entre os educadores e os educandos, quer dizer as intenções e objetivos da educação exigem a cooperação dos alunos. O estado pode e deve colocar a disposição os recursos materiais e pessoais, as escolas podem organizar o ensino e *transmitir* conhecimentos e normas, mas são os alunos como sujeitos que tem que *assimilar* os conteúdos. Em consequência, a causalidade entre a intervenção pedagógica e os resultados dos processos de aprendizagem é contingente e as possibilidades de um melhoramento através de uma direção do sistema educacional baseado em sistemas de indicadores e avaliações permanentes são limitadas. Indicadores, avaliações, instrumentos de administração gerencial, *benchmarking*, concorrência, toda a retórica disseminada sob o sigilo da «nova administração pública» encontra na realidade educacional seus limites

nas estruturas, na inércia das tradições históricas das escolas e na complexidade dos processos operativos. Aumentar a eficiência e eficácia do sistema educacional mediante de indicadores e métodos gerenciais é uma promessa arriscada e difícil a cumprir como mostram as experiências e resultados insatisfatórios em muitos países do mundo nos últimos anos.

## 6. Considerações finais

A importância crescente e o uso permanente de indicadores e índices na esfera educacional é um fenômeno ambíguo especificamente quando se trata de medir um construto complexo como a qualidade educacional. Um componente deste conceito multi-dimensional são certamente as competências e os conhecimentos dos alunos, mas outras variáveis, as vezes nem quantificáveis, são igualmente importantes. Mais ainda, a qualidade de um sistema educacional depende do contexto sócio-histórico, seu nível de desenvolvimento e suas necessidades concretas, que não são necessariamente idênticos para todos os países. Sendo assim, não surpreende que até o momento não existe um índice de desenvolvimento ou progresso educacional internacionalmente reconhecido.

A força sedutiva dos indicadores e índices consiste na sua imagem de objetividade, transparência, plausibilidade e simplicidade. Os indicadores atendem a demanda por informações quantitativas articulados em parâmetros e gráficos que impressionam e que prometam facilitar a tomada de decisões. Mas eles aparentam muitas vezes uma racionalidade e um nível de precisão e validade que de verdade não existe, mais ainda se eles carecem de um sistema teórico elaborado justificando seu uso.

Portanto, a tendência atual de atribuir aos indicadores aspectos normativos é metodologicamente injustificável. Medir é um ato descritivo, avaliar (latino: *valere*) é um ato normativo e confundindo os dois é um caminho errado.

## Referências bibliográficas

- Beltrán, José e Villar, Alicia (2010): “Avaliação em Educação: algumas considerações a respeito dos indicadores”. *Educação & Linguagem*, 13 (21), 132-149.
- Cicourel, Aaron Victor (1964). *Method and measurement in sociology*. New York: Free Press of Glencoe.
- Cicourel, Aaron Victor (2004). “I Am NOT Opposed to Quantification or Formalization or Modeling, but I Do Not Want to Pursue Quantitative Methods That Are Not Commensurate With the Research Phenomena Addressed”. <http://www.qualitative-research.net/index.php/fqs/article/view/549/1186>, consultado el 18 de marzo de 2017.
- Felbinger, Sabine (2010). *Messtheoretische Grundlagen*. LMU München: Institut für Statistik. <http://thomas.userweb.mwn.de/Lehre/wise1011/SeminarSozi/SabineFelbinger.pdf>, consultado el 02 de marzo de 2017.
- Hand, David J. (1996): “Statistics and the theory of measurement”. *Journal of the Royal Statistical Society, Series A*, 159 (3), 445-492.
- Kaube, Jürgen (2014): “Literarische Bestenlisten. In Genie hatte Goethe achtzehn Punkte”. *Frankfurter Allgemeine Zeitung*, 8 de octubre de 2014.
- Meyer, Heinz-D. e Bemavot, Aaron (eds.) (2013). *PISA, Power, and Policy: the emergence of global governance*. Oxford: Symposium.

- Meyer, John W. e Rowan, Brian (1977): "Institutionalized Organizations: Formal Structure as Myth and Ceremony". *American Journal of Sociology*, 83 (2), 340-363.
- Meyer, John W. e Rowan, Brian (1978). "The Structure of Educational Organizations" en Marshall W. Meyer (ed.): *Environments and Organizations*. San Francisco: Jossey-Bass.
- Ministério de Educação. Ideb - Apresentação. <http://portal.mec.gov.br/secretaria-de-educacao-basica/programas-e-acoes?id=180>, acessado em 18 de março de 2017.
- OECD (2016a). Education at a Glance 2016: OECD Indicators, OECD Publishing, Paris. <http://dx.doi.org/10.1787/eag-2016-en>, acessado em 18 de março de 2017.
- OECD (2016b). PISA 2015 Results (Volume I): Excellence and Equity in Education, OECD Publishing, Paris. <http://dx.doi.org/10.1787/9789264266490-en>, acessado em 18 de março de 2017.
- Parreira do Amaral, Marcelo (2011). *Emergenç eines Internationalen Bildungsregimes? International Educational Governance und Regimetheorie*. Münster et al.: Waxmann.
- Pasquali, Luiz (2009): "Psicometria". *Revista da Escola de Enfermagem da USP*, 43, 992-999. <http://www.revistas.usp.br/recusp/article/view/40416>, acessado em 15 de março de 2017.
- Pfeiffer, Dietmar K. e Püttmann, Carsten (2011). *Einführung in die Methoden empirisch-pädagogischer Forschung*<sup>4</sup>. Hohengehren: Schneider.
- Pfeiffer, Dietmar K. (2011): "Daten und Diskurse in der aktuellen Bildungspolitik" en Christian Tarnai (ed.): *Sozialwissenschaftliche Forschung in Diskurs und Empirie*. Münster, New York: Waxmann.
- Psacharopoulos, George (1995): "Evaluation of Education and Training: What room for the Comparative Approach?". *International Review of Education*, 40 (1), 5-17.
- Radtke, Franz-Olaf (2009): "Evidenzbasierte Steuerung – Der Aufmarsch der Manager im Erziehungssystem" en Rudolf Tippelt (ed.): *Steuerung durch Indikatoren*. Opladen: Budrich.
- Rohwer, Anja e Hülsewig, Oliver (2011): "Länderrankings: Zu welchen Ergebnissen kommt die neuere Forschung?" *Wirtschaftsdienst*, 91 (11), 735-759. DOI: 10.1007/s10273-011-1295-6, acessado em 19 de março de 2017.
- Rost, Jürgen (2005): "Messen wird immer einfacher". *ZA-Information*, 56, 6-7.
- Sampieri, Roberto Hernández et al., (2013). *Metodologia de pesquisa*<sup>5</sup>. Porto Alegre: Penso
- Siche, Raúl et al., (2007): "Índices versus indicadores: Precisoões conceituais na discussão da sustentabilidade de países". *Ambiente & Sociedade*, X (2), 137-148.
- Soligo, Valdecir (2012): "Indicadores: Conceito e complexidade do mensurar em estudos de fenômenos sociais". *Estudos em Avaliação Educacional*, 23 (52), 15-25
- Spoerhase, Carlos (2014): "Das Maß der Potsdamer Garde. Die ästhetische Vorgeschichte des Rankings in der europäischen Literatur-und Kunstkritik" en Wilfried Barner et al., (eds): *Jahrbuch der Deutschen Schillergesellschaft*. Berlin: de Gruyter.
- Vedder, Paul (1994): "Global Measurement of the Quality of Education: A Help to Developing Countries?". *International Review of Education*, 41 (3-4), 259-284.



Wolter, Andrä (2009): “Hochschulindikatoren in der nationalen Bildungsberichtserstattung” en Rudolf Tippelt (ed.): *Steuerung durch Indikatoren*. Opladen: Budrich.

### Nota biográfica

**Dietmar K. Pfeiffer** es Doctor en Ciencias Sociales. Profesor investigador y miembro del Centro Brasileiro de la Universidad de Münster (Alemania). Consultor internacional en las áreas de economía educacional, desarrollo institucional y evaluación. Entre sus publicaciones recientes están «Análise de dados verbais» en M.Tavares y R. Jarry R. (Org.) (2015), *Metodologias Qualitativas*. Curitiba: CRV; «Das Bildungssystem Brasiliens unter besonderer Berücksichtigung des Hochschulwesens» en V. Oelsner y C. Richter (eds.) (2015): *Bildung in Lateinamerika. Strukturen, Entwicklungen, Herausforderungen*, Münster/New York: Ed. Waxmann. Coautor en «A evasão escolar no contexto da expansão do ensino superior», en M.D. Sobrinho; R. Ennafa; E. Chaleta (eds.) (2016), *La educación superior, el estudiantado y la cultura universitaria*. València: Ed. Neopatria.

# Para discutir la hegemonía epistémica evaluativa eurocéntrica: un enfoque exploratorio

## Discussing Eurocentric evaluative epistemic hegemony: an exploratory approach

Almerindo Janela Afonso<sup>1</sup>

### Resumen

En el presente artículo pretendo, de una forma aún exploratoria, reunir algunos presupuestos de las teorías poscoloniales/descoloniales para elaborar un pensamiento crítico sobre la hegemonía de determinadas políticas, procesos e instrumentos de evaluación (o de medición educativa), que son claramente eurocéntricos (o euroamericanos) y que integran la actual *agenda educativa global*. Sin pretender contribuir para subsanar cualquier posible laguna pero reconociendo que, a pesar de las conocidas excepciones, el campo de las políticas de evaluación se encuentra relativamente saturado por abordajes repetitivos y previsibles, me resulta llamativa la capacidad hermenéutica de los presupuestos poscoloniales/descoloniales, cuya presencia en el campo del pensamiento crítico es cada vez más inevitable. Entre otros aspectos, quisiera llamar la atención sobre la urgencia en la construcción de relaciones y condiciones más justas y democráticas entre los diferentes países del sistema-mundial, especialmente en lo que se refiere a la producción del conocimiento, a través de la evaluación. En este sentido, si los países periféricos o semi-periféricos del sistema-mundial, en diferentes zonas del globo, adoptan de manera acrítica los indicadores definidos por el centro o no son capaces de resistir a la dominación cognitiva, o bien no cuestionan la integración subordinada en programas de evaluación comparativa internacional (como PISA y el AHELO), en cuanto instrumentos al servicio de una hegemonía epistémica eurocéntrica y euroamericana, la agenda educativa global continuará consolidándose únicamente como la expresión de los intereses del capitalismo, ampliando así, cada vez más, la colonialidad del conocimiento científico y evaluativo.

### Palabras clave

Sistema mundial, eurocentrismo, presupuestos descoloniales, hegemonía evaluativa eurocéntrica.

### Abstract

In this article I assemble, in an exploratory manner, some assumptions of postcolonial/decolonial theories so as to critically consider the hegemony of certain policies, processes and evaluation tools (or academic measurement) that make up the current global educational agenda. Without claiming to contribute to overcoming any possible gap, yet recognising that, in spite of known exceptions, the field of assessment policies is pretty much saturated with repetitive and predictable approaches, I feel called upon by the hermeneutic capacity of postcolonial/decolonial assumptions, whose presence in the field of critical thinking is increasingly inescapable. One aspect to which I would like to draw attention is the urgency of building fairer and more democratic relations and conditions between the various nations of the world system, particularly with respect to the production of knowledge, and its evaluation. With this in mind, if the peripheral and semi-peripheral countries of the world system in different parts of the world uncritically adopt the indicators established by the centre, or are unable to resist cognitive domination, or agree to participate (even in a subordinate situation) in international comparative evaluation programmes (such as PISA and AHELO) as instruments in the service of Eurocentric and Euro-American epistemic hegemony, the global education agenda will continue to be consolidated solely as an expression of the interests of capitalism, thereby increasingly expanding the *coloniality* of evaluative and scientific knowledge.

### Key words

World system, eurocentrism, decolonial assumptions, eurocentric evaluative hegemony.

Recibido: 22-03-2017  
Aceptado: 02-05-2017

<sup>1</sup> Profesor y miembro del Centro de Investigación en Educación (CIED), Universidad de Minho, Portugal; [ajafonso@je.uminho.pt](mailto:ajafonso@je.uminho.pt)

## 1. Introducción

Numerosos autores cuestionan de manera bastante contundente las formas de conocimiento hegemónico, o incluso llegan a defender un corte radical con la *dependencia epistémica* moderna, occidental y eurocéntrica. Como recuerda Walter Mignolo, «hoy, la descolonización ya no es un proyecto de liberación de las colonias, con el objetivo de dar lugar a la formación de estados nación independientes, sino más bien el proceso de descolonización epistémica y de socialización del conocimiento». (Mignolo, 2003:631).

Partiendo de este y de otros presupuestos de las teorías poscoloniales/descoloniales, pretendo elaborar un pensamiento crítico sobre la hegemonía de ciertos instrumentos y procesos de evaluación (o de medición educativa) que son claramente eurocéntricos (o euroamericanos) y se encuentran conectados con la agenda educativa global<sup>2</sup>. Se emplean diferentes designaciones en el campo de los estudios poscoloniales, aunque muchos de los presupuestos acaben siendo convergentes. Sin embargo, y en cualquiera de los casos, se trata de presupuestos cuyo conocimiento resulta insuficiente en determinadas áreas del campo educativo<sup>3</sup>. Sin pretender contribuir para subsanar cualquier posible laguna pero reconociendo que, a pesar de las conocidas excepciones, el campo de las políticas de evaluación se encuentra relativamente saturado por abordajes repetitivos y previsible, me resulta llamativa la capacidad hermenéutica de los presupuestos poscoloniales, cuya presencia en el campo del pensamiento crítico es cada vez más inevitable. Como afirma Grosfoguel, «hoy día resulta inexcusable para pensadores del Norte Global seguir produciendo teoría crítica sin diálogo, con sordera hacia las epistemologías del sur». (Grosfoguel, 2011:97).

Pretendo de esta forma en este texto, y de una manera aún exploratoria, dar seguimiento a un diálogo *inter-epistémico* que me parece muy sugerente y prometedor (también) para los abordajes al *campo* de las políticas educativas, en las que preferencialmente me sitúo. Más concretamente, realizaré una breve aproximación a componentes teórico-conceptuales poscoloniales, cuyos presupuestos políticos, culturales y epistemológicos pueden, desde mi punto de vista, aportar nuevas perspectivas al campo de la evaluación en educación, buscando una mayor comprensión sobre las políticas y las agendas actualmente hegemónicas en el contexto mundial.

## 2. De la agenda nacional al giro hacia la agenda global: el papel de las organizaciones internacionales

Efectivamente, si estas políticas empezaron teniendo un primer desarrollo (nacional) en países centrales, el impulso de la globalización y de la creciente internacionalización del capitalismo (donde se encuadra la acción de poderosas agencias e instancias internacionales y supranacionales) acabaron por reconfigurar estas agendas, ampliándolas también a países semi-periféricos y periféricos del sistema mundial.

Si consideramos las actuales políticas de evaluación (inicialmente impulsadas por los EE. UU. y por Inglaterra, y rápidamente adoptadas por diferentes países capitalistas occidentales), constataremos que estamos hoy en una fase sustancialmente diferente de aquella que Guy Neave (1988) inicialmente denominó

---

2 Ensayé inicialmente esta perspectiva en una conferencia de apertura que realicé en [desanonimizar]. el II Coloquio Cabo-Verdiano de Educação – Políticas e Práxis da Educação nas Perspetivas e em Contextos Pós-coloniais. Praia: Universidade de Cabo Verde, 24 de abril de 2015 y en una comunicación oral presentada en el congreso *From decolonisation to postcolonialism: A global approach*. Oporto: Facultad de Letras de la Universidade do Porto, 11-13 noviembre de 2015. La participación en estos eventos así como la actual publicación fueron apoyados por los fondos nacionales de la FCT- Fundação para a Ciência e Tecnologia – en el ámbito del proyecto PEst-OE/CED/UI1661/2014, del Centro de Investigação em Educação da Universidade do Minho (Cied/UM).

3 En lo que dice respecto a la investigación portuguesa en educación (sobre todo en el dominio de la sociología de la educación y de las políticas educativas), se hace una referencia escasa a los trabajos de Ramón Grosfoguel, Walter Mignolo, Enrique Dussel, Alejandro Moreno, Aníbal Quijano, Stuart Hall, Edward Said, y muchos otros, a excepción de María Paula Menezes y de otros investigadores e investigadoras del Centro de Estudos Sociais de la Universidad de Coimbra, donde destaca Boaventura Sousa Santos como un autor de referencia obligatoria en este campo.

como siendo la del contexto de emergencia del Estado-evaluador<sup>4</sup>. Efectivamente, la primera fase inicial del Estado-evaluador se circunscribe al contexto nacional, habiendo sido la evaluación un instrumento esencialmente de control y supervisión adoptado por diferentes gobiernos en el ámbito de la redefinición del papel del Estado –redefinición ésta que ha dado protagonismo a la evaluación como instrumento político, en articulación con otros instrumentos, dispositivos y modos de acción por entonces emergentes–. Éste es el caso, por ejemplo, de las formas emergentes de *gobernanza* que se fueron construyendo a través de rearticulaciones diversas entre pilares de regulación social (tradicionales) y nuevos actores colectivos, también sustentadas en redes de relaciones e interdependencias pluriescales, con distribución descentralizada o desconcentrada de responsabilidades político-administrativas. Es también el caso de la diversificación y amplificación de formas de privatización; de la proliferación de discursos y modelos de *accountability* conservadores y tecnocráticos, así como de las agendas de la nueva gestión pública (*new public management*), entre otros.

En el sector público, considerado en sentido amplio, «la necesidad de desarrollar y perfeccionar herramientas de control, especialmente orientadas a la evaluación de los logros o resultados obtenidos, se asocia en la literatura a la nueva gestión pública, con la introducción de conceptos, prácticas y técnicas procedentes del sector privado». (García Sánchez, 2007:44). En el sector de la educación, más específicamente, la nueva gestión pública tuvo su expresión en «políticas como la autonomía de las escuelas, la profesionalización de los directores de las escuelas, la evaluación estandarizada y la rendición de cuentas de los profesores, y siendo ampliamente difundida por organizaciones internacionales como la OCDE» (Verger y Curran, 2014:253).

En lo que respecta a la evaluación en educación, sobre todo en el contexto de sus países miembros, la acción de la OCDE ha pasado a sentirse de forma más regular a partir del inicio de los años noventa (aunque haya sido ya esta misma organización la que haya conducido, muchos años antes, los famosos estudios de sistemas educativos nacionales). Así, si aún en la primera fase del Estado-evaluador varias decenas de estados, independientemente de la orientación político-ideológica de sus gobiernos, crearon, desarrollaron o reactualizaron sistemas nacionales de evaluación, empiezan ya a atisbarse las líneas de la transición hacia una nueva fase cuando, paralelamente a las pruebas y evaluaciones externas nacionales, estos estados van aumentando su participación en estudios internacionales basados en metodologías e indicadores esencialmente cuantitativos. Se convierten en práctica frecuente, por ejemplo, los informes periódicos *Education at a Glance*. La influencia de estos informes, tal como indica A. Teodoro, está bien presente en las políticas educativas de los estados miembros de la OCDE, desde la década de 1990. Se trata, según este autor, de «una influencia que se manifiesta no por un mandato explícito, sino por la necesidad de responder a una agenda global basada en la comparación y, sobre todo, en la competición de performance de los sistemas educativos». (Teodoro, 2012:28).

Por este motivo, organizaciones internacionales y supranacionales cada vez más pasan a incentivar dispositivos de evaluación comparativa internacional a gran escala, no solamente por la definición, perfeccionamiento y uso de indicadores cuantitativos, sino también por el uso de tests estandarizados. Los resultados de estos estudios comparativos han estado al servicio de diversos objetivos, entre los que se encuentran la construcción de una «narrativa legitimadora» para las políticas en educación adoptadas en el contexto interno (entre otros, Lingard, 2016:611).

---

4 Adopto la expresión Estado-evaluador para situar la problemática de la evaluación en su relación con el Estado y otras instancias de regulación y, en este sentido, es un uso más amplio que aquel que se hace por parte de Guy Neave, que acuñó la expresión esencialmente para análisis de la enseñanza superior. Para una perspectiva más amplia, especialmente en relación a la formación docente, ver también John Elliott (2002).

En esta segunda fase del Estado-evaluador, además de informes como el *Education at a Glance*, el programa PISA (Programme for International Student Assessment) se convirtió en una referencia y ahora tiene «pretensiones normativas» (Beltrán Llavador, 2016: 130). Este programa ha sido una piedra angular en «la construcción de un espacio conmensurable de medición educativa a nivel global» (Rizvi y Lingard, 2010:133). Además de ésta, una otra contribución se está realizando a través del programa AHELO (Assessment of Higher Education Learning Outcomes), cuyo estudio de viabilidad ha sido recientemente concluido, también por la OCDE. Con el programa PISA se comparan países, mientras que con AHELO se compararán instituciones de enseñanza superior. Se trata, según R. Shahjahan, de un nuevo impulso para «desterritorializar la política de la enseñanza superior mediante la implementación de un régimen de tests transnacionales». También en este caso, al buscar «construir un espacio global de equivalencia para la enseñanza y el aprendizaje en la enseñanza superior» (Shahjahan, 2013:683), AHELO «perpetúa la colonialidad» del conocimiento (y su respectiva regulación), mediante una «retórica imperial» que sugiere que la educación superior global se encuentra en profunda crisis y que solamente se puede «salvar» incorporando realidades locales en modelos euroamericanos que, por su parte, al transformarse en universales, constituyen importantes instrumentos «para un nuevo orden mundial neoliberal vinculado al mercado de trabajo global». (Shahjahan, 2013:689).

Con el impacto que han tenido (y que es previsible que continúen a tener) los sucesivos estudios, recomendaciones y programas, no hay duda de que la OCDE es en la actualidad «el principal *think tank* mundial de la globalización hegemónica». (Teodoro, 2012:30).

Además del protagonismo de organizaciones como la OCDE, existen otras organizaciones económicas internacionales que se disputan la educación, incluso de una forma más agresiva y radical, obsesionadas con las agendas de la privatización y de la mercantilización de la educación a escala mundial. En este sentido, la acción de éstas últimas permite caracterizar una tercera fase (que he venido a denominar como de post-estado evaluador), en la que se hace cada vez más evidente la presión originada en instancias cuyo ejemplo paradigmático es la OMC – Organización Mundial del Comercio (Afonso, 2013).

No es casualidad que la OMC, después de considerar la educación como un servicio de elevado valor mercadeable y rentable, haya recibido contribuciones de algunos países donde surge la idea de extender los acuerdos de GATS a la evaluación y a los tests en educación –lo que, desde mi punto de vista, será inevitable que ocurra, dadas las potencialidades evidentes que los tests y pruebas estandarizadas tienen para ampliar los procesos de comercialización y de privatización y, en consecuencia, servir para el proceso de *acumulación* capitalista–.

Tal vez sea esta organización, más que ninguna otra, la que está radicalizando procesos que conducen de manera inevitable a la imposición (y no solo a la inducción) de una epistemología eurocéntrica en los países semi-periféricos y periféricos. Ya no se trata solamente del incentivo creciente a regímenes de *franchising*, sino también a la valorización de instrumentos estandarizados de evaluación, porque éstos son congruentes con la transferencia de modelos curriculares (originados en el Norte Global), los cuales, por lo tanto, se sobrepondrán o tenderán a ser ajenos a la autonomía relativa de los sistemas educativos nacionales y sus realidades históricas, políticas y culturales. Como si esto no bastase son ya muy visibles, en especial en países periféricos y semi-periféricos del sistema mundial, los efectos de la exportación y utilización de *marcas* creadas en países centrales. Estos efectos vienen asociados a la creación de monopolios privados de organizaciones de enseñanza (nacionales e internacionales), que resultan en la venta y fusión

de escuelas, facultades y universidades, altamente lucrativos, y en muchos casos ya contados en las bolsas de valores (Bianchetti y Sguissardi, 2017).

### 3. Colonialidad y comparativismo evaluador en países periféricos y semi-periféricos

Como la ampliación de estas agendas se confunde con el movimiento creciente de internacionalización de las políticas de evaluación (y de la mayor frecuencia de uso de indicadores y otros instrumentos de medición), éstas tenderán a extenderse por todos los países, con independencia de su posición en el sistema mundial. En este sentido, el comparativismo evaluador se encuentra hoy presente no solamente en los países capitalistas centrales, sino también en los países semi-periféricos, en el contexto europeo (como es el caso de Portugal y España), o en los países semi-periféricos y periféricos de una parte del África subsahariana, de Asia y de América Latina.

Sin embargo, la densidad y el impacto de estas agendas es aún relativamente bajo, y muy diferenciado en lo que respecta especialmente a países de Asia y del África subsahariana. Por ejemplo, hasta el momento el informe PISA no ha salido completamente del ámbito de los países miembros (o asociados e invitados) de la OCDE (Sellan y Lingard, 2013). En su lugar, en lo que se refiere a la educación, tienen un peso mucho mayor los condicionantes y exigencias que ejercen organizaciones financieras tales como el Banco Mundial (BM) o el Fondo Monetario Internacional (FMI). Pero no son solamente éstos los aspectos que podrán explicar la (aún) inexpresiva participación en evaluaciones internacionales comparativas, aunque haya alguna excepción relativa al hecho de que algunos países del África subsahariana hayan participado en los estudios del *Trends in International Mathematics and Science Study* (por ejemplo, para el caso del TIMSS 2011, ver Bofah y Hannula, 2015). En el caso de África del Sur, después de muchas décadas de «aislamiento político y social, y de un legado de inequidad» (Howie, 2012:88), fue necesario dar prioridad a políticas contra la segregación racial y reorganizar un sistema educativo que estaba extremadamente fragmentado. Esto podrá explicar, al menos en parte, el hecho de que África del Sur solamente haya empezado a participar en estudios comparativos a partir del TIMSS 1995. Por otra parte, y al menos hasta 2004, este país se ha resistido a suscribir compromisos comerciales en el área de la educación, distanciándose de los acuerdos del GATS. En ese momento, además, se puede escribir: «Se trata de una posición significativa dado que este país exporta servicios educativos al resto del continente. Lo hace, en todo caso, en el ámbito de acuerdos bilaterales y en un marco de beneficio mutuo para los países implicados, y precisamente fuera del régimen de la política comercial». (Santos, 2004:37).

Más recientemente, con la creación del PISA-D (PISA para el desarrollo), al que algunos países semi-periféricos y periféricos, de África, América Latina y Asia ya se han unido, es de esperar que, de una forma más sistemática, se amplíen a estos y otros países de diferentes continentes los efectos globales de este programa de evaluación. Como indica el Director de Educación y Competencias de la OCDE, se trata de ofrecer «datos útiles para ayudar» a tomar decisiones relativas a los respectivos sistemas educativos y, por otro lado, «igualmente importante es que los países participantes en el PISA para el desarrollo se beneficiarán de intercambios entre pares con otros miembros de la comunidad global PISA». (Schleicher, 2016:1). Desde mi punto de vista, el deseo de contribuir para la creciente hegemonía del conocimiento científico y eurocéntrico relativo a la evaluación (y medición) en educación se encuentra aquí claramente presente en la idea, aparentemente benigna, de la afirmación de una «comunidad global PISA<sup>5</sup>».

---

5 Una visión aún más optimista se desprende de la afirmación de M. Lockheed, cuando escribe: «El aumento del número de países en desarrollo que participan en las evaluaciones internacionales es indiscutible, gracias a la gran participación de los organismos donantes internacionales [...] una creciente necesidad de evidencias

La cuestión, sin embargo, es mucho más compleja porque, tal como revela esta breve incursión en la literatura, las realidades nacionales están muy diferenciadas y, en muchos casos, poco o nada permeables a la adhesión a estas agendas. En el caso de muchos países del África subsahariana, por ejemplo, hay que tener en consideración que

*«Las preocupaciones centrales en los contextos post-coloniales de educación son la identidad y el idioma, la formulación y la reformulación de políticas y regímenes post-coloniales, términos post-coloniales de contestación discursiva, cambios en la naturaleza del Estado y nuevas teorías de derechos [...]. A pesar de esto, muchos [autores] postulan agendas de investigación, aberturas conceptuales e indicadores para estrategias políticas de cambio educativo que expresen elevados niveles de aspiración a un proyecto post-colonial que aún no se ha realizado pero que, de algún modo, se encuentra inmanente en las ideas, relaciones sociales y contestaciones que tienen lugar en la actualidad».* (Unterhalter, 2012:14).

En lo que se refiere, más específicamente, a la evaluación de los aprendizajes, hay países africanos que son más favorables a los exámenes, y otros que valorizan más la evaluación continua. Curiosamente, «la educación centrada en el alumno es una de las ideas educativas más difundidas en el África subsahariana». Como consecuencia de ello, las políticas educativas han intentado «disminuir el peso de los exámenes y aumentar la importancia de la evaluación continua como una forma de estimular pedagogías centradas en el alumno». (Chisholm y Leyendecker, 2012:54). En este contexto, habrá con certeza tensiones y contradicciones que es necesario entender y problematizar con mayor profundidad<sup>6</sup>.

De cualquier modo, en la perspectiva de algunos autores, esto es indicador de que el globo está siendo reconstruido como un espacio único de medición comparativa o como un espacio conmensurable del desempeño de los sistemas escolares, observándose lo que se puede denominar como «una forma de *panoptismo global*, con el ojo global funcionando en una capacidad reguladora a través y dentro de los estados nacionales». (Lingard *et al.*, 2013:540).

Por este motivo, desde mi punto de vista la cuestión fundamental subyacente a este movimiento se enmarca claramente en el ámbito de una epistemología eurocéntrica o del Norte Global, y en la creencia en un único conocimiento universal. Como indica Grosfoguel, «El eurocentrismo se ha perdido en el camino del universalismo [...] que diluye todos los particulares en lo universal». Y, en la estela de Césaire, aquel autor añade aún que «el universalismo abstracto es lo que, a partir de un particularismo hegemónico, pretende constituirse como un proyecto global imperial para el mundo entero y que, al presentarse como desincorporado (“disembodied”), esconde el lugar epistémico de la enunciación». (Grosfoguel, 2012:100).

Este supuesto conocimiento universal es congruente con las lógicas de la internacionalización del capitalismo y con los (supuestos) proyectos de modernización, a los que sirven, directa o indirectamente,

---

de la “eficacia de la ayuda”, ha impulsado la demanda de más países, particularmente países en desarrollo, para participar en evaluaciones de carácter internacional. [...] Este aumento se atribuye en parte al posicionamiento de los resultados del PISA en publicaciones dirigidas a economistas. [...] la participación en todos los tipos de evaluaciones internacionales ha beneficiado a los países en desarrollo tanto indirecta como directamente. Indirectamente, la participación ha influenciado las normas para el currículo y la práctica docente, llevando a una reforma curricular y dirigiendo la atención al desarrollo profesional de los maestros. Directamente, la participación ha fortalecido la capacidad de evaluación de sus organismos nacionales de evaluación, aportando formación relacionada con la evaluación, así como experiencia práctica sobre el proceso de elaboración y ejecución de evaluaciones a gran escala». (Lockheed, 2013:163).

6 Las tensiones y contradicciones no tienen lugar solamente en situaciones en que la *colonialidad* del poder se expresa en las relaciones Norte/Sur o en las relaciones centro/semiperiferia/periferia/centro, sino también en el interior del propio centro, de la propia semiperiferia o de la propia periferia. Las contribuciones que se vienen produciendo movilizándolo las teorías post-coloniales en torno a la problematización de las relaciones pedagógicas en la escuela pública de masas (predominantemente frecuentada por las clases populares) en un país semiperiférico como Brasil muestran, por ejemplo, que lo que se considera éxito o fracaso puede ser visto como expresión de la colonialidad del poder, siendo una de las principales funciones de la evaluación «naturalizar la diferencia colonial, que silenciosamente está presente en el cotidiano escolar, para justificar la selección y exclusión de sujetos, conocimientos y culturas». (Esteban 2010:51).

las políticas de educación y de evaluación. Sobre este último aspecto, ciertas formas de evaluación comparativa (como el informe PISA) tienen, por lo menos, el apoyo tácito de las élites de países periféricos y semi-periféricos que continúan sin cuestionar muchos de los supuestos objetivos de *modernización*, seducidas ideológicamente por ímpetus de progreso (o de desarrollo) solamente susceptibles de ser realizados en conexión con agendas educativas de países capitalistas centrales o de elevado desarrollo (a pesar de que estas agendas parezcan neutras o benévolas, al ser transmitidas por organizaciones internacionales de prestigio tales como la OCDE o el Banco Mundial)<sup>7</sup>.

La creencia de que no hay alternativas más allá del capitalismo liberal es congruente no solo con el a-historicismo vigente en el occidente europeizado y americanizado, sino que también es congruente con la adopción de orientaciones políticas, culturales y económicas que parecen presuponer un único patrón de referencia –el de los países capitalistas centrales y/o países más desarrollados y competitivos, y supuestamente más cultos y más *educados*–. Dicho de otra forma, estos discursos sugieren que los mismos niveles de educación y conocimiento científico, y las mismas oportunidades de inserción ventajosa en una economía global de elevado carácter competitivo resultan accesibles a todos los países, siendo para esto necesario que las evaluaciones comparativas nacionales o los tests estandarizados internacionales indiquen (objetivamente) los déficits a superar y los caminos a seguir.

Pero las desigualdades entre países no se superan con evaluaciones comparativas basadas en tests estandarizados. A pesar de que existan fuertes tendencias para que estos instrumentos encuentren su justificación a la luz de la evolución de las metodologías y técnicas de evaluación (ellas propias producidas por el conocimiento científico hegemónico), pienso que es pertinente apelar aquí (recontextualizándola) a la observación de Immanuel Wallerstein cuando apuntaba que a los países situados en posiciones periféricas y semiperiféricas en la jerarquía del sistema mundial «se les incita a superar el hándicap de su bajo *status*» asumiendo valores supuestamente universales. Y añade: «Haciéndose competitivos en el mercado, los individuos y los grupos pueden obtener lo que otros ya tienen, y así alcanzarán algún día la igualdad; mientras tanto, la desigualdad resulta inevitable». (Wallerstein, 2004:263)<sup>8</sup>.

Si el ansia ilimitada por el lucro y la acumulación forma parte del *ethos* del sistema capitalista, es importante cuestionarnos por qué razón los países centrales (o las organizaciones internacionales más poderosas en la construcción de la agenda global) están interesadas en la subida de los niveles educativos de los países periféricos y semi-periféricos. Una respuesta posible, entre otras muchas, puede remitir para el hecho de que sea deseable el desarrollo de estos países, siempre que este desarrollo tenga lugar como condición para la continua expansión capitalista, especialmente en la preparación e integración de nuevos consumidores, lo antes posible, en el mercado global de la enseñanza (ahora ya asumido y sin ningún escrúpulo, a modo de bien transaccionable y de elevado valor mercantil) así como también en el mercado de trabajo global.

Sabiendo que hay vínculos frecuentemente convergentes (aunque no siempre claros) entre los intereses del capitalismo y los objetivos de muchas organizaciones internacionales, la acción de éstas últimas no puede pensarse aceptando de manera acrítica los discursos que ellas mismas elaboran y las misiones que

---

7 Para un interesante abordaje sobre el papel de las élites en un país africano, ver por ejemplo Sumich (2008).

8 Aunque el lugar de los diferentes países en el sistema mundial pueda modificarse (y se modifique), la tendencia es que, una vez definida la división internacional del trabajo, esa situación se reproduzca: «Esta posición puede ser sintetizada en la idea de que existe efectivamente una división internacional del trabajo, que ha sido establecida por un recorrido histórico de consolidación del capitalismo como un sistema global a lo largo de aproximadamente doscientos años; más aún, que esta división internacional del trabajo se reproduce indefinidamente, a menos que se alteren aspectos estructurales». (Almeida Filho, 2011:37).



anuncian. En este sentido, podemos y debemos interrogarnos en relación al exacerbado (y aparentemente benévolo) protagonismo de organismos internacionales, tales como la OCDE y el Banco Mundial que, de forma creciente, se aceptan como incuestionables y supuestamente imprescindibles en la estructuración de políticas educativas nacionales y en la formulación de muchas otras decisiones. También por este motivo, pensar las políticas de evaluación en educación de una forma crítica puede encontrar, en las perspectivas poscoloniales, contribuciones originales que justifiquen atravesar las habituales fronteras de la investigación y reflexión en este campo.

Desde mi punto de vista, es precisamente esta relación de dominación cognitiva la que continúa sustentando muchas de las opciones de la política educativa y muchos de los modelos de evaluación adoptados en diferentes sistemas educativos, sobre todo los de países que tienen una posición periférica o semiperiférica en el contexto mundial. En el fondo, lo que está en cuestión es saber si es posible (y deseable) pensar en la evaluación poniendo de lado la referencia al conocimiento científico moderno y las teorías dominantes, es decir, fuera de cualquier dependencia epistémica occidental u occidentalizada.

Desde mi punto de vista sí que es posible y deseable, a pesar de que la evaluación en educación y, sobre todo aquella que se presenta como científica y técnicamente avanzada (como es el caso de los tests estandarizados) continúe siendo un instrumento de la supremacía epistémica evaluativa eurocéntrica, que se mantendrá seguramente de esta forma durante más tiempo como uno de los instrumentos actuales de la globalización capitalista o incluso de *neocolonialismo* en un contexto de post-colonización, hasta el momento en que se puedan fortalecer agendas alternativas. Además, más allá del currículo, pocas áreas serán tan universalizables y homogeneizables como la evaluación. Y es esta posibilidad de universalización y de homogeneización la que la ha difundido como uno de los virus del neoliberalismo en educación.

De nuevo, una vuelta al cientifismo neopositivista le queda bien a las lógicas capitalistas y las llamadas políticas basadas en evidencias (Afonso, 2015). No es por casualidad que estas políticas indujeron e inducen al uso de indicadores cuantificables y medibles, que han sido una de las estrategias privilegiadas por las organizaciones internacionales y supranacionales. Es más, si vamos atrás en el tiempo, tal como llama la atención uno de los autores que escribe a partir de los postulados de las teorías poscoloniales, las decisiones basadas en la evidencia tienen sus raíces en los procesos (iluministas) de burocratización, estandarización y normalización que tuvieron lugar de forma concomitante con el comercio de esclavos y el sistema de plantaciones, los cuales constituyeron experiencias de disciplinación e ingeniería social, de producción en serie, de sistematización de la vida humana y de institucionalización de la estandarización de procesos (Shahjahan, 2011). En el contexto actual y de una forma más genérica, «la objetivación numérica de fenómenos sociales» sirve también para «despolitizar las cuestiones políticas al naturalizar determinadas interpretaciones de la realidad en detrimento de visiones alternativas». (Erkkilä y Piironen, 2014:344).

Por este motivo, la cuestión que se levanta una vez más es la de saber quién tiene el poder de definir los indicadores que cuentan para la comparación entre países; qué países participan en la definición de estos indicadores; y si no deberíamos ser capaces de pensar indicadores alternativos a los indicadores actualmente hegemónicos. Pero hay ya una importante reflexión que va más lejos, cuestionando «si no deberíamos comenzar a pensar no tanto, o no solo, en la construcción de “indicadores alternativos”, sino más bien en una alternativa a los indicadores». (Beltrán Llavador, 2012:49). En cualquier caso, sabemos que no todos los países participan (o tienen el poder de interferir) en la definición y construcción de

indicadores, pero cuando las decisiones políticas y administrativas están basadas en cuantificaciones numéricas, sus premisas no podrían dejar de estar «visibles, accesibles y abiertas a la deliberación pública» (Erkkila y Piironen, 2014:346).

Tal como en el caso de los tests estandarizados, los indicadores pueden ser vehículos de un «nuevo imperialismo que refleja los intereses de las naciones occidentales y, más ampliamente, del capitalismo global». (Nguyen *et al.*, 2009:110). Y si los países periféricos o semi-periféricos del sistema-mundial, en diferentes zonas del globo, adoptan de manera acrítica los indicadores definidos por el centro, o bien no son capaces de resistir a la dominación cognitiva, o no cuestionan la integración subordinada en programas de evaluación comparativa internacional (tales como PISA y AHELO), en cuanto instrumentos al servicio de una hegemonía epistémica eurocéntrica y euroamericana, la agenda educativa global continuará consolidándose como la expresión de los intereses del capitalismo, ampliando así, cada vez más, la *colonialidad* del conocimiento científico y evaluativo.

## Referencias bibliográficas

- Afonso, Almerindo J. (2013): “Mudanças no Estado-avaliador: comparativismo internacional e teoria da modernização revisitada”. *Revista Brasileira de Educação*, 18 (53), 267-284.
- Afonso, Almerindo J. (2015): “Recuo ao cientificismo, paradoxos da transparência e corrupção em educação”. *Educação e Pesquisa*, 41 (nº especial), 1313-1326.
- Almeida Filho, Niemeyer (2011): “Os países em desenvolvimento estão se tornando desenvolvidos? Crítica e Sociedade. *Revista de cultura política*, 1 (2), 35-50.
- Beltrán Llavador, José (2012): “¿Indicadores alternativos o alternativa a los indicadores? Alguns reflexiones sobre política educativa” en António Teodoro y Edineide Jezine (eds.): *Organizações internacionais e modos de regulação das políticas de educação. Indicadores e comparações internacionais*. Brasília: Liber Livro.
- Beltrán Llavador, José (2016): “Medida y valor de la escuela. Desde la rendición de cuentas hasta la rendición de sentido”. *Profesorado – Revista de curriculum y formación del profesorado*, 20 (3), 119-143.
- Bianchetti, Lucídio y Sguissardi, Valdemar (2017). *Da universidade à commoditycidade*. Campinas: Mercado das Letras.
- Bofah, Emmanuel Adu-tutu y Hannula, Markku S. (2015): “TIMSS data in an African comparative perspective: Investigating the factors influencing achievement in mathematics and their psychometric properties”. *Large-Scale Assessments Education*, 3 (4), 1-36.
- Chisholm, Linda y Leyendecker, Ramon (2012): “A reforma do currículo na África subsariana: quando o local encontra o global” en Robert Cowen *et al.*, (eds.): *Educação Comparada – panorama internacional e perspectivas*. Brasília: CAPES/UNESCO.
- Elliott, John (2002): “The paradox of educational reform in the evaluatory State: Implications for teacher education”. *Prospects*, 32 (3), 273-287.
- Erkkila, Tero; Piironen, Ossi (2014): “(De)politicizing good governance: the World Bank Institute, the OECD and the politics of governance indicators”. *Innovation: The European Journal of Social Science Research*, 27 (4), 344-360.

- Esteban, M. Teresa (2010): “Diferença, aprendizagens e avaliação: perspectiva pós-colonial e escolarização” en M. Teresa Esteban y Almerindo J. Afonso (eds): *Olhares e Interfaces. Reflexões críticas sobre a avaliação*. São Paulo: Cortez.
- García Sánchez, Isabel María (2007): “La nueva gestión pública: evolución y tendencias”. *Presupuesto y Gasto Público*, 47, 37-64.
- Grosfoguel, Ramón (2011): “La descolonización del conocimiento: diálogo crítico entre la visión descolonial de Frantz Fanon y la sociología descolonial de Boaventura de Sousa Santos” en *Formas-Otras: Saber, nombrar, narrar, hacer*. Barcelona: CIDOB Edicions.
- Grosfoguel, Ramón (2012): “Decolonizing Western Uni-versalisms: Decolonial Pluri-versalism from Aimé Césaire to the Zapatistas”. *Transmodernity. Journal of Peripheral Cultural Production of the Luso-Hispanic World*, 1 (3), 88-104.
- Howie, Sara (2012): “High-stakes testing in South Africa: friend or foe?” *Assessment in Education: Principles, Policy & Practice*, 19 (1), 81-98.
- Lingard, Bob (2016): “PISA: fundamentações para participar e acolhimento político”. *Educação & Sociedade*, 37 (136), 609-627.
- Lingard, Bob, Martino, Wayne, Rezai-Rashti, Goli (2013): “Testing regimes, accountabilities and education policy: commensurate global and national developments”. *Journal of Education Policy*, 28 (5), 539-556.
- Lockheed, Marlane (2013): “Causes and consequences of international assessments in developing countries” en Heinz-Dieter Meyer y Aaron Benavot (eds): *PISA, Power, and Policy the emergence of global educational governance*. Oxford: Symposium Books.
- Mignolo, Walter (2003): “Os esplendores e as misérias da «ciência»: colonialidade, geopolítica do conhecimento e pluri-versalidade epistémica” en Boaventura S. Santos (comp.): *Conhecimento prudente para uma vida Decente: Um discurso sobre as ciências revisitado*. Porto: Afrontamento.
- Neave Guy (1988): “On the cultivation of quality, efficiency and enterprise: an overview of recent trends in higher education in Western Europe, 1986-1988”. *European Journal of Education*, 23 (1/2), 7-23.
- Nguyen, Phuong-Mai, Elliott, Julian, Terlouw, Cees, Pilot, Albert (2009): “Neocolonialism in education: Cooperative learning in Asian context”. *Comparative Education*, 45 (1), 109-130.
- Rizvi, Fazal y Lingard, Bob (2010). *Globalizing education policy*. New York: Routledge.
- Santos, Boaventura S. (2004). *A universidade no século XXI. Para uma reforma democrática e emancipatória da universidade*. São Paulo: Cortez.
- Schleicher, Andreas (2016). “Prólogo”. En OCDE (2016). *Pisa para el Desarrollo*. [https://www.oecd.org/pisa/pisa-for-development/PISA-D\\_brochure\\_2016\\_ESP.pdf](https://www.oecd.org/pisa/pisa-for-development/PISA-D_brochure_2016_ESP.pdf)
- Sellan, Sam y Lingard, Bob (2013): “The OECD and global governance in Education”. *Journal of Education Policy*, 28 (5), 710-725.
- Shahjahan, Riyad (2011): “Decolonizing the evidence-based education and policy movement: revealing the colonial vestiges in educational policy, research, and neoliberal reform”. *Journal of Education Policy*, 26 (2), 181-206.

- Shahjahan, Riyad (2013): “Coloniality and a global testing regime in higher education: unpacking the OECD’s AHELO initiative”. *Journal of Education Policy*, 28 (5), 676-694.
- Sumich, Jason (2008). “Construindo uma nação: ideologias de modernidade da elite moçambicana”. *Análise Social*, 43 (2), 319-345.
- Teodoro, António (2012): “Os novos modos de regulação transnacional das políticas de educação: a regulação pelos resultados e o papel das comparações internacionais” en António Teodoro y Edineide Jezine (eds.): *Organizações Internacionais e Modos de Regulação das Políticas de Educação. Indicadores e comparações internacionais*. Brasília: Liber Livro.
- Unterhalter, Elaine (2012): “Introdução editorial da secção” en Robert Cowen *et al.*, (eds.): *Educação Comparada – panorama internacional e perspectivas*. Brasília: CAPES/UNESCO.
- Verger, Antoni y Curran, Marta (2014): “New public management as a global education policy: its adoption and re-contextualization in a Southern European setting”. *Critical Studies in Education*, 55 (3), 353-271.
- Wallerstein, Immanuel (2004). *Capitalismo histórico y movimientos antisistémicos. Un análisis de sistemas-mundo*. Madrid: Akal.

## Nota biográfica

**Almerindo Janela Afonso** es sociólogo y profesor de la Universidade do Minho, Portugal, donde imparte clase desde 1985. Ha investigado y ejercido docencia en las áreas de sociología de la educación y políticas educativas, entre otras. Ha ejercido varios cargos de gestión científico-pedagógica y es director del Departamento de Ciencias Sociales de la Educación y miembro del comité directivo del Doctorado en Ciencias de la Educación. Ha recibido el premio «Rui Gracio», otorgado por la Sociedad Portuguesa de Ciencias de la Educación y la Fundación Calouste Gulbenkian. Entre otros trabajos, es autor de *Avaliação Educacional: Regulação e Emancipação*. São Paulo: Cortez Editora, 2014; coorganizador de *Olhares e Interfaces. Reflexões Críticas sobre a Avaliação*. São Paulo: Cortez, 2013; *Fragments de Escrita Pública*. Porto: Profedições, 2013.

# Competências de Estudo dos Estudantes Universitários: Estudo comparativo entre uma universidade portuguesa e brasileira

Competencias de los estudiantes universitarios: estudio comparativo entre una universidad portuguesa y brasileña

Rubia Salheb Fonseca, Joaquim Escola, Armando Loureiro y Amâncio Carvalho<sup>1</sup>

## Resumo

Para se compreender as competências de estudo e autorregulação no ensino superior, há que se colocar o estudante como centro do seu processo de aprendizagem. Neste estudo analisou-se a relação entre o perfil de competências de estudos dos estudantes de universidades de dois países (Brasil e Portugal). Trata-se de um estudo comparativo, transversal, de abordagem quantitativa, com 1240 estudantes universitários, aos quais foi aplicado um inquérito por questionário. O tratamento de dados foi efetuado com recurso ao SPSS (23.0).

## Palavras-chave

Estudantes Universitários, universidade, competência em Estudo, processo de Ensino-Aprendizagem.

## Resumen

Para comprender las competencias de estudio y autorregulación en educación superior, el estudiante debe situarse en el centro de su proceso de aprendizaje. Este estudio analizó la relación entre el perfil de competencias de estudios de los estudiantes universitarios en dos países (Brasil y Portugal). Se trata de un estudio comparativo, transversal, cuantitativo, con 1240 estudiantes universitarios, sobre los que se realizó un sondeo por encuesta. El tratamiento de los datos fue realizado mediante el programa SPSS (23.0).

## Palabras clave

Estudiantes universitarios, universidad, competencias de estudio, proceso de enseñanza-aprendizaje.

## Abstract

In order to understand the competences of study and self-regulation in higher education, the student must be placed at the center of his learning process. This study analyzed the relationship between the proficiency profile of studies of students from universities in two countries (Brazil and Portugal). This is a cross-sectional, cross-sectional, quantitative study involving 1240 college students, to whom a questionnaire survey was applied. Data processing was done using SPSS (23.0).

## Key words

University Students, university, study skills, teaching-learning process.

Recibido: 03-03-2017  
Aceptado: 02-05-2017

<sup>1</sup> Universidade de Trás-os-Montes e Alto Douro (UTAD), Portugal, Rubia Salheb Fonseca, rubiasalf@yahoo.com.br; Joaquim Escola, jescola@utad.pt; Armando Loureiro, aloureiro@utad.pt; Amâncio Carvalho, amancioc@utad.pt

## 1. Introdução

A obtenção do diploma de curso superior é uma meta para grande parte dos jovens como se pode notar no crescimento das taxas de inscrição no Ensino Superior (ES) em Portugal. De acordo com a Pordata (2017) o crescimento do número de inscritos permitiu a gradual qualificação da população portuguesa, tendo a percentagem de população residente com um grau superior aumentado de 0.5% em 1960 (49.065 indivíduos) para 12% em 2011 (1.244.742 indivíduos).

Ainda fazendo menção à Pordata (2014), a população residente em Portugal no ano de 2014 registrou o número de 10.401.1, tendo como população residente de 15 e mais anos com ensino superior 16.5% e, em 2015, 17.1%. O número de alunos diplomados era de 72.906 e o total dos que efetuaram a matrícula pela primeira vez no ES foi de 104.255. Destes, 46.120 eram homens e 57.518 mulheres, demonstrando mais uma vez a permanência deste quadro de ascensão das mulheres na história dos alunos no ES em Portugal.

No Brasil em 2004, a parcela de jovens de 18 a 24 anos no Ensino Superior era de 32,9% e cresceu para 58,5% em 2014 de acordo com o Instituto Brasileiro de Geografia e estatística- IBGE (2015) na Síntese de Indicadores Sociais (SIS) de 2015. Segundo dados do IBGE (2015), no Brasil, o sistema de Educação Superior ampliou-se expressivamente nos últimos anos, gerando mudanças significativas não somente na dinâmica de funcionamento das instituições de ensino superior (IES) mas também no perfil dos estudantes que nelas ingressam (Almeida, Marinho-Araujo, Amaral & Dias, 2012). Em termos de quantidade de estudantes, a Educação Superior brasileira era a maior da América Latina e a quinta maior do mundo em 2010, com 6,7 milhões de matrículas (UNESCO, 2011). O sistema chegou a registrar um total de 2.391 instituições (87,4% privadas e 12,6% públicas) e 7.305.977 alunos de graduação matriculados em cursos presenciais e a distância no ano de 2013. No atual cenário da Educação Superior brasileira, no censo de 2014, quanto ao gênero, a maior participação é feminina, que corresponde a 54,7% dos ingressantes, 55,5% dos estudantes matriculados e 59% dos concluintes (INEP, 2014). A idade modal dos estudantes matriculados é de 21 anos nos cursos de graduação presenciais e 32 anos nos cursos a distância. No Brasil em média há 2,5 alunos matriculados na rede privada para cada aluno matriculado na rede pública em cursos presenciais. Mais de 3,1 milhões de alunos ingressaram em cursos de educação superior de graduação. Desse total 82,4% em instituições privadas. O País tem projetos que facilitam o acesso de alunos e professores à educação superior, ajudando a melhorar a qualidade do ensino.

A partir desse quadro, encontra-se uma diversidade de contextos que, por sua vez, permitem aos estudantes diferentes experiências. Ingressar na universidade acarreta grandes e novos desafios afetivos, cognitivos e sociais. Os anos que os estudantes passam na universidade são importantes, tanto para o desenvolvimento pessoal, quanto para a formação profissional, refletindo no desenvolvimento da própria sociedade onde irão atuar quando graduados. Dada essa diversidade, é importante conhecer qual é o perfil universitário do século XXI. Como estabelece a UNESCO (2011), «os alunos não são o objeto da educação, mas sujeitos com direito a uma educação que potencie ao máximo o seu desenvolvimento como pessoas, e lhes permita inserir-se e influir na sociedade em que estão imersos» (p. 10). As exigências de qualificação profissional e de aprendizagem contínua (Jenschke, 2003; Soares, 2000), somadas à expansão e à democratização do acesso ao Ensino Superior (Ministério da Educação, 2007) têm estimulado o ingresso de um número cada vez maior de estudantes nas universidades. Como consequência disso, constata-se a heterogeneidade dos estudantes universitários em termos de idade, classe social, procedência geográfica e competências de estudo (Macedo, Trevisan, Trevisan & Macedo 2005; Soares, 2002; Zago, 2006). Ao mesmo tempo que se observa a ampliação do sistema de educação superior, verifica-se

a necessidade de apoio e orientação aos universitários, no decorrer de seus anos de formação, a fim de facilitar sua aprendizagem, sucesso escolar e desenvolvimento psicossocial (Almeida & Soares, 2004). Nesse sentido, surge o interesse em conhecer em maior profundidade as competências de estudo, dos alunos (Ferreira, Almeida, & Soares, 2001).

Os estudos de Nardelli *et al.*, (2013) afirmam que poucas vezes, durante a formulação do processo ensino-aprendizagem, dados relativos aos alunos têm sido considerados relevantes, porém a criação de um perfil do ingressante pode ajudar a universidade no estabelecimento de políticas que interfiram positivamente na vida do estudante, bem como proporcionar subsídios para uma melhor interação entre professor-aluno e para o processo de ensino-aprendizagem. Os estudos de Sarriera *et al.*, (2012) também apontaram para a importância da identificação das características, pessoais e institucionais, para a melhor integração dos alunos nos seus contextos universitários. Para o perfil de competências de estudo, Silva e Silva (2015) defendem que a criação de estratégias institucionais que favoreçam os hábitos de estudo dos estudantes é fundamental para melhorar o desempenho acadêmico no ES. Segundo Méndez (1999), os métodos de estudo são considerados como um utensílio vantajoso para aqueles estudantes que o utilizam, o qual, os pode ajudar a aperfeiçoar os seus resultados escolares, ao mesmo tempo que podem aprender e criar objetivos, a planificar as suas atividades e distribuir seu tempo. No mesmo seguimento, Competência de estudo, designado na literatura como Study Skills, tem sido objeto de específica atenção por parte de pesquisadores na área da Educação e na Psicologia Educacional (Almeida, 2002; Credé & Kuncel, 2008). Essas competências (de estudo) são definidas por Credé e Kuncel (2008) como a capacidade de apropriação das estratégias de estudo e método que o aluno tem, capacidade de gestão de tempo e de outros recursos, visando responder questões acadêmicas.

Para se compreender as competências de estudo no ensino superior, há que se colocar o estudante como centro do seu processo de aprendizagem e como ator principal. É nesse sentido que Zimmerman (2000; 2008) e Zimmerman e Martinez-Pons (1988), precursores da teoria da autorregulação, colocam como centrais os processos de autorregulação da aprendizagem, definindo-os como autodirecionados e de crenças pessoais que possibilitam ao estudante transformar competências cognitivas em competências de desempenho. Veiga Simão (2004) define a autorregulação da aprendizagem como o processo em que o aluno estabelece metas que se misturam com suas expectativas, desenvolve estratégias para alcançá-las, criando condições para que se efetive a aprendizagem. As estratégias vinculadas à autorregulação são planejamento, monitoramento e autoavaliação. (Zimmerman, 2000)

O planejamento das atividades refere-se à definição de objetivos e metas, alguns exemplos de estratégias de planejamento são: passar os olhos rapidamente no texto antes de ler, formular questões antes de ler um texto, fazer uma análise prévia das informações (Pintrich, 1999, 2004).

As estratégias de monitoramento alertam o estudante quanto a possíveis problemas relativos à sua atenção ou compreensão que podem ser reparados com alguma estratégia de regulação. Dentre as atividades de controle, pode-se destacar o monitoramento de atenção ao ler um texto por meio de questionamentos sobre o material para verificar a compreensão, além do monitoramento de velocidade de leitura a fim de ajustar o tempo disponível, de controlar os comportamentos, o ambiente físico e seus próprios processos internos. (Pintrich, 1999, 2004; Pintrich & *et al.*, 1993).

A autoavaliação consiste «no juízo de valor que o aluno faz do rendimento educativo que obteve» (Carrasco, 1989:22). Segundo Ferreira (2007:111), é «através da reflexão que o aluno modifica a sua ma-

neira de atuar, atribuindo um sentido ao seu processo de aprendizagem». Portanto, a regulação das atividades ou autoavaliação, ocorre após o desempenho, refere a um ajuste contínuo das ações cognitivas que se realizam em função de um controle prévio (Rojas, 2008; Schreiber, 2005).

Os estudantes do ensino superior desconhecem muitas vezes as regras de funcionamento das instituições que escolheram, desconhecem os cursos e o seu planeamento curricular, enfrentando novas disciplinas e uma diversidade de métodos de avaliação (Pinheiro, 2003). Na atual Sociedade da Informação e do Conhecimento é importante aceder, analisar e utilizar a informação de forma adequada. Para isso, segundo as diretrizes do Espaço Europeu de Educação Superior (EEES), é necessária uma série de competências e habilidades relacionadas com a pesquisa, avaliação, gestão, uso e difusão da informação (Lopes & Pinto, 2011; Pinto, 2010, 2011, 2012). As dificuldades dos estudantes universitários no uso e gestão de informação é uma realidade presente nos nossos dias. Acumulam muitos dados, mas não sabem como estruturar e interrelacionar o conhecimento. De acordo com Read (1999), para que um aluno da universidade seja bem-sucedido na aprendizagem académica, deve ser capaz de interpretar e compreender a informação escrita, de ter raciocínios abstratos, de refletir e de generalizar conceitos, de colocar questões e esclarecer dúvidas bem como expor conhecimentos oralmente, por escrito e de forma numérica (*cit. in* Santos, 2000). Cabe destaque que, se entende por competência de estudo como a capacidade do estudante em se apropriar de alguns recursos como, por exemplo, estratégias e métodos de estudo e capacidade de gestão de tempo, com o objetivo de realizar as tarefas académicas (Credé & Kuncel, 2008). Robbins *et al.*, (2004) acrescentam que as competências de estudo são comportamentos que estão diretamente relacionados com um desempenho produtivo e determinante do sucesso académico.

No dizer de Ferreira *et al.*, (2001:8) «a educação universitária deve promover o desenvolvimento de competências académicas, cognitivas e pessoais. Estas competências devem ser promovidas através de atividades curriculares e extracurriculares, tendo em vista a preparação dos alunos para a vida ativa, considerando as coordenadas histórico-socio-culturais e geográficas em que vivem». Ainda segundo Ferreira *et al.*, (2001: 8) a Universidade deve encarar o «sucesso académico dos seus estudantes para além dos resultados obtidos em cada disciplina, devendo tomar isso em consideração desde o 1º ano dos seus cursos». Ao reduzir o sucesso académico dos estudantes às suas classificações curriculares torna-se, muitas vezes, o desenvolver competências nos alunos em apenas reproduzir informação, o que conduz a enfatizar pouco a preparação dos estudantes para que possam, no futuro, se integrarem em ambientes profissionais e sociais. Zabalza (2004) destaca como função do Ensino Superior a qualificação de pessoas, integrando na formação superior as dimensões que o estudante pode desenvolver e regular como consequência das próprias experiências e aponta os seguintes caminhos formativos: a) aprender a aprender, b) aprender a fazer, c) aprender a ser, d) aprender a (con)viver. Na atualidade, a universidade é considerada uma *Multi-purpose University*, de acordo com os documentos da OCDE (1981, *cit. in* Oliveira, 2006), permitindo não só a reprodução de conhecimento, mas igualmente a criação, a difusão e a inovação.

Segundo Maestro (2014), os académicos deste século, nativos da geração Y (nascidos entre 1980 a 2000), têm mais facilidade em lidar com tecnologias, ditam as regras, têm ideias inusitadas, inovam, porém nada os satisfaz, são imediatistas e de pouca paciência. Nesse contexto estão os universitários de hoje: criativos e em constante busca pelo conhecimento, também característica do mercado de trabalho atual. Nunca se produziu tanto conhecimento como nesta era, e somente estarão no mercado aqueles que detêm de conhecimento, Maestro (2014) ressalta ainda as mudanças no mercado de trabalho, apontando a



forte ascensão na prestação de serviço e não mais na agricultura e nas fábricas, a exemplo do início do século passado, mostrando que no mundo pós-industrial o trabalho não é mais como obrigação opressora, mas, sobretudo como um prazer criativo e estimulante. Educar é um ato complexo, onde não se pode menosprezar o fato de que o aluno é o objeto de estudo, este também com todo o seu contexto, sua vivência e complexidade e diante de tais, a educação deve levar em consideração as dimensões deste ser em sala de aula. É no âmbito desta problemática que este estudo tem como objetivo geral analisar a relação entre o perfil de competências de estudos, dos estudantes universitários das duas universidades dos dois países (Brasil e Portugal), participantes no estudo.

## 2. Aspectos metodológicos

Desenvolveu-se um estudo comparativo, transversal, de abordagem quantitativa (Fortin *et al.*, 2013). A amostra incluiu 1240 alunos que frequentavam 10 cursos (Biologia, Ciências do Desporto, Comunicação e Multimédia, Educação Básica, Enfermagem, Engenharia Civil, Engenharia Eletrotécnica e de computadores, Engenharia Informática, Línguas, Literatura e Cultura e Psicologia) de uma Universidade de Portugal e uma Brasileira, que estavam presentes no momento de recolha de dados e aceitaram participar no estudo. Estabelecemos como critérios de seleção, ser aluno dos cursos selecionados e frequentar o 1º, 2º ou 3º ano dos cursos.

Para a recolha de dados foi aplicado um questionário de autopreenchimento, com questões abertas e fechadas, de escolha múltipla, com questões abertas construídas e validadas para o efeito, através de um teste piloto. A primeira parte inclui um conjunto de questões de caracterização sociodemográfica e a escala de Graffar (1956), e a última, integra a escala de competências de estudo de Almeida & Joly (2013) e questões acerca do estudo autónomo. A escala de competências de estudo pretende avaliar os métodos de estudo e as abordagens à aprendizagem dos estudantes no ensino superior. É composta por 16 itens do tipo Likert com seis pontos, de discordo totalmente (1 ponto) até concordo totalmente (6 pontos), agrupando-se em 3 fatores.

O fator 1, designado *Comportamentos Estratégicos de Planeamento* contém cinco itens que se referem a decisões prévias relativas à forma como o estudante organiza o seu estudo, integrando os itens 3, 7, 8, 9 e 14. O fator 2, designado *Comportamentos Estratégicos de Monitoração*, inclui cinco itens, objetiva realizar a automonitoração pela auto-observação do desempenho durante a execução de atividades de estudo é composto pelos itens 4, 6, 11, 15 e 16. Por último, o fator 3, designado *Comportamentos Estratégicos de Auto-avaliação* contém seis itens, reportando-se a comportamentos e a preocupações dos alunos no sentido de viabilizarem um estudo competente e bem sucedido, sendo por isso dependentes da existência de uma autoreflexão por parte do estudante acerca de seu desempenho, quer seja antes, durante ou depois de estudar, sendo composto pelos itens 1, 2, 5, 10, 12 e 13.

No dizer de Almeida & Joly (2013) a escala revela boas qualidades psicométricas. A análise fatorial exploratória revelou que os itens se agruparam nos três fatores revelando cargas fatoriais bastante altas que variaram de 0,49 a 0,79 e comunalidade acima de 35%. O fator 1 explica 17,7% da variância, o fator 2 explica 17,68% de variância e o fator 3 explica 17,9% da variância. A análise da precisão foi realizada por meio da consistência interna dos itens (alfa de Cronbach) para cada um dos três fatores. Constataram-se bons índices de precisão, sendo de 0,75 para o fator 1, 0,77 para o fator 2 e em 0,76 para o fator 3. Antes da aplicação do questionário foi realizado um pedido de autorização para a realização do estudo à comissão de ética das duas universidades, que nos deram o seu parecer favorável (nº 7/2016) Portugal e plata-

forma Brasil (nº 1.901.179). De seguida foi agendada por email a recolha de dados com os professores de cada um dos cursos. A investigadora deslocou-se às salas de aula, tendo informado os estudantes sobre o objetivo do estudo, e realizada a entrega e recolha dos questionários em sala de aula. Para o tratamento dos dados utilizou-se o software estatístico *Statistical Package for the Social Sciences* (SPSS), versão 22.0 para Windows, no qual foi construída uma base de dados e os mesmos foram editados. Recorremos à estatística descritiva, utilizando as frequências absolutas e relativas, a média e o desvio-padrão da idade e à estatística inferencial com os testes não paramétricos de  $\chi^2$ , Mann-Whitney e Kruskal-Wallis. O nível de significância adotado foi de 5% (Marôco, 2014).

### 3. Resultados e Discussão

O perfil sociodemográfico abrangeu as variáveis sexo, grupo etário, estado civil, coabitação em tempo de aulas, tipo de habitação, meio de deslocação para a universidade, manutenção financeira, nível socioeconómico (NSE), curso e ano frequentado pelo estudante. Do total da amostra (n=1240), sendo da universidade portuguesa (n= 533) e da universidade brasileira (n=707), a maioria dos estudantes era do sexo feminino (54.9%). A maior parcela dos alunos inquiridos enquadrava-se no grupo etário dos 19-20 anos (39.4%), o mínimo foi de 17 anos e o máximo 56 anos, ou seja, estamos perante estudantes relativamente jovens que ambicionam adquirir conhecimentos e que procuram uma melhor qualidade de vida, ou condição financeira. Sobre o estado civil 94.1% eram solteiros e 95.0% não tinham filhos. Salienta-se que 6.6% dos estudantes viviam sozinhos e quanto à coabitação dos alunos em tempo de aulas, verificamos que 47.5% residem com os amigos em tempo de aulas, numa casa/apartamento (46.3%), mantidos pela família através de mesada (41.8%), embora 41,2% asseguravam a sua manutenção financeira através de bolsa de estudo ou atividade académica. A maioria dos estudantes do presente estudo é pertencente à classe média alta (46.7%). No que diz respeito à deslocação dos estudantes para a universidade, a maioria deslocava-se de transporte público (autocarro/ônibus) (37.1%) e 36.5% deslocavam-se a pé.

### 4. Competências de estudo

Relativamente à questão «*Fora de sala de aula em média quantas horas faz de estudo diário?*», a grande maioria da amostra em estudo respondeu 1 hora de estudo (26,6%), logo seguido pelo grupo de estudantes que assinala estudar 2h/dia (24,4%). Na categoria de maior número de horas de estudo diário (Mais de 5h/dia), apenas se enquadravam 2,9% dos estudantes. Este cenário repete-se nas duas universidades, com o maior grupo de estudantes tanto na universidade brasileira (26,1%), como na universidade portuguesa (27,3%), a assinalar que em média estudam 1H/dia, sendo a proporção da categoria «Mais de 5h/dia», de novo a que obteve menores percentuais, 2,4% e 3,5%.

No estudo comparativo do tempo de estudo dos alunos entre as universidades dos dois países, o teste Mann-Whitney revelou diferenças estatísticas muito significativas  $p \leq 0,014$ , sendo que a média de ordenação da universidade brasileira (611,69) foi mais elevada do que a da universidade portuguesa (573,70), significando que os estudantes brasileiros estudam mais tempo fora da sala de aula. Na comparação das proporções de estudantes das duas universidades enquadrados nas diferentes categorias de tempo de estudo (opções de resposta) também se observaram diferenças estatísticas muito significativas ( $\chi^2$ :  $p \leq 0,007$ ), sendo que a proporção de estudantes da universidade brasileira que se enquadram na categoria de 3 horas é superior à da Universidade Portuguesa (21,2% versus 13,4%), ou seja os estudantes brasileiros estudam durante mais tempo.

Na questão *quantos livros leem por ano?* a maior percentagem dos alunos da Universidade Portuguesa são dos que leem 2 livros por ano (22%), embora em segundo lugar a maior percentagem é um dado alarmante, sendo dos alunos que não leem nenhum livro por ano (20,6) em relação aos alunos da universidade brasileira a maior percentagem são dos que leem de 4 a 6 livros por ano. No teste T-Student apareceram diferenças altamente significativas  $T = p \leq 0,000$ , onde na universidade portuguesa a média de ordenação foi de (3,77), e na universidade brasileira a média de ordenação foi de (6,28) demonstrando uma leitura maior.

Quando questionados sobre *Que tipo de livro lê?*

No tipo de livro de ficção, nos testes  $\chi^2 p \geq 0,352$  não houve diferenças estatísticas significativas.

No tipo de livro de autoajuda, houve diferenças estatísticas altamente significativas  $\chi^2 p \leq 0,000$ , onde na universidade brasileira aparece mais 9,4 casos que o esperado.

No tipo de livro de não ficção, houve também diferenças estatísticas altamente significativas  $\chi^2 p \leq 0,000$ , onde na universidade brasileira aparecem mais 7,2 casos que o esperado.

Os alunos que assinalaram que não leem *nenhum tipo de livro*, houve também diferenças estatísticas altamente significativas  $\chi^2 p \leq 0,000$ , onde na universidade portuguesa apareceram mais 6,65 casos que o esperado. Analisando a estatística descritiva o tipo de livro mais escolhido tanto na universidade portuguesa (60,2%) e brasileira (57,6%) foi o tipo de ficção.

Quanto às medidas de tendência central e de dispersão da pontuação do total da amostra (duas universidades) dos três fatores que constituem a ECE (Almeida & Joly, 2013), a maior média foi obtida pelo F3, que também obteve uma moda e um valor máximo mais elevados, significando que os estudantes investem mais nos comportamentos e nas preocupações no sentido de viabilizarem um estudo competente e bem-sucedido. Quando analisamos estas medidas em cada uma das universidades em estudo, observamos que quer a média, a mediana e a moda dos três fatores são mais elevados na universidade Brasileira, o que quer dizer que estes estudantes têm mais competências de estudo. No entanto essas diferenças não são estatisticamente significativas.

**Tabela I. Medidas de tendência central e de dispersão dos fatores da Escala de Competências de Estudo**

Universidade Brasileira					
F1	22,07±7,350	24,0	25,0	0	30
F2	21,80±7,243	23,0	25,0	0	30
F3	27,97±8,460	30,0	36,0	0	36
Universidade Portuguesa					
F1	21,86±5,995	23,0	20,0	0	30
F2	21,38±5,930	22,0	20,0	0	30
F3	27,53±6,902	29,0	30,0	0	36
Total da amostra					
F1	21,98±6,799	24,0	25,0	0	30
F2	21,62±6,711	23,0	25,0	0	30
F3	27,78±7,829	30,0	30,0	0	36

Fonte: Realização Própria

No que diz respeito à comparação entre as duas universidades em termos das competências de estudo, constatamos que as médias do fator *Comportamentos Estratégicos de Planeamento* das duas universidades não diferem estatisticamente (t de Student:  $p \geq$ ). O mesmo sucedeu no que se refere às médias do fator *Comportamentos Estratégicos de Monitoração* das duas universidades (t de Student:  $p \geq$ ). Repetiram-se os mesmos resultados relativamente ao fator *Comportamentos Estratégicos de Autoavaliação* (t de Student:  $p \geq$ ). Ou seja as competências de estudo dos estudantes Portugueses e dos estudantes Brasileiros não diferem significativamente, sendo bastante semelhantes entre si. Tais resultados estão na mesma direção dos estudos de Prates (2011) com sua amostra de 814 universitários, onde afirma que «leva à conclusão de que os universitários não apresentaram grande variabilidade em suas respostas». Estes resultados também corroboram com os estudos de Dias & Joly (2013) que verificaram que sua amostra de 653 universitários que de uma forma geral, declararam utilizar de comportamentos estratégicos de estudo, tanto com relação ao planeamento, ao monitoramento e autoavaliação, porém ainda assim afirmaram que «aventa-se a hipótese de que os universitários mesmo relatando a utilização de comportamentos estratégicos de autorregulação, podem não estar sabendo aplicá-las adequadamente quando encontram dificuldades em seu estudo ou problemas que precisam resolver para dar continuidade para atingir a aprendizagem de um conteúdo». (p.8).

Numa ótica geral entre as duas universidades e as 16 questões da escala, a questão mais pontuada (49,1%) foi *quando me aparece importante faço anotações nas aulas*. E a questão com mais discordância (4,7%) foi *Procuro anotar aspectos que não compreendo para depois pedir ajuda ou voltar a eles e superar as dificuldades*. Estes dados foram exatamente os mesmos nos estudos de Prates (2011), onde apresenta a questão menos pontuada e a mais pontuada.

Estudos realizados têm demonstrado a autorregulação como um aspecto decisivo para o desempenho melhor e conseqüente sucesso acadêmico dos estudantes universitários (Lindner & Harris, 1993; Pintrich, Smith, Garcia & Mckeachie, 1993). Porém, as pesquisas têm revelado que os estudantes ingressam na universidade com poucas competências para assumirem uma abordagem mais profunda na sua aprendizagem e para autorregular o seu estudo de forma eficaz (Almeida, Guisande, *et al.*, 2009; Pintrich & Zusho, 2002).

## 5. Considerações Finais

Tendo em conta os objetivos delineados para este estudo e analisando os resultados obtidos podemos retirar um conjunto de conclusões que passamos a descrever de seguida.

O perfil dos alunos que frequentam a universidade Portuguesa e que participaram neste estudo, pode considerar-se ser um estudante do sexo feminino, pertence ao grupo etário de 19 a 20 anos, coabita com amigos numa casa/apartamento mantido pela família, recebe mesada da família e pertence ao NSE médio alto, pelo que a universidade aparenta ser um espaço elitista. O perfil dos estudantes que frequentavam a Universidade Brasileira em estudo, pode considerar-se do sexo feminino, pertencente ao grupo etário de 18 e 19 anos, coabita com os colegas numa casa/apartamento mantido pela família, porém, sua manutenção financeira é proveniente de bolsa de estudos e/ou atividade académica e pertence ao NSE classe média.

Apesar dos estudantes Brasileiros despendem em média mais tempo em estudo fora da sala de aula, e lerem mais livros do que os estudantes Portugueses, concluímos que em relação às competências de estudo, quanto ao planeamento, monitoramento e autoavaliação, não se constataram diferenças significativas entre os estudantes das duas universidades, pelo que os seus comportamentos, em termos de estratégias de estudo são muito semelhantes, no tipo de livro de leitura os alunos das duas universidades prefeririam mais a ficção.

Os resultados obtidos poderão contribuir para que os docentes que lecionam, possam ficar a conhecer melhor os seus estudantes e adequar melhor as suas estratégias de ensino-aprendizagem, melhorando o processo de ensino aprendizagem, o sucesso académico dos estudantes, bem como o incentivo à leitura e estudo. Neste mundo global em que vivemos e com a proposta da internacionalização e mobilidade, este estudo oferece um recorte, uma breve visão das semelhanças e diferenças sociodemográficas e de competências de estudo entre dois países.

## Referências bibliográficas

- Almeida, Leandro (2002): “Facilitar a aprendizagem: ajudar os alunos a aprender e a pensar”. *Psicologia Escolar e educacional*, 6 (2), 155-165.
- Almeida, Leandro S. & Soares, Ana Paula (2004): “Os estudantes universitários: Sucesso escolar e desenvolvimento psicossocial” in Mercuri, E & Polydoro, S. (eds.): *Estudante universitário: Características e experiências de formação*. Taubaté: Cabral Editora e Livraria Universitária.
- Almeida, Leandro *et al.*, (2009): “Escala de competências de estudo (ECE-SUP): fundamentos e construção”. Actas do X Congresso Internacional Galego-Português de Psicopedagogia. Braga: CIED Editora.
- Almeida, Leandro S. *et al.*, (2012): “Democratização do acesso e do sucesso no Ensino Superior: Uma reflexão a partir das realidades de Portugal e do Brasil”. *Avaliação*, 17 (3), 899-920.
- Almeida, Leandro S. & Joly, M. C. R. A. (2013): “Escala de competências de estudo - ECE (S&H)”. Teste em construção. Universidade do Minho (Portugal)/Universidade de Brasília (Brasil).
- Carrasco, J. B. (1989). *Como avaliar a aprendizagem. Coleção Práticas Pedagógicas*. Porto: Edições Asa.
- Credé, Marcus & Kuncel, Nathan (2008): “Study Habits, Skills, and Attitudes: The Pillar Supporting Collegiate Academic Performance”. *Perspectives on Psychological Science*, 3 (6), 425-453.
- Dias, Anelise Silva & Joly, Maria Cristina Rodrigues Azevedo (2013): “Habilidades cognitivas para a competência de estudo”. Congresso Internacional Galego-Português de Psicopedagogia. Braga: CIED Editora.
- Ferreira, Joaquim Armando; Almeida, Leandro S. & Soares, Ana Paula (2001): “Adaptação Acadêmica em estudante do 1º ano: Diferenças de género, situação de estudante e curso”. *PsicoUSF*, 6 (1), 01-10.
- Ferreira, Carlos Alberto (2007). *A avaliação no Quotidiano da sala de aula*. Porto: Porto Editora.
- Fortin, Marie Fabienne; Côté, José & Filion, F. (2013). *Fundamentos e etapas do processo de investigação*. Loures: Lusodidacta.
- Graffar, M. (1956): “Une methode de classification sociale d'échantillons de population”. *Courrier*, 6 (8), 455-459.
- Instituto Brasileiro De Geografia e Estatística (2015). Disponível em: <http://www.ibge.gov.br/home>, acesso em 20 de dezembro 2016.
- Instituto Nacional de Estudos e Pesquisas Educacionais (2014). Censo da educação superior: Resumo técnico. Disponível em: [http://download.inep.gov.br/download/superior/censo/2012/resumo\\_tecnico\\_censo\\_educacao\\_superior\\_2012.pdf](http://download.inep.gov.br/download/superior/censo/2012/resumo_tecnico_censo_educacao_superior_2012.pdf), acesso em 18 dezembro 2016.

- Jenschke, B. (2003): “A cooperação internacional: Desafios e necessidades da orientação e do aconselhamento em face das mudanças mundiais no trabalho e na sociedade”. *Revista Brasileira de Orientação Profissional*, 4, 35-55.
- Lindner, Reinhard W. & Harris, Bruce (1993): “Self-regulated learning: its assessment and instructional implications”. *Educational Research Quarterly*, 16, 29-37.
- Lopes, Carlos Alberto & Pinto, Maria (2010): IL-HUMASS – Instrumento de avaliação de competências em literacia da informação: um estudo de adaptação à população portuguesa (Parte I) in 10º Congresso de Bibliotecários, Arquivistas e Documentalistas, Guimarães. Lisboa: BAD. Disponível em <http://repositorio.ispa.pt/handle/10400.12/200>, acesso em 3 janeiro 2017.
- Macedo, A. R., *et al.*, (2005): “Educação superior no século XXI e a reforma universitária brasileira”. *Ensaio: Avaliação de Políticas Públicas em Educação*, 13 (47), 127-148.
- Maestro, Valther (2014). Palestra trata do perfil universitário no século XXI. <http://uaumais.com.br/palestra-trata-do-perfil-universitario-no-seculo-xxi>, acesso em 8 novembro.
- Marôco, João (2014). *Análise estatística com o SPSS Statistics*. Pêro Pinheiro: Report Number.
- Méndez, F. (1999). *Aprender a estudar*. Madrid: Ediciones Pirámide.
- Ministério da Educação. (2007). O plano de desenvolvimento da educação: Razão, princípios e programa. Disponível em: <http://portal.mec.gov.br/arquivos/pdf/livromiolov4.pdf>, acesso em 5 Novembro 2016.
- Nardelli, Giovanna. Gaudenci *et al.*, (2013): “Perfil dos alunos ingressantes dos cursos da área da saúde de uma universidade federal”. *REAS*, 2 (1), 3-12.
- Oliveira, J. B. (2006): Aprendizagem auto-regulada, envolvimento e ganhos académicos em estudantes do ensino superior. Tese de doutoramento não publicada, Universidade de Aveiro. <https://ria.ua.pt/bitstream/10773/4752/1/2007001149.pdf>, acesso em 2 setembro 2016.
- Pinheiro, Maria do Rosário Manteigas e Moura (2003). Uma época especial: Suporte social e vivências académicas na transição e adaptação ao ensino superior. Tese de doutoramento não publicada, Faculdade de Psicologia e Ciências de Educação, Universidade de Coimbra. <http://hdl.handle.net/10316/988>, acesso em 3 novembro 2016.
- Pinto, M. (2010): “Design of the IL-HUMASS survey on information literacy in higher education: a selfassessment approach”. *Journal of Information Science*, 38 (1), 86-103.
- Pinto, Carlos Lopes Maria (2011): “An approach to the internal facet of information literacy using the IL-HUMASS survey: design of the IL-HUMASS survey”. *Journal of Academic Librarianship*, 37 (2), 145-154.
- Pinto, Carlos Lopes Maria (2012): “Information literacy perceptions and behaviour among history students”. *Aslib Proceedings*, 64 (3), 304-327.
- Pintrich, Paul R. *et al.*, (1993): “Reability and predictive validity of the motivated strategies for learning questionnaire (MSLQ)”. *Educational and Psychological Measurement*, 53, 801-813.
- Pintrich, Paul R. (1999): “The role of motivation in promoting and sustaining self-regulated learning”. *International Journal of Educational Research*, 459-470.

- Pintrich, Paul R. & Zusho, A. (2002): “A. The development of academic self-regulation: The role of cognitive and motivational factors” in Allan Wigfield & Jacquelynne Eccles (eds.): *Development of achievement motivation*. San Diego: Academic Press.
- Pintrich, P. R. (2004): “A conceptual framework for assessing motivation and selfregulated learning in college students”. *Educational Psychology Review*, 16 (4), 385-407.
- PORDATA Base de Dados Portugal contemporâneo. 2014. Disponível em: <http://www.pordata.pt/Subtema/Portugal/Alunos+Matriculados+do+Ensino+Superior-74>, acesso em 4 janeiro 2017.
- PORDATA Base de Dados Portugal contemporâneo. 2017. Disponível em: <http://www.pordata.pt/Subtema/Portugal/Alunos+Matriculados+do+Ensino+Superior-74>. Acesso em 23 janeiro 2017.
- Prates, Eli Andrade Rocha (2011): “Estudo de validade da escala de competência em estudo – ECE-SUP (S&H) pela correlação com a motivação de universitários”. Tese de mestrado. Itatiba: Universidade São Francisco.
- Robbins, S *et al.*, (2004): “Do Psychosocial and study skill factors predict college outcomes? A meta-analysis”. *Psychological Bulletin*, 130 (2), 261-288.
- Rojas, Héctor Lamas (2008): “Aprendizaje autorregulado, motivación y rendimiento académico”. *Libera-bit. Revista de Psicología*, 14, 15-20.
- Santos, Luísa (2000). Vivências académicas e rendimento escolar: Estudo com alunos universitários do 1º ano. Dissertação de mestrado não publicada, Universidade do Minho, Braga. <http://hdl.handle.net/1822/206>, acesso em 20 Dezembro 2016.
- Sarriera, Jorge Castellá, Paradiso, Ángela Carina, *et al.*, (2012): “Estudo comparativo da integração ao contexto universitário entre estudantes de diferentes instituições”. *Revista Brasileira de Orientação Profissional*, 13 (2), 163-172.
- Schreiber, F. J. (2005): “Metacognition and self-regulation in literacy” em S. E. Israel, C. C. Block, K. L. Bauserman & K. Kinnucan-Welsch (orgs.): *Metacognition in literacy learning: theory, assessment, instruction and professional development*. Philadelphia: Lawrence Erlbaum.
- Silva, Katia Regina Xavier & Silva, Douglas Bordalo (2015): “Estratégias de autorregulação da aprendizagem no curso de educação física do UNIABEU Centro Universitário”. *Gestão & Sociedade: Revista de Pós-Graduação da Uniabeu*, 4 (1), 56-71.
- Soares, D. H. P. (2000): “As diferentes abordagens em orientação profissional” in M. D. Lisboa & D. H. P. Soares (eds.): *Orientação profissional em ação: Formação e prática de orientadores*. São Paulo.
- Soares, M. (2002). “Novas práticas de leitura e escrita: letramento na cibercultura”. *Educação & Sociedade*, 23 (81), 143-160.
- UNESCO (2011). “Institute for Statistics. Global Education Digest: Comparing education statistics across the world”. Quebec: Author. Disponível em: <http://www.uis.unesco.org/Education/GED%20Documents%20C/GED-2011-Book-EN-web2.pdf>, acesso em 3 agosto 2016.

- Veiga Simão, Ana Margarida (2004): “O conhecimento estratégico e a auto-regulação da aprendizagem. Implicações em contexto escolar” in Adelina Lopes da Silva, António Manuel Duarte, Isabel Antunes de Sá & Ana M. Veiga Simão: *Aprendizagem autorregulada pelo estudante: perspectivas psicológicas e educacionais*. Porto: Porto Editora.
- Zago, N. (2006): “Do acceso a permanência no ensino superior: percursos de estudantes universitários de camadas populares”. *Revista Brasileira de Educação*, 11 (32), 226-237.
- Zabalza, Miguel (2004). *O ensino universitário: Seu cenário e seus protagonistas*. Porto Alegre: Artmed.
- Zimmerman, Barry (2000): “Attainment of self-regulation: a social cognitive perspective” in Monique Boekaerts, Moshe Zeidner, Paul R Pintrich: *Handbook of self-regulation*. San Diego: Academic Press.
- Zimmerman, Barry & Martinez-Pons, Manuel (1988). “Construct validation of a strategy model of student self-regulated learning”. *Journal of Educational Psychology*, 80, 284-290.
- Zimmerman, Barry (2008). “Investigating self-regulation and motivation: historical background, methodological developments, and future projects”. *American Educational research Journal*, 45 (1), 166-183.

## Notas biográficas

**Rubia Salheb Fonseca** es doctoranda en Ciencias de la Educación en la Universidade de Trás-Os-Montes e Alto Douro (UTAD), en Portugal. Máster en Ciencias de la Educación. Pedagoga y psicopedagoga. Miembro del grupo de Investigación del Centro de Investigación en Estudos da Criança (CIEC) en la Universidade do Minho.

**Joaquim José Jacinto Escola** es Doctor en Educación. Profesor en el Departamento de Educação e Psicologia de la Universidade de Trás-Os-Montes e Alto Douro (UTAD), en Portugal. Miembro del Grupo de Investigación reconocido en la Universidad de Salamanca «Educación Comparada y Políticas Educativas (ECPES)» y del grupo de investigación IF de la Universidade do Porto.

**Amâncio Carvalho** es Doctor en Estudios del Niño, con especialización en Salud Infantil. Profesor en la Escuela de Salud de la Universidade de Trás-os-Montes e Alto Douro (UTAD), en Portugal. Pertenece al Centro de Investigación en Estudios del Niño (CIEC).

**Armando Loureiro** es Doctor en Educación. Profesor en el Departamento de Educação e Psicologia la Universidade de Trás-Os-Montes e Alto Douro (UTAD), en Portugal. Pertenece al grupo de investigación KIDE- Conhecimento, Inovação e Diversidades em Educação de la Universidade do Porto.



# The Performance Evaluation of Polytechnic Higher Education Institutions: Tensions and Reconfigurations Resulting from the Adoption of Specific Metrics

La Evaluación del desempeño de las instituciones de educación superior politécnica: tensiones y reconfiguraciones resultantes de la adopción de métricas específicas

Sandra Isabel Gonçalves Saúde, Sandra Cristina Santana Lopes, Carlos Manuel Lopes Borralho e Isidro Lourenço Rodrigues Góis Féria<sup>1</sup>

## Abstract

Aligned with the binary structure of the higher education system in Portugal, divided into universities and polytechnics, the Government acknowledged, in 2014, the need to identify distinct performance indicators/metrics aiming at a more accurate assessment of the effect and quality of the action developed by Higher Education Institutions (HEIs). For the Polytechnic Higher Education Institutions (PHEIs) it was necessary to define «(...) performance indicators that address the production, transfer and diffusion of knowledge» (FCT, 2014:1), aiming to measure «(...) the applied research and the cultural production activities as well as their impact on the region where they [PHEIs] are located (...)» (FCT, 2014:1).

Founded on the presentation of specific indicators to be considered when evaluating the performance of PHEIs (developed under a research project funded by FCT: Fundação para a Ciência e a Tecnologia - portuguese national funding agency for science, research and technology), we now focus our reflection on the tensions, reconfigurations and gains for the organizational and educational management that these new metrics can induce.

By successive aggregation, the proposed evaluation indicators allow for the typification and characterization of the separate performance (faculty member/Department/HEI) at the level of applied research, knowledge transfer, population qualification and contribution to the territorial dynamics, which account for the distinct profile of Polytechnics.

## Keywords

Performance Evaluation, higher education, performance evaluation indicators/metrics.

## Resumen

En línea con la estructura binaria existente en el sistema de educación superior en Portugal, dividido en universidades y escuelas superiores politécnicas, fue identificada, en 2014 por el Ministerio de Educación Superior, la necesidad de identificar los diferentes indicadores/métricas de desempeño que pueden evaluar mejor el efecto y la calidad de la acción desarrollada. En concreto, para las Instituciones de Educación Superior Politécnica (IESP) fue asumido como necesario definir «(...) indicadores de desempeño para la producción, la transferencia y difusión del conocimiento» (FCT, 2014:1), es decir, que permitan evaluar «das actividades de investigación aplicada, de creación cultural y su impacto en las regiones en las que operan (...)» (FCT 2014:1).

Teniendo en cuenta la presentación de indicadores específicos que deben ser considerados en la evaluación de desempeño de las IESP, desarrollados en el ámbito de un proyecto de investigación financiado por la FCT (Fundación para la Ciencia y Tecnología: agencia portuguesa de financiación nacional de la ciencia, investigación y tecnología), centramos, en este artículo, la reflexión sobre el valor añadido y las tensiones que la introducción de nuevas métricas de desempeño, para docentes e instituciones, pueden inducir en la dinámica educativa y organizacional. Los indicadores de evaluación propuestos permiten tipificar y caracterizar la *performance* de la institución de educación superior en las áreas de la investigación aplicada, de la transferencia de conocimiento y de la contribución para la dinámica socioeconómica del territorio en el que opera.

## Palabras clave

Evaluación del desempeño; educación superior; indicadores/métricas de evaluación del desempeño.

Recibido: 14-02-2017

Aceptado: 02-05-2017

<sup>1</sup> Interdisciplinary Centre of Social Sciences, University of Évora - CICS.NOVA.UÉvora, Polytechnic Institute of Beja, Portugal, Sandra Isabel Gonçalves Saúde, [ssaude@ipbeja.pt](mailto:ssaude@ipbeja.pt); Sandra Cristina Santana Lopes, [slopes@ipbeja.pt](mailto:slopes@ipbeja.pt); Polytechnic Institute of Beja, Carlos Manuel Lopes Borralho, [cborralho@ipbeja.pt](mailto:cborralho@ipbeja.pt); Isidro Lourenço Rodrigues Góis Féria, [iferia@ipbeja.pt](mailto:iferia@ipbeja.pt).

## 1. Introduction

The project Performance Indicators for Polytechnic Higher Education Institutions we now present was carried out between October 2014 and February 2016, funded by FCT – Fundação para a Ciência e a Tecnologia (Portuguese national funding agency for science, research and technology) as a result of a program launched in 2014 by the Portuguese Ministry of Education and Science. It was developed as a partnership between researchers of the Polytechnic Institute of Beja, the University of Aveiro and the Centre for Research in Higher Education Policies (CIPES - Centro de Investigação em Políticas do Ensino Superior). This study involves the validation and consensualisation of performance evaluation indicators organized around five essential dimensions of the mission of Higher Education Institutions (HEIs), specifically the Polytechnics (PHEI):

- a) services provided,
- b) knowledge transfer,
- c) scientific and artistic production,
- d) collaborative research,
- e) societal impact.

The validation of the evaluation indicators for each dimension included the hearing of the heads of PHEIs, through e-Delphi, in a first moment, then ratified by case studies, in a second phase. In this paper we will share the main results of this process.

## 2. Performance evaluation of Higher Education Institutions

### 2.1. Performance evaluation of Higher Education Institutions in Portugal – recent trends

Reflection around performance indicators for Higher Education Institutions (HEIs) in Portugal starts with the discussion and publication of the revised formula for Higher Education funding in Law 37/2003. This Law entails the termination of the standard budget concept (existing since 1997), by establishing that the financing of HEIs should be based on a reference budget, calculated by a funding formula covering objective criteria of quality and excellence, standard values and performance indicators (Law 37/2003).

Since the relation between performance and funding has not been consolidated through time, the reflection on the importance and viability of estimating the outputs generated by HEIs was buried in doubts of a political nature in what concerns the real interest and practicability of this process.

It has been hard indeed to assume, at the political level, that it is necessary to substitute a paradigm that associates public funding with institutional history and input indicators (namely the number of students) with one that focuses on performance indicators and the outputs generated by the institution. Besides the reflection around the most suitable indicators, there is also the absence of a standardized methodology to collect the data on some domains of the activity of HEIs and the need for greater clarity on the part of the Ministry about a framework of principles, objectives and purposes for the performance evaluation of HEIs.

At the moment, the performance evaluation of HEIs is mainly evident in the aims and scope of the Agency for Evaluation and Accreditation of Higher Education (A3ES – Portuguese acronym),

which, for purposes of evaluation and accreditation of the educational offer of HEIs, evaluates a set of performance indicators particularly associated with pedagogical expertise and the capacity of scientific production versus knowledge transfer, within the domain of the course under scrutiny. Over the past four years, while implementing the audit and certification of internal systems which guarantee the quality of HEIs, the A3ES has given a valuable contribution to consolidate a culture of evaluation, demonstrating that the path to implement and collect a systematic set of performance indicators is under way, in Portugal.

## 2.2. Polytechnic Higher Education and its performance evaluation metrics

Based on international studies and evaluation programs, developed by the OECD<sup>2</sup> and ENQA<sup>3</sup>, in 2006, the Portuguese government in effect, aware of the need to boost a reform of the sector, issued a new Legal Regime for Institutions of Higher Education (RJIES – Portuguese acronym). Law 62/2007 confirmed the binary structure of Portuguese higher education, divided into the university and polytechnic subsystems.

Under this legal framework, each of the subsystems must comprehend specific ontological codes and curricula. This way, university education must aim at «the provision of sound academic training, combining the efforts and skills of both teaching and research units» (RJIES, 2007, article 6<sup>th</sup>: 6359), while polytechnic education, essentially practical and experimental in nature, should especially concentrate on «vocational and advanced technical training, both professionally-oriented» (RJIES, 2007, article 7<sup>th</sup>: 6359).

This legal framework is reinforced by the latest manifestations of both the Parliament and the two last governments, evidenced in recent legal and political publications. Parliament Resolution 77/2013 urges the Government to: «1. Clearly strengthen the binary model, clarifying the specific missions of university and polytechnic higher education; 2. Promote the regional articulation of synergies within the scope of the training provided, thus discouraging the mimetism between subsystems (...)» (Resolução da Assembleia da República 77/2013:3234). The State Reform Guide, published on October 30<sup>th</sup> 2013, on the other hand, establishes the need for «(...) a reform of Higher Education, aiming at improving the quality of university education and consolidating the polytechnic educational offer (...)» (Guião da Reforma do Estado, 2013:76).

More recently, the current Minister for Science, Technology and Higher Education announced his interest in «(...) differentiating the evaluation of universities and polytechnic institutions» (MCTES, June 8<sup>th</sup> 2016: 6) and highlighted «(...) the need to increase the distinction between universities and polytechnics, from aims and contents to their individual missions», in a debate promoted among all the actors of higher education (Online Newspaper *Observador*, July 18<sup>th</sup> 2016).

It is consensual that to evaluate the performance of PHEIs, the simple application of traditional indicators is far from enough. Indicators such as the number of papers produced, the number of researchers belonging to research centres (in the Portuguese situation, those accredited by the FCT) or the number of citations in scientific papers by teacher/researcher, generally used to evaluate teaching staff performance in the university system, are rather insufficient and even inappropriate in what polytechnic higher education is concerned.

---

2 Organisation for Economic Co-operation and Development.

3 European Association for Quality Assurance in Higher Education.

It seems therefore important to contribute to the adjustment of evaluation metrics to the characteristics of the outputs generated by PHEIs. This was the key objective of the work developed within this project, namely: to identify and develop performance indicators for the production, transfer and diffusion of knowledge of PHEIs.

### 3. The Project

#### 3.1. Methodology

The challenge to identify performance indicators for PHEIs implied the development of a process which occurred in five stages (see Table 1), from the selection and construction of more adjusted indicators, according to the existing scientific literature and the suggestions of the key informants consulted, to, at a final stage, their validation and applicability test.

In *Stage I* the team collected existing indicators for the evaluation of the performance of PHEIs, with special focus on applied research, cultural production and societal impact. Several reference works, produced both in the national and international contexts, were consulted and analysed, including:

- i) the indicators applied by the Agency for Evaluation and Accreditation of Higher Education (A3ES) when evaluating and accrediting the courses,
- ii) projects *U-Map e U-Multirank*, funded by the European Union, and project *E3M – European Indicators and Ranking Methodology for University Third Mission*,
- iii) the work developed by CHEPS – *Measuring performance of applied R&D. A study into performance of applied R&D in the Netherlands and some other countries*, and by UASnet – *The EDUPROF project: developing indicators of applied research. Final Report*. Full list of works consulted to be found in footnote<sup>4</sup>.

In addition, we carried out a review of the literature on higher education, focusing on the coexistence of two subsystems – typical of some European countries, including Portugal – and the extent to which the differentiation of missions can be depicted by the preponderance of different research modalities associated with each of the subsystems: fundamental research in the university subsystem and applied research in the polytechnic context.

---

4 Bornman, Lutz (2012). «Measuring the societal impact of research. Research is less and less assessed on scientific impact alone – we should aim to quantify the increasingly important contributions of science to society». *EMBO reports*, 13, 673- 676. DOI 10.1038/embor.2012.99

Deen, Jarno y Vossensteyn, Hans (2006). *Measuring performance of applied R&D. A study into performance of applied R&D in the Netherlands and some other countries*. Enschede: Center for Higher Education Policy Studies (CHEPS)/Universiteit Twente. Retrieved from <https://www.utwente.nl/bms/cheps/publications/publications%202006/engreport06measuringperformance.pdf> on 18<sup>th</sup> September 2016.

Finne, Håkon *et al.*, (2011). *A composite indicator for knowledge transfer. Report from the European Commission's expert group on knowledge transfer indicators*. Luxembourg: European Union Publications. Retrieved from <https://ec.europa.eu/research/innovation-union/pdf/kti-report-final.pdf> on 18<sup>th</sup> September 2016.

Healy, Adrian *et al.*, (2014). *Measuring the impact of university-business cooperation. Final Report*. Luxembourg: European Union Publications. Retrieved from [http://ec.europa.eu/dgs/education\\_culture/repository/education/library/study/2014/university-business\\_en.pdf](http://ec.europa.eu/dgs/education_culture/repository/education/library/study/2014/university-business_en.pdf) on 18<sup>th</sup> September 2016.

REF 2014 (2011). *Assessment Framework and guidance on submissions*. Retrieved from <http://www.ref.ac.uk/media/ref/content/pub/assessmentframeworkandguidanceonsubmissions/GOS%20including%20addendum.pdf> on 18<sup>th</sup> September 2016.

UASnet (2011). *The EDUPROF project: developing indicators of applied research. Final Report*. Retrieved from [http://www.scienceguide.nl/media/700624/eduprof\\_report\\_november\\_2011.pdf](http://www.scienceguide.nl/media/700624/eduprof_report_november_2011.pdf) on 18<sup>th</sup> September 2016.

**Table I. Project stages, techniques and objectives by stage**

STAGES	TECHNIQUES	OBJECTIVES
Stage I	Documental analysis	<ul style="list-style-type: none"> <li>(i) to analyse and reflect upon the state of the art on performance evaluation of HEIs</li> <li>(ii) to ensure that the indicators to be constructed relate to existing indicators used by PHEIs</li> <li>(iii) to draft the theoretical framework supporting the composition of the indicators to propose</li> </ul>
Stage II	Semi-structured interviews and Focus Group	<ul style="list-style-type: none"> <li>(i) to listen to the positions of the PHEI representatives on the proposed dimensions and performance indicators, according to the specific characteristics of their institutions</li> <li>(ii) to promote the discussion between the representatives of the various institutions in order to assess the most appropriate way to measure institutional performance regarding the proposed dimensions</li> <li>(iii) to discuss, justify and ratify, with the project consultants, the indicators previously identified through documental analysis</li> </ul>
Stage III	Triangulation of data resulting from Stages I and II	<ul style="list-style-type: none"> <li>(i) to analyse and gather consensus on the final list of indicators to be validated with the PHEIs, through the e-Delphi method</li> </ul>
Stage IV	e-Delphi	<ul style="list-style-type: none"> <li>(i) to validate the final version of the analysis dimensions and respective indicators, with the heads of the PHEIs, both public and private</li> </ul>
Stage V	Direct observation and Structured interviews applied to selected cases	<ul style="list-style-type: none"> <li>(i) to test the feasibility and applicability of the final list of validated indicators on which consensus was reached, by applying the e-Delphi method, in 4 PHEIs selected according to particular key characteristics, namely: size, geographic location, mono or multidisciplinary profile, and area covered</li> </ul>

Source: the authors.

In *Stage II*, various stakeholders of the polytechnic higher education system were surveyed, through a set of semi-structured interviews, first with the state secretary, on the scenarios for the use of indicators, and, later, with heads of some PHEIs, trying to validate, together with them, a proposal of dimensions and indicators which made it possible to estimate applied research, cultural production and societal impact. With this latter group of interviewees we tried to include some diversity and randomness aiming to obtain representativity of the sample, regarding the size, geographic location and organizational characteristics (public or private) of the PHEIs. In effect semi-structured interviews were applied to six heads of polytechnic institutes and schools, both public and private.

After the interviews, a workshop was held to which the heads of all Portuguese PHEIs were invited (of the public and private sectors, both polytechnic institutions and higher schools integrated in universities). The objective of this meeting was, once again, to promote a discussion among the representatives of the various institutions in order to assess the most accurate way of measuring the performance of PHEIs in applied research, cultural production and societal impact activities. The workshop was attended by 18 representatives of 15 PHEIs.

Within this second stage, three thematic meetings took place, focusing on the performance evaluation of HEIs, with a team of consultants from the University of Alcalá de Henares and the University of Lleida, specially chosen for this purpose for their recognized experience in HEI administration and the development of indicators to evaluate the effects generated by HEIs. During these meetings, various indicators were discussed, theoretically justified and chosen, both deductively and inductively. By resorting to critical analysis and consensus decision-making, it was possible to construct and reconstruct a substantiated provisional list of indicators adjusted to our purpose.

In *Stage III* it was necessary to reach consensus on the performance measurement dimensions and indicators. It was relatively easy to find consensus around a final list of 29 indicators, which would eventually be submitted to validation through the e-Delphi method.

The e-Delphi technique was used in *Stage IV*, with the objective of reaching consensus and a decision on the final validation of the performance indicators, among the heads of the PHEIs, belonging to the polytechnic subsystem, both public and private. This technique is a qualitative research tool that seeks to gather consensus from a panel of experts on a theme or future event. Such consensus is based on the principle that the intuitive judgment from a structured group of experts, previously selected, is far more accurate than that of an individual or an unstructured group with no expertise (Hsu y Sandford, 2007).

Finally, in *Stage V*, aiming to test the feasibility and applicability of the list of consensualized and validated indicators through the e-Delphi technique, four PHEIs were selected to carry out a set of exploratory studies. The selection aimed to guarantee the diversity of institutions (public, private, polytechnic institutions, higher schools integrated in universities and non-integrated higher schools, mono and multi-disciplinary, located in different regions and with different dimensions in terms of number of students). The interaction with the PHEIs (through direct observation and semi-structured interviews) provided global and specific assessment for each indicator concerning the difficulty to collect the data, the applicability of the calculation formula, its feasibility, as well as the identification of specific resistance against the calculation of the final indicators).

### 3.2. The results: final list of indicators obtained through e-Delphi

To collect the opinion of the heads of the PHEIs on the performance indicators proposed, a questionnaire survey was designed, following the principles and procedures of the e-Delphi technique. The 29 performance evaluation indicators initially identified were organized around 5 dimensions within the range of activities of polytechnic institutions. The evaluation of each indicator used a likert scale of 6 points (from Totally inadequate to Totally adequate) supported by six analysis criteria, namely:

- Relevance: the degree to which the indicator is considered adequate to assess the performance of a PHEI.
- Clarity: degree to which it may be easily understood by everyone.
- Measurability: degree to which it can be measured.
- Traceability: degree to which it is considered to be auditable.
- Aggregation/Disaggregation: degree to which it is considered to be subject to aggregation or disaggregation from the individual level (faculty) to that of the PHEI.
- Efficiency of data collection: how easy it is to collect and calculate the data in terms of time and resources used.

The first round of questionnaires obtained 33 valid answers: 16 from public PHEIs and 17 from private ones, corresponding to a return rate of 48.5%.

The second round focussed on 25 indicators and was addressed only to the heads of the PHEIs who had answered in the first round. In this second round a total of 25 valid answers was obtained, 13 from public institutions and 12 from private ones. This corresponds to a percentage of 72.7.

The list of final indicators that obtained greater consensus among the heads of the PHEIs is as follows:

**Table II. Final List of Performance Indicators**

**DIMENSION 1: PROVISION OF SERVICES**

**Indicator 1: Percentage of income generated from Services provided**

**Objective:** To measure the relative weight of the income generated by the PHEI from activities related to consultancy, provision of services and professional development courses or specialization programs not integrated in undergraduate or master's degrees or CTSP (short-cycle tertiary education).

**Calculation Formula**

[Revenue from Services provided by the PHEI, including continuous education programs]/[Operating budget of the PHEI]

(on 31/12 of N-1)

**DIMENSION 2: SCIENTIFIC AND ARTISTIC PRODUCTION**

**Indicator 2: Percentage of income generated by Research**

**Objective:** To measure the relative weight of the revenue generated by the PHEI from Research.

**Calculation Formula**

[Revenue obtained by the PHEI from Research]/[Operating budget of the PHEI]

(on 31/12 of N-1)

**Indicator 3: Research effort rate by full-time faculty member**

**Objective:** To measure the dedication and involvement of the teaching staff regarding research activities.

**Calculation Formula**

[Total % of time devoted to research by the PHEI faculty staff]/[Number of full-time faculty staff of the PHEI]

(on 31/12 of N-1)

**Indicator 4: Peer-reviewed scientific production by full-time faculty member**

**Objective:** To measure the scientific production by full-time faculty member.

**Calculation Formula**

[Average number (in the last 3 years) of peer-reviewed publications indexed to SCOPUS, in which at least one author is a member of the PHEI and the other authors possess an academic institutional address]/[Average number of full-time faculty staff in the last 3 years]

(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 5: Artistic production by full-time faculty member**

**Objective:** To measure the artistic production by full-time faculty member.

**Calculation Formula**

[Average number (in the last 3 years) of artistic outputs, in which at least one author is a member of the PHEI and the other authors possess an academic institutional address]/[Average number of full-time faculty staff in the last 3 years]

(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 6: Impact of the scientific production by full-time faculty member**

**Objective:** To measure the impact of the scientific production by full-time faculty member.

**Calculation Formula**

[Average number (in the last 3 years) of citations from publications indexed to SCOPUS, in which at least one author is a member of the PHEI and the other authors possess an academic institutional address]/[Average number of full-time faculty staff in the last 3 years]

(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 7: Revenue for academic research by full-time faculty member**

**Objective:** To estimate the global amount of funding obtained from research projects approved by national and international agencies, non-profit institutions, foundations, among other, weighted by the number of full-time academic staff.

**Calculation Formula**

[Revenue from research carried out by the PHEI]/[Number of full-time faculty staff of the PHEI]

(on 31/12 of N-1)

### DIMENSION 3: COLLABORATIVE RESEARCH

#### Indicator 8: Percentage of the income devoted to collaborative research

**Objective:** To quantify the research developed in collaboration with companies or other non-academic organizations, public or private.

**Calculation Formula**

[Income of the PHEI from research projects resulting from partnerships with external non-academic entities]/[Revenue from the PHEI research]  
(on 31/12 of N-1)

#### Indicator 9: Co-patents per PHEI

**Objective:** To measure the capacity to produce knowledge with commercial application potential, in partnership with external non-academic entities.

**Calculation Formula**

[Average number (in the last 3 years) of patents of the PHEI, in which at least one of the holders belongs to an external non-academic organization]/[Average number of full-time faculty staff in the last 3 years]  
(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

#### Indicator 10: Scientific production with non-academic partners by full-time faculty member

**Objective:** To measure the scientific production in collaboration with external non-academic organizations.

**Calculation Formula**

[Average number (in the last 3 years) of academic publications indexed to SCOPUS, by the PHEI academic staff, in which at least one author has a non-academic institutional address]/[Average number of the PHEI full-time faculty staff in the last 3 years]  
(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

#### Indicator 11: Artistic Production with external non-academic entities by full-time faculty member

**Objective:** To measure the artistic production in collaboration with external non-academic entities.

**Calculation Formula**

[Average number (in the last 3 years) of artistic outputs of the PHEI with external non-academic entities]/[Average number of full-time faculty staff of the PHEI in the last 3 years]  
(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

#### Indicator 12: Global funding allocated to R&D for collaborative research by full-time faculty member

**Objective:** To measure the funding or co-funding amounts allocated to research projects with companies or other non-academic organizations, public or private, by full-time faculty member.

**Calculation Formula**

[Revenue from collaborative research developed by the PHEI]/[Number of full-time faculty staff of the PHEI]  
(on 31/12 of N-1)

#### Indicator 13: Post-graduate students involved in collaborative research carried out with non-academic partners

**Objective:** To measure the capacity of the PHEI to promote research developed by the students in companies or other non-academic organizations, public or private.

**Calculation Formula**

[Number of post-graduate students of the PHEI co-funded and/or co-supervised by non-academic organizations]/[Number of post-graduate students of the PHEI]  
(on 31/12 of N-1)

### DIMENSION 4: KNOWLEDGE TRANSFER

#### Indicator 14: Patents by PHEI

**Objective:** To measure the capacity to produce knowledge with commercial application potential.

**Calculation Formula**

[Average number (in the last 3 years) of patents of the PHEI, in which all the holders belong to academic institutions]/[Average number of full-time faculty staff in the last 3 years]  
(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)

#### Indicator 15: Revenue from royalties and licensing agreements

**Objective:** To estimate the PHEI capacity to generate income from knowledge transfer activities.

**Calculation Formula**

[Average income (in the last 3 years) from royalties and intellectual property licensing agreements of the PHEI]/[Average number of full-time faculty staff in the last 3 years]  
(average in the last 3 years calculated from 31/12 of N-3 to 31/12 of N-1)



**DIMENSION 5: SOCIETAL IMPACT**  
(indexed to the impact generated in the regions where the PHEI is located)

**Indicator 16: Contribution to the qualification of the active population of the NUTS (Nomenclature of Territorial Units for Statistics) 3<sup>5</sup>**

**Objective:** To measure the contribution of the PHEI to the qualification of the active population of the region where it is located (NUTS 3).

**Calculation Formula** 
$$\frac{\text{[Number of working students enrolled in the PHEI]}}{\text{[Working-age population with no higher education qualifications of the NUTS 3 where the PHEI is located]}}$$
  
(numerator on 31/12 of N-1 and denominator on the last census)

**Indicator 18: Contribution to the qualification of the NUTS 3 resident population**

**Objective:** To measure the contribution of the PHEI to meet one of the targets of Europe 2020 strategy at the regional level: 40% of the population aged 30-34 having a higher education diploma.

**Calculation Formula** 
$$\frac{\text{[Average number (in the last 3 years) of graduates of the PHEI, aged 30-34 residing in the NUTS 3]}}{\text{[Average number (in the last 3 years) of all residents of the NUTS 3, aged 30-34]}}$$
  
(Average numbers in the last 3 years, calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 19: Capacity to attract a young population to the NUTS 3**

**Objective:** To measure the capacity of the PHEI to attract a young population to the region where it is located (NUTS 3).

**Calculation Formula** 
$$\frac{\text{[Number of students of the PHEI aged 20-24 coming from outside the NUTS 3]}}{\text{[Number of residents in the NUTS 3 aged 20-24]}}$$
  
(on 31/12 of N-1)

**Indicator 20: Capacity to renovate / requalify the workforce of the NUTS 3**

**Objective:** To measure the contribution of the PHEI to the renovation/requalification of the workforce of the region where it is located.

**Calculation Formula** 
$$\frac{\text{[Average number (in the last 3 years) of graduates of the PHEI (bachelor's and master's graduates), employed in the NUTS 3]}}{\text{[Average number (in the last 3 years) of the employed population of the NUTS 3]}}$$
  
(Average numbers in the last 3 years, calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 21: Capacity to retain the NUTS 3 human capital**

**Objective:** To measure the contribution of the PHEI to retain its graduates in the region where it is located (NUTS 3).

**Calculation Formula** 
$$\frac{\text{[Average number (in the last 3 years) of graduates of the PHEI (bachelor's and master's graduates), employed in the NUTS 3]}}{\text{[Average number (in the last 3 years) of all graduates of the PHEI (bachelor's and master's graduates)]}}$$
  
(Average numbers in the last 3 years, calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 22: Business practices and job placement of graduates**

**Objective:** To measure the adjustment between the training provided by the PHEI and the labor market demands of the region where it is located.

**Calculation Formula** 
$$\frac{\text{[Average number (in the last 3 years) of academic internships of the PHEI undergraduates which led to professional internships and/or actual job placements]}}{\text{[Average number (in the last 3 years) of academic internships of the PHEI undergraduates]}}$$
  
(Average numbers in the last 3 years, calculated from 31/12 of N-3 to 31/12 of N-1)

**Indicator 23: Contribution to social dynamics**

**Objective:** To measure the contribution of the PHEI to the regional social dynamics.

**Calculation Formula** 
$$\frac{\text{[Number of PHEI faculty who participate in scientific committees, juries of local awards, local and regional councils or as members of governing bodies of local institutions]}}{\text{[Number of full-time faculty staff]}}$$
  
(on 31/12 of N-1)

Source: the authors

### 3.3. Applicability test on the list of indicators

In the last stage of this project it was fundamental to apply a complementary test on the applicability of the list of performance indicators with 4 PHEIs. The objective of this step was to assess the practicality of the calculations and to obtain some final considerations on the relevance of the final proposal.

<sup>5</sup> Specific information at: <http://ec.europa.eu/eurostat/web/nuts/overview>

The PHEIs were selected according to certain differentiating characteristics, as follows:

- the sector to which they belong – public or private
- geographic location – dispersed across the country, located in large or small urban areas
- training profile – mono or multidisciplinary
- size of the academic community – students, teaching and non-teaching staff
- organizational structures with different degrees of complexity

The heads of the PHEIs were surveyed through semi-structured interviews on: (i) global and specific assessment for each indicator; (ii) main difficulties felt when applying the calculation formulae; (iii) possible difficulties to collect the data; (iv) challenges the model presents as a performance evaluation tool for PHEIs, as well as a management support mechanism. The following are some of the highlights of this global assessment:

- *Traceability, comparability and clarity of indicators*

According to the heads of the PHEIs surveyed, the indicators respect the fundamental criteria for performance evaluation and meet the criteria for *traceability, clarity and comparability*. Besides, they can be easily calculated by resorting to official sources.

- *Validation of the dimensions and recognition of their effective contributions to prospective institutional management*

The recognition of the existence of distinct institutional profiles through the positioning of each PHEI regarding the five dimensions analysed is unanimous and supports the comments made during the e-Delphi rounds. The institutional perception of the place where the PHEI is located, when crossed with its performance in each of the dimensions, may contribute to the design of a more efficient development strategy in future, since the governing body can opt for investing more in a particular dimension. The proposed evaluation tool was, therefore, considered relevant in the perspective of institutional performance monitoring metrics, at the level of management, at the micro, meso or macro levels, in the field of organizations.

- *Recognition of the importance of performance evaluation metrics as an aid to management, rather than a tool for a state funding formula*

In general, everyone acknowledged the merit of the indicators proposed as an aid for management; however, it must not be used as the basis for a funding formula for PHEIs. This debate is related to the fact that the list of indicators includes several variables that are not controllable or controlled by PHEIs, and, if used, may contribute to a deeper gap between institutions with very different contexts and profiles - resulting, as was mentioned before, from different strategic options in each of the dimensions, which account for their specific standard profiles.

It was also pointed out by some of the participants that the adoption of such a model for evaluating the performance of PHEIs would imply a necessary institutional adjustment in accordance with the specificities of each PHEI.

- *Need to include other bibliographic databases besides SCOPUS*

The use of *SCOPUS* as the only source for counting publications was considered questionable and limited. It was suggested that a more comprehensive list of databases (with an associated qualitative evaluation grid) be created, including, not only the publications indexed to *SCOPUS*,

but also, for example, those indexed to *SciELO* or *Latindex*, as well as books and/or chapters of books existing in scientific repositories.

- *Impossibility to include projects of a different nature, which don't generate a financial return*

When evaluating a PHEI, there should be a way of including projects of high societal impact even if they don't necessarily generate a financial return. For some of the representatives of the PHEIs included in this stage of the project, the focus on the financial return of projects/activities is considered a form of pressure which may divert the focus of the PHEIs' public mission (mainly in what concerns education and training). Also associated with financial management, the PHEIs argue that there is no autonomy to strategically allocate their own revenues. According to some of the interviewees, a greater degree of autonomy would allow them to strengthen their institutional profile, by deciding on the investment or the allocation of resources to a particular sector of the institutional dynamics.

- *Recognition that the societal impact of some indicators must be measured beyond the mere indexation to the NUTS 3*

The NUTS 3, as geographic indexation regions, were considered a limitation, due to the real societal impact of some PHEIs, since the profile of their educational offer provides qualification well beyond the NUTS 3. Therefore, widening the regional scope to include the NUTS 2 and/or NUTS 1 was suggested, especially for employability indicators. It was also considered that the capacity of the PHEI to attract students should remain focused on the NUTS 3. Even so, it is expected that PHEIs located in more developed areas and in large urban centres have a competitive advantage over those located in interior regions (and islands), more desertified and characterized by a lower supply of services, housing, transports and cultural or leisure offer.

## 4. Conclusions

According to the results obtained, the list of indicators proposed was accepted as a good model for monitoring the performance of PHEIs and their faculty staff in relation to key indicators and dimensions, not properly covered in the existing organizational evaluation instruments so far. It should be regarded as an important tool to support the strategic management of the institution, since it identifies the dimensions in which the PHEI needs to improve.

Even though these indicators are mainly of a quantitative nature, it is argued that the management of the organizational performance must not be limited to this aspect. This quantitative approach should be complemented by parameters of a qualitative nature, sustained, for example, in success stories, which may show differentiating trajectories of PHEIs.

## References

- ENQA (2006). Quality Assurance of Higher Education in Portugal An Assessment of the Existing System and Recommendations for a Future System. Retrieved from <http://www.enqa.eu/pubs.lasso>, on 18<sup>th</sup> September 2016.
- FCT (2014). Public notice for funding of projects aiming at the development and validation of performance indicators for the production, transfer and diffusion of knowledge of PHEIs. Retrieved from <http://www.fct.pt/apoios/outros/indicadores/docs/Edital.pdf>, on 14<sup>th</sup> February 2017.

- Guião da Reforma do Estado (State Reform Guide) (2013). Retrieved from [https://pt.scribd.com/document/180306383/Guiao-da-Reforma-de-Estado#download&from\\_embed](https://pt.scribd.com/document/180306383/Guiao-da-Reforma-de-Estado#download&from_embed) on 14<sup>th</sup> February 2017.
- Hsu, Chia-Chien & Sandford, Brian A. (2007): "The Delphi Technique: Making Sense Of Consensus. Practical Assessment." *Research & Evaluation*, 12, 1-8.
- MCTES (Ministério da Ciência, Tecnologia e Ensino Superior) (2016). Intervenção do Ministro da Ciência, Tecnologia e Ensino Superior no Debate temático no Conselho Nacional de Educação sobre ciência, tecnologia e ensino superior no âmbito da apresentação do Plano Nacional de Reformas. Retrieved from <http://www.portugal.gov.pt/media/19856265/20160608-mctes-plano-nacional-reformas.pdf> on 14<sup>th</sup> February 2017.
- OCDE (2006). Tertiary Education in Portugal. Background Report. Retrieved from <http://www.oecd.org/portugal/39710472.pdf>, on 18<sup>th</sup> September 2016.
- Resolução da Assembleia da República 77/2013 (Parliament Resolution 77/2013). Diário da República, 1.<sup>a</sup> série — N.º 110, 3234. Retrieved from <https://dre.pt/application/file/496653> on 14<sup>th</sup> February 2017.
- RJIES. Regime jurídico das instituições de ensino superior (Legal system of higher education institutions) (2007). Lei n.º 62/2007 de 10 de setembro. Diário da República, 1.<sup>a</sup> série — N.º 174, 6358 – 6389. Retrieved from [https://sigarra.up.pt/fpceup/pt/legislacao\\_geral.ver\\_legislacao?p\\_nr=316](https://sigarra.up.pt/fpceup/pt/legislacao_geral.ver_legislacao?p_nr=316) on 14<sup>th</sup> February 2017.

## Notas biográficas

**Sandra Saúde** es Profesora del Departamento de Educación, Ciencias Sociales y del Comportamiento –Instituto Politécnico de Beja e Investigadora del CICS. Nova (Centro Interdisciplinario de Ciencias Sociales CICS. NOVA - FCSH/UNL), Portugal. Principales intereses de investigación: planificación y estrategia organizacional, evaluación de resultados organizacionales y de impactos socioeconómicos. Publicaciones recientes: Saúde, S., *et al.*, (2014). *Os impactos Socioeconómicos do Ensino Superior: um retrato a partir de estudos de caso de Portugal e Espanha*. Lisboa: Edições Sílabo; Saúde, S., *et al.*, (2017). *Indicadores de Desempenho para as Instituições de Ensino Superior Politécnico. Investigação Aplicada, Criação Cultural e Impacto regional*. Lisboa: Edições Sílabo.

**Sandra Lopes** es Profesora del Departamento de Educación, Ciencias Sociales y del Comportamiento - Instituto Politécnico de Beja, Portugal. Principales intereses de investigación: demografía, cultura y territorio y planificación estratégica. Publicaciones recientes: Saúde, S., *et al.*, (2014). *Os impactos Socioeconómicos do Ensino Superior: um retrato a partir de estudos de caso de Portugal e Espanha*. Lisboa: Edições Sílabo; Saúde, S.; Borralho, C.; Féria, I.; Lopes, S. *et al.*, (2017). *Indicadores de Desempenho para as Instituições de Ensino Superior Politécnico. Investigação Aplicada, Criação Cultural E Impacto regional*. Lisboa: Edições de Sílabo.

**Carlos Borralho** es Profesor Invitado del Departamento de Ciencias Empresariales - Instituto Politécnico de Beja, Portugal. Áreas de investigación actual: gobierno electrónico, negocio electrónico y CRM (gestión de la relación con el cliente), y evaluación de impactos socioeconómicos de organizaciones o eventos. Publicaciones recientes: Borralho, C.; Pires, C. e Basílio, M. (2016). *Casos Empresariais de Sucesso no*

*Baixo Alentejo y Alentejo Litoral*. Lisboa: Chiado Editora.; Saúde, S.; Borralho, C.; Féria, I.; Lopes, S. *et al.*, (2017). *Indicadores de Desempenho para as Instituições de Ensino Superior Politécnico. Investigação Aplicada, Criação Cultural E Impacto regional*. Lisboa: Edições de Sílabo.

**Isidro Féria** es Profesor del Departamento de Ciencias Empresariales - Instituto Politécnico de Beja, Portugal. Principales intereses de investigación: modelos de financiación para los sistemas de educación superior y de control presupuestario; evaluación de los resultados organizacionales y evaluación de impactos socioeconómicos. Publicaciones recientes: Saúde, S.; Borralho, C.; Féria, I. y Lopes, S. (2014). “El impacto de la Educación Superior en el Desarrollo Socioeconómico: El estudio de caso del Instituto Politécnico de Beja, Portugal”. *Copernican Journal of Finance & Contabilidade*. Polonia: Universidad Nicolás Copérnico, Torún. Vol. 3, n.º 2; Saúde, S.; Borralho, C.; Féria, I. y Lopes, S. (2013). *O impacto socioeconómico do IPBeja no concelho de Beja*. Porto: IPBeja/Norprint.

# Measuring, Monitoring, and Managing for Productive Learning? Australian Insights into the Enumeration of Education

¿Medición, seguimiento y gestión al servicio del aprendizaje productivo? Enseñanzas australianas respecto a la enumeración de la educación

Ian Hardy<sup>1</sup>

## Abstract

This article reveals how ongoing teacher learning in schools is heavily influenced by the increased measurement of student learning outcomes. It draws upon a broad range of literature on the practice of education, measuring learning, and its effects on ongoing teachers' learning at the school site. The paper analyses the meeting transcripts of an ongoing, long-term teacher learning initiative in a school in northern Queensland, Australia, and indicates how processes of measurement influenced such learning. Teachers' learning is revealed as heavily influenced by processes of «measuring [student] learning», the subsequent «monitoring» of student learning through students' data profiles, and 'managing' the learning of lower performing students. The article cautions that while these resultant teacher learning practices may be beneficial for students' learning, this is a contested point, and requiring much more careful analysis of the effects of the measurement of education on such practices.

## Keywords

Quantification of learning; testing, standardized testing, standardized learning, teacher learning, teacher professional development.

## Resumen

Este artículo muestra cómo la formación continua de los maestros en las escuelas está muy influida por el incremento de la medición de los resultados del aprendizaje de los estudiantes. Se basa en un amplio repertorio de literatura sobre la práctica educativa, la medición del aprendizaje y sus efectos sobre la formación continua de los docentes en la escuela. El texto analiza las transcripciones de una serie de reuniones en el marco de una iniciativa de formación continua a largo plazo de docentes en una escuela en el norte de Queensland, Australia, e indica cómo los procesos de medición influyeron en esa formación. Se constata que la formación de los maestros está muy influida por los procesos de «medición del aprendizaje de los estudiantes», el seguimiento posterior del aprendizaje estudiantil través de sus perfiles de datos y la «gestión» del aprendizaje de los estudiantes de bajo rendimiento. El artículo advierte que, aunque estas prácticas resultantes de la formación del profesorado pueden ser beneficiosas para el aprendizaje del estudiantado, esta es una cuestión a debate, y requiere un análisis mucho más cuidadoso de los efectos que la medición de la educación ejerce sobre tales prácticas.

## Palabras clave

Cuantificación del aprendizaje; pruebas; pruebas estandarizadas; aprendizaje estandarizado; formación del maestro; desarrollo profesional docente.

Recibido: 31-03-2017  
Aceptado: 28-04-2017

<sup>1</sup> University of Queensland, (Australia); [ian.hardy@uq.edu.au](mailto:ian.hardy@uq.edu.au)

## 1. Introduction: Learning in an era of accountability

This article explores how teacher learning is construed under current conditions of increased accountability for educational outcomes. These processes are part of a broader push away from focusing upon educational attainment as a marker of educational quality within the educational system, to measuring outcomes of schooling practices. As Carnoy argues, «[t]he new emphasis on educational quality has been accompanied and promoted by the rapid spread of testing and measurement» (Carnoy, 2016:34). Measuring and comparing school outcomes have taken on increasing significance in nation-states, as educational policy-makers have sought to deploy such practices to ascertain the «success» of their respective educational systems. These processes have not occurred in isolation but have been fostered through the «soft power» (Nye, 2004) of the International Association for Evaluation of Educational Attainment (IEA), the Organization for Economic Co-operation and Development (OECD), the World Bank, Asian Development Bank, and Inter-American Bank. Such bodies promote a conception of education as equivalent to economic competitiveness, and of improved educational outcomes as key to increased economic productivity. Such pressures are felt at the nation-state level in the push to ensure some way of tracking, of keeping account, of these outcomes over time. This includes monitoring progress on international tests of achievement, such as the IEA's Trends in International Mathematics and Science Survey (TIMSS), and the OECD's Programme for International Student Assessment (PISA).

These processes of keeping account of educational outcomes are also expressed at the national level through national standardized tests of achievement, most notably in the areas of literacy, mathematics and science. In the Australian context, this relates to the national literacy and numeracy tests orchestrated through the National Assessment Program – Literacy and Numeracy (NAPLAN).

Within this context, it is important to ascertain how such broader processes and foci influence the nature of the learning supported within schools. This includes teachers' learning. In this paper, I draw upon transcripts of professional development meetings of teachers in one school in north Queensland, Australia, to reveal how measuring learning outcomes has influenced what is considered important teacher and student learning. The meetings, described as «Inquiry Cycles», were day-long events involving teachers at each year level meeting together to inquire into the nature of their students' learning.

## 2. The challenge of conceptualising education and learning: «Practical» accountability

To ascertain how the measurement of learning is expressed in schooling settings, it is firstly necessary to consider the purposes of education more broadly. Education is an inherently «practical» activity, insofar as it is grounded in the everyday practices of teachers and students. Drawing upon Aristotelian theorising, Biesta (2013) frames this as very much a problem of how to respond to ongoing, practical problems. In trying to understand what education is, Biesta (2013) calls us to consider a number of binaries about the nature of the practice of education. If we are to take up the call to consider education as an open system of meaning (rather than «a causal, deterministic system of inputs and outcomes»), as focused upon interpretive encounter (rather than «an anonymous machine»), and as «event-like» (rather than «production-like») in its intent, then we are called to enact a conception of education as one which is premised upon particular decisions at particular moments (Biesta, 2013:686). Furthermore, these decisions are never straightforward, but always open to consideration, and determined by context; as Schwab (2013/1970) argues, such decisions are inherently «practical». The practical in this instance is always something immediate and grounded; it «is always something taken as *concrete* and *particular*

and treated as indefinitely susceptible to *circumstance*, and therefore highly liable to unexpected *changes* (Schwab, 2013/1970:593; emphasis added). On this rendering of education, accountability pertains to ensuring conditions for the fostering of deliberative encounters with practical problems in consultation and collaboration with others. Understanding education is neither inductive nor deductive, but inherently «deliberative», and focused upon choosing «not the *right* alternative, for there is no such thing, but the *best* one», under the circumstances (Schwab, 2013/1970:618; emphasis original). Such deliberation resists precise calculation and is dependent upon accountability in the form of teacher judgment.

### 3. Quantifying learning in an era of enumerative accountability

However, there are myriad forms of accountability within schooling systems, and more «practical» conceptions, based on teacher judgment, struggle for recognition within current educational policy discourses of increased measurement of learning. This measurement of learning is a product of the increased reliance upon numbers in education, itself a manifestation of perceptions of a relative lack of status, or relatively poor standing within which this call to numbers is rendered. Porter (1995) refers to a «trust in numbers» to try to capture perceptions of the benefit and value of numbers. Numbers are seen to give an air of gravitas, a sense of «objectivity» to the particular practices to which they pertain. Indeed, in more complex environments, characterized by conflicting perspectives about how best to proceed – such as educational settings – more «objective» modes of reasoning are able to exert disproportionate influence; Porter refers to numbers as associated with «mechanical objectivity», and foregrounds how complex environments encourage such mechanical objectivity: «Where a consensus of experts is hard to reach, or where it does not satisfy outsiders, mechanical objectivity comes into its own» (Porter, 1995:4). Mathematical and quantitative accounts are seen as particularly effective in justifying specific positions. This is the case even as such figures are themselves never able to capture the nature of the actual practices to which they pertain. Even as such quantified information is in a sense artificial, such information helps constitute a «reality» which may be difficult to deny. Under these circumstances, «rival measures are... at a great disadvantage» (Porter, 1995:5-6)

The dominance of such measures is also resonant in the ways in which various «state-numbers» or statistics actually help to constitute the phenomena to which they purport to relate; in this sense, such numbers are complicit in the ways in which particular phenomena come to be understood and «known» (Desrosières, 1998). Sætnan, Lomell and Hammer (2011) summarise this succinctly: «society and the statistics that measure and describe it are mutually constructed» (p. 1). Various forms of statistics are both productive of specific and actual circumstances, at the same time as they represent constructs of the worlds to which they relate (Desrosières, 1998:3). Through the act of counting, we do not exist beyond that which we seek to measure but rather «enter into it, redefine it, change the stakes that affect it» (Sætnan, Lomell & Hammer, 2011:1). The power of such measures is a response to efforts to develop more «objective measures» of particular phenomena, but in so doing, such measures construct knowledge in their own image.

### 4. An infrastructure of accountability

Globally in schooling, the power of numbers is particularly evident via what Lingard, Martino, Rezai-Rashti and Sellar (2016) refer to as forms of «globalizing, top-down, test-based modes of accountability» (p. 14). Such renderings of accountability are part of a broader global data infrastructure that construes particular kinds of measures of student learning, such as OECD PISA results, as more significant than



others. While bodies such as the OECD are not able to exert direct influence over educational practices in sovereign nation-states, through support from member nations for it to undertake its work, the OECD has been able to adopt a position of «soft power» (Nye, 2004) over (and beyond) member states. Through processes of comparison enabled by such numbers, politicians have been able to use such data as evidence of the need for further interventions in education (Novoa & Yariv-Mashal, 2003).

These processes of enumerating and comparing learning are enabled by what Anagnostopoulos, Rutledge and Jacobsen (2013) refer to as an «infrastructure of accountability», which simultaneously constitutes and «collects» the data it «records». This infrastructure includes intricate, large-scale information systems, operating to collect data across schools, regions and educational authorities, and purports to be able to collect, analyse and disseminate information about the nature of student learning, and the effects of teachers' teaching practice. In English and European contexts, Lawn (2013) refers to the «rise of data» in education systems to try to capture the nature of these enumerative technologies of control that enable particular kinds of collection, visualisation and use of data. Williamson (2016) also refers to how numbers and associated forms of computational data provided through schooling systems and associated bodies provide important visual representations of practice, constituting the forms of educational governance processes that arise, including through various forms of profiles, summaries and comparisons of performance.

## 5. «Data use» and Teachers' learning

In the arena of teachers' learning, this quantification of education is often expressed in relation to various forms of «data use». Such data use is evident in efforts to shift the focus from engaging in various forms of professional development oriented towards how to engage with data, to the use of data as a form of professional development (Vanhoof & Schildkamp, 2014). This is often described in literature on teachers' learning as «data-based decision making». Schildkamp and Lai (2013) define data use, or «data-based decision making» as processes whereby 'schools make decisions about students, about instruction, and about school and system functioning based on a broad range of evidence, such as scores on students' assessments and observations of classroom teaching' (p. 1). Relatedly, Lai and Schildkamp (2013) take a broad-based approach to data, which they understand as any «information that is collected and organized to represent some aspect of schools» (p. 10). Such definitions suggest a broad conception of data including, but not limited to, more quantitative conceptions articulated earlier. Vanhoof and Schildkamp (2014) argue effective use of data enables educational processes and fosters enhanced student achievement. However, actually engaging substantively with such data is challenging work; Schildkamp, Karbautzki and Vanhoof's (2014) research into how data was actually used in schools in the UK, Germany, Poland, Lithuania and the Netherlands revealed that even as data were collected for multiple purposes, including for policy development, school development/improvement, teacher professional development, and broader accountability purposes, the actual use of data was a much more superficial process, and underdeveloped.

In the Australian context, Klenowski (2016) questions the validity of the varied uses of NAPLAN data. Reflecting upon the OECD's (2013) emphasis upon the importance of a variety of formative and summative assessment tasks, and O'Neill's (2013) differentiation between primary and secondary uses of test data at school and systemic levels, Klenowski (2016) argues the dominance of secondary uses of data calls into question the validity of such approaches; she questions «whether the policy can deliver accountability without negatively impacting high quality and high-equity teaching and learning» (p. 44).

At the same time, as Groundwater-Smith and Mockler (2009) reveal, even as teachers' work and learning are influenced by more reductive approaches to curriculum, teaching and testing, they simultaneously endeavour to engage in more practice-based approaches, rather than simply «complying» with more reductive accounts. More localized conditions and curricula are not simply ignored, even as globalizing and transnational policy processes exert influence (Sivesind & Wahlström, 2016). And testing practices and outcomes can be used for more diagnostic purposes, for determining areas of need, and for improvement, rather than for more performative accountability purposes (Hamilton *et al.*, 2013).

## 6. Contextualising testing: The Australian experience

In the Australian context, the enumeration of education is most obvious in the form of the National Assessment Program – Literacy and Numeracy (NAPLAN), a national testing regime administered by Australian Curriculum, Assessment and Reporting Authority (ACARA), a national authority endorsed by the federal and state Ministers of Education. State and territory ministers argued for the development of national tests in 2007, and the subsequently elected Rudd Labor Government took up this call (Lingard, Thompson & Sellar, 2016). Interestingly, national testing preceded the development of a national curriculum in Australia, and it is only during 2017 that it is expected that the curriculum (and other reforms) will better articulate with national testing.

The NAPLAN testing regime comprises a census-style test, undertaken every year by all students throughout the country in Years 3, 5, 7 and 9. Feedback is provided to parents and students as an individual report which indicates which «band» or level (out of a 9 point scale) students attained. This is further complemented by a report delivered at the school level on students' performance. This includes patterns in the item-by-item responses, including in comparison with «like» (similar) schools. Perhaps the most overt form of monitoring and visualisation of national testing data occurs through the display of such data via a public website to enable public scrutiny of student results, at the individual school level. The *MySchool* website was introduced in 2010 on the assumption it would provide enhanced transparency into school outcomes, read as literacy and numeracy outcomes on the NAPLAN test. The website compares school performance against 60 statistically similar schools throughout Australia, and national average performance. The «like» schools are denoted through the ICSEA index – The Index of Community Socio-educational Advantage – developed by ACARA as a means of acknowledging the effects of socio-economic status on student performance.

## 7. Conceptualising learning in context: Penn Primary and the «Inquiry Cycles»

To help understand the nature of these processes in context, and how enumeration processes play out in practice, the research draws on recently collected data drawn from an ongoing 4-year study into the nature of teacher learning practices at one school in north Queensland, Australia. The school serves a lower to middle socio-economic community in a regional city in the northern half of the state. Students come from a range of backgrounds in which both parents work, some parents work, or where parents may be unemployed. Some students live in single-parent families. The school has approximately 850 students, with roughly equal numbers of boys and girls, and 10% of students are Indigenous.

The data comprise a series of meeting transcripts of teachers deliberating together about the nature of the curriculum they were teaching, and the effectiveness of their teaching, in light of evidence of student work. These meetings, known as «Inquiry Cycle», were stimulated by the work of literacy educa-

tor Helen Timperley (2011), who advocated teachers engage in a cycle of investigation into their work, with an emphasis upon evidence of student learning. A key focus of attention during these meetings were various forms of student data collected over the course of students' schooling. This included data in the form of student work in response to formative and summative assessment tasks, students' bookwork, NAPLAN data, as well as other forms of data collected at the regional and school level. Regional data included standardized literacy and numeracy data collected through standardized «Progressive Assessment Test – Reading (PAT-R)», and «Progressive Assessment Test – Mathematics (PAT-M)» which students sat at the beginning and end of each year (with the difference between the two results purportedly reflecting students' learning over that year). At the same time, at the school level, teachers were also required to collect various forms of «levelled» reading data, known as «PM Benchmark data»; these data were generated through students' engagement with a set of standardized reading books (graded from 1 to 30), and developed to ascertain students' reading fluency and comprehension capabilities.

## 8. The data

The data comprise the most recent set of Inquiry Cycle meetings collected up to this point, in July 2016. This set of meetings was selected to provide detailed insights into the nature of teachers' deliberations about data, in the context of a strong push for data-based decision making in the Queensland context. As part of a process of «combating threats to validity» (Maxwell, 2013), these data are also recognized as part of a larger corpus of data of 10 sets of meetings attended between 2013 and 2016, the collection of which represent ongoing and detailed engagement in the work of the Inquiry Cycles over time.

The data comprised written transcripts of the most recent set of Inquiry Cycle meetings. Meetings involved all year level teachers, alongside senior members of staff (deputy principal), and specialist staff (e.g. special needs teachers), an external critical friend, and the researcher (author). All names reported are pseudonyms. Each year level met for one day per term to discuss the nature of the curriculum they were teaching their students, and to develop a better understanding of the curriculum. The meetings also entailed teachers analysing evidence of their students' learning – work samples from students' books, samples of work from assessment tasks, as well as standardized test results. Each meeting was recorded, and transcribed remotely.

Within these data, various forms of numeric measures were most explicitly expressed within the Year 2, Year 4 and Year 5 Inquiry Cycles. These data were purposively selected (Yin, 2015) to understand how such numbers and various forms of measurement were deployed during the Inquiry Cycles; these data were analysed in light of literature and theorising into the nature of education, the increased enumeration and datafication of schooling, and teachers' uses of data. The composition of each year level group is outlined below in Table 1.

**Table 1. Composition of year level groups in which numeric measures were expressed explicitly**

Year Level	Year 2	Year 4	Year 5
Composition of Group	Facilitator (deputy-principal; female)	Facilitator (deputy-principal; female)	Facilitator (deputy-principal; female)
	Acting head of Curriculum (female)	Acting head of Curriculum (female)	Acting head of Curriculum (female)
	Critical friend (female)	Critical friend (female)	Critical friend (female)
	6 classroom teachers (all female)	6 classroom teachers (3 male; 3 female)	5 classroom teachers (2 male; 3 female)
	Author (male)	Author (male)	Author (male)
		1 student-teacher (male)	

These teachers were part of a larger teaching workforce of 43 classroom teachers in the school (6 male; 37 female), with the composition outlined in Table 2 below.

**Table II. Composition of teachers across all year level groups**

Year Level	Prep	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
Composition of Group (number of teachers & gender)	7	5	6	6	3	3	5
	females	females	females	females	females	females	females
					3	2	1
				males	males	male	

## 9. Findings: The enumeration of education at Penn Primary

An emergent thematic analysis (Shank, 2002) of the transcripts revealed three key themes associated with the quantification of education at Penn Primary. These were related to various processes of trying to «measure» student learning within individual classes, «monitoring» such learning across class profiles of data, and «managing» the student learning that subsequently ensued, particularly amongst students deemed as under-performing.

### 9. 1. The primacy of measuring learning at Penn Primary

Processes of «measuring» learning were clearly evident during the meetings in a variety of ways. This included how teachers were asked by the facilitator of the meeting («Felicity») to indicate the percentage of students who were not yet achieving a pass grade in their year level:

*«Paddy: Alright I'll go, I've got 20% ...*

*Lydia: Okay I have 36% ...*

*Ted: Yes, percent failing, it's 19% failing. ...*

*Lucy: Yeah, I got 20% failing. (Year 4 Inquiry Cycle)*

*Felicity: How many Phil?*

*Phil: 30%, 8 out of 26.*

*Tess: ... I'm 8 as well, so I'm 30%» (Year 5 Inquiry Cycle).*

In Year 2, the identification of these numbers was a challenging experience for some teachers, including a first-year teacher who was struggling to engage his class. The significant proportion of students identified as failing in his class provided the stimulus for a discussion about the importance of ensuring that the activities asked of students were reflected against the criteria against which they were to be assessed, and that anomalies were perhaps the result of not giving the students the opportunity to understand the nature of the task:

*«Chris: So in mine I've got 63% as failing. I guess just obviously what I'm doing is not working in general, they're not getting what I'm trying to teach them, the way I'm going about it obviously is not working.*

*Felicity: ... So we've got to be careful, we don't want to mark these kids down because we didn't teach the task well, and this is your first time in Year 4 and probably the first time dealing with a whole year of C2C [new curriculum] and not just English. So we want to have a look at this closely before this*

*term, get the tasks right and so the kids who are able to demonstrate a “C” are actually getting one; that’s going to shift your data a lot». (Year 4 Inquiry Cycle).*

The facilitator’s focus on not «want[ing] to mark these kids down», and emphasis upon «teach[ing] the task well» reflects a more educative, «practical», approach to the particular circumstances at play – including a new teacher struggling to make sense of the curriculum. At the same time, this conversation was made possible by beginning with the nature of the marks themselves; the numbers provided a sense of authority against which it was hard to argue (Porter, 1995).

These numbers, including the percentage of students failing, were seen as especially dramatic in Year 2 more broadly, which was seen as a cohort of students achieving at a particularly low level. There was considerable «consistency» across each of the classes in this regard, and with these students’ results from the previous year:

*«Felicity: So just for all us to think about because we all own all of the data as well, but we would expect to see some consistency in the classes because they came through fairly consistently apparently, as good as – as reliable as the data was. So now we want to have a look – okay so have you done [class] A, what’s [class] A, “D” or below?*

*Jenny: 52[%]...*

*Felicity: And B?*

*Linley: ... 29%...*

*Felicity: Okay 29% yep. C?...*

*Eleanor: I’m C – 50%...*

*Cindy: E?...*

*Felicity: So that’s going to make yours 50% so it was 10 out of 21...*

*Felicity: ... who’s D?...*

*Lisetta: I got 57% D and below...*

*T: E?*

*Cindy: Me? ... 29%...*

*Felicity: And then F is?*

*Jenny [for absent teacher]: 45% failing». (Year 2 Inquiry Cycle).*

The power of these numbers was evident in the way in which these figures were considered particularly stark for the cohort as a whole, and evidenced in the way the facilitator began this account with a collective comment about all teachers needing to share the results of the cohort as a whole («we all own all of the data»). Mention of the expectation of «consistency» of high proportions of lower performing students between classes also reflects the power of these numbers in the construction of these students as particularly problematic.

The primacy, but also complexity, of measurement was also evident in the way standardized forms of data were privileged, and other forms of data – including numeric data seen as dependent upon teacher judgment alone – critiqued. What seemed to be valued most were those forms of more standardized quantitative data that seemed to provide «hard evidence» of student achievement – particularly NAPLAN and PAT results; other forms of data, generated by teachers, including the numeric PM Benchmark data, were seen as less reliable – as subject to «just teacher judgement»:

*«Felicity: But what's written in here really could be anything, other than the NAPLAN, all the rest is just teacher judgement – maybe not the PAT-R. But knowing that, my point there is knowing that Lauretta is a 24 – what does that mean? You know, it doesn't really tell me anything, so getting into her data [daily bookwork, formative and summative classroom assessment] will tell you more and then you can talk to it more». (Year 4 Inquiry Cycle).*

Reflecting the power of NAPLAN in the Australian context, this data was accorded primacy; while the validity of the use of NAPLAN has been questioned more broadly in the Australian context (Klenowski, 2016), in these discussions, there was a sense in which NAPLAN data was ascribed an authority to which other forms of data could only aspire.

Even as there were concerns about some forms of data (including PM numeric data), there was still a valuing of more standardized, numeric forms of evidence. This was expressed in different ways, including in the expectation that teachers would undertake PM Benchmark reading tests with students (using a «running record») if students' results were deemed too low on PAT-Reading tests (e.g. below 5 (half way) on a 10-point («stanine») scale):

*«Felicity: Here as well, the stanine 4 – anyone stanine 4 and below should have a PM in there for Term 1 and Term 2.» (Year 5 Inquiry Cycle).*

Teachers' accounts of their students' performances were heavily peppered with reference to numeric measures of achievement – whether in relation to PMs, PAT results or NAPLAN – even as they recognized the particular, specific circumstances of students in relation to these measures, such as failing to attend school consistently:

*«Felicity: I noticed in your's, ... is Lesley Bruce, ... what's her PM level?*

*Ted: 19-ish, 19...*

*Felicity: ... if you look at Joel Coolamon ... his NAPLAN's pretty average...*

*Lucy: Lou Smith ... like her NAPLAN is not too bad at all ... She wasn't too bad on NAPLAN. Lou's big issue is her attendance, I think it's really impacting on her grades...*

*Felicity: That makes sense, ... she's a little low on NAPLAN but not disastrous, the same with the PMs and the PAT-R; but, yeah, attendance – okay.» (Year 4 Inquiry Cycle).*

In this way, specific numeric, and often standardized measures of attainment were foregrounded, even as teachers were simultaneously aware of the limitations of these numbers, and of the personal circumstances which influenced their students' learning.

## 9. 2. Monitoring learning through class profiles

From the outset of the Inquiry Cycle meetings, it was clear that teachers were to be engaged in using these measures to monitor learning using various summaries/class profiles of data at Penn. This was an active undertaking in which teachers were asked to consider the nature of their students' results as a whole, and to identify lower performing students whom they believed could improve, and attain a passing grade for the year. This involved scrutiny of individual teachers' summaries of student data (LoA data, PMs, PAT results and NAPLAN) in their «class profile»:

*«Felicity: I want to just get into a process to look at your class profile and say, “Okay; who are the kids that we’re going to try and shift?” (Year 4 Inquiry Cycle).*

*Felicity: Over in the data [in class profile] ... look at that as well because that’s telling a whole other story. But also look at this and think, yeah, but which kids, with just a bit of tweaking are going to get there; because the more that you can catch up to year level this year, the more the [Year] 3s can catch up, after that.» (Year 2 Inquiry Cycle).*

This monitoring, through the visualisation of data (Williamson, 2016) via the class profiles, was also evident in the way teachers compared results across different data sets, drawing upon more standardized and teacher judgment measures to bolster arguments about coherence between the different data sets:

*«Frances: I’m just looking now, and ... I’m matching [Levels of Achievement data] across to the PAT-R and- (Year 5 Inquiry Cycle).*

*Ted: She’s the main one that where the data didn’t really line up, the rest of them kind of, like if you look at Liza up the top – the [PAT-R] stanine reflects her LOA [Level of Achievement A to E] data, reflects the PM, and the NAPLAN are pretty much lined up.*

*Felicity: Yeah it does.» (Year 4 Inquiry Cycle).*

Reflecting the co-constitutive nature of the numbers and society (Sætnan, Lomell & Hammer, 2011) – in this case, in relation to students whose results deemed they did not require further scrutiny – there was a sense in which it was not necessary to closely monitor those students whose «indicators» of learning appeared to be in keeping with one another on the class profile documents, and who were deemed achieving at an «adequate» level.

However, this monitoring of students' learning was not straightforward, and involved close scrutiny of an array of data collected about individual students, and sometimes disparate results within individual students' profiles. Teachers were called upon to justify specific students' outcomes across this array of data, and to explain why there may be anomalies between different data sets within the profiles:

*«Felicity: So what we’re going to do first is get “down and dirty” and interrogate the data, so I’ve printed your class profiles off, and that, first of all, we’ll look at. (Year 4 Inquiry Cycle).*

*Felicity: I’ll just fire questions when I see something that’s going to impact on you – and I can see one here, Toby Jones. He’s got a [PAT-R] stanine 1 but you think, are you confident that he’s a [PM] level 25?*

*Tess: Yeah, yeah-*

*Felicity: Why is he a D then?*

*Tess: I don't know what happened in that stanine 1.» (Year 5 Inquiry Cycle).*

Quantitative measures of data were referred to explicitly in an effort to try to justify why students were achieving as they were according to Level of Achievement data, Pat-R data (stanine 1), and PM Benchmark data (level 25). Monitoring *across* these data sets was construed as a useful means of ascertaining anomalies in terms of specific students' performance. However, this also involved questioning the extent to which certain measures were the «best tool to use»:

*«Felicity: So what's her strengths according to this set of data?*

*Dulcie: She can read but she doesn't understand what she's reading. When I listen to her PMs, she's gone from a 26 to a 25, so when I listen to her read . . . , so I've just listened to her last term read [PM] 25, she's got 96% accuracy but her comprehension was only 2 out of 8, so she should even be lower.*

*Felicity: So how is she getting a [PAT-R] stanine 4 at the beginning of the year?*

...

*Dulcie: Good guessing. . . . So I think that might just be a stroke of luck.*

*Felicity: So, well what about the NAPLAN then? Because her reading was okay and her PAT-R is suggesting that she – it's not brilliant – but it's suggesting that she is reading okay and she's got some comprehension.» (Year 4 Inquiry Cycle).*

Such dialogue reveals the Inquiry Cycle process as a vehicle for the simultaneous critique and affirmation of standardized data, and enumerative measures of attainment. Even as there were concerns expressed about the extent to which specific standardized measures were valid (PMs), other forms of data were promoted as potentially helpful for shedding light upon the nature of evidence of actual student learning (NAPLAN); such expression provides evidence of the co-constitutive nature of various forms of numbers and the nature of the teacher learning to which they related (cf. Sætnan, Lomell & Hammer, 2011).

### 9. 3. Managing learning: «Shifting» and engaging lower performing students

The enumeration processes that underpinned the measuring and monitoring of student learning subsequently lent themselves to the «management» of students' learning, particularly in relation to lower performing students. As alluded to above, such management was evident in efforts to not only monitor students who were currently achieving at a D level or below, but also how to «shift» these students to ensure a greater level of success into the future. This was an «active» process, and seen as only possible in light of the collection (measurement) and scrutiny (monitoring) of student data:

*«Felicity: So we'll go through this process of looking at your data, working out which kids are failing at this point in time, because a lot of those won't be getting D's by the end of the year, hopefully; that's our goal. It's not to make a judgement on where they're at now, but it's to look at what they currently know and what we need to do next to shift these kids to a "C".» (Year 4 Inquiry Cycle; emphasis added).*



In this way, there was a proactive process operating within the school that sought to foster teacher learning for student engagement and improvement, and monitoring of various forms of student data were construed as central to this work. As part of this dialogue about data, there was a continued focus upon students deemed likely to perform below year level benchmarks:

*«Ted: There's a few of those ones that are 'C's though; they're the ones that I focused on in the data conversations [held once per term, with a senior member of staff] that could slip back to 'D's.» (Year 4 Inquiry Cycle).*

Where there were gaps in the data, there was an expectation that teachers would address these by collecting the relevant information:

*«Felicity: So check that everything's there that should be. Just put a circle around anything that you haven't put in and attend to it as soon as you can.» (Year 5 Inquiry Cycle).*

Managing learning also entailed moving beyond reductive conceptions of data, and foregrounding opportunities to engage more proactively with efforts to foster student learning, and progress, and to recognize students' individual gains:

*«Ted: And that's where Jessica fell down, I gave her a "D" ... [I gave her] a lot of the structure of explaining the language features and what they show you. There was a lot of writing frames involved, and things like that to help her out.*

*Felicity: Yeah and that's great, that's great to give her that success and that's exactly what she needs. (Year 4 Inquiry Cycle).*

*Felicity: More importantly and you guys know this ... more importantly than whether they're getting a "D" or not is – are they progressing?*

*Jenny: That's what I – I don't like to get bogged down with, "Oh, is it a 'C' or a 'D'?" I just want to look at individual progress and I know that those children, even though they're a "D", they have made individual gains; and I just don't like having those targets.» (Year 2 Inquiry Cycle).*

The focus on «individual progress» and «individual gains» reflects much more contextualised, «practical» (Biesta, 2013) conceptions of practice, focused upon the peculiar needs of particular students, and how to address these. The targets – in this case, the proportion of students achieving above a “D” on Level of Achievement (summative curriculum) results – were seen as restrictive of the broader notions of learning that were seen as evidence of real progress on the part of students.

This management of student learning also entailed teachers considering the nature of the pedagogies they should enact to help some students remain focused upon their work, and not under-perform. This was evident in the discussion in the Year 5 Inquiry Cycle in relation to one student whose data was disparate, but who was recognized as capable of better learning (and improved outcomes):

*«Felicity: Look at Mitch Speith. Stanine 2 at the beginning of the year. Apparently he's reading at a level 30; I'd question his comprehension on a whole lot of things for other reasons*

*Tess: But he got stanine 5 last year*

*Felicity: But he got a D ...*

*Felicity: Right now he's getting a D. He's capable, competent. Why is he getting a D*

*Tess: Yep*

*Felicity: Not because he's lazy*

*Tess: No – he just 'checks out' [becomes distracted in class]*

*Felicity: We've got to look at ... well then – what pedagogies might keep him "in" class, okay? That's the kind of thing.» (Year 5 Inquiry Cycle).*

Such responses reflect the multifaceted nature of engagement with and through these numbers, particularly in relation to lower performing students. Critiques of the accuracy of standardized measures of learning, and emphases upon specific pedagogies for specific students, resonate with calls for more practical applications and understandings of education (Biesta, 2013); the very creation of such numbers, potentially, had the capacity to generate conversations and actions about what to do next for particular students in particular circumstances. However, that such conversations as vehicles for teachers' learning were always framed within this paradigm of particular measures of student learning simultaneously frustrates the scope of such potentiality. Better delineating the nature of this quandary requires further inquiry.

## 10. Discussion and conclusion: Measuring, monitoring, and managing for productive learning?

In many ways, processes of measuring, monitoring and managing educational practices via particular forms of standardized numbers represent forms of «fast policy» (Peck & Theodore, 2015) that have come to characterise processes of statecraft under neoliberal conditions. In educational settings, such processes are associated with not only the enumeration of education, but as evident in the data presented here, processes of constant comparison (Novoa & Yariv-Mashal, 2003) with such measures to «validate» forms of learning. It is the dominance of such measures and monitoring which give pause for thought in relation to the potentially educative effects of the uses of enumerative data in schooling settings, including for fostering teachers' learning. Learning at Penn was conceptualised in light of multiple measures of learning, and the subsequent monitoring and managing practices that these measures enabled. It would not have been possible to engage in the sorts of conversations that characterized the Inquiry Cycles in these year levels if it had not been for the generation of these data, and the various «infrastructure of accountability» (Anagnostopoulos *et al.*, 2013), including the class profiles, that enabled this process.

Arguably, the engagement with data through the Inquiry Cycle conversations helped to cultivate an understanding of the nature of these data, and how they might inform teachers' learning for enhanced student learning. The use of numbers to identify students who were underperforming, reveals the power of numbers; this was evident in the way in which various measures were foregrounded, particularly the percentage of students deemed underperforming in their Level of Achievement (A to E) data. Such measurement served as a precursor to closer identification and subsequent scrutiny/monitoring of students deemed as requiring interventions to enhance their learning outcomes, and how to best orchestrate learning opportunities for these students. These numbers could be deployed for productive purposes, as a starting point for more detailed cultivation of student and teacher learning. In part, this was the case in relation to the Inquiry Cycles more broadly. It could be argued, therefore, that these

teachers were not simply constrained by broader conditions, within what has been termed an «age of compliance» (Groundwater-Smith & Mockler, 2009), with subsequently reductive effects upon their learning; after all, these teachers were engaged in conversations in the Inquiry Cycle process that did enable more «practical» (Schwab, 2013), practice-based approaches to their learning (Groundwater-Smith & Mockler, 2009). Even as the numbers that peppered their dialogue can be seen as evidence of more globalizing and transnational policy processes, more localized conditions and curriculum were being addressed, and taken into account (Sivesind & Wahlström, 2016). Particular students' needs, in light of the curriculum as taught at the school, were highlighted, and flagged for intervention; such data were being used to try to question the pedagogies in action, and to foster enhanced student achievement (Vanhoof & Schildkamp, 2014), even as this was also challenging (cf. Schildkamp *et al.*, 2014). And there was at least some broad evidence of the use of various PAT and NAPLAN test results for more diagnostic purposes, and not simply for more performative reasons (cf. Hamilton *et al.*, 2013).

However, and at the same time, understandings of student learning were heavily mediated by a focus upon particular kinds of data – data that could be measured, monitored and perceived as useful for «managing» the vicissitudes of learning. That such data existed in so many formats – in this case, NAPLAN, Pat-results, PM Benchmark results, and the proportion of students performing below a pass grade («D») in the regular curriculum – is indicative of how educational quality has indeed been associated with the «rapid spread of testing and measurement» (Carnoy, 2016:34). Significantly, such data were construed as providing important evidence under broader conditions in which teachers' judgment is increasingly under scrutiny. Even as the Inquiry Cycles were explicitly designed to effect enhanced teacher judgment, and involved much discussion about the specific aspects of students' work that needed to be attended to, and interrogation of curriculum to improve teachers' understandings to improve their teaching, the visualisation of enumerative data (Williamson, 2016) and the subsequent focus upon this data within the discussions, also reflects how such judgement alone was considered insufficient, and needed to be complemented by the gravitas that attends numeric representations of student learning; rival «measures» (Porter, 1995) to these numbers were at a distinct disadvantage during these discussions.

As the «pinnacle data» within the Australian national context, NAPLAN data clearly influenced how student learning was construed in the context of the Inquiry Cycles. While individual NAPLAN data were deployed by teachers to provide some insights into the nature of their students' learning, that this work was occurring within a broader context of scrutiny of students' results at the national level in light of both national and international data reveals how the dialogue about data was not simply a «local» phenomenon, but also heavily influenced by the broader policy conditions within which it unfolded. Under such conditions, even as teacher judgement was recognized as important at Penn Primary, there was something of a hierarchy in terms of perceived validity, and «objectivity» (Porter, 1995) of the numbers associated with student learning. NAPLAN was perceived as most authoritative, followed by PAT data; these modes of data were perceived as less subject to teacher judgment than the PM Benchmark data, or teachers' qualitative accounts of student progress, and therefore more useful for «monitoring» student learning. This was the case even as these numbers were recognized as inherently limited and limiting, as needing to be subject to scrutiny on the part of teachers, and as not providing solutions to the problems which they constructed. This was the domain of teacher judgment, even as such judgment was simultaneously critiqued and criticized.

The numbers referred to here were clearly productive of teachers' learning, but the cultivation of teacher judgment more broadly was also simultaneously hemmed in by the parameters of the standardized measures with which they were associated. The way in which such numbers/measures are deployed can potentially limit the sorts of context-relevant, «practical» conceptions of education to which we might aspire as educators (Biesta, 2013). Such numbers are complicit in how particular kinds of knowledge become «known» (Desrosières, 1998), and constitutive of the conditions within which they are generated and deployed (Sætnan, Lomell & Hammer, 2011). In the context of the Inquiry Cycle approach, the potential for these numbers to help inform teachers' understandings of their students' capabilities was evident. However, this potential existed alongside a broader set of conditions which foregrounded particular kinds of numbers as valid and meaningful, and teacher judgment as problematic, even as efforts were made to cultivate such judgment. Such tensions reflect the conflicted and contested nature of the enumeration of education, and ongoing school-based teacher learning, and of the challenges which subsequently attend the measurement, monitoring and management of student learning.

## References

- Anagnostopoulos, Dorothea; Rutledge, Stacey & Jacobsen, Rebecca (2013): "Mapping the information infrastructure of accountability" in Dorothea Anagnostopoulos, Stacey Rutledge & Rebecca Jacobsen. (eds.): *The Infrastructure of Accountability: Data Use and the Transformation of American Education*. Cambridge: Harvard Education Press.
- Biesta, Gert (2013): "Knowledge, judgement and the curriculum: on the past, present and future of the idea of the Practical". *Journal of Curriculum Studies*, 45 (5), 684-696,
- Carnoy, Martin (2016): "Educational policies in the face of globalization: Whither the nation state?" in Karen Mundy, Andy Green, Bob Lingard, & Antoni Verger (eds.): *The Handbook of Global Education Policy*. Chichester: Wiley-Blackwell.
- Desrosières, Alain (1998). *The Politics of Large Numbers: A History of Statistical Reason*. Cambridge: Harvard University Press.
- Groundwater-Smith, Susan & Mockler, Nicole (2009). *Teacher Professional Learning in an Age of Compliance: Mind the Gap*. Rotterdam: Springer.
- Hamilton, Laura; Schwartz, Heather; Stecher, Brian & Steele, Jennifer. (2013): "Improving accountability through expanded measures of performance". *Journal of Educational Administration*, 51 (4), 453-475.
- Klenowski, Val (2016): "Questioning the validity of the multiple uses of NAPLAN data" in Bob Lingard, Greg Thompson, & Sam Sellar (eds.): *National Testing in Schools: An Australian Assessment*. Abingdon: Routledge.
- Lai, Mei Kuin & Schildkamp, Kim (2013): "Data-based decision-making: An overview" in Kim Schildkamp, Mei Kuin Lai & Lorna Earl. (eds.): *Data-Based Decision Making in Education: Challenges and Opportunities*. Springer: Dordrecht.
- Lawn, Martin (2013): "Introduction: The rise of data in education" in Martin Lawn (ed.): *The Rise of Data in Education Systems*. Oxford: Symposium.
- Lingard, Bob; Martino, Wayne; Rezai-Rashti, Goli & Sellar, Sam (2016). *Globalizing Educational Accountabilities*. New York: Routledge.

- Lingard, Bob; Thompson, Greg & Sellar, Sam (2016): "National testing from an Australian perspective" in Bob Lingard, Greg Thompson & Sam Sellar (eds.): *National Testing in Schools: An Australian Assessment*. Abingdon: Routledge.
- Maxwell, Joseph (2013). *Qualitative Research Design: An Interactive Approach* (3<sup>rd</sup> ed.). Thousand Oaks: Sage.
- Novoa, António & Tali Yariv-Mashal (2003): "Comparative research in education: A mode of governance or a historical journey?". *Comparative Education*, 39 (4), 423-438.
- Nye Jr., Joseph S. (2004). *Soft Power: The Means to Success in World Politics*. New York: Public Affairs.
- O'Neill, Onora (2013): "Intelligent accountability in education". *Oxford Review of Education*, 39 (1), 4-16.
- Peck, Jamie & Theodore, Nik (2015). *Fast Policy: Experimental Statecraft at the Thresholds of Neoliberalism*. Minneapolis: University of Minnesota Press.
- Porter, Theodore (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Sætnan, Ann; Lomell, Heidi & Hammer, Svein (2011): "Introduction: By the very act of counting – The mutual construction of statistics and society" in Ann Sætnan, Heidi Lomell, & Svein Hammer (eds): *The Mutual Construction of Statistics and Society*. New York: Routledge.
- Schildkamp, Kim; Karbautzki, Louisa & Vanhoof, Jan (2014): "Exploring data use practices around Europe: Identifying enablers and barriers". *Studies in Educational Evaluation*, 42, 15-24.
- Schildkamp, Kim & Lai, Mei Kuin. (2013): "Introduction" in Kim Schildkamp, Mei Kuin Lai & Lorna Earl. (eds.): *Data-Based Decision Making in Education: Challenges and Opportunities*. Springer: Dordrecht.
- Shank, Gary (2002). *Qualitative Research: A Personal Skills Approach*. Upper Saddle River: Pearson.
- Schwab, Joseph (2013/1970): "The practical: a language for curriculum." *Journal of Curriculum Studies*, 45 (5), 591-621.
- Sivesind, Kirsten & Wahlström, Ninni (2016): "Curriculum on the European policy agenda: Global transitions and learning outcomes from transnational and national points of view." *European Educational Research Journal*, 15 (3), 271-278.
- Timperley, Helen (2011). *Realizing the Power of Professional Learning*. New York: Open University Press.
- Vanhoof, Jan & Schildkamp, Kim (2014): "From 'professional development for data use' to 'data use for professional development'." *Studies in Educational Evaluation*, 42, 1-4.
- Williamson, Ben (2016): "Digital education governance: data visualization, predictive analytics, and 'real-time' policy instruments". *Journal of Education Policy*, 31 (2), 123-141.
- Yin, R. (2015). *Qualitative Research from Start to Finish* (2<sup>nd</sup> ed.). New York: The Guilford Press.

## Nota biográfica

Ian Hardy es *Senior Lecturer* en The University of Queensland (Australia). Ha analizado las condiciones laborales y de aprendizaje de los docentes con colegas de Australia, Estados Unidos y Canadá. La investigación del Dr. Hardy se centra en la relación entre la educación y la sociedad, y su investigación actual explora las respuestas de los educadores a los contextos sociopolíticos más amplios en los que desarrollan su trabajo. Desde 2004 es miembro del consorcio internacional de investigación «Pedagogía, Educación y Praxis», con la participación de investigadores de Australia, Reino Unido, Suecia, Noruega y Finlandia. Ha publicado como autor *The politics of teacher professional development: Policy, research and practice*. New York, USA: Routledge, 2012; y como coautor: *Changing practices, changing education*. Singapore: Springer, 2014.

# Permanência na educação superior no Brasil: Construção de uma escala de medida

## Permanence in Higher Education in Brazil: Construction of a measuring scale

Uyguaciara Veloso Castelo Branco, Paulo Hideo Nakamura y Edineide Jezine<sup>1</sup>

### Resumo

O presente trabalho integra o projeto «Políticas de Educação Superior: os desafios da inclusão, acesso e permanência no contexto da expansão das Universidades Federais» (Edital Universal CNPq, N.º 14/2013) e tem por objetivo apresentar a construção e validação da Escala para Avaliação da Permanência Discente (Nakamura; Castelo Branco; Jezine, 2016), que intenciona avaliar os fatores que possibilitam a permanência de alunos na educação superior brasileira. Utilizamos, como suporte teórico, os trabalhos de alguns autores, a exemplo de Ronsoni (2014); Sguissardi (2009); Ristoff (1999; 2011); Neves; Raizer; Fachinetti (2007); Baggi; Lopes (2011); Zago; Pereira; Paixão (2015); Zago (2006); Polydoro (2000); e Bittar; Oliveira; Morosini (2013). São apresentados os passos seguidos para submissão do instrumento a um grupo de expertises e o uso da Lógica Paraconsistente (Da Costa, 1999; Abe, 1992; Subrahmanian, 1987) como critério de validação dos resultados obtidos. Na primeira etapa do processo de validação, a semântica, constatou-se que a análise de pertinência e relevância apontou que a escala é adequada e coerente, para a obtenção de informações relativas à permanência do estudante em determinado curso superior, obtendo mediana igual a 5. Na segunda etapa, a de consistência, constatou-se, a partir do processo de validação aplicado, a lógica paraconsistente anotada, LPA2v, que a escala construída é consistente e que as proposições não produziram contradições e foram suficientes, não sendo detectada falta de informação.

### Palavras-chave

Permanência, motivação, educação superior, validação de escala, lógica paraconsistente.

### Resumen

El presente trabajo integra el proyecto «Políticas de Educación Superior: los desafíos de la inclusión, acceso y permanencia en el contexto de la expansión de las Universidades Federales» (Edital Universal CNPq, N.º 14/2013) y tiene como objetivo presentar una construcción y validación de la Escala de Evaluación de Permanencia Discente (Nakamura; Castelo Branco; Jezine, 2016), que pretende evaluar los factores que posibilitan la permanencia de estudiantes de educación superior brasileña. Utilizamos, como apoyo teórico, los trabajos de algunos autores, por ejemplo de Ronsoni (2014); Sguissardi (2009); Ristoff (1999; 2011); Neves; Raizer; Fachinetti (2007); Baggi; Lopes (2011); Zago; Pereira; Paixão (2015); Zago (2006); Polydoro (2000) y Bittar; Oliveira; Morosini (2013). Se presentan los pasos seguidos para probar el instrumento a un grupo de expertos y el uso de la Lógica Paraconsistente (Da Costa, 1999; Abe, 1992; Subrahmanian, 1987) como criterio de validación de los resultados obtenidos. En la primera etapa del proceso de validación, la semántica, se constató que el análisis de pertinencia y de relevancia confirmó que la escala es adecuada y coherente para la obtención de informaciones relativas a la permanencia de los estudiantes en determinado curso superior, obteniendo una mediana igual a 5. En la segunda etapa, la de consistencia, se constató, a partir del proceso de validación aplicado, la lógica paraconsistente anotada, LPA2v, que la escala construída es consistente y que las proposiciones no produjeron contradicciones y fueron suficientes, no siendo detectada falta de información.

### Palabras clave

Permanencia, motivación, educación superior, validación de escala, lógica paraconsistente.

<sup>1</sup> Universidade Federal da Paraíba, Brasil, Uyguaciara Veloso Castelo Branco, [uyguaciara@gmail.com](mailto:uyguaciara@gmail.com); Paulo Hideo Nakamura, [nkm.paulo@gmail.com](mailto:nkm.paulo@gmail.com); Edineide Jezine, [edjezine@gmail.com](mailto:edjezine@gmail.com)

## Abstract

The present work integrates the project «policies of Higher Education: the challenges of inclusion, access and permanence in the context of expanding federal universities» (Universal Project CNPq. No 05/2013) and aims to present the construction and validation of the Scale for Assessment of Student Permanence (Nakamura; Castelo Branco; Jezine, 2016), which intends to evaluate the factors that make the permanence of students in higher education in Brazil. We used, as theoretical support, the work of some authors, the example of Ronsoni (2014); Sguissardi (2009); Ristoff (1999; 2011); Neves; Raizer; Fachinetti (2007); Albany; Lopes (2011); Zago; Pereira; Passion (2015); Zago (2006); Polydoro (2000); and Bittar; Oliveira; Morosini (2013). You are introduced to the steps followed for submission of the instrument to a group of expertise and the use of the Paraconsistent Logic (Da Costa, 1999; Abe, 1992; Subrahmanian, 1987) as a criterion for validation of results obtained. In the first stage of the validation process, the semantics, it was verified that the pertinence and relevance analysis pointed out that the scale is adequate and coherent, to obtain information regarding the student's stay in a certain higher course, obtaining a median equal to 5. In the second step, the one of consistency, it was verified, from the applied validation process, the annotated paraconsistent logic, LPA2v, that the constructed scale is consistent and that the propositions did not produce contradictions and were enough, not being detected a lack of information.

## Key words

Permanence, motivation, higher education, scale validation, paraconsistent logic.

Recibido: 04-03-2017  
Aceptado: 02-05-2017

## 1. Introdução

O tema Educação Superior, no contexto das políticas públicas atualmente em vigor no Estado Brasileiro, traz, em seu bojo, a necessidade de contemplar, em nível de debate ou reflexão, um tripé de categorias de análise extremamente importante: o acesso, a *permanência* e a conclusão ou *sucesso* dos estudos. Face às várias estratégias e programas políticos adotados no contexto pós-LDB (Brasil, 1996), a categoria acesso parece contemplada, plenamente, em virtude da ampliação do acesso ao ensino superior para as camadas populacionais de baixa renda, antes excluídas do sistema educacional formal, desde à educação básica até o topo da pirâmide educacional, na educação superior. Os dados mostram que o crescimento estatístico da participação de alunos em situação de vulnerabilidade nas estatísticas educacionais, no Brasil, se tornou uma realidade inquestionável (Castelo Branco, Nakamura, 2013).

Entretanto, se a discussão sobre o acesso parece bem aprofundada, não há como ignorar que, para atingirmos a compreensão sobre o processo de inclusão na educação superior, haveremos que incorporar um debate necessário: incluir em nossas reflexões a relação existente entre acesso – permanência – sucesso (Santos, 2005:68). Por isso, elegemos a categoria *permanência* como o centro de nossas investigações, o que repercutirá em trazê-la como ferramenta fundamental para a gestão de Instituições de Ensino Superior (IES), que intencionem garantir o empoderamento de sujeitos em situação de vulnerabilidade, através da ampliação de capital cultural, social e simbólico. Tal ampliação é necessária à mudança de padrão social e econômico dessas camadas sociais, e é possibilitada pela permanência no ensino formal, garantindo-lhes uma formação com qualidade e sucesso como egresso da educação superior, para posterior inserção no mercado de trabalho.

Como já afirmamos em estudo anterior, «O acesso, sem as condições para a permanência, não se constituiria em um engodo, uma vez que as políticas de acesso à educação superior, em sua materialização, não se configuram em sinônimo de permanência no sistema?» (Castelo Branco, Jezine, Nakamura, 2015:755).



Tal questionamento somente pode ser respondido a partir da sistematização de pesquisas empíricas que oportunizem aos alunos das IES afirmarem como está ocorrendo o desenrolar de seus estudos no ensino superior, quais as dificuldades enfrentadas e como as instituições tem respondido, objetivamente, às suas carências, necessidades, problemáticas enfrentadas e desafios surgidos. Objetivando responder a essa gama de situações, intencionamos auferir os fatores que contribuem para a permanência discente, construindo um instrumento de fácil aplicação e análise, no formato de Escala do Tipo Likert, que passaremos a apresentar.

Para a construção do instrumento, necessitamos de aprofundamento teórico, a partir dos trabalhos de alguns autores, a exemplo de Ronsoni (2014); Sguissardi (2009); Ristoff (1999; 2011); Neves; Raizer; Fachineto (2007); Baggi e Lopes (2011); Zago; Pereira; Paixão (2015); Zago (2006); Polydoro (2000); Bittar, Oliveira; Morosini (2013).

## 2. A categoria permanência como centro de nossas reflexões

A observação dos alarmantes índices de evasão/abandono nos últimos anos, na educação superior brasileira, não pode ser tratada como mera contemplação, mesmo que não sejam encontradas medidas efetivas de acompanhamento ou avaliação por parte das IES, para garantir a formação da grande massa de alunos que tiveram seu acesso ampliado, mas que não tiveram garantias financeiras, sociais, culturais ou mesmo psicológicas de manter-se no sistema de ensino (Velloso; Cardoso, s.d; Borges, 2011).

Nessa perspectiva, os estudos de Castelo Branco; Jezine (2013); Zago (2006); Catani; Hey; Gilioli (2006); Felicetti; Morosini (2009) apontam o caráter ainda incipiente de discussão sobre a categoria permanência, necessitando de uma base teórica mais robusta que aponte para a compreensão e mensuração das dimensões desta categoria, para a identificação de seus indicadores e fatores que revelem as dificuldades do aluno em permanecer na universidade, face às exigências e contingências sociais mais amplas, constituindo-se em importante instrumento de gestão e avaliação da educação superior.

Em estudo recente (Castelo Branco; Jezine; Nakamura, 2016:261), apontamos que

*«[...] a mão (do sistema educacional) que, aparentemente, acolhe, é a mesma que expulsa o aluno dos seus bancos escolares, não lhe dando as mínimas condições de acolhimento, de permanência e de formação, sobretudo àqueles cujo perfil socioeconômico não lhes favorece a conclusão do curso».*

Contrariando a lógica dos estudos, optamos por estudar os que determinantes potenciais da futura evasão ou interrupção do ciclo de estudos do aluno, ao invés de analisar as estatísticas oficiais de abandono, tido como realidade ou constatação fatal. Para tanto, tratar a categoria permanência é discutir o processo que vai do acesso ao sistema de ensino e tem seu fechamento quando o aluno conclui o curso, mesmo tendo que encarar a discussão da retenção de alunos que se encontram no sistema de ensino *ad infinitum*. Compreender esse processo constitui-se em fonte de pesquisa fundamental para abortar a perda de alunos do sistema educacional, que abandonam ou simplesmente passam a não mais figurarem nos bancos escolares, partindo para outras opções de vida, deixando um vazio e um sentimento de dever não cumprido em si mesmos e nos gestores responsáveis, por não terem cumprido com a função social da educação: formar com sucesso!

Alguns autores apontam alguns fatores que contribuem para a saída do sistema educacional que, entendemos, são também definidores para a permanência do aluno nesse mesmo sistema: a falta de recursos financeiros (Zago, 2006); questões de ordem acadêmica, que contemplam expectativas em relação ao

curso ou instituição que podem motivar ou desmotivar o aluno, tornando a conclusão do curso uma prioridade ou não para sua vida (Silva Filho *et al.*, 2007); a falta de aquisição de «capital cultural» não apenas na vida escolar, desde à educação básica, mas ao longo de sua «trajetória» de vida (Gisi, 2006; Baggi; Lopes, 2011); ou a junção de múltiplos fatores de natureza pessoal, sociocultural e econômica (Kira, 1998; Braga; Peixoto; Bogutchi, 2003). Há, ainda, autores que, além dos fatores próprios aos estudantes (externos à instituição), tais como variáveis econômicas, sociais, culturais ou individuais, acrescentam fatores relacionados às instituições ou que possuem um caráter mais interno a estas, tais como a estrutura e dinâmicas de cada curso (Ronsoni, 2014). E afirmam:

*«Estes fatores podem ser classificados como os que se relacionam ao próprio estudante, os relacionados ao curso e à instituição e, finalmente, os fatores socioculturais e econômicos externos. Não é possível analisá-los de forma dissociada, pois fatores de uma ordem interferem necessariamente nos demais».* (Ronsoni, 2014:26).

Intencionando analisar esses múltiplos fatores, alguns estudos mostram que, ao lado da maior presença de alunos oriundos de camadas populares (mais de 90%), as instituições passam a conviver com o desafio em fazê-los permanecer, diante das dificuldades dos alunos em conseguir acompanhar ou ter prosseguimento nos cursos, em conciliar trabalho e estudo, e, finalmente, em sobreviver diante das tímidas políticas de assistência estudantil de suas instituições (Zago; Pereira; Paixão, 2015).

Sem o reconhecimento dessas condições de desigualdade,

*«[...] haverá muitos alunos incluídos no sistema escolar, mas poucos irão realmente se apropriar do conhecimento que o processo de ensino e aprendizagem exige. A permanência do aluno na IES, portanto, também depende do suporte pedagógico disponibilizado por esta, no entanto encontramos muitas instituições particulares e públicas despreparadas para estes desafios. A implementação e o acompanhamento de políticas públicas educacionais, tendo por base a igualdade de oportunidades de acesso, é uma condição necessária, mas não é a única para que ocorra a democratização efetiva nas IES particulares para combater a evasão».* (Baggi; Lopes, 2011:357).

Corroborando com as ideias das autoras, não se pode falar em equidade na educação superior considerando apenas a ampliação do acesso, sem tornar a permanência e a aquisição de conhecimentos como determinantes para a viabilização da apropriação de capital cultural, para a consequente conclusão do curso com a qualidade desejada.

*«Neste sentido, a equidade representa a intensidade em que os alunos podem se beneficiar da educação e da formação, através de acesso, oportunidades, acompanhamentos e resultados».* (Oliveira; Silveira, 2011:3).

### 3. A construção do instrumento «escala para avaliação da permanência discente»

Para a construção da Escala para Avaliação da Permanência Discente (Nakamura; Castelo Branco; Jezine, 2016) levamos em consideração, como critérios para definição dos itens, os elementos apontados por Zago (2006), Silva Filho *et al.*, (2007), Gisi (2006), Baggi; Lopes (2011), Kira (1998), Braga; Peixoto; Bogutchi (2003), Ronsoni (2014), Zago; Pereira; Paixão (2015), Baggi; Lopes (2011), Oliveira; Silveira (2011), dividido em seis dimensões, abrangendo desde aspectos antecedentes ao ingresso dos alunos na educação superior, até aspectos de natureza acadêmica, incluindo programas de bolsas de pesquisa e programas assistenciais, ofertados pela instituição.

O instrumento de coleta de dados consiste, portanto, de um questionário elaborado pelos autores e é composto por itens que visam traçar o perfil do estudante cujas respostas são variáveis do tipo nominais, ordinais e escalares, e por itens que visam avaliar a permanência, elaborados com escalas do tipo Likert.

As variáveis escolhidas para o perfil do estudante são: idade, sexo, cor/raça, estado civil, tipo de escola em que cursou o ensino fundamental, tipo de escola em que cursou o ensino médio, turno em que cursou o ensino médio, ano em que concluiu o ensino médio, renda familiar, trabalho remunerado, ocupação do pai, ocupação da mãe, profissão do pai, profissão da mãe, nível de instrução do pai, nível de instrução da mãe, desempenho escolar, frequência a cursinho pré-vestibular e dificuldade de acesso ao ensino superior.

As proposições consistem de afirmações e as respostas em escalas do tipo Likert, muito utilizadas para mensurar atitudes e comportamentos. Tanto as atitudes quanto as opiniões carecem de ação, e isto dificulta a sua métrica: atribuição de números a parâmetros descritores de objetos, ou acontecimentos ou situações, de acordo com certa regra (Kaplan,1975:182). Likert (1976, apud Sanches; Meireles; Di Sordi, 2011:2) afirma que o instrumento de medida proposto por ele pretende «verificar o nível de concordância do sujeito com uma série de afirmações que expressem algo favorável ou desfavorável em relação a um objeto psicológico».

As proposições foram elaboradas visando aferir as seguintes dimensões ou fatores: externa ou convicção prévia, permanência, estrutura, programas acadêmicos, programas de assistência e conclusão do curso. Para cada proposição, é solicitado que o estudante assinale o grau de concordância com 4 níveis, ou seja, discordância total, discordância parcial, concordância parcial e concordância total. O ponto de neutralidade foi suprimido para evitar as respostas com a tendência central, nem discorda nem concorda, forçando a resposta para uma das posições. É permitido deixar a resposta em branco, caso não seja pertinente ou caso não se sinta em condições de se posicionar.

A dimensão externa ou convicção prévia é composta de proposições que se referem às motivações e convicções anteriores ao ingresso no curso, composta das seguintes afirmações:

- Eu escolhi o curso porque era mais fácil entrar.
- Eu escolhi o curso porque me identifico com o curso.
- Eu escolhi o curso porque me prepara para o mercado de trabalho.
- Eu escolhi o curso porque me proporciona prestígio social.
- Eu estou no curso que sempre desejei cursar.

A dimensão permanência é composta de proposições que se referem às situações atuais no curso que apontam para a sua permanência e continuidade do curso, como seguem:

- Eu pretendo continuar matriculado no curso.
- Eu continuo no curso porque a família ajuda.
- Eu continuo no curso porque estou num programa acadêmico de bolsas de estudo.
- Eu continuo no curso porque estou inserido no programa de assistência estudantil.
- Eu continuo no curso por força de vontade.
- Eu continuo no curso porque me proporciona boa perspectiva profissional
- Eu continuo no curso porque proporcionará minha ascensão social.

- Os professores do curso motivam o aluno a permanecer e concluir o curso
- De maneira geral, eu estou satisfeito com o curso.

A dimensão estrutura é composta de proposições que se referem à estrutura da instituição que afeta diretamente o dia a dia do estudante, como seguem:

- As salas de aula são confortáveis
- Os laboratórios são bem equipados e confortáveis
- Sempre encontro o que preciso nas bibliotecas da Universidade/Faculdade

A dimensão programas acadêmicos é composta de proposições que se referem a bolsas de pesquisas no sentido da permanência, devendo ser submetidos as proposições apenas os que estão ou foram contemplados com este recurso, como seguem:

- Minha participação no programa acadêmico contribuiu ou contribuirá para qualificação profissional
- Minha participação no programa acadêmico contribuiu ou contribuirá para produção acadêmica
- Minha participação no programa acadêmico contribuiu ou contribuirá para formação crítico social
- Minha participação no programa acadêmico contribuiu ou contribuirá para dar continuidade nos estudos
- O aspecto que desmotivou ou desmotiva a minha participação em programa acadêmico é o valor da bolsa de estudos.
- O aspecto que desmotivou ou desmotiva a minha participação em programa acadêmico é a orientação inadequada.
- O aspecto que desmotivou ou desmotiva a minha participação em programa acadêmico é a complexidade no desenvolvimento da pesquisa.
- O aspecto que desmotivou ou desmotiva a minha participação em programa acadêmico é a carga horária excessiva.

É solicitado, ainda, que o estudante registre o tipo do programa acadêmico e a situação atual no programa.

A dimensão programas de assistência é composta de proposições que se referem aos programas assistenciais oferecidos pela instituição, tais como, alimentação, moradia, transporte, etc., devendo, também, ser submetidos às proposições apenas os que estão contemplados com um ou mais destes recursos, ou ainda os que tenham se candidatado ou submetido à seleção para o recurso.

- Os programas de assistência estudantil são amplamente divulgados.
- É fácil e simples candidatar-se a um programa de assistência.
- Os recursos fornecidos pelos programas são eficientes.
- Os recursos fornecidos pelos programas são suficientes.

É solicitado, ainda, que o estudante registre o(s) tipo(s) do(s) programa(s) de assistência em que está contemplado.

A dimensão conclusão do curso é composta de proposições que visam identificar as condições principais que permitiram chegar ao final do curso, devendo ser submetidos a essas proposições apenas os que estão nos últimos períodos do curso.

- A minha permanência no curso foi favorecida pela ajuda da família.
- A minha permanência no curso foi favorecida por estar num programa acadêmico de bolsa de estudos.
- A minha permanência no curso foi favorecida por ter sido contemplado com assistência estudantil.

A validação da escala foi realizada por dois critérios: a semântica em que o instrumento é posto para ser avaliado por uma equipe de especialistas que analisam os itens quanto à coerência, pertinência e relevância e a consistência que é feita, neste trabalho, utilizando a teoria da lógica paraconsistente anotada de dois valores (LPA2v).

Da Costa *et al.*, (1999:37) afirmam que

*«A lógica paraconsistente pode ser aplicada para modelar conhecimentos por meio de procura de evidências, de tal forma que os resultados obtidos são aproximados do raciocínio humano. [...] A lógica paraconsistente pode modelar o comportamento humano e assim ser aplicada em sistemas de controle, porque se apresenta mais completa e mais adequada para tratar situações reais, com possibilidades de, além de tratar inconsistências, também contemplar a indefinição».*

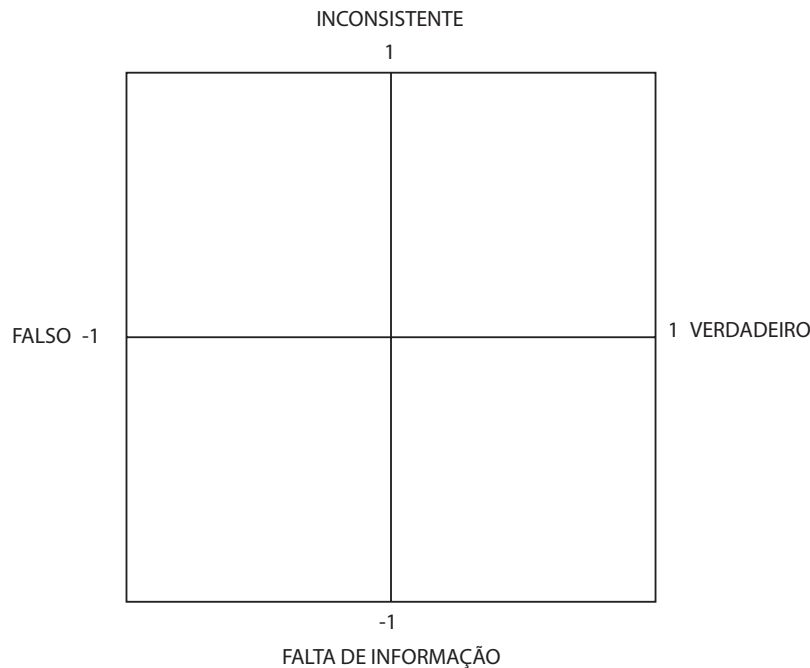
Este tipo de análise confere ao trabalho robustez demonstrativa. Em Da Costa (1999) tem-se que na lógica paraconsistente (LP) «as anotações são representativas de graus de crença e descrença atribuídos à proposição, dando-lhe conotações de valoração». Nesse sentido,

*«O método consiste em estabelecer as proposições e parametrizá-las de forma a poder “isolar os fatores de maior influência nas decisões e, por meio de especialistas, obter anotações para esses fatores, atribuindo-lhes um grau de crença ( $\mu_c$ ) e um grau de descrença ( $\mu_d$ )”, é importante observar que esses valores são independentes e podem variar de 0 a 1».* (Carvalho, 2002).

As Lógicas Paraconsistentes nasceram da necessidade de se encontrar meios de dar tratamento às situações contraditórias e a Lógica Paraconsistente Anotada de anotação com dois valores-LPA2v mostra-se adequada quando se utiliza o conhecimento de evidências para ajudar a resolver conflitos, quando várias inferências inconsistentes ocorrem.

De acordo com Da Costa (1999: 19), as lógicas anotadas constituem uma classe de lógicas paraconsistentes e acham-se relacionadas a um reticulado completo denominado QUPC- Quadrado Unitário do Plano Cartesiano, dividido em quatro partes, duas na vertical e duas na horizontal, como mostra a figura 1, conforme Silva Filho, J. I. (2006: 21). Na vertical, a parte superior indica inconsistência decorrente de informação conflituosa e a parte inferior indeterminação por falta de informação. Neste caso, a faixa central é a ideal, denotando informação não-inconsistente e não-conflitante. Na horizontal, o plano cartesiano divide-se à esquerda e à direita de um eixo central: a esquerda denota a ocorrência de falsidade e a direita a ocorrência de verdade.

**Figura 1. Quadrado unitário do Plano Cartesiano (QUPC)**

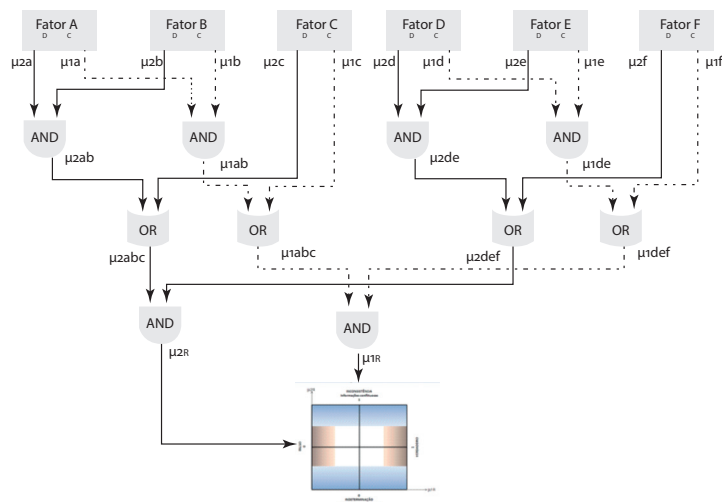


Fonte: Silva Filho, J. I. (2006).

No QUPC, estarão dispostos valores do grau de certeza  $G_1 = \mu_{1R} - \mu_{2R}$  e grau de contradição  $G_2 = \mu_{1R} + \mu_{2R} - 1$  onde  $\mu_{1R}$  é o grau de crença resultante e  $\mu_{2R}$  o grau de descrença resultante, obtidos através dos graus de crença e descrença,  $\mu_1$  e  $\mu_2$ , de cada um dos fatores ou dimensão, usando a rede lógica com conectivos OR e AND, como descrita por Da Costa *et al.*, (1999).

O grau de certeza  $G_1$  próximo de 1 indica que as respostas às proposições foram concordantes, levando a resultado verdadeiro do fenômeno em estudo e o contrário, próximo a -1, a resultado falso ou discordantes. O grau de contradição  $G_2$ , próximo a 1, indica grau muito alto de contradição e próximo a -1, indeterminação ou falta de informação. Em relação ao grau de contradição, um valor próximo a zero indica consistência e adequação das proposições.

**Figura 2. Redes lógicas para a conversão de crença  $\mu_1$  e descrença  $\mu_2$  em um grau de certeza ( $G_1 = \mu_{1R} - \mu_{2R}$ ) e grau de contradição ( $G_2 = \mu_{1R} + \mu_{2R} - 1$ ) com conectivos OR e AND**



Fonte: Da Costa *et al.*, (1999).

Os graus de certeza e de contradição são registrados no QUPC, permitindo a interpretação dos dados como um todo.

Para a validação semântica, o questionário foi submetido a especialistas que trabalham o tema da permanência no ensino superior, no subgrupo 5 da Rede Universitas/BR. Os pesquisadores são da Universidade Federal do Mato Grosso (UFMT), Universidade Federal da Grande Dourados (UFGD), da Universidade Estadual do Mato Grosso (UNEMAT) e Universidade Católica Dom Bosco (UCDB).

A equipe foi composta pelos seguintes especialistas: Profa. Dra. Tereza Christina M.A. Veloso, professora do Curso de Graduação em Nutrição; Profa. Dra. Patrícia Simone Nogueira, professora do Curso de Graduação em Nutrição; Prof. Luiz Zardo, professor do Curso de Estatística; Jhessica Kirch, estudante no 7º semestre do curso de Estatística; Roseli Nunes, técnica do Sistema de Gerenciamento Acadêmico, todos da UFMT; Profa. Dra. Giselle Cristina Martins Real, doutora em Educação; Profa. Dra. Eugênia Portela de Siqueira Marques, doutora em Educação; Ms. Jonas de Paula Oliveira, doutorando em Ensino de Ciências e Ms. Aline Anjos da Rosa, mestre em Educação da UFGD; Prof. Dr. Fernando Cezar Vieira Malange, doutor em Engenharia Elétrica; Profa. Dra. Elizeth Gonzaga dos Santos Lima, doutora em Educação, Profa. Dra. Heloisa Salles Gentil, doutora em Educação da UNEMAT e Dra. Ana Luisa Alves Cordeiro, doutora em Educação pela UCDB.

As avaliações solicitadas constaram de três itens para cada uma das questões e proposições: clareza, pertinência e relevância. Solicitou-se ao avaliador que atribuisse uma pontuação com valores de 0 a 5 para cada item e observações ou sugestão que pudessem contribuir para uma melhor formulação ou mesmo a retirada da questão se for o caso.

Recebidas as análises, procederam-se as adequações e reformulações que foram sugeridas, sendo que as proposições que obtiveram avaliações inferiores a 3 nos quesitos pertinência ou relevância foram retiradas e as que necessitavam maior clareza na formulação foram reescritas, atendendo às sugestões.

A mediana da pontuação atribuída aos itens que compuseram o formato final do questionário, nos quesitos pertinência e relevância, foi 5 pontos, que permite concluir que o instrumento é adequado quanto à coerência e válido para a obtenção das informações acerca da permanência do estudante no curso.

Para a análise da consistência, usou-se dados de um pré-teste com 102 questionários aplicados a estudantes de 3 cursos: Arquivologia e Biblioteconomia da UFPB e Nutrição da UFMT<sup>2</sup>.

As proposições foram formuladas com a finalidade de fornecer uma medida da intensidade da pretensão de permanecer e concluir o curso. Algumas proposições foram formuladas de forma invertida, ou seja, a discordância levaria a uma medida maior de permanência e a concordância, ao contrário. Por esta razão, para facilitar o entendimento, substituiu-se os níveis de concordância ou discordância pela condição favorável ou não à permanência.

Os itens que compõem a escala foram tabulados e os resultados estão dispostos na tabela 01 que segue, composta dos seguintes valores: número de respostas para cada nível, discordante a concordante de cada proposição, renomeada por desfavorável a favorável, total de respostas, a resposta mediana,  $D_p$  = discordante da proposição e  $F_p$  = concordante da proposição e na última linha das proposições que compõem o fator ou dimensão, tem-se  $D_f$  = discordantes do fator,  $F_f$  = Concordantes do fator,  $\mu_1$  =

---

<sup>2</sup> Os dados da UFPB foram coletados pela Profa. Ms. Edileuda Soares Diniz (doutoranda em Educação – PPG/UFPPB) e os dados da UFMT, pela Profa. Dra. Patrícia Simone Nogueira.

crença de que as proposições do fator ou dimensão como um todo sejam verdadeiras e  $\mu_2$  = descrença de que as proposições do fator ou dimensão como um todo sejam verdadeiras.

**Tabela I. Distribuição das respostas por dimensão**

Externa ou prévia	Desfavorável	Parcialmente desfavorável	Parcialmente favorável	Favorável	Total	Mediana	Dp	Fp
Eu escolhi o curso porque era mais fácil entrar.**	6	8	24	63	101	F	14	87
Eu escolhi o curso porque me identifiquei com o curso.	4	13	40	45	102	PF	17	85
Eu escolhi o curso porque me prepara para o mercado de trabalho.	9	24	38	30	101	PF	33	68
Eu escolhi o curso porque me proporciona prestígio social.**	13	11	33	44	101	PF	24	77
Eu estou no curso que sempre desejei cursar.	22	29	24	26	101	PD	51	50
<b>** escala invertida</b>	<b>Df = 139</b>		<b>Ff = 367</b>		<b>506</b>		$\mu_2 = 0,27$	$\mu_1 = 0,73$

Permanência	Desfavorável	Parcialmente desfavorável	Parcialmente favorável	Favorável	Total	Mediana	Dp	Fp
Eu pretendo continuar matriculado no curso.	2	4	12	84	102	F	6	96
Eu continuo no curso porque a família ajuda.	22	21	18	40	101	PF	43	58
Eu continuo no curso porque estou num programa acadêmico de bolsas de estudo.	76	10	7	5	98	D	86	12
Eu continuo no curso porque estou inserido no programa de assistência estudantil.	89	3	1	4	97	D	92	5
Eu continuo no curso porque me proporciona boa perspectiva profissional.	8	15	39	40	102	F	23	79
Eu continuo no curso porque proporcionará minha ascensão social.	14	31	35	22	102	PF	45	57
Os professores do curso motivam o aluno a permanecer e concluir o curso	6	30	45	21	102	PF	36	66
De maneira geral, eu estou satisfeito com o curso	5	14	43	40	102	PF	19	83
	<b>Df = 350</b>		<b>Ff = 456</b>		<b>806</b>		$\mu_2 = 0,43$	$\mu_1 = 0,57$

Estruturais	Desfavorável	Parcialmente desfavorável	Parcialmente favorável	Favorável	Total	Mediana	Dp	Fp
As salas de aula são confortáveis.	35	37	26	4	102	PD	72	30
Os laboratórios são bem equipados e confortáveis.	30	39	24	9	102	PD	69	33
Sempre encontro o que preciso nas bibliotecas da UFMT.	19	30	40	13	102	PD	49	53
<b>** escala invertida</b>	<b>Df = 190</b>		<b>Ff = 116</b>		<b>306</b>		$\mu_2 = 0,62$	$\mu_1 = 0,38$

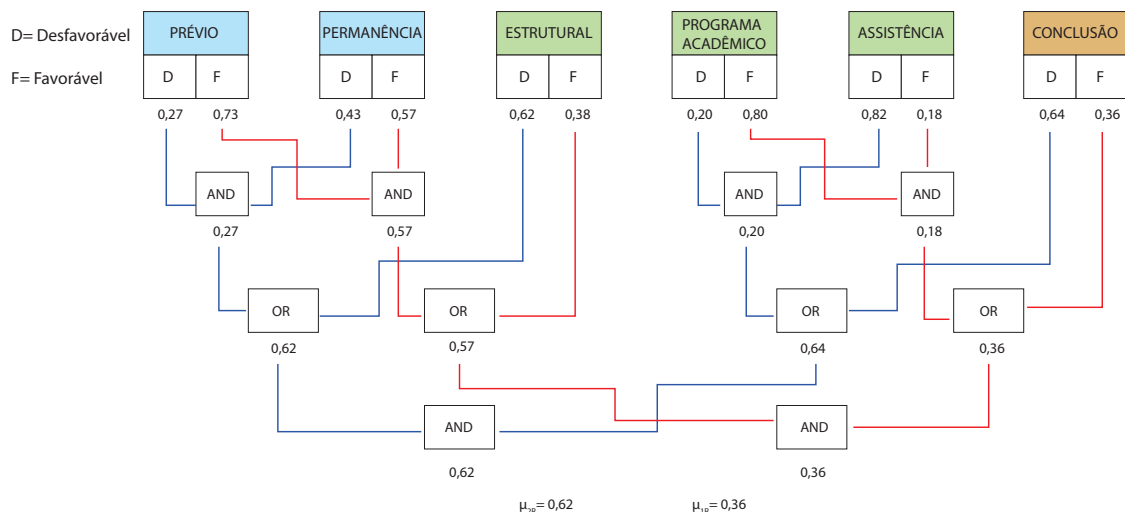


Programas Acadêmicos	Desfavorável	Parcialmente desfavorável	Parcialmente favorável	Favorável	Total	Mediana	Dp	Fp
Minha participação no programa acadêmico contribuiu ou contribuirá para qualificação profissional.	5	3	15	44	67	F	8	59
Minha participação no programa acadêmico contribuiu ou contribuirá para produção acadêmica.	4	6	17	39	66	F	10	56
Minha participação no programa acadêmico contribuiu ou contribuirá para formação crítico social.	4	4	14	44	66	F	8	58
Minha participação no programa acadêmico contribuiu ou contribuirá para dar continuidade nos estudos.	4	3	17	4	65	F	7	58
O aspecto que desmotivou ou desmotivou a minha participação em programa acadêmico é o valor da bolsa de estudos.**	8	8	23	28	67	PF	16	51
O aspecto que desmotivou ou desmotivou a minha participação em programa acadêmico é a orientação inadequada.**	6	14	15	33	68	PF	20	48
O aspecto que desmotivou ou desmotivou a minha participação em programa acadêmico é a complexidade no desenvolvimento da pesquisa.**	5	9	24	30	68	PF	14	54
O aspecto que desmotivou ou desmotivou a minha participação em programa acadêmico é a carga horária excessiva.**	4	19	30	15	68	PF	23	45
<b>** escala invertida</b>	<b>Df = 106</b>		<b>Ff = 429</b>		<b>535</b>		$\mu_2 = 0,20$	$\mu_1 = 0,80$
Programas de Assistência	Desfavorável	Parcialmente desfavorável	Parcialmente favorável	Favorável	Total	Mediana	Dp	Fp
Os programas de assistência estudantil são amplamente divulgados.	24	18	12	7	61	PD	42	19
É fácil e simples candidatar-se a um programa de assistência.	35	21	4	1	61	D	56	5
Os recursos fornecidos pelos programas são eficientes.	17	30	8	1	56	PD	47	9
Os recursos fornecidos pelos programas são suficientes.	27	21	6	2	56	PD	48	8
<b>** escala invertida</b>	<b>Df = 193</b>		<b>Ff = 41</b>		<b>234</b>		$\mu_2 = 0,82$	$\mu_1 = 0,18$
Conclusão	Desfavorável	Parcialmente desfavorável	Parcialmente favorável	Favorável	Total	Mediana	Dp	Fp
A minha permanência no curso foi favorecida pela ajuda da família.	7	9	16	43	75	F	16	59
A minha permanência no curso foi favorecida por estar num programa acadêmico de bolsa de estudos.	50	7	7	4	68	D	57	11
A minha permanência no curso foi favorecida por ter sido contemplado com assistência estudantil.	63	0	3	2	68	D	63	5
<b>** escala invertida</b>	<b>Df = 136</b>		<b>Ff = 75</b>		<b>211</b>		$\mu_2 = 0,64$	$\mu_1 = 0,36$

Fonte: Dados do estudo-piloto realizado pelos autores.

Aplicando a rede lógica com 6 fatores conforme a figura 2 onde, de acordo com Carvalho (2002), no conectivo OR, a saída é o maior valor das duas entradas e, no conectivo AND, a saída é o menor valor das duas entradas, tem-se os graus de crença resultante  $\mu_{1R}$  e o de descrença resultante  $\mu_{2R}$ , como apresentado na figura 3.

Figura 3. Rede lógica



Fonte: Figura construída pelos autores com base em Da Costa *et al.*, (1999).

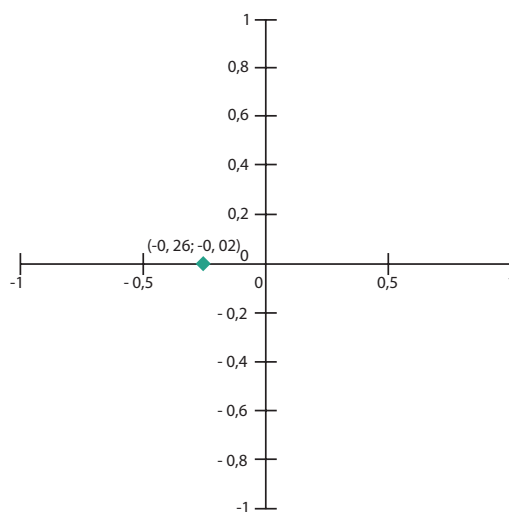
O grau de crença resultante  $\mu_{1R}$  encontrado foi 0,36 e o de descrença resultante  $\mu_{2R}$ , 0,62. Têm-se então os respectivos graus de certeza  $G_1$  e o de contradição  $G_2$ , como seguem:

$$G_1 = \mu_{1R} - \mu_{2R} = 0,36 - 0,62 = -0,26$$

$$G_2 = \mu_{1R} + \mu_{2R} - 1 = 0,36 + 0,62 - 1 = -0,02$$

A figura 4 mostra estes valores dispostos no QUPC.

Figura 4. . QUPC – Quadrado Unitário do Plano Cartesiano para interpretação do resultado



Fonte: Figura construída pelos autores com base em Da Costa *et al.*, (1999).

Conforme descrito anteriormente, o grau de contradição  $G_2$  próximo de 0 indica que as proposições não produziram contradições e também foram suficientes, ou seja, não se detectou falta de informação. Assim, a escala proposta, segundo a análise por LPA2v, é consistente.

O grau de certeza  $G_1$  pode ser interpretado como tendendo a falso, ou seja, no sentido da discordância com as proposições. Para este grupo de estudantes que participaram do pré-teste, predomina a discordância das proposições, o que indica para a situação não favorável à permanência.

#### 4. Considerações finais

A partir dos dados apresentados, construímos um percurso teórico-metodológico importante para a futura, em futuro próximo, de estudos de caso, aplicados em IES brasileira e estrangeiras, dentro da perspectiva da educação comparada, que permitam aprofundar a compreensão da categoria *permanência*, apresentando indicadores eficientes para subsidiar ações e intervenções da gestão universitária e das práticas docentes, no sentido de produzir políticas de acolhimento e inclusão para alunos em situação de vulnerabilidade, que lhes permitam ingressar e concluir seus estudos.

Lembramos, ainda, que a Escala construída integrará o projeto «Políticas de Educação Superior: os desafios da inclusão, acesso e permanência no contexto da expansão das Universidades Federais», projeto este inserido na Rede Universitas/BR, iniciando sua fase de aplicação em várias universidades brasileiras que integram a Rede.

A Escala para Avaliação da Permanência Discente (Nakamura; Castelo Branco; Jezine, 2016) foi construída, tomando-se como critérios para definição dos itens, os elementos apontados por Zago (2006), Silva Filho *et al.*, (2007), Gisi (2006), Baggi; Lopes (2011), Kira (1998), Braga; Peixoto; Bogutchi (2003), Ronsoni (2014), Zago; Pereira; Paixão (2015), Baggi; Lopes (2011), Oliveira; Silveira (2011), dividido em seis dimensões que abrangem desde aspectos antecedentes ao ingresso dos alunos até aspectos de natureza acadêmica, incluindo programas de bolsas de pesquisa e assistenciais, ofertadas pela instituição.

O processo de validação efetuado, em relação à análise de pertinência e relevância, foi realizado com a análise de 13 especialistas qualificados (*expertises*), e considerou a Escala adequada e coerente, com *mediana* igual a 5 (nível mais alto da escala), para a obtenção de informações relativas à permanência do estudante em cursos de graduação, assim como poderá ser adaptada para outros contextos da educação, em níveis básico ou médio de estudos.

Em relação ao uso da Lógica Paraconsistente (Da Costa, 1999; Abe, 1992; Subrahmanian, 1987), para análise da consistência do instrumento, aplicamos o estudo-piloto a 102 estudantes da UFPB e da UFMT, abrangendo três cursos de graduação. A análise de consistência constatou que a escala proposta é consistente, segundo a LPA2v, ou seja, as proposições da escala não produziram contradições e foram suficientes para avaliar a permanência, não tendo sido considerada como insuficiente em informações apresentadas.

Após esse processo de validação, iniciaremos a fase de estudos comparativos entre as IES, podendo investigar a categoria *permanência* em instituições do sistema público e privado nacional e internacional, o que, certamente, contribuirá para uma construção de um banco de dados robusto, permitindo a ampliação da compreensão teórico-metodológica acerca da categoria *permanência* na educação.

Espera-se, com este instrumento, oferecer condições e informações às IES para auxiliar os gestores do sistema educacional e a comunidade acadêmica a envidar esforços para minimizar ou até mesmo sanar os enormes obstáculos existentes para a permanência de alunos em situação de vulnerabilidade social e econômica.

## Referências bibliográficas

- Abe, Jair M. (1992). *Fundamentos da lógica anotada*. Tese (Doutorado em Filosofia). Ciências e Letras da Universidade de São Paulo. Departamento de Filosofia da Faculdade de Filosofia.
- Baggi, Cristiane Aparecida dos Santos; Lopes, Doraci Alves (2011): “Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica”. *Avaliação, Campinas*, 16 (2), 355-374.
- Bittar, Mariluce; Oliveira, João Ferreira de; Morosini, Marília (orgs.) (2008). Educação superior no Brasil - 10 anos pós-LDB. Brasília: Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Disponível em: <<http://www.oei.es/pdf2/educacao-superior-brasil-10-anos.pdf>>, acesso em 08 abril de 2013.
- Borges, Sandra Marques (2011). *Fatores determinantes da evasão escolar no ensino superior: o estudo de caso do ILES/ULBRA de Itumbiara*. Dissertação de Mestrado, Dissertação de Mestrado Profissional em Desenvolvimento Regional, Goiânia, 2011.
- Braga, M. M.; Peixoto, M. C. L.; Bogutchi, T. F. (2003): “A evasão no ensino superior brasileiro: o caso da UFMG”. *Avaliação, Campinas*, 8 (1), 161-189.
- Brasil (2014). Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Coletiva Censo da Educação Superior 2013. Brasília. Disponível em: <[http://download.inep.gov.br/educacao\\_superior/censo\\_superior/apresentacao/2014/coletiva\\_censo\\_superior\\_2013.pdf](http://download.inep.gov.br/educacao_superior/censo_superior/apresentacao/2014/coletiva_censo_superior_2013.pdf)>, acesso em 22 abril de 2014.
- Brasil (2012). Análise sobre a expansão das universidades federais 2003 a 2012. Brasília: Ministério da Educação. Disponível em: <[http://portal.mec.gov.br/index.php?option=com\\_docman&task=doc\\_download&gid=12386&Itemid=>](http://portal.mec.gov.br/index.php?option=com_docman&task=doc_download&gid=12386&Itemid=>), acesso em 25 setembro de 2014.
- Brasil (1997). *Diplomação, retenção e evasão em cursos de graduação em Instituições de Ensino Superior Públicas*. Brasília: Secretaria da Educação Superior.
- Brasil (1996). Lei n. 9.394, de 20 de dezembro de 1996. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/L9394.htm](http://www.planalto.gov.br/ccivil_03/leis/L9394.htm)>, acesso em 05 janeiro de 2017.
- Carvalho, Fábio Romeu (2002). *Lógica paraconsistente aplicada em tomadas de decisão*. São Paulo: Aleph.
- Castelo Branco, Uyguciara Veloso; Jezine, Edineide (2013): “A expansão da (na) UFPB: avaliando o REUNI (2008 a 2012)”. *Temas em Educação (UFPB)*, 22 (2), 60-82.
- Castelo Branco, Uyguciara Veloso; Jezine, Edineide; Nakamura, Paulo Hideo (2016): “Alguns indicadores de permanência/abandono na educação superior: elementos para o debate” in Chaves, Vera Lúcia Jacob; Maués, Olgaíses Cabral; Hage, Salomão. Mufarrej (orgs.): *Expansão privado-mercantil da educação superior no Brasil*. Campinas: Mercado das Letras.

- Castelo Branco, Uyguciara Veloso; Nakamura, Paulo Hideo (2013): “Reflexões sobre a expansão de vagas na educação superior. A UFPB e os desafios da inclusão” in Jezine, Edineide; Bittar, Mariluce (orgs.): *Políticas de Educação Superior no Brasil: expansão, acesso e igualdade social*. João Pessoa: UFPB.
- Catani, Afrânio Mendes; Hey, Ana Paula; Gilioli, Renato de Sousa Porto (2006): “ProUni: democratização do acesso às instituições de Ensino Superior?”. *Educar, Curitiba*: UFPR, 28, 125-140.
- Da Costa, N.C.; Abe, Jair M.; Murolo, A.C.; Silva Filho, J.I.; Leite, C.F.S. (1999). *Lógica paraconsistente anotada*. São Paulo: Atlas.
- Felicetti, Vera Lucia; Morosini, Marília Costa (2009): “Equidade e iniquidade no ensino superior: uma reflexão”. *Ensaio: Avaliação e Políticas Públicas em Educação*, 17, (62), 9-24.
- Gisi, Maria Lourdes (2006): “A Educação Superior no Brasil e o caráter de desigualdade do acesso e da permanência”. *Diálogo Educacional*, 6, (17), 97-112.
- Kaplan, Abraham (1975): *A conduta na pesquisa: metodologia para as ciências do comportamento*. São Paulo: EDUSP.
- Kira, L.F. (1998): *A evasão no ensino superior: o caso do curso de pedagogia da Universidade Estadual de Maringá (1992-1996)*. Dissertação (Mestrado em Educação) – Programa de Pós-graduação em Educação da Universidade Metodista de Piracicaba, Piracicaba.
- Likert, Rensis (1976): “Una Técnica para la Medición de Actitudes (A technique for the measurement of attitudes, Archives of Psychology, 140, 1-50, 1932)” in Weinerman, Catalina H.: *Escala de Medición en Ciencias Sociales*. Buenos Aires: Nueva Visión.
- Nakamura, Paulo Hideo; Castelo Branco, Uyguciara Veloso; Jezine, Edineide (2013): “UFPB nos anos da expansão (2008 a 2012): quantos e quem somos e o que mudou no perfil dos nossos alunos?”. Anais do XXI Seminário Nacional Universitas/Br: Políticas Públicas para a Educação Superior, Expansão e Internacionalização. São Carlos: UFSCAR.
- Neves, Clarissa Eckert Baeta; Raizer, Leandro; Fachinetto, Rochele Fellini (2007): “Educação superior para todos? Acesso, expansão e equidade: novos desafios para a política educacional”. *Sociologias*, 9 (17), 124-157.
- Oliveira, Adriana Rivoire Menelli; Silveira, André Stein da (2011): “Acesso e permanência – desafios à problemática da evasão na educação superior do Brasil”. Revista UTP, 8 p. Disponível em: <<http://revistas.utp.ac.pa/index.php/clabes/article/view/1229/1247>>, acesso em 29 março de 2017.
- Polydoro, Soely Aparecida Jorge (2000). *O trancamento de matrícula na trajetória acadêmica no universitário: condições de saída e de retorno à instituição*. Tese (Doutorado em Educação). Universidade Estadual de Campinas.
- Ristoff, Dilvo I. (1999): *Universidade em foco: reflexões sobre a educação superior*. Florianópolis: Insular.
- Ristoff, Dilvo I. (2011): *Construindo outra educação: tendências e desafios da educação brasileira*. Florianópolis: Insular.
- Ronsoni, M. L. (2014): “Permanência e evasão de estudantes da UFFS Campus Erechim” in Pereira, T. I. (org.): *Universidade pública em tempos de expansão: entre o vivido e o pensado*. Erechim: Evangraf.

- Sanches, Cida; Meirelles, Manue; De Sordi, José Osvaldo (2011): “Análise Qualitativa Por Meio da Lógica Paraconsistente: Método de Interpretação e Síntese de Informação obtida Por Escalas Likert”. III Encontro de Ensino e Pesquisa em Contabilidade e Administração. João Pessoa, 20 a 22 novembro.
- Santos, Boaventura de Sousa (2005): *A Universidade no século XXI. Para uma reforma democrática e emancipatória da Universidade*. São Paulo: Cortez.
- Sguissardi, Valdemar (2009): *Universidade brasileira no século XXI: desafios do presente*. São Paulo: Cortez.
- Silva Filho, Roberto L. Lobo *et al.*, (2007): “A evasão no ensino superior brasileiro”. *Cadernos de Pesquisa*, 37 (132), 641-659.
- Silva Filho, João Inácio (2006): “Métodos de Aplicações da Lógica Paraconsistente Anotada de anotação com dois valores-LPA2v”. *Revista Seleção documental*, 1 (1) 18-25.
- Subrahmanian, V.S. (1987): “On the semantics of quantitative Logic programs”. Proc. 4 th. IEEE Symposium on Logic Programming, Computer Society press, Washington D.C.
- Velloso, Jacques; Cardoso, Claude Batista (2015): “Evasão na Educação Superior: alunos cotistas e não cotistas na Universidade de Brasília”. ANPED, s.d. Disponível em: <<http://www.anped11.uerj.br/31/jacquesveloso.pdf>> acesso em 11 agosto de 2015.
- Veloso, Tereza Christina Mertens Aguiar ; Silva, Maria das Graças Martins; Beraldo, T. M. (2011). “Expansão no ensino superior noturno em Mato Grosso: um processo democrático?” in Silva, Maria das Graças Martins (org.): *Políticas educacionais: faces e interfaces da democratização*. Cuiabá: Ed. UFMT.
- Zago, Nadir (2006): “Do acesso a permanência no ensino superior: percursos de estudantes universitários de camadas populares”. *Revista Brasileira de Educação*, 11 (32), 226-237.
- Zago, Nadir; Pereira, T. I.; Paixão, L. P. (2015): “Expansão do Ensino Superior: problematizando o acesso e a permanência de estudantes em uma nova Universidade Federal”. 37ª Reunião Nacional da ANPED, UFSC – Florianópolis, 04 a 08 outubro de 2015.

## Notas biográficas

**Uyguaciara Veloso Castelo Branco** es graduada en Psicología por la Universidade Federal da Paraíba (1991). Master en Educación por la Universidade Federal da Paraíba (1995) y Doctora en Historia por la Universidade Federal de Pernambuco (2004). Profesora de la Universidade Federal da Paraíba e investigadora en el área de Historia de la Educación, con énfasis en Educación Superior, acceso y permanencia.

**Paulo Hideo Nakamura** es graduado en Estadística por la Universidade Estadual de Campinas – UNICAMP (1979). Master en Estadística por UNICAMP (1986) y Doctor en Ciencias de la Educación por la Universidad Autónoma de Asunción - PY (2014). Profesor de la Universidade Federal da Paraíba e investigador en Educación Superior, con énfasis en temas de acceso y permanencia.

**Edineide Jezine** es graduada en Pedagogía por la Universidade Federal do Amazonas (1988). Master en Educación por la Universidade Federal da Paraíba (1997), Doctora en Sociología por la Universidade Federal de Pernambuco (2002) y posdoctorado por la Universidade Lusófona de Humanidades e Tecnologias – Lisboa - PT/Universitat de València - ES (2011). Profesora de la Universidade Federal da Paraíba e investigadora en Educación Superior, con énfasis en temas de acceso y permanencia.

# La medida de la investigación en educación y su impacto social: las revistas de educación de Iberoamérica en los índices bibliométricos internacionales<sup>1</sup>

## The Measure of Research in Education and its Social Impact: The Iberoamerican Journals of Education in the International Bibliometric Indices

Armando Alcántara Santuario y Alejandro Márquez Jiménez<sup>2</sup>

### Resumen

Las revistas académicas han sido uno de los vehículos más importantes en la difusión del conocimiento científico, humanístico y artístico. Últimamente han adquirido mayor preeminencia por la globalización y el desarrollo de las tecnologías de la información y la comunicación (TIC). En los años 60 del siglo pasado comenzaron los primeros trabajos cuantitativos relacionados con el análisis numérico y estadístico de las revistas científicas, dando paso a la cienciometría, la infometría y la bibliometría. Estas disciplinas aportaron importantes conclusiones sobre los aspectos medibles de la ciencia. La información disponible comenzó a utilizarse para evaluar la producción, el desempeño y la pertinencia de la política científica de los países. Indicadores como el total de artículos publicados, el número de citas de cada artículo y el «índice de impacto» de las revistas, pasaron a ser considerados criterios de calidad del trabajo científico. Con ello se transformaron las funciones sociales que inicialmente cubrían las revistas científicas (la rápida difusión del conocimiento científico, la protección de la autoría de los investigadores y de la veracidad de los nuevos descubrimientos) y se volvieron instrumentos para medir la productividad de los investigadores, las instituciones y los países. En este artículo se presentan los resultados preliminares de un proyecto para analizar los efectos que produce en las revistas académicas de los países de Iberoamérica el nuevo uso que se asigna a la información bibliométrica. Los hallazgos muestran que son muy pocas las revistas científicas en educación publicadas en Iberoamérica que están incluidas en los principales índices internacionales.

### Palabras clave

Educación, iberoamérica, conocimiento, investigación educativa, bibliometría, revistas científicas.

### Abstract

Academic journals have been one of the most important vehicles in the dissemination of scientific, humanistic and artistic knowledge. Recently they have gained more prominence due to globalization and to the development of information and communication technologies (ICT). The 1960's was the beginning of the first quantitative work related to the numerical and statistical analysis of scientific journals, giving way to scientometry, infometrics and bibliometrics. These disciplines provided important insights on the measurable aspects of science. The available information began to be used to assess the production, performance and relevance of the countries' scientific policy. Indicators such as the total number of articles published, the number of citations of each article and the «impact index» of the journals, became important criteria for the quality of scientific work. These new elements transformed the social functions that initially covered the scientific journals (the rapid diffusion of scientific knowledge, protection of the researchers' authorship (copy rights) and the veracity of new discoveries), and became instruments for measuring the productivity of researchers, institutions and countries. This article presents the preliminary results of a project aimed at analyzing the effects of the new use of bibliometric information in the academic journals of Iberoamerica countries. These findings show that very few scientific journals in education published in Iberoamerica are included in the main international indexes.

### Key words

Education, iberoamerica, knowledge, educational research, bibliometrics, scientific journals.

Recibido: 12-03-2017  
Aceptado: 02-05-2017

1 La elaboración de este texto contó con el apoyo del Proyecto PAPIIT IN302116 («La producción científica en educación en Iberoamérica y su impacto social: un análisis de las revistas de investigación educativa»). Una primera versión de este artículo fue presentada en la Conferencia SPCE-SEC 2016. «La Educación Comparada más allá de los números: contextos locales, realidades nacionales y procesos transnacionales», celebrada en Lisboa, del 25 al 27 de enero. Los autores agradecen el apoyo invaluable de Daniel A. Márquez, estudiante de Sociología de la FCPS de la UNAM en la captura y procesamiento estadístico de las bases de datos de la Web of Science, SCOPUS y SciELO.

2 Instituto de Investigaciones sobre la Universidad y la Educación. Universidad Nacional Autónoma de México, aralsantuario@gmail.com, alejandro\_marquez08@yahoo.com.mx

*No hay elección científica —elección del área de investigación, elección de los métodos empleados, elección del lugar de publicación, elección que describe Hagstrom entre una publicación rápida de los resultados parcialmente verificados o la publicación tardía de resultados plenamente controlados— que que no sea, por uno de sus aspectos, al menos confesado y el menos confesable, una estrategia política de ubicación al menos objetivamente orientada hacia la maximización del beneficio propiamente científico, es decir, al reconocimiento susceptible de ser obtenido de los pares-competidores.*

**Pierre Bourdieu**

## 1. Introducción

Uno de los fenómenos más apreciables de la época actual es el crecimiento explosivo de las actividades científicas y tecnológicas. Los datos disponibles señalan que hoy los ritmos de crecimiento no tienen precedente en la historia humana, si bien las desigualdades se siguen manteniendo. De un modo semejante a lo que ocurre en el terreno económico, también en el mundo de las disciplinas científicas existen enormes concentraciones de recursos y talentos en unos pocos países e instituciones que dominan la generación y difusión del conocimiento. En los procesos anteriores, la creciente globalización y el desarrollo de las nuevas tecnologías de la información y la comunicación han servido como poderosas plataformas. En este contexto, los medios para difundir los productos de la investigación científica y humanística también han mostrado un crecimiento exponencial.

Las formas convencionales de difusión a través de las revistas impresas se han revolucionado con la digitalización que ha incrementado su presencia en el ciber espacio en un grado mucho mayor. Así, un artículo impreso al que solo tenían acceso unos cuantos individuos, al circular por todo el mundo a través de los medios electrónicos, ha aumentado de manera casi ilimitada las posibilidades de acceso a cualquier persona en cualquier lugar del planeta.

Es evidente también que el proceso de elaboración y gestión se ha vuelto más complejo pues requiere de la participación de personal especializado en el diseño de materiales digitales. Asimismo, con frecuencia los recursos financieros para sostener los gastos de la publicación electrónica son elevados y para que una revista pueda sostenerse necesita de apoyos económicos constantes y suficientes, que solo pueden dar las instituciones editoriales. A esta situación se agrega el que las revistas de mayor prestigio en la mayoría de los campos científicos se editan en inglés, la *lingua franca* del mundo moderno y que para publicar en esas revistas —las de mayor prestigio y circulación— los académicos deben enviar sus trabajos escritos en ese idioma, escritos correctamente y cumplir además con los estrictos procesos de dictaminación.

En este trabajo se presentan los primeros resultados de un proyecto de investigación que pretende mostrar los efectos sociales de las publicaciones científicas en educación editadas en Iberoamérica y también el lugar que ocupan en los índices internacionales como el *Web of Science*, *Scopus* y *SciELO*. Para ello, el texto se ha dividido en cuatro apartados. En el primero se ofrece un panorama actualizado del desarrollo científico mundial y la manera en que las revistas científicas se insertan en él. La segunda parte se dedica a revisar los antecedentes y la evolución de la bibliometría entre las disciplinas científicas. La tercera sección incluye la discusión de algunas implicaciones sociales del uso de la bibliometría en la evaluación de las labores científicas. Finalmente en el cuarto apartado se presenta un análisis de las revistas iberoamericanas de educación dentro de los índices internacionales mencionados arriba.



## 2. Las publicaciones científicas en el contexto del desarrollo científico mundial

En los primeros tres lustros del siglo XXI las tendencias del desarrollo científico y tecnológico muestran diversos niveles de desarrollo en las capacidades de investigación de las naciones. Este ha sido el caso de los diversos indicadores como la inversión en investigación y desarrollo (I+D), el número de personas dedicadas de tiempo completo a las labores científicas y las publicaciones científicas. En el panorama de algunos países de América Latina, España y Portugal es posible apreciar progresos incipientes y altibajos, situaciones que con frecuencia tienen su explicación en las variaciones de sus economías y también en las políticas públicas que deciden el monto de las inversiones en actividades de ciencia y tecnología. En este apartado se describe el comportamiento de algunos de los indicadores mencionados de acuerdo con datos recientes de la UNESCO, publicados en el más reciente informe sobre la ciencia, cuyo horizonte se traza en el año 2030 (UNESCO, 2015b).

Dentro del contexto del comportamiento económico de los años recientes, el reporte señala que al haberse terminado el auge de los productos básicos, América Latina está en busca de nuevas estrategias de crecimiento. Aunque en la última década la región ha conseguido reducir sus elevados niveles de desigualdad económica, con la caída de la demanda internacional de materias primas las tasas de crecimiento latinoamericanas han comenzado a estancarse e incluso, en algunos casos, a contraerse. En el caso de los países europeos, los niveles de desarrollo han sido mayores y más consistentes aunque el problema del desempleo sigue siendo un factor preocupante. Los países de América Latina no carecen de iniciativas políticas ni de estructuras institucionales sofisticadas para promover la ciencia y la investigación. Estos países han dado grandes pasos en cuanto al acceso a la enseñanza superior, la movilidad científica y la producción científica. No obstante, pocos de ellos parecen haber aprovechado el auge de los productos básicos para volcarse en una competitividad impulsada por la tecnología.

Un caso que llama la atención dentro de este panorama es el de Brasil, quien después de haber conseguido altos niveles de crecimiento económico al final de la primera década de este siglo, en 2014 entró en recesión. Pese a que el país incrementó el acceso a la educación superior en los últimos años y creció el gasto social, la productividad de la mano de obra siguió en niveles reducidos. Lo anterior implica que hasta hoy Brasil no ha conseguido aprovechar la innovación para el crecimiento económico. A ello se añaden la grave crisis política por la que atraviesa desde 2016 con motivo de la polémica destitución (*impeachment*) de la presidenta Dilma Rousseff. Sin embargo, el reporte de la UNESCO considera que, frente a los retos del futuro la región latinoamericana podría estar bien posicionada para desarrollar un modelo de excelencia científica capaz de apuntalar el conocimiento ecológico, combinando para ello sus ventajas naturales en el ámbito de la diversidad biológica en sus principales puntos fuertes en materia de conocimiento indígena (UNESCO, 2015b).

Volviendo al tema de los indicadores del desarrollo científico y tecnológico, a nivel global la distribución de la inversión en conocimiento es desigual. Los Estados Unidos (EE. UU.) ocupan una posición predominante (28% de la inversión global en investigación y desarrollo, I+D). China ocupa el segundo lugar con el 20%, por delante de la Unión Europea (UE) (19%) y el Japón (10%). El resto de los países, que representa el 67% de la población mundial, realiza solo el 23% de la inversión mundial en I+D. Para algunos países iberoamericanos como Argentina, Brasil, México, España y Portugal, el panorama de los indicadores anteriores es el que aparece enseguida:

**Tabla I. Indicadores de población, PIB y GBID de países de Iberoamérica**

País	Población en millones 2013	PIB en millones de USD 2005	GBID como proporción del PIB 2013
Argentina	41.4	802.2	0.61
Brasil	200.4	2.596.5	1.24
México	122.3	1.593.6	0.50
España	47.13	1.157	1.26
Portugal	10.4	197.3	1.33

Fuentes: UNESCO (2015b) Informe de la UNESCO sobre la ciencia: Hacia 2030 y Banco Mundial (2015).  
GBID (Gasto Bruto en Investigación y Desarrollo; incluye la inversión pública y privada).

En este cuadro puede apreciarse que la inversión en desarrollo científico y tecnológico de los países considerados es baja comparada con los países punteros, como es el caso de Brasil, España y Portugal y más reducida aún en el caso de Argentina y México. Una consecuencia de este indicador sería que a una baja inversión correspondería un bajo número de investigadores y una producción relativamente pequeña en el contexto mundial de la producción de conocimientos. Sin embargo, se ha observado un aumento considerable en el número de investigadores y en las publicaciones científicas, como se verá más adelante.

Uno de los fenómenos más destacables de los tiempos recientes es el aumento en el número de personas dedicadas a la realización de actividades científicas y tecnológicas. Según datos de la UNESCO, en 2015, aproximadamente 7.8 millones de científicos e ingenieros trabajaban en actividades de investigación en todo el mundo. Entre 2007 y ese año, el número de investigadores había crecido en un 21%. La Unión Europea (UE) es el líder mundial en el número de investigadores (22.2%), seguido de China (19%) y EE. UU. (16.7%). Japón decreció en estas cifras, del 10.7% en 2007 al 8.5% en 2013, al igual que la Federación Rusa que pasó del 7.3% al 5.7% en el mismo periodo. El cuadro siguiente muestra el comportamiento de este indicador en el caso de los países seleccionados para este artículo.

**Tabla II. Número de investigadores (equivalente a jornada completa) y por millón de habitantes en países seleccionados 2007-2013 y 2000-2010**

	Investigadores (miles)				Investigadores por millón de habitantes	
	2007	2009	2011	2013	2000	2010
<b>Mundo</b>	6.400.9	6.901.9	7.350.4	7.758.9	1.083,273	1.282,147
Argentina	38.7	43.7	50.3	51.6	713	1.121
Brasil	116.3	129.1	138.7		420	698
México	37.9	43	46.1		216	312
España	123.2	133.8	134.6	123.2	1.881	2.889
Portugal	28.1	39.8	44	37.8	1.628	3.923

Fuentes: Para Argentina, Brasil y México: UNESCO (2015b) y Banco Mundial (2015). Para España y Portugal: *España en Cifras* (2015) y Pordata (2017), respectivamente.

Tan importante incremento se ha visto reflejado en un aumento explosivo del número de publicaciones científicas. De acuerdo con cifras de la UNESCO (2015a), en el periodo 2008-2014, el número de artículos científicos incluidos en el índice de citas científicas de la plataforma *Web of Science de Thomson Reuters* aumentó en un 23%, al pasar de 1.029.471 a 1.270.425. El incremento porcentual registrado más

elevado (94%) correspondió a los artículos científicos procedentes de economías de medianos ingresos, debido básicamente al fuerte incremento (151%) de las publicaciones procedentes de China. Estados Unidos fue el país de donde provino la gran mayoría de artículos científicos publicados en 2013 en el mundo: 412.542. Tal cantidad equivale al 25,3% del total de artículos editados a nivel mundial, aunque es menor a la registrada en 2008 (28,1%). En contraste, el porcentaje representado por las publicaciones científicas de China en el total mundial mostró un gran incremento, al pasar del 9,9% al 20,2% en el periodo 2008-2014. Tomados en conjunto, las 28 naciones de la Unión Europea han sido las más productivas del mundo. Los científicos de esta región publicaron un total de 432.195 artículos científicos, lo que equivale a algo más de un tercio (34%) del total mundial y un promedio de 847 publicaciones científicas por cada millón de habitantes. (UNESCO, 2015a). En el caso de los países seleccionados en este trabajo, en el siguiente cuadro se presenta el número de artículos publicados en 2013, así como con el total mundial y los publicados en China y EE. UU.

**Tabla III. Artículos en publicaciones científicas y técnicas en países seleccionados 2013**

Mundo	2.184.420
Argentina	8.053
Brasil	48.662
México	13.112
España	53.342
Portugal	13.556
China	401.135
EE. UU.	412.542

Fuente: World Bank (2015). *World Development Indicators: Science and Technology*.

Como puede observarse, la producción de artículos de los países iberoamericanos es considerablemente baja comparada con los grandes volúmenes publicados por los científicos chinos y estadounidenses. La suma de lo publicado por las cinco naciones Iberoamericanas es un poco mayor a la tercera parte de lo publicado en estos dos últimos países. También es destacable que España y Brasil lideran a las naciones de Iberoamérica, así como que el número de artículos publicados en Portugal supera ligeramente al de México, pese a tener un número de investigadores ligeramente menor.

### 3. Antecedentes y desarrollo de la bibliometría en la investigación científica

Durante siglos, las revistas académicas han sido uno de los vehículos más importantes en la difusión del conocimiento en las ciencias, las humanidades y las artes. Las primeras de carácter científico hicieron su aparición en 1665 (Nishikawa, 2006; Mendoza y Paravic, 2006) y desde ese momento se fueron convirtiendo en uno de los principales instrumentos para la comunicación formal de los nuevos descubrimientos. Por ello, quienes se dedicaron a cultivar las disciplinas científicas se fijaron como objetivo de su actividad la aparición de los resultados de su trabajo de investigación en este tipo de publicaciones (Nishikawa, 2006; Mendoza y Paravic, 2006).

Los procesos mediante los cuales se instauraron las revistas académicas de investigación como elemento central de la difusión científica fueron largos y no exentos de problemas. Sin embargo, pese a las dificultades lograron institucionalizar una serie de funciones sociales vinculadas con la difusión de los nuevos conocimientos científicos; por ejemplo, además de asegurar la rápida difusión de los conocien-

tos, se constituyeron en salvaguarda de la autoría y de la veracidad de los nuevos descubrimientos (López y Cordero, 2005; citados en Mendoza y Paravic, 2006).

No obstante, en las últimas décadas las revistas científicas han adquirido una preeminencia todavía mayor debido a factores como la globalización y el desarrollo de las tecnologías de la información y la comunicación (TIC), las cuales han contribuido a internacionalizar la difusión de la ciencia, así como a aumentar el interés por la medición de la producción científica (Aguado *et al.*, 2003). La computadora, el internet y otros avances tecnológicos, han permitido la expansión de las revistas académicas de investigación en formato electrónico favoreciendo su consulta a nivel mundial; asimismo, han facilitado la elaboración de grandes bases de datos que concentran la información existente sobre la producción científica, principalmente la referida a artículos de investigación publicados en revistas académicas que cubren cierto tipo de criterios académicos.

En la década de los 60 del siglo xx se empezaron a desarrollar trabajos cuantitativos relacionados con el análisis numérico y estadístico de las revistas científicas, dando paso a las denominaciones de cienciometría, infometría y bibliometría (Mendoza y Paravic, 2006; De Bellis, 2009; 2014). El físico e historiador de la ciencia Derek J. De Solla Price ([1963], 1986) fue el pionero en los estudios de lo que inicialmente se llamó «cienciometría», al analizar de manera sistemática la expansión de los recursos dedicados al avance de la ciencia y el significado social de dicho fenómeno. Price consideró como antecedente de este campo de estudio los trabajos de medición sistemática de Galton y de William Perry, cuyo estudio sistemático de los «*bill*s» (listas, conteos) de mortalidad en el siglo xviii dio origen a lo que después se denominó como «aritmética política». Asimismo, Price también recuperó de Robert Boyle, el célebre científico e inventor inglés del siglo xvii, el término «colegio invisible» para designar los colectivos informales de científicos que trabajan de manera estrecha, generalmente de tamaño reducido, a fin de poder manejarse mediante relaciones interpersonales. Estos colegios invisibles son formaciones sociales y cognitivas que de forma significativa desarrollan diversos campos de investigación (Price, 1986).

Posteriormente la evolución de esas disciplinas permitió a sus autores y usuarios sacar conclusiones sobre los aspectos medibles de la ciencia, es decir, la información disponible empezó a ser usada también para evaluar la producción, el desempeño y la pertinencia de la política científica en varios países (De Bellis, 2009; 2014). Las naciones desarrolladas fueron pioneras en la implementación de estos procesos; aunque, como señala Cantoral (2007), los países de América Latina no resultan ajenos a los impactos de las nuevas formas de difusión y medición del conocimiento. Es más, el autor considera que su impacto fue mayor en los países de la región debido a que se conjugaron con la aplicación de nuevas políticas de evaluación del trabajo académico (implementadas a finales de los años 80), lo cual redundó en la necesidad de los académicos universitarios por publicar en revistas consideradas como de «reconocido prestigio académico.»

En concordancia con lo anterior, los sistemas de indicadores de la producción científica más reconocidos internacionalmente adquirieron cada vez mayor importancia, tales como la *Web of Science (WoS)*<sup>3</sup> y *Scopus*<sup>4</sup>. Son bases de datos que concentran una gran cantidad de revistas científicas, sirviendo de sustento

---

3 La *Web of Science* es una plataforma en línea que contiene enormes bases de datos de información bibliográfica y recursos de análisis de la información que permiten evaluar el rendimiento de la investigación. Su finalidad no es proporcionar el texto completo de los documentos que alberga sino proporcionar herramientas de análisis que permitan valorar su calidad científica. Es propiedad de la empresa *Thomson Reuters*.

4 *Scopus* es la base de datos más grande de citas y resúmenes de literatura arbitrada, cuenta con herramientas inteligentes para rastrear, analizar y visualizar trabajos de investigación. Pertenece a la empresa *Elsevier*.

para que sus indicadores sean utilizados ya no solo para difundir el conocimiento científico, sino también para evaluar la producción científica. De esta forma, indicadores tales como el total de artículos publicados, el número de citas que recibe cada artículo y el denominado «índice de impacto»<sup>5</sup> de las revistas, han pasado a ser utilizados como criterio de calidad (SCImago, 2012).

#### 4. Algunas implicaciones sociales del uso de la bibliometría

Puede decirse que al asociar la noción de calidad a los indicadores bibliométricos, se transformaron las funciones sociales que inicialmente cubrían las revistas científicas. Es decir, de ser instrumentos que servían para la rápida difusión del conocimiento científico, la protección de la autoría de los investigadores y de la veracidad de los nuevos descubrimientos; ahora al asumirse como un buen reflejo de la productividad y calidad científica, pasaron sin más a ser utilizados como instrumentos para evaluar tanto la productividad como la calidad de la investigación científica, ya sea de países, universidades, áreas o entidades de investigación y de los propios investigadores (Bordons y Gómez, 1997; Sancho, 1990; Brunner y Salazar, 2009). El auge de los *rankings* y los sistemas de estímulos del desempeño académico que proliferan por todos lados parece ser una clara muestra de ello. Más aún, dado que se privilegia la publicación del trabajo científico en revistas internacionales y los índices de impacto provienen en general de un pequeño número de revistas producidas por enormes conglomerados editoriales, así como que los temas de investigación siguen siendo establecidos por los principales centros de generación de conocimientos, las comunidades científicas de los países periféricos parecen enfrentar a un tipo de dependencia científica y cultural.

No obstante, la afirmación anterior debe tomarse con cautela, pues como señala Altbach (2014:7), «Los más importantes centros de creación y difusión del conocimiento no buscan imponer sus valores o metodologías a los otros; la influencia ocurre orgánicamente y sin conspiraciones». En un sentido semejante, Biegel (2015) postula que las jerarquías del Sistema Académico Mundial (SAM) son el resultado de un proceso de acumulación originaria que fue separando áreas geográficas, lenguas y disciplinas más dotadas, respecto de espacios, lenguas y disciplinas cada vez más carentes de recursos materiales y de capital académico. No obstante, la dependencia académica en el SAM no es una subordinación vertical mediante la cual una metrópoli académica somete a una periferia pasiva, meramente receptora de conocimientos producidos en los países más avanzados. De esta forma, el hecho de que el conocimiento producido en la periferia tenga una baja circulación en los circuitos académicos centrales no implica que ese conocimiento sea el resultado de la importación masiva de conocimientos extranjeros.

Con el propósito de superar el estereotipo que compara centralidad con autonomía, Biegel plantea la existencia de una distribución desigual tanto de las capacidades de investigación como del reconocimiento científico internacional. Adicionalmente considera que la *periferialidad* científica constituye un fenómeno cada vez más complejo, que no es posible explicar con el modelo economicista de la dependencia. Para ello propone reformular el concepto de «campo científico» de Bourdieu, que atienda el peso creciente de los circuitos de publicación académica. Para demostrar la complejidad mencionada con anterioridad, Beigel cita el ejemplo de China, que recientemente alcanzó el segundo lugar, después de los EE. UU. en la cantidad de artículos publicados en revistas indexadas en el ISI-Web of Science (más de 400 mil) (véase la tabla 3), pero que no fue acompañado de un aumento comparable en las estadísticas de artículos citados en dicho circuito (1.95/1 000, de China, frente al 70.15/1 000 de los EE. UU.). De esta manera, el

---

5 El «factor de impacto» de una revista es el número de veces que se cita por término medio un artículo publicado en dicha revista.

ingreso a los circuitos principales no necesariamente se traduce en prestigio internacional, si bien otorga reconocimiento local (Beigel, 2015).

Conviene señalar, sin embargo, que la utilización de los indicadores bibliométricos está produciendo efectos «perversos» en algunas comunidades científicas; ya que por un lado, no se puede inferir una asociación directa entre las publicaciones y el progreso laboral de los investigadores; y por otro, porque la dinámica de publicación de los artículos científicos en las revistas académicas presenta diversos sesgos que favorecen a determinados países, idiomas, campos y temas científicos; y además porque existen diversos problemas en la elaboración de sus índices (Spinak, 1996; Prat, 2001; Dong, Loh y Mondry, 2005; Cantoral, 2007; Ordorika, 2015).

Los efectos «perversos» usualmente se manifiestan a través de las prácticas que realizan los actores vinculados con la producción científica (ya sean tomadores de decisiones gubernamentales e institucionales, las propias revistas académicas y los investigadores), debido al afán de participar de los estímulos asociados a su presencia (prestigio o recursos económicos, por ejemplo) en la producción integrada a las bases de datos más prestigiosas (*WoS* y *Scopus*). En este sentido conviene destacar la situación planteada en 2001 por Ana María Prat, jefa en ese entonces del Departamento de Información de la Comisión Nacional de Investigación Científica y Tecnológica (CONICYT) de Chile, quien señaló el peligro de utilizar los indicadores bibliométricos del entonces ISI (ahora *WoS*) como único criterio para valorar el desempeño académico. Mencionó el caso de una universidad chilena cuya fortaleza era la antropología y por ello tenía un amplio reconocimiento internacional. Sin embargo, como los antropólogos no publicaban sus artículos en revistas indizadas en ISI, las autoridades les comunicaron que su universidad había sido juzgada únicamente con base en un artículo publicado por un físico en una revista de corriente principal, desdeñando así la producción que por diversas razones no estaban incluidas en el *ISI*. Así, la funcionaria resalta el problema que representa el sustentarse en los indicadores bibliométricos para distribuir los apoyos destinados al desarrollo de la ciencia y la tecnología, pues, prosigue, tomar en cuenta solo los criterios bibliométricos puede conducir al descuido de áreas de excelencia académica, las que, además, eventualmente pudieran ser prioritarias para el desarrollo del país (Prat, 2001).

Con base en las consideraciones anteriores, procedemos ahora a examinar en la siguiente sección de este artículo los resultados obtenidos del análisis de las revistas iberoamericanas de educación en los índices internacionales *Web of Science*, *Scopus* y *SciELO*<sup>6</sup>.

## 5. Las revistas iberoamericanas de educación en los índices bibliográficos internacionales

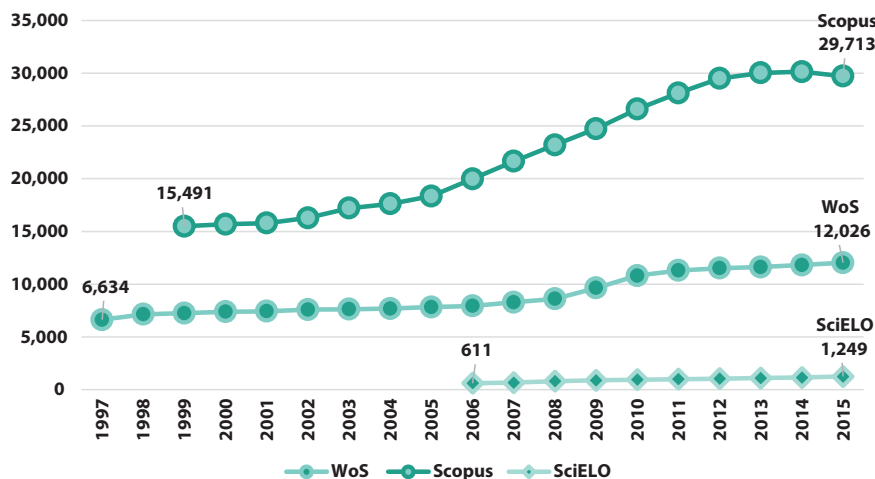
Un primer aspecto que llama la atención al apreciar la participación de las revistas de investigación de los países de Iberoamérica, en un área particular de conocimiento como es la educativa, es la baja representación que tienen en las tres bases de datos a pesar del crecimiento experimentado por el número de las revistas incorporadas en cada una de ellas. En 2015, las revistas de educación representaban en el *WoS* el 1.92% del total de revistas indizadas en dicha base de datos; en *Scopus* el 3.59% y en *SciELO* el 4.56%. De ellas, las que correspondían a los países iberoamericanos representaban en el *WoS* solo el 0.07% (9 revistas) con respecto al total de revistas, en *Scopus* el 0.20% (58 revistas) y en *SciELO*, que es una base

---

<sup>6</sup> En su página electrónica *SciELO* (*Scientific Electronic Library on Line*) señala que «es un modelo para la publicación electrónica cooperativa de revistas científicas en internet. Especialmente desarrollada para responder a las necesidades de la comunicación científica en los países en desarrollo y particularmente de América Latina y el Caribe, el modelo proporciona una solución eficiente para asegurar la visibilidad y el acceso universal a su literatura científica, contribuyendo para la superación del fenómeno conocido como 'ciencia perdida'. Además, el modelo *SciELO* contiene procedimientos integrados para la medida del uso y del impacto de las revistas científicas».

regional, el 4.56% (57 revistas). Como se puede apreciar a través de estos datos, la participación de las revistas iberoamericanas de educación ocupa una posición muy marginal. (Ver gráfico 1 y tabla IV).

**Gráfico 1. Número total de revistas indizadas en WoS, Scopus y SciELO, 1997-2015**



Fuentes: Web of Science (2017), SCImago (2017) y SciELO (2017).

**Tabla IV. WoS, Scopus y SciELO: total de revistas, revistas de educación y revistas de educación de los países iberoamericanos, 2015**

	Total	Educación	Educación Iberoamérica	%	%	%
	(a)	(b)	(c)	b/a	c/a	c/b
WoS	12,026	231	9	1.92	0.07	3.90
Scopus	29,713	1,066	58	3.59	0.20	5.44
SciELO*	1,249	57	57	4.56	4.56	100.00

Fuentes: Web of Science (2017), SCImago (2017) y SciELO (2017).

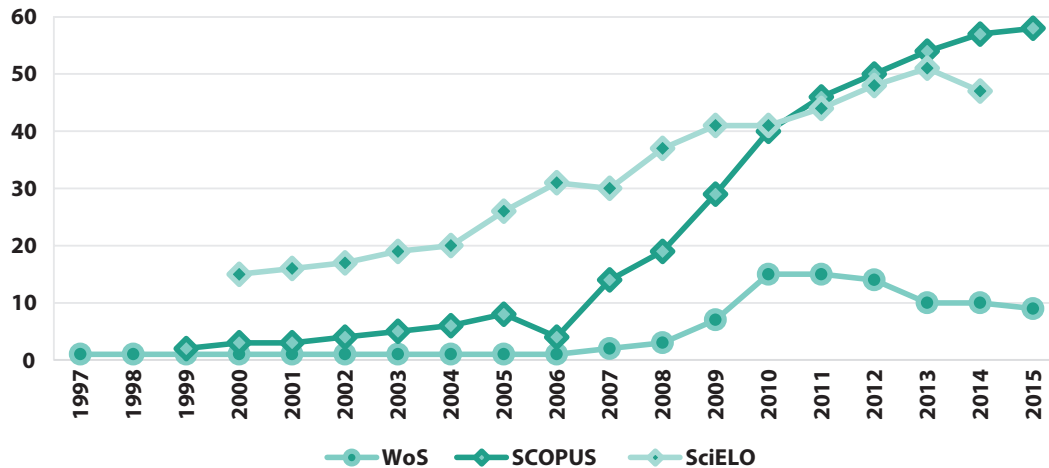
A pesar de que la base de datos *Scopus* y la regional *SciELO*, han mejorado un poco la situación en términos de la participación de las revistas dedicadas a la investigación educativa en los países de Iberoamérica, cabe destacar que esto es un proceso reciente que se circunscribe a la última década.

Entre el año 2006 y 2015, la participación de revistas que abordan desde distintas perspectivas el tema educativo (política educativa, enseñanza en disciplinas específicas, etcétera) pasó de 1 a 9 en el caso del *WoS*, de 4 a 58 en el caso de *Scopus* y de 31 a 57 en el caso de *SciELO*. Si bien, en el *WoS* hay una tendencia errática pues crece en los años 2010 y 2011, pero después tiende a disminuir. En *Scopus*, las revistas de educación presentan un rápido crecimiento entre 2006 y 2010, pero a partir de ese año disminuye su tasa de crecimiento. En *SciELO* se presenta un crecimiento más regular pero también decrece ligeramente hacia el año 2014. (Ver gráfico 2 y tabla V).

Un aspecto que conviene destacar es que en 2015 la participación de revistas de los países iberoamericanos se concentra en España y Brasil. En el *WoS* las revistas españolas concentran el 66.7% del total de revistas. En *Scopus* a las revistas de España les corresponde el 41.4% del total y a las de Brasil el 31%, en conjunto ambos países concentran más del 70%. En la base de datos *SciELO*, la situación es claramente diferente, hay una baja participación de las revistas españolas (2.1%), pero en este caso la participación

se concentra en tres países que representan el 76% del total de revistas, Brasil 48.9%, México 14.9% y Colombia 12.8%. (Ver tabla V).

**Gráfico 2. Revistas Iberoamericanas de Investigación Educativa incorporadas en WoS, Scopus y SciELO, 1997-2015**



Fuentes: Web of Science (2016), SCImago (2016) y SciELO (2016).

**Tabla V. Países de Iberoamérica con mayor número de Revistas de Investigación Educativa incorporadas en Web of Science, Scopus y SciELO, 2006-2015**

País	WoS		SCOPUS				SciELO					
	2006	%	2015	%	2006	%	2015	%	2006	%	2014	%
España	0	0.0	6	66.7	2	50.0	24	41.4	1	3.2	1	2.1
Brasil	0	0.0	1	11.1	1	25.0	18	31.0	15	48.4	23	48.9
México	0	0.0	1	11.1	1	25.0	5	8.6	3	9.7	7	14.9
Portugal	1	100.0	1	11.1	0	0.0	3	5.2	2	6.5	2	4.3
Venezuela	0	0.0	0	0.0	0	0.0	2	3.4	4	12.9	1	2.1
Colombia	0	0.0	0	0.0	0	0.0	2	3.4	3	9.7	6	12.8
Subtotal	1	100.0	9	100.0	4	100.0	54	93.1	28	90.3	40	85.1
Otros	0	0.0	0	0.0	0	0.0	4	6.9	3	9.7	7	14.9
<b>Total</b>	<b>1</b>	<b>100.0</b>	<b>9</b>	<b>100.0</b>	<b>4</b>	<b>100.0</b>	<b>58</b>	<b>100.0</b>	<b>31</b>	<b>100.0</b>	<b>47</b>	<b>100.0</b>

Fuentes: Web of Science (2016), SCImago (2016) y SciELO (2016).

Es importante señalar que el total de revistas de investigación educativa de los países iberoamericanos identificadas en las tres bases de datos asciende a 94 en total; sin embargo, algunas de ellas no se han logrado mantener. El WoS presenta el patrón más excluyente pues de 17 revistas que lograron ingresar a la base de datos en el periodo 1997 a 2015, solo 15 estuvieron de manera simultánea en los años 2010 y 2011, pero en 2015 ya solo había 9 revistas. La participación de las revistas en Scopus es más estable, de 58 revistas iberoamericanas incluidas en esta base durante el periodo 1999-2015, todas se encuentran en la base de datos al final del periodo. No obstante, algunas de ellas salieron en el transcurso del periodo analizado, pero posteriormente volvieron a incluirse. Una salió por nueve años y otras tres lo hicieron solo por el transcurso de un año. En este sentido, se puede señalar que la participación en Scopus tiende a ser más estable y mantener un



patrón acumulativo de las revistas iberoamericanas que logran incorporarse a la base. No obstante, este no es el mismo caso en la base *SciELO*, pues en el periodo comprendido entre 2000-2014, se identificaron en total 57 revistas de investigación educativa, pero de ellas solo 47 (82.5%) se mantuvieron hacia el final del periodo. (Ver tabla VI).

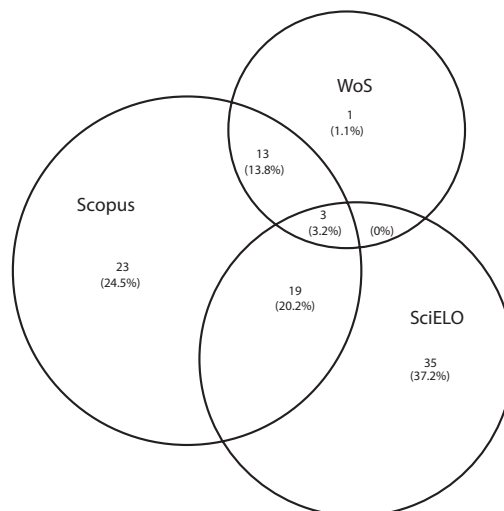
**Tabla VI. Número de años y situación de permanencia de las revistas de investigación educativa de los países de Iberoamérica en *WoS*, *Scopus* y *SciELO***

Núm. de años	WoS 1997-2015			Scopus 1999-2015			SciELO 2000-2014		
	Permanece	No permanece	Total	Permanece	No permanece	Total	Permanece	No permanece	Total
1								1	1
2		4	4	3		3	5		5
3		1	1	4		4	5	1	6
4		1	1	5		5	3		3
5	1	2	3	6		6	1	2	3
6	5		5	10		10	4	1	5
7	1		1	11		11	4	2	6
8				5		5			0
9	1		1	7		7	3	1	4
10				1		1	5	1	6
11				1		1	1	1	2
12				1		1	1		1
13									0
14				1		1			0
15				1		1	15		15
16				2		2			
17									
18									
19	1		1						
<b>Total</b>	<b>9</b>	<b>8</b>	<b>17</b>	<b>58</b>	<b>0</b>	<b>58</b>	<b>47</b>	<b>10</b>	<b>57</b>
<b>%</b>	<b>52.9</b>	<b>47.1</b>	<b>100.0</b>	<b>100.0</b>	<b>0.0</b>	<b>100.0</b>	<b>82.5</b>	<b>17.5</b>	<b>100.0</b>

Fuentes: Web of Science (2016), SCImago (2016) y SciELO (2016).

Otro aspecto a destacar, es que la participación de las revistas iberoamericanas es diferente en cada una de las bases de datos. De las 94 revistas identificadas en las tres bases de datos, solo hay tres revistas que llegaron a participar simultáneamente en las tres bases de datos; si bien, en 2015 dos de ellas ya no participaban en el *WoS*. Por su parte, *Scopus* es la base de datos que comparte el mayor número de revistas con las otras dos bases de datos: 13 revistas con el *WoS* y 19 revistas con *SciELO*. A su vez, *SciELO* es la base de datos que contiene el mayor número de revistas iberoamericanas que no han participado en las otras bases de datos durante el periodo de análisis, con un total de 35; le sigue *Scopus* con 23 revistas y finalmente el *WoS*, solo tiene una revista que no participa en las otras bases de datos. (Ver figura 1).

**Figura 1. Participación de las revistas iberoamericanas de investigación educativa en WoS, Scopus y SciELO, 1997-2015**



Fuentes: Web of Science (2016), SCImago (2016) y SciELO (2016).

## 6. Consideraciones finales

En este trabajo se ha buscado mostrar el modo en que las actividades científicas han tenido un crecimiento sin precedentes. Así lo demuestran los crecientes niveles de inversión en materia de investigación y desarrollo, en el número de personas dedicadas a las labores científicas y tecnológicas y en la cantidad de publicaciones aparecidas en los últimos años. No obstante que continúan y en algunos casos se acentúan las desigualdades en casi todos los países, el crecimiento es indiscutible. Aunque los recursos y los principales proyectos se siguen concentrando en pocos países e instituciones, lo cierto es que la ciencia en sus múltiples facetas es hoy más que nunca una empresa global.

En el caso específico de las publicaciones en el campo educativo de los países iberoamericanos también se ha podido apreciar un crecimiento considerable, aunque muy lejano de los países que invirtieron grandes recursos financieros y que cuentan con mayores contingentes de personas dedicadas de tiempo completo a las labores de investigación.

Coincidiendo con García-Pereira y Quevedo-Blasco (2015), se observó que es muy reducido el número de las revistas que se incluyen en los principales índices bibliográficos y menor aún el de las que se encuentran en los primeros dos cuartiles de esas clasificaciones. No obstante, y a pesar de que las perspectivas no son muy promisorias, lo que es innegable es que el panorama actual ofrece enormes oportunidades de colaboración, no solo para fortalecer su presencia en los índices internacionales, sino para consolidarse como órganos eficaces de difusión del conocimiento educativo generado en sus respectivos países y regiones. Compartir experiencias y conocimiento acumulado en la gestión editorial pueden ser mecanismos iniciales que conduzcan a beneficios comunes y reducción de asimetrías entre publicaciones e instituciones (VV. AA., 1999). Como parte del carácter cada vez más colaborativo en la generación y difusión del conocimiento en el mundo, las revistas científicas de Iberoamérica en el campo de la educación no pueden dejar de avanzar en esta dirección.

## Referencias bibliográficas

- Aguado López, Eduardo; Eduardo Sandoval Forero y Salvador Chávez Ávila (2003): “La ciencia perdida y las nuevas tecnologías de divulgación del conocimiento: el proyecto REDALyC”. *Educación y Ciencia*, Nueva época, 7 (13), 11-39.
- Altbach, Philip (2014): “MOOCs as neocolonialism: who controls knowledge?”. *International Higher Education*, 75, Spring, 5-7.
- Beigel, Fernanda (2015): “Culturas (evaluativas) alteradas”. Política Universitaria. Instituto de Estudios y Capacitación. *Federación Nacional de Docentes Universitarios*, 2, Agosto, 12-21.
- Bordons, María y Isabel Gómez Caridad (1997): “La actividad científica española a través de indicadores bibliométricos en el periodo 1990-93”. *Revista General de Información y documentación*, 7 (2), 69-86.
- Brunner, José Joaquín y Felipe Salazar (2009): *La investigación educacional en Chile: Una aproximación bibliométrica no convencional*. Documento de Trabajo CPCE N° 1. Chile: Centro de políticas Comparadas de Educación, Universidad Diego Portales.
- Bueno Sánchez, Ángel y Antonio Fernández Cano (2003): “Análisis cientimétrico de la productividad en la revista de investigación educativa (1983-2000)”. *Revista de Investigación Educativa*, 21 (2), 507-532.
- Cantoral, Ricardo (2007): “¿Publicar o perecer, o publicar y perecer?”. *Revista Latinoamericana de Investigación en Matemática Educativa*, 10 (3), 311-313 (en línea) <http://www.scielo.org.mx/pdf/relime/v10n3/v10n3a1.pdf>, consultado el 10 de febrero de 2015.
- De Bellis, Nicola (2014): “History and evolution of (Biblio)Metrics” en Blaise Cronin and Cassidy R. Sugimoto (ed.): *Beyond Bibliometrics. Harnessing Multidimensional Indicators and Scholarly Impact*. Cambridge, Massachusetts, London, England: The MIT Press.
- De Bellis, Nicola (2009). *Bibliometrics and citation analysis. From the Science Citation Index to Cybermetrics*. United Kingdom: The Scarecrow Press, Inc.
- Dong, Peng; Marie Loh y Adrian Mondry (2005): “The ‘impact factor’ revisited”. *Biomedical Digital Libraries*, 2 (7) (en línea) <http://www.bio-diglib.com/content/2/1/7>, consultado el 27 de marzo de 2016.
- García-Pereira, S. y Quevedo-Blasco, R. (2015): “Análisis de las revistas iberoamericanas de Psicología y de Educación indexadas en el Journal Citation Reports (JCR) del 2013”. *European Journal of Education and Psychology*, 8, 85-96. (en línea) <http://elsevier.es>, consultado el 16 de marzo de 2017.
- Instituto Nacional de Estadística (2015): España en Cifras 2015 (en línea) [www.ine.es/prodyser/espaa\\_cifras/2015/](http://www.ine.es/prodyser/espaa_cifras/2015/), consultado el 29 de marzo de 2017
- Mendoza, Sara y Tatiana Paravic (2006): “Origen, clasificación y desafíos de las revistas científicas”. *Investigación y Posgrado*, Vol. 21 (1), 49-76.
- Nishikawa Aceves, Antonieta Kiyoko (2006): “Breve diagnóstico de una publicación electrónica”, *Revista de la UABC*, 53, pp. 44-49 (En línea) <http://132.248.129.5/cursosOJS/index.php/uabc/article/view/825>, consultado el 19 de diciembre de 2014.
- Ordorika, Imanol (2015): “Rankings universitarios”, *Revista de la Educación Superior*, vol. XLIV (173), 7-9.

- Pordata (2017). Base de Datos Portugal Contemporáneo (en línea) [www.pordata.pt/Portugal](http://www.pordata.pt/Portugal), (en línea), consultado el 03 de abril de 2017.
- Prat, Anna Maria (2001): “Evaluación de la producción científica como instrumento para el desarrollo de la ciencia y la tecnología”, *ACIMED*, 9 (4), 111-114.
- Price, Derek John de Solla ([1963]1986). *Little Science, Big Science and Beyond*. [Foreword by Robert K. Merton]. New York: Columbia University Press.
- Sancho, Rosa (1990): “Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica”. *Revista Española de Documentación Científica*, 13 (3-4), 842-865.
- SciELO (2017). Scientific Electronic Library Online, Sobre el SciELO (en línea) <http://www.scielo.org/php/level.php?lang=es&component=44&item=26>, consultado el 10 de marzo de 2017.
- SciELO (2016). Scientific Electronic Library Online, Sobre el SciELO (en línea) <http://analytics.scielo.org/w/reports>, consultado el 30 de junio de 2016.
- SCImago (2017). SJR – Scimago Journal & Country Rank (en línea) <http://www.scimagojr.com/journal-rank.php?category=3304>, consultado el 10 de marzo de 2017.
- SCImago (2016). SJR – Scimago Journal & Country Rank, (en línea) <http://www.scimagojr.com/journal-rank.php>, consultado el 13 de junio de 2016.
- SCImago (2012). En síntesis SCImago, *Formación Universitaria*, 5 (5), (en línea) <http://www.scielo.cl/pdf/formuniv/v5n5/art01.pdf>, consultado el 19 de diciembre de 2014.
- Spinak, Ernesto (1996). “Los análisis cuantitativos de la literatura científica y su validez para juzgar la producción latinoamericana”. *Boletín de la Oficina Sanitaria Panamericana*, 120 (2), 139-146.
- UNESCO (2015a): UNESCO Science Report. Informe de la UNESCO sobre la Ciencia: Hacia 2030. Hechos y Cifras. París: UNESCO.
- UNESCO (2015b): Informe de la UNESCO sobre la ciencia: Hacia 2030. París: UNESCO (en línea) [unesdoc.unesco.org/images/0023/002354/2354075](http://unesdoc.unesco.org/images/0023/002354/2354075), consultado el 03 de abril de 2017.
- VV. AA. (1999): “Publishing in developing countries: problems and solutions CBE”. *Views*, vol. 22 (6), 198.
- Web of Science (2017). InCites “Journal Citation Report” (en línea) <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action?SID=B2-R9Cx2Bgkl0MxxBd4pm7Mnn9VIpjI6BR-x2Fj5-18x2dDFwIIAx2BHEoseA24DqhpckAx3Dx3DFgSiwsxx1THx2FyTrBz2rx2FJ4gx3Dx3D-9vvmzcn dpRgQCGPd1c2qPQx3Dx3D-wx2BJQh9GKVmtdJw3700KssQx3Dx3D&SrcApp=IC2LS&Init=Yes>, consultado el 10 de marzo de 2017.
- Web of Science (2016): InCites “Journal Citation Report” (en línea) [http://apps.webofknowledge.com/UA\\_GeneralSearch\\_input.do?product=UA&search\\_mode=GeneralSearch&SID=3AeBXL0Fh1ceOp5sIO2&preferencesSaved=](http://apps.webofknowledge.com/UA_GeneralSearch_input.do?product=UA&search_mode=GeneralSearch&SID=3AeBXL0Fh1ceOp5sIO2&preferencesSaved=), consultado el 23 de junio de 2016.
- World Bank/Banco Mundial (2015). World Development Indicators: Science & Technology (en línea) [wdi.worldbank.org/table/5.1](http://wdi.worldbank.org/table/5.1), consultado el 03 de abril de 2017.

## Notas biográficas

**Armando Alcántara Santuario** es Doctor en Educación por la Universidad de California, Los Ángeles. Sus principales líneas de investigación se centran en políticas de la educación superior, política científica, educación superior internacional y comparada, desigualdad y educación superior. En 2014 ha publicado «Globalización y redes de investigación en políticas educativas. La experiencia de la RIAIPE» en María L. Chavoya y Norma G. Gutiérrez (coord.). *Producción de conocimiento y procesos de internacionalización en México y países de América*. México: Universidad de Guadalajara.

**Alejandro Márquez Jiménez** es Doctor en Educación por la Universidad Autónoma de Aguascalientes, México. Sus principales líneas de investigación se centran en financiamiento, planeación y políticas de la educación superior. Ha publicado recientemente «La desigualdad social y las reformas educativas», *Perfiles Educativos*, 154, vol. XXXVIII, octubre-diciembre, 2016.

# El aprendizaje en la base de la pirámide: Restricciones, comparabilidad y política en países en vías de desarrollo

Learning at the bottom of the pyramid:  
Constraints, comparability and policy in developing countries

Daniel A. Wagner y Nathan M. Castillo<sup>1</sup>

## Resumen

Los objetivos para el desarrollo de las Naciones Unidas han asignado, de manera sostenida, una alta prioridad a la calidad educativa –y a la del aprendizaje–. Esto ha llevado a avances sustanciales en ayuda para el desarrollo internacional hacia la educación, y también a una mayor atención, a nivel mundial, a la importancia del aprendizaje de los niños. Sin embargo, tales metas son principalmente normativas: tienden a representar promedios entre naciones, brindando limitada atención a las variaciones dentro de dichos países. La presente investigación proporciona un análisis de las tensiones científicas en la comprensión del aprendizaje en poblaciones pobres y marginadas: aquellos en la base de la pirámide (BdP). Si bien agencias internacionales como UNESCO y OCDE a menudo invocan a estas poblaciones como el «objetivo» de sus inversiones y evaluaciones, continúan los debates importantes sobre la ciencia empírica involucrada tanto en la investigación como en las políticas. El presente análisis concluye que los objetivos para el desarrollo post-2015 de las Naciones Unidas tienen que tomar en cuenta la necesidad crítica de enfocarse en el aprendizaje de los pobres a fin de abordar adecuadamente las inequidades sociales y económicas<sup>2</sup>.

## Palabras clave

Aprendizaje, países en vías de desarrollo, poblaciones pobres y marginalizadas, resultados de aprendizaje, educación comparativa, política educativa.

## Abstract

United Nations development goals have consistently placed a high priority on the quality of education – and of learning. This has led to substantive increases in international development assistance to education, and also to broader attention, worldwide, to the importance of children's learning. Yet, such goals are mainly normative: they tend to be averages across nations, with relatively limited attention to variations within countries. This review provides an analysis of the scientific tensions in understanding learning among poor and marginalized populations: those at the bottom of the pyramid. While international agencies, such as UNESCO and OECD, often invoke these populations as the «target» of their investments and assessments, serious debates continue around the empirical science involved in both research and policy. The present analysis concludes that the UN post-2015 development goals must take into account the critical need to focus on learning among the poor in order to adequately address social and economic inequalities..

## Key words

Learning, low-and middle-income countries, poor and marginalized populations, learning outcomes, constraints comparability, education policy

Recibido: 27-03-2017  
Aceptado: 28-03-2017

<sup>1</sup> University of Pennsylvania, Daniel A. Wagner, [wagner@literacy.upenn.edu](mailto:wagner@literacy.upenn.edu); Nathan M. Castillo, [ncast@gse.upenn.edu](mailto:ncast@gse.upenn.edu)

<sup>2</sup> Una versión anterior de este artículo apareció en Wagner y Castillo (2014). Agradecemos a Jazmín Mora por su ayuda con la traducción.

## 1. Introducción

La Conferencia Mundial sobre Educación para Todos en Jomtien, Tailandia, fue un momento decisivo en la educación y desarrollo internacionales. Llevada a cabo en 1990, la Conferencia abarcó dos desafíos clave: aumentar significativamente el acceso a la educación para los niños de países pobres y promover la calidad de aprendizaje en la educación. Una década más tarde, en la conferencia Educación para Todos (EPT) de 2000 en Dakar, estos dos mismos desafíos fueron expandidos en una lista detallada de seis objetivos de educación en el Marco de Acción de EPT en Dakar. Los objetivos eran promover la atención y el cuidado de la primera infancia, hacer obligatoria la enseñanza primaria, abordar las necesidades de aprendizaje para todos, promover la alfabetización de adultos, reducir las disparidades de género y desarrollar mediciones de calidad de los resultados del aprendizaje (UNESCO 2003:28). Estos fueron reforzados de nuevo en los Objetivos de Desarrollo del Milenio (ODM) de las Naciones Unidas para el 2015, donde la educación primaria universal se convirtió en el segundo de ocho objetivos principales (Naciones Unidas, 2000). Estos esfuerzos globales llevaron a las agencias de desarrollo internacional a aumentar significativamente su asistencia a la educación y también llevó al público en general a prestar más atención al aprendizaje de los niños en una escala global.

Existe una gran y diversa base de investigación empírica en el área del aprendizaje humano. Sin embargo, gran parte de la información disponible se encuentra sustancialmente limitada por restricciones de diversos tipos. La restricción más prominente es la limitada capacidad de generalizar desde hallazgos en contextos de una población a contextos de una población distinta. Del mismo modo, los métodos de investigación pueden variar de gran manera entre conjuntos de estudios, lo que dificulta discernir si los resultados varían debido a los métodos o a otros factores. Estos son, por supuesto, problemas clásicos en las ciencias sociales.

En este artículo, analizamos las tensiones científicas en la comprensión del aprendizaje de las poblaciones pobres, aquellas que Prahalad (2006) llamó la base de la pirámide (BdP). Si bien las agencias internacionales a menudo sostienen que invierten en las poblaciones más necesitadas, continúan los serios debates sobre la ciencia empírica necesaria para implementar las políticas apropiadas, con consecuencias inevitables para la implementación efectiva en países de ingreso bajo y medio.

## 2. Restricciones

### El aprendizaje en dos aulas sudafricanas:

La Escuela Primaria Shayandime, con edificios de paredes de adobe y techos de zinc, se encuentra ubicada en una pequeña aldea en la provincia septentrional de Limpopo, Sudáfrica. A tan solo unas docenas de millas de la frontera con Zimbabue, el área está salpicada con casas tradicionales llamadas *rondaveles*, una versión adaptada de la cabaña del estilo de África meridional. Los babuinos rondan por el recinto escolar libremente y son conocidos por deslizarse por el espacio entre las paredes de barro rojo y el techo corrugado y vandalizar las aulas por la noche. A pesar de la ocasional ventana rota, a la escuela no le faltan recursos. Es uno de los muchos establecimientos en la región que recibió una donación de antiguas computadoras de escritorio en las cuales los estudiantes de primaria superior pasan tiempo trabajando en actividades básicas de mecanografía. No obstante, los alumnos desfavorecidos con escasa competencia en inglés no tienen acceso a las computadoras, ya que las mismas carecen de programas en el idioma local, *Venda*. En el aula, los estudiantes pasan la mayoría de su tiempo copiando oraciones de la pizarra y rara vez participan de actividades estimulantes que apoyen la creatividad y las habilidades de pensamiento crítico.

En contraste, a cuatro horas de distancia, en la capital provincial de Polokwane, se sitúa la Escuela Primaria Central. Con veredas de ladrillo alrededor del perímetro y con un laboratorio informático de vanguardia, la misma tiene monitores de pantalla plana y una pizarra inteligente con proyector. El laboratorio de informática, parecido al de la universidad local, fue adquirido en parte a través de las ganancias obtenidas al alquilar el salón de eventos de la escuela a la comunidad. No hay ventanas rotas, los profesores presentan planes de estudio estructurados y los padres conforman una parte integral de la cultura escolar. Teniendo en cuenta su atractivo ambiente de aprendizaje, los funcionarios provinciales exhiben orgullosamente esta escuela urbana a los planificadores educativos nacionales e internacionales en sus visitas. Muchos estudiantes tienen teléfonos móviles y dan la impresión de estar motivados a aprender y estar conectados con el futuro de Sudáfrica<sup>3</sup>.

Comparaciones de contextos rurales y urbanos en países en vías de desarrollo a menudo consisten en estos tipos de observaciones de características sociales y de infraestructura. La Evaluación Nacional Anual de Sudáfrica (ANA, por sus siglas en inglés), administrada al final de cada año escolar, mide el progreso en los logros de aprendizaje de los grados primero al sexto y al noveno (DBE, 2013). Esta tiende a confirmar la perspectiva subjetiva presentada más arriba. Las escuelas sudafricanas se clasifican según un índice de pobreza que está basado en la riqueza o pobreza relativa de la comunidad y se agrupan en quintiles. La Escuela Primaria rural Shayandime pertenece al quintil más bajo. Cuando allí se llevó a cabo la ANA, solo cinco estudiantes en el tercer grado obtuvieron puntajes por encima del promedio nacional, mientras que la gran mayoría obtuvo resultados en el 10% inferior, creando así una distribución sesgada. En contraste, la Escuela Primaria Central se sitúa en el (tercer) quintil medio con resultados normalmente distribuidos; los mismos se encuentran en cierto modo por debajo de los promedios urbanos nacionales para las competencias en matemáticas y la alfabetización en la lengua materna (DBE, 2013).

El contraste en el desempeño de las pruebas ANA entre estas dos escuelas plantea dos preguntas relacionadas: ¿Por qué Shayandime tiene algunos estudiantes de alto rendimiento, con el resto de sus estudiantes agrupados alrededor del continuo de rendimiento más bajo? Y, dentro de la Escuela Primaria Central, ¿Por qué los puntajes se distribuyen de manera más normal, aunque por debajo del promedio nacional? Los estudios a menudo señalan la función que las influencias sociales y familiares tienen en predecir los resultados de aprendizaje, principalmente en términos de lo que llamamos poder, padres, y privilegio. Por ejemplo, varios autores han estudiado el impacto de la estratificación social en los resultados escolares en términos de tales factores (Benedict y Hoag, 2004; Buchmann y Hannum, 2001; Korinek y Punpuing, 2012; Lu y Treiman, 2011). En muchos casos, los estudiantes con la correcta combinación de estas influencias son los que tienden a sentarse más cerca del frente en aulas grandes y también tienen mayor concentración y motivación para aprender.

Estos hallazgos tienen implicaciones serias. Nuestro entendimiento de la ciencia psicológica del aprendizaje deriva principalmente de los datos de los países ricos de la OCDE, donde los resultados educativos, incluyendo aquellos de exámenes estandarizados son, gracias a su diseño, normalmente distribuidos. Sin embargo, la variación alrededor del mundo, y especialmente en contextos de la BdP como hemos observado en las zonas rurales de Sudáfrica, puede ser sesgada: unos cuantos puntajes altos y muchos estudiantes con puntajes bajos. En lugares donde el contraste es tan dramático, y donde el foco se encuentra en el extremo inferior de la curva normal, nos vemos obligados a reconsiderar las nociones de la

---

<sup>3</sup> Nota del autor: los nombres de las escuelas han sido cambiados, y estos perfiles combinan detalles de varias escuelas



normalidad estadística, y con ello, la idea que los contextos de la BdP son simplemente una extensión de la típica curva normal.

La idea de que el aprendizaje dentro y fuera de la escuela puede variar considerablemente entre culturas no es nueva (Cole, Gay, Glick y Sharp, 1971; Wagner, 1993, 2014). No obstante, el aumento continuo de la presión para globalizar la recopilación de datos sobre la educación ha llevado a los investigadores y a los responsables de las políticas públicas a ignorar, o minimizar tales diferencias (Benavot y Tanner, 2007). Creemos que la falta de consideración al aprendizaje en la base de la pirámide distorsiona la realidad y puede llevar a intervenciones desconsideradas a favor de los estudiantes pobres.

Más allá de Sudáfrica, abundan los ejemplos de estilos de aprendizaje distintivos en la literatura de investigación. Parte de este trabajo se originó hace décadas en los Estados Unidos con estudios pioneros sobre las diferencias individuales en el aprendizaje (Kagan, Moss y Sigel, 1963; Witkin, Moore, Goodenough y Cox, 1977). Hoy en día, los estudiantes de las escuelas pobres de países de ingreso bajo y medio a menudo se esfuerzan para entender el lenguaje de instrucción y el lenguaje de lectura. En tales situaciones, los profesores hacen énfasis en el aprendizaje mecánico y la memorización; aunque esta sigue siendo una estrategia de aprendizaje muy común en el mundo, es ridiculizada por los pedagogos modernos (occidentales) (Wagner, 1983). Otro ejemplo de estos contrastes es la manera en la cual el aprendizaje es construido en varias sociedades, como las que están fuertemente influenciadas por el Confucionismo (Li, 2003).

El aprendizaje es omnipresente y toma diferentes formas en la vida cotidiana. En la educación, el aprendizaje se mide con instrumentos que puedan estimar de forma fiable tanto los procesos como los resultados —o las evaluaciones de aprendizaje—. Si es necesario que una evaluación sea representativa de la población entera de un país, o válida en varios países en un marco comparativo, entonces probablemente costará más en términos de tiempo y recursos. Hasta ahora, los investigadores han controlado el tiempo y los recursos delimitando el rango de *habilidades* a ser evaluadas (la muestra de habilidades), y limitando la *población* a ser incluida (la muestra de la población). Es importante comprender estas dos formas de límites en términos de requerimientos técnicos y estadísticos, así como también los requerimientos y resultados de las políticas. Cada uno de estos temas plantea desafíos empíricos y estadísticos.

### Muestreos de habilidades y evaluaciones

Es ampliamente aceptado que los humanos aprenden a través del muestreo de sus entornos, comenzando por el uso de sus sentidos incorporados desde el momento del nacimiento. Claramente, ningún niño, joven ni adulto podría sobrevivir asimilando la totalidad de la información disponible en el entorno. En otras palabras, los sistemas humanos están diseñados para discriminar información para así tomar muestras de la información que mejor les ayude a manejar los desafíos del aprendizaje (Kahne-man, 2011). De hecho, los padres normalmente preparan a los niños pequeños a adaptarse, aprender y sobrevivir precisamente exponiéndolos a la gama de situaciones que encontrarán en sus vidas. Por supuesto, no todos estos entornos de aprendizaje pueden ser igualmente adaptables para el futuro del niño en ambientes educativos.

Cuando se trata de la investigación científica sobre el aprendizaje, nos va mejor a los seres humanos cuando tomamos muestras de nuestro ambiente informativo, ya sea de instituciones educativas, verbalmente o, cada vez más, a través de buscadores de internet como Google. Esta simple observación es muy relevante aquí: uno de los problemas más exasperantes que encuentran los investigadores al estudiar y evaluar el aprendizaje es como generalizar de una muestra de habilidades a otra. Por lo tanto,

muestrear una colección finita de habilidades y comprender las situaciones contextuales en las cuales se utilizan son elementos claves de todas las evaluaciones de aprendizaje.

Al diseñar estrategias para la investigación y evaluación del aprendizaje, los investigadores toman decisiones altamente complejas: eligen variables contextuales y demográficas (por ejemplo edad del niño, año escolar, sexo, situación socioeconómica), y eligen las habilidades a ser evaluadas y el tipo de metodología de investigación a ser utilizada. Cada opción está ligada a varias suposiciones y concesiones, y las selecciones incluidas en el diseño final influenciarán la validez, fiabilidad y viabilidad práctica del enfoque elegido (Braun y Kanjee, 2006; Wagner, 2010, 2011a). Además, la investigación debe ser diseñada de manera que pueda adaptarse a los cambios dinámicos a lo largo del tiempo. Y, a medida que las expectativas de alfabetización, aritmética y las capacidades de orden superior se adaptan a los cambios en los entornos sociales y económicos, los métodos de medición también deben ser adaptados para que se alineen con los objetivos educativos (Wagner, Murphy y de Korne, 2012).

### El aprendizaje y muestras de la población

El muestreo de población también es importante. Por ejemplo, cerca del 5% de la población mundial reside en los Estados Unidos, pero alrededor del 95% (Arnett, 2008) de las publicaciones científicas sobre el desarrollo psicológico son basadas en poblaciones que son OEIRD (occidentales, educadas, industrializadas, ricas y democráticas; Heinrich, Heine y Norenzayan, 2010), y viviendo principalmente en países OCDE. Además, de las investigaciones sobre desarrollo psicológico conducidas en los Estados Unidos, aproximadamente el 80% pertenece a grupos étnicos «mayoritarios» (de ascendencia europea), por más de que estos grupos representan solo el 50% de la población actual de los Estados Unidos (Arnett, 2008). Parece obvio que los investigadores deben abordar explícitamente las cuestiones de representatividad y validez externa, pero a menudo no lo hacen. Estas críticas también se aplican a la investigación internacional, ya que gran parte de la investigación disponible sobre el aprendizaje está importantemente restringida por bases de datos científicos y estudios de investigación extraídos de muestras de poblaciones que viven en países más desarrollados. Afortunadamente, esta tendencia está empezando a cambiar (Wagner, 2014).

En las evaluaciones educativas internacionales a gran escala (LSEAs por sus siglas en inglés) partes claves de las poblaciones de la BdP pueden verse excluidas o sub-representadas en muestras que se consideran nacionales (Engel y Feuer, 2014; Wagner, 2011b). El género ha sido una de las principales razones por la cual los niños no asisten a la escuela en los países en vías de desarrollo, aunque en las últimas décadas se han registrado avances significativos. Aun así, en los países más pobres hay menos niñas que niños presentes en escuelas en los dos momentos donde normalmente los datos están recogidos: el comienzo de la escuela primaria y la secundaria. La exclusión sistemática de niñas en países pobres tiende a resultar en una menor cantidad de niñas adolescentes que asisten a la escuela; aquellas que sí asisten a menudo obtienen resultados más bajos en evaluaciones nacionales en comparación con los varones. Por ejemplo, según los resultados del sexto grado de la evaluación regional del Consorcio Africano Meridional y Oriental para el Seguimiento de la Calidad Educativa (SACMEQ por sus siglas en español), llevada a cabo en el 2007, Saito (2011) encontró que, a base de un promedio de 15 países africanos, los niños generalmente superaban a las niñas en matemáticas, mientras que las niñas superaban a los niños en la alfabetización. Sin embargo, las diferencias nacionales en las disparidades de género varían ampliamente tanto en alfabetización como en matemáticas. Tendencias semejantes surgen en las evaluaciones nacionales que sobre-muestran las zonas urbanas de fácil acceso en países

pobres. Además, en algunos países en vías de desarrollo, la dificultad de rastrear a niños nómadas puede ser onerosa y costosa como para que las autoridades educativas puedan incluirlos en las escuelas (UNESCO, 2010).

Otro problema que existe en casi todos los países es la variación de idioma entre los grupos étnicos. Muchos de estos grupos, a veces denominados minorías etnolingüísticas, están bien integrados en una «mezcla» nacional, como en Suiza; pero en otras situaciones estas variaciones pueden contribuir a conflictos civiles. En Latinoamérica, con más de 500 idiomas indígenas, es una región donde la educación intercultural bilingüe se está expandiendo para promover el cambio social; hasta la fecha, 12 gobiernos han institucionalizado la pedagogía multilingüe (Cortina, 2014). A menudo, las fuerzas sociales y políticas intentan ayudar a resolver las diferencias, generalmente incluyendo decisiones políticas que resultan en una jerarquía de idiomas aceptables para el uso en escuelas y estructuras de gobierno. En tales situaciones, ya sea en países de la OCDE o en países en vías de desarrollo, no es raro que los niños que hablan idiomas minoritarios sean excluidos de la investigación y evaluaciones de aprendizaje.

Este proceso de exclusión también ocurre en regiones donde los conflictos civiles o las crisis económicas provocan una migración transfronteriza sustancial, donde los grupos inmigrantes (y sus hijos) son tratados como transitorios, y donde los niños reciben poca o ninguna educación (Pigozzi, Carrol, Hayden y Ndaruhutse, 2014). El Informe de Monitoreo Global del 2010 describe como la marginación puede amenazar el logro educativo, ya que estos niños enfrentan muchos desafíos. En general los alumnos más marginados del mundo se enfrentan a «desigualdades, estigmatización y discriminación vinculadas a la riqueza, el género, la etnicidad, el idioma, la ubicación y la discapacidad» (UNESCO, 2010:5). La magnitud en que estos grupos están, o no están, incluidos en muestras de la población tiene implicaciones graves a medida que los investigadores desarrollan normas para los resultados del aprendizaje. La mayoría de los que forman parte de la población de interés pueden tratar a los «otros» como un grupo inferior que «no puede aprender». Irónicamente, en Sudáfrica, donde los pobres están en la mayoría, son los estudiantes rurales y pobres los que se sienten más marginados y desamparados (Babson, 2010). En resumen, tanto las muestras de habilidades como de población varían, así como también los procesos de aprendizaje que los individuos implementan y los contextos en los cuales estos tienen lugar.

Finalmente, debemos considerar agentes interesados que realizan el muestreo. Ya sea se trate de aquellos que diseñan las políticas públicas, psicómetras o profesores locales, todos estos llegan a la tarea de muestrear habilidades y poblaciones desde sus propios puntos de vista y experiencias. Las decisiones sobre que habilidades muestrear, entre que poblaciones, y en que idiomas y contextos, también contribuyen a un posible sesgo o a un ya complejo juego de asuntos de muestreo. Para hacer frente a estos sesgos, los investigadores pueden usar un rango de métodos incluyendo muestreo a medida y diseños de submuestras; las muestras de concordancia, el sobre-muestreo de poblaciones marginadas y diseños de métodos mixtos. Las consecuencias de estas varias restricciones pueden tener un impacto importante en la política y práctica pedagógica, y en la gobernanza mundial de la educación (Meyer y Benavot, 2013).

### 3. Credibilidad Metodológica

La investigación que puede convertirse en política depende de su credibilidad, lo que significa que especialistas debidamente capacitados deben llegar a un consenso sobre los méritos de un conjunto particular objetivo de resultados, incluso si no están de acuerdo con la interpretación de dichos resultados.

Las dos dimensiones más citadas de la credibilidad en la investigación del aprendizaje son la validez y la fiabilidad.

La validez de cualquier herramienta de medición de aprendizaje se puede determinar de varias maneras. En primer lugar, la validez interna está determinada por el grado en que los hallazgos pueden relacionarse de manera creíble con la lógica conceptual de la intervención minimizando el error sistemático, o sesgo. Por ejemplo, ¿Se relacionan realmente las preguntas en una prueba de opción múltiple con la capacidad del niño de leer o con la capacidad de recordar lo que él o ella ha leído antes? La validez puede variar significativamente según el contexto y la población, ya que una prueba que puede ser válida en Londres puede tener poca validez en Lahore. Del mismo modo, una prueba efectiva para un grupo lingüístico de hablantes de lengua materna puede ser inapropiada para niños que hablan ese mismo idioma como segunda lengua. Esta segunda forma de validez se denomina apropiadamente validez externa: se refiere a la replicabilidad de los hallazgos a través de distintos contextos. Si las estadísticas continúan acumulándose sin tener en cuenta el contexto local, las evaluaciones pueden malinterpretar a los alumnos en contextos de la BdP.

Se ha planteado un tercer tipo de preocupación por la validez con respecto a los países en vías de desarrollo: ¿hasta qué punto son válidas las opciones de los ítems de prueba y cuán apropiado es su contenido cuando son aplicadas a las culturas y a los sistemas escolares locales? Si bien muchas investigaciones de aprendizaje toman la forma de pruebas cuantitativas, los métodos cualitativos y etnográficos también pueden contribuir, particularmente con respecto a las variaciones culturales en el proceso de aprendizaje en contextos diversos.

A menudo, la fiabilidad se mide de dos maneras cuantitativas. En general, la confiabilidad es el grado en que los resultados de un individuo en una prueba están constantemente relacionados en tiempos adicionales en los que el individuo toma la misma prueba (o equivalente). Generalmente, la alta fiabilidad significa que el orden de rango de individuos que toman una prueba sería muy similar en una segunda ocasión que la tomen. Una segunda manera más fácil de medir la fiabilidad es mirar la función interna de los ítems de la prueba: ¿Tienen los ítems en cada parte de la evaluación una fuerte asociación entre sí? Esto es conocido como la confiabilidad entre-ítem, medida por la *estadística alfa* de Cronbach. Por supuesto, la fiabilidad implica poco sobre la validez del instrumento: el consenso de los investigadores de que el instrumento es relevante a los resultados educativos.

Desde el punto de vista cualitativo, la fiabilidad se alcanzaría cuando los etnógrafos con sensibilidad al contexto, por ejemplo, se pusieran de acuerdo en un conjunto de observaciones de procesos de aprendizaje que hayan recolectado de manera independiente en un contexto particular. Este es un ejemplo de «etnografía de equipo» que se está utilizando cada vez más en la investigación educativa en los Estados Unidos y Europa (Bartlett y Garcia, 2011; Blackledge y Creese, 2010). Además, el uso de experimentos controlados aleatorios (ECA) es visto como una manera importante de aumentar la credibilidad de los resultados de la investigación, comparando intervenciones con otros grupos de control. Evaluaciones hechas por Kremer y Holla (2009), Banerjee y Duflo (2011) y Bruns, Filmer y Patrinos (2011) apoyan el uso de ECA para mejorar la credibilidad de las investigaciones en el desarrollo internacional, mientras que otros (por ejemplo, Castillo y Wagner, 2014) sugieren algunas limitaciones graves en el uso de ECA para el diseño de la política educativa.

La diversidad de los resultados de aprendizaje se resume a menudo en términos de un rango promedio o normal que puede ser aplicado a lo largo de las dimensiones predecibles de una curva campaniforme

(Gurn, 2010). Como muchos han indicado, la conclusión que el comportamiento humano cae a lo largo de la curva normal, con la mayoría de las observaciones concentradas alrededor de un promedio discernible, simplifica excesivamente el rango y la diversidad de las experiencias humanas (Dudley-Marling y Gurn, 2010). La sociedad y la cultura influyen casi todos los aspectos de la condición humana, desde la inteligencia hasta la altura y el peso, de muchas maneras no aleatorias. En el ámbito del aprendizaje y desarrollo internacional, la sobre dependencia en interpretar los hallazgos a través de una perspectiva de la distribución normal contiene sesgos inherentes.

¿Por qué nos importa? Es potencialmente engañoso basar afirmaciones sobre el aprendizaje humano, y hacer predicciones al respecto, que estén basadas en una suposición de distribuciones normales. Hemos sostenido que, para los estudiantes en la BdP, el aprender ciencia puede ser sustancialmente diferente de aquellas poblaciones más favorecidas. Una estrategia útil podría ser enfocarse en lo que los estudiantes en la BdP contribuyen al proceso del aprendizaje en lugar de enfocarse en lo que les falta. Por ejemplo, Harper (2012) enmarca la investigación anti-déficit en contextos de educación de los Estados Unidos, además Moll, Amanti, Neff y González (1992) describen una orientación que examina los «fondos del conocimiento» o activos. Cada uno de estos enfoques respalda la noción de que es necesario centrarse más en qué y cómo se lleva a cabo el aprendizaje en la base de la pirámide.

#### 4. Comparabilidad de resultados de aprendizajes entre contextos

La comparabilidad es central en las bases de datos globales de educación, tales como la colección de datos a gran escala realizada por el Instituto de Estadística de UNESCO (UIS) y la OCDE. Sin embargo, si la meta principal es la comparabilidad, se puede prestar menos atención a la validez local y cultural de las definiciones y clasificaciones del aprendizaje. Además, los datos pueden volverse menos significativos y posiblemente menos aplicables a nivel local. Esta es una tensión natural y esencial entre los enfoques de la medición universalistas *etic* y los dependientes del contexto *emic*, y es particularmente relevante para el estudio de las poblaciones de la BdP. En un ejemplo conocido, los enfoques *emic* son aquellos que están conscientemente centrados en relevancia cultural y local, tales como palabras o descripciones locales para una persona «inteligente». Los enfoques *etic* son aquellos que pueden definir «inteligencia» como un concepto universal y tratar de medir a individuos entre culturas con ese único concepto o definición.

¿Pueden la comparabilidad y la sensibilidad al contexto estar balanceadas apropiadamente en la investigación del aprendizaje? ¿Deberían ser examinados los países con resultados por debajo del promedio usando las mismas escalas que países con resultados mucho más altos que el promedio? Si algunos países, o grupos de estudiantes, se encuentran en el «piso» de la escala, algunos dirían que la solución es bajar la escala a un nivel menor de dificultad. Otros podrían decir que la escala misma es defectuosa y que hay diferentes habilidades que podrían evaluarse mejor, especialmente si las variaciones son evidentemente causadas por variables culturales, étnicas, lingüísticas y variables relacionadas que llevan a uno a cuestionar la prueba tanto o más que el mismo grupo que está siendo examinado. Aun así, algunos dicen que tener diferentes escalas para diferentes grupos o naciones es un compromiso inaceptable de los puntos de referencias que buscan los formuladores de políticas internacionales, como el grupo operativo de aprendizaje métrico (Learning Metrics Task Force, Brookings Institution, 2013) o la Iniciativa Mundial para la Educación ante Todo de las Naciones Unidas (GEFI por sus siglas en inglés, 2014). Si el objetivo más importante es mejorar el aprendizaje en la BdP, ¿cuán creíbles son los hallazgos en la cola de la distribución de las evaluaciones internacionales (o incluso nacionales)?

En la medida en que se puede lograr la comparabilidad (y ninguna evaluación de aprendizaje afirma tener una comparabilidad perfecta), los resultados permiten a los responsables de la formulación de políticas considerar su propia situación nacional o regional en relación con otras. Esto parece tener más mérito cuando las opciones a ser tomadas se aplican a situaciones proximales, más que a las distales. Por ejemplo, consideren un país de África que ha adoptado un programa particular de educación bilingüe que parece funcionar bien en la escuela primaria. Si el ministro de educación de un país vecino cree que el caso es lo suficientemente parecido a su contexto nacional, entonces tiene sentido comparar los resultados en, por ejemplo, las pruebas de alfabetización en la escuela primaria. Una comparación distal podría ser observar que un cierto programa bilingüe de educación en Canadá parece ser eficaz, pero dudar de la posibilidad de aplicarlo en un contexto muy diferente en África. Pero la proximidad no siempre es la característica más pertinente; por ejemplo, la competición entre los resultados educativos y los sistemas económicos de los Estados Unidos y Japón han sido el motivo de discusiones y debates prominentes durante muchos años (Stevenson and Stigler, 1982). En un ejemplo más reciente, los altos funcionarios de Botsuana estaban interesados en saber cómo Singapur llegó a conseguir los mejores puntajes en matemáticas en varios LSEAs (Gilmore, 2005; ver también Sjoberg, 2007).

La cuestión clave aquí es el grado en que es necesario tener comparabilidad completa en los resultados del aprendizaje, con todos los individuos y grupos en las mismas escalas de evaluación. O si se elige no «forzar» las concesiones necesarias para una sola escala unificada, ¿cuáles son las ganancias y las pérdidas en términos de comparabilidad? ¿Pueden mantenerse los objetivos internacionales y las estadísticas proporcionales como estables y fiables si se eligen enfoques localizados sobre la comparabilidad internacional? Las respuestas a estas preguntas han llevado a situaciones donde algunos países en vías de desarrollo pueden estar tentados a participar en evaluaciones internacionales de aprendizaje, pero dudan en hacerlo porque sus resultados pueden resultar ser muy bajos en comparación. O bien, pueden concluir que el costo de participar no aumenta suficiente valor a la toma de decisiones a nivel nacional (Greaney y Kellaghan, 1996). Otros pueden participar porque no quieren ser vistos como que tienen índices de referencia inferiores a aquellos usados en países de la OCDE; para ver la reciente discusión de algunos de estos asuntos, ver OCDE (2014) y Bloem (2013).

Finalmente, la investigación internacional sobre el aprendizaje requiere alguna forma de comparabilidad, pero tal vez de formas más variadas de las que se consideran generalmente hoy en día. Por ejemplo, las evaluaciones internacionales y regionales pretenden específicamente a la comparabilidad entre países, mientras que las evaluaciones híbridas (Wagner, 2011b) están más enfocadas en contextos locales y en una mayor validez. Estas últimas intentan combinar aspectos de evaluaciones a gran escala con las de escala local y pueden considerarse más pequeñas, rápidas y baratas. Una de las primeras evaluaciones híbridas fue el proyecto de evaluación de la alfabetización de UNESCO (ILI, 2002); versiones posteriores pueden verse en las evaluaciones de alfabetización temprana que han crecido en popularidad (Gove y Wetterberg, 2011). Las evaluaciones híbridas ofrecen una validez localizada que las evaluaciones a gran escala no ofrecen y pueden brindar resultados más específicos para mejorar el aprendizaje y las intervenciones entre las poblaciones pobres y marginadas. La clase de comparabilidad de mayor importancia depende de las metas políticas de interés, así como también de las consideraciones del calendario y recursos.

## 5. ¿Qué funciones desempeñan las partes interesadas?

Muchos funcionarios interesados –incluyendo los formuladores de políticas, los ministros de educación, los líderes comunitarios en las aldeas rurales, los profesores, los padres y los especialistas en educación– deberían de ser responsables por el contenido y el proceso de aprendizaje de los niños. Los editores de revistas académicas y las universidades pueden desempeñar un papel al requerir que los investigadores ofrezcan más explicaciones intencionales de la representación y las implicaciones inherentes de las muestras que incluyen en sus estudios publicados. Sin embargo, incluso hoy, los especialistas en educación y los estadísticos en la mayoría de los países han sido los «guardianes» de los procesos de aprendizaje y su importancia para el éxito escolar y económico. Una de las principales razones de este acceso restringido al conocimiento sobre el aprendizaje es la complejidad de la ciencia empírica del aprendizaje, como se indicaba anteriormente.

Una segunda razón es la falta de conocimiento –y a veces creencias erróneas– entre padres e hijos sobre cuán importante es el aprendizaje y la educación para sus oportunidades de vida. Mucha evidencia, de muchas sociedades, sugiere que las personas en comunidades pobres subestiman el valor del aprendizaje y la educación escolar; por ejemplo, Stevenson y Stigler (1982) compararon las creencias de los padres en Estados Unidos, China y Japón. Hoy en día, es más importante que nunca involucrar a múltiples actores en la toma de decisiones educativas. En muchos países, el público se ha interesado más por el aprendizaje de los niños y los logros escolares en perspectiva comparativa, probablemente debido a la creciente globalización, a la influencia de las agencias internacionales, a los esfuerzos de las organizaciones no gubernamentales (ONGs), al mayor activismo comunitario y al interés de los padres. Algunos estudios de campo han involucrado un fuerte compromiso comunitario que ha llevado a los gobiernos a incorporar los hallazgos para el cambio de políticas; véase Bhattacharjea, Wadhwa, y Banerji (2011) en India, y Piper y Korda (2010) en Liberia.

Este tipo de intercambio de información a varios niveles es otra manera de vincular la ciencia con la responsabilidad y las expectativas. ¿A quién debe preocuparle si un niño, maestro, escuela, distrito o nación no se está desempeñando en un nivel dado de aprendizaje? De hecho, ¿cómo siquiera se construyen tales expectativas? ¿Cuáles expectativas se deben de tomar en cuenta? El conocimiento sobre la importancia del aprendizaje –y como este puede lograrse en contextos formales e informales y de manera estructurada e informal– tiene la posibilidad de abrir nuevos caminos en la investigación, el desarrollo de políticas, la participación comunitaria y familiar, y la apropiación local. No hay lugar donde esto sea más evidente que en la base de la pirámide, donde recientemente los padres y comunidades están tomando conciencia del papel que el aprendizaje puede desempeñar en la vida de sus hijos.

## 6. Conclusiones Políticas

En principio, las investigaciones sobre cómo mejorar el aprendizaje en países de bajos recursos y en comunidades pobres y marginadas –poblaciones de la BdP– no son más difíciles de llevar a cabo que las investigaciones semejantes en países más desarrollados. Sin embargo, teniendo en cuenta donde la mayoría de los recursos científicos (humanos y fiscales) se encuentran (en su mayoría en países de la OCDE), puede ser mucho menos conveniente para aquellos con la formación avanzada que se necesita para realizar el trabajo. Ese hecho, entre otros, es por qué aún queda mucho por saber sobre el aprendizaje en contextos de la BdP.

La manera en la cual el aprendizaje es estudiado en los países en vías de desarrollo, y específicamente en poblaciones de la BdP, podría tener una gran importancia científica para los investigadores y los funcionarios de la educación. A medida que avanzamos desde Jomtien, Dakar, y los ODM de las Naciones Unidas hacia los objetivos de desarrollo posteriores al 2015, queda claro que las inequidades sociales y económicas persistirán a menos que mantengamos un enfoque serio en el aprendizaje entre los pobres. En su libro seminal sobre nuevos enfoques para llegar a los mercados de consumo de la BdP, C.K. Prahalad (2006) desafió a las corporaciones a adoptar una nueva filosofía de prestación de servicios para las poblaciones históricamente ignoradas. Al transformar la manera en que el aprendizaje es entendido en contexto de la base de la pirámide, podemos empezar a entender cómo mejorar la promoción de políticas para aumentar la calidad educativa e incrementar los resultados del aprendizaje entre aquellos que son más difíciles de acceder.

## Referencias bibliográficas

- Arnett, Jeffrey Jensen (2008): "The neglected 95%: Why American psychology needs to become less American". *American Psychologist*, 63 (7), 602-614.
- Babson, A. (2010). *The place of English in expanding repertoires of linguistic code, identification and aspiration among recent high school graduates in Limpopo Province, South Africa*. Ph.D. dissertation: University of Michigan, Ann Arbor.
- Banerjee, Abhijit V. y Duflo, Esther (2011). *Poor economics: A radical rethinking of the way to fight poverty*. New York: Public Affairs.
- Bartlett, Lesley y Garcia, Ofelia (2011). *Additive schooling in subtractive times: Bilingual education and Dominican immigrant youth in the Heights*. Nashville: Vanderbilt University Press.
- Benavot, A. y Tanner, E. (2007). The growth of national learning assessments in the world, 1995–2006. Background paper prepared for EFA Global Monitoring Report 2008. Paris: UNESCO.
- Benedict, Mary Ellen y Hoag, John (2011): "Seating location in large lectures: Are seating preferences or location related to course performance?" *The Journal of Economic Education*, 35 (3), 215-239.
- Bhattacharjea, Suman; Wadhwa, Wilima y Banerji, Rukmini (2011). Inside primary schools: A study of teaching and learning in rural India. New Delhi: ASER. [http://img.asercentre.org/docs/Publications/Inside\\_Primary\\_School/Report/tl\\_study\\_print\\_ready\\_version\\_oct\\_7\\_2011.pdf](http://img.asercentre.org/docs/Publications/Inside_Primary_School/Report/tl_study_print_ready_version_oct_7_2011.pdf)
- Blackledge, Adrian y Creese, Angela (2010). *Multilingualism: A critical perspective*. London: Continuum.
- Bloem, S. (2013): "PISA in Low and Middle Income Countries". *OECD Education Working Papers*, 93.
- Braun, Henry y Kanjee, Anil (2006): "Using assessment to improve education in developing nations" en Joel E. Cohen, David E. Bloom, & Martin B. Malin (eds.): *Educating all children: A global agenda*. Cambridge, MA: MIT Press.
- Brookings Institution (2013). Toward universal learning: Recommendations from the Learning Metrics Task Force. Washington, DC: Brookings Institution Center for Universal Education. <http://www.brookings.edu/~media/Research/Files/Reports/2013/09/learning%20metrics%20task%20force%20universal%20learning/LTMF%20RecommendationsReportfinalweb.pdf>



- Bruns, Barbara; Filmer, Deon y Patrinos, Harry Anthony (2011). *Making schools work: New evidence on accountability reforms*. Washington: World Bank.
- Buchmann, Claudia y Hannum, Emily (2001): "Education and stratification in developing countries: A review of theories and research". *Annual Review of Sociology*, 27, 77-102.
- Castillo, Nathan M. y Wagner, Daniel A. (2014): "Gold standard? The use of randomized controlled trials for international educational policy". *Comparative Education Review*, 58 (1), 166-173.
- Cole, M.; Gay, J.; Glick, J. y Sharp, D. (1971). *The cultural context of learning and thinking*. New York: Basic Books.
- Cortina, Regina (ed.) (2014). *The education of indigenous citizens in Latin America*. Clevedon: Multilingual Matters.
- DBE [Department of Basic Education], South Africa (2013). Report on the Annual National Assessment of 2013: Grades 1 to 6 & 9. Pretoria: DBE. <http://www.education.gov.za/LinkClick.aspx?fileticket=Aiw7HW8ccic%3D&tabid=36>
- Dudley-Marling, Curt y Gurn, Alex (eds.) (2010). *The myth of the normal curve*. New York: Peter Lang.
- Engel, L. C. & Feuer, M. J. (2014): "Five myths about international large-scale assessments". *Quality Assurance in Education*, 22, 18-21.
- GEFI [Global Education First Initiative] (2014). Global Education First Initiative: An initiative of the United Nations Secretary-General. [http://www.globaleducationfirst.org/files/GEFI\\_Brochure\\_ENG.pdf](http://www.globaleducationfirst.org/files/GEFI_Brochure_ENG.pdf)
- Gilmore, A. (2005). *The impact of PIRLS (2001) and TIMSS (2003) in low - and middle-income countries: An evaluation of the value of World Bank support for international surveys of reading literacy (PIRLS) and mathematics and science (TIMSS)*. New Zealand: IEA.
- Gove, Amber y Wetterberg, Anna (2011). *The early grade reading assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI.
- Greaney, Vincent y Kellaghan, Thomas (1996). *Monitoring the learning outcomes of education systems*. Washington: World Bank.
- Gurn, Alex (2010): "Conclusion: Re/visioning the ideological imagination in (special) education" en Curt Dudley-Marling & Alex Gurn (eds.): *The myth of the normal curve*. New York: Peter Lang.
- Harper, Shaun R. (2012). *Black male student success in higher education: A report from the National Black Male College Achievement Study*. Philadelphia: University of Pennsylvania, Center for the Study of Race and Equity in Education.
- Heinrich, Joseph, Heine, Steven y Norenzayan, Ara (2010): "The weirdest people in the world?" *Behavioral Brain Science*, 33 (2-3), 85-135.
- Kagan, J., Moss, H.A. y Sigel, I.E. (1963): "Psychological significance of styles of conceptualization". *Monographs of the Society for Research in Child Development*, 28 (4), 73-112.
- Kahneman, Daniel (2011). *Thinking, fast and slow*. New York: Macmillan.

- ILI [International Literacy Institute] (2002). Towards guidelines for the improvement of literacy assessment in developing countries: Conceptual dimensions based on the LAP Project. Philadelphia: ILI/UNESCO.
- Korinek, Kim & Punpuing, Sureeporn (2012): "The effect of household and community on school attrition: An analysis of Thai youth". *Comparative Education Review*, 56 (3), 474-510.
- Kremer, Michael y Holla, Alaka (2009): "Improving education in the developing world: What have we learned from randomized evaluations?" *Annual Review of Economics*, 1, 513-542.
- Li, Jin (2003): "U.S. and Chinese cultural beliefs about learning". *Journal of Educational Psychology*, 95, 258-267.
- Lu, Yao y Treiman, Donald J. (2011): "Migration, remittances and educational stratification among Blacks in Apartheid and Post-Apartheid South Africa". *Social Forces*, 89 (4), 1119-1143.
- Meyer, Heinz-Dieter y Benavot, Aaron (eds) (2013). *PISA, power and policy: The emergence of global educational governance*. Oxford: Symposium Books.
- Moll, Luis C.; Amanti, Cathy; Neff, Deborah y Gonzalez, Norma (1992): "Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms". *Theory into Practice*, 31 (2), 132-141.
- OECD (2014). PISA for development. Paris: OECD. <http://www.oecd.org/pisa/aboutpisa/pisafordevelopment.htm>
- Pigozzi, Mary Joy; Carrol, Bidemi, Hayden, Jacqueline y Ndaruhutse, Susy (2014): "Fragile and conflict-affected situations" en Daniel A. Wagner (ed.): *Learning and education in developing countries: Research and policy for the post-2015 UN development goals*. New York: Palgrave Macmillan.
- Piper, Benjamin y Korda, Medina (2010). Early Grade Reading Assessment (EGRA) Plus: Liberia. Program evaluation report: Prepared for USAID/Liberia. Research Triangle Park, NC: RTI International.
- Prahalad, C.K. (2006). *The fortune at the bottom of the pyramid: Eradicating poverty through profits*. Upper Saddle River: Wharton School Publishing.
- Saito, M. (2011). Gender equality in education: Looking beyond parity. Paris: UNESCO-IIEP. [http://www.sacmeq.org/sites/default/files/sacmeq/publications/7\\_1\\_1\\_5\\_mioko\\_saito\\_eng\\_version\\_2012\\_02\\_15.pdf](http://www.sacmeq.org/sites/default/files/sacmeq/publications/7_1_1_5_mioko_saito_eng_version_2012_02_15.pdf)
- Sjøberg, Svein (2007). "PISA and 'real life challenges': Mission impossible?" en S. T. Hopmann, G. Brinek & M. Retzl (eds.): *PISA according to PISA: Does PISA keep what it promises?* Vienna: LIT Verlag. <http://folk.uio.no/sveinsj/Sjoberg-PISA-book-2007.pdf>.
- Stevenson, Harold W. y Stigler, James W. (1982). *The learning gap: Why our schools are failing and what we can learn from Japanese and Chinese education*. New York: Summit.
- UNESCO (2003). *Education for All Global Monitoring Report: Gender and Education for All - The leap to equality*. Paris: UNESCO.
- UNESCO (2010). *Education for All Global Monitoring Report: Reaching the marginalized*. Paris: UNESCO.
- United Nations (2000). United Nations millennium declaration. Resolution adopted by the General Assembly. United Nations A/RES/55/2. <http://www.un.org/millennium/declaration/ares552e.htm>

- Wagner, Daniel A. (1983): Rediscovering 'rote': Some Cognitive and Pedagogical Preliminaries en S. Irvine y J. W. Berry (eds.): *Human assessment and cultural factors*. New York: Plenum.
- Wagner, Daniel A. (1993). *Literacy, culture and development: Becoming literate in Morocco*. New York: Cambridge University Press.
- Wagner, Daniel A. (2010): "Quality of education, comparability, and assessment choice in developing countries". *COMPARE: A Journal of Comparative and International Education*, 40 (6), 741-760.
- Wagner, Daniel A. (2011a): "What happened to literacy? Historical and conceptual perspectives on literacy in UNESCO". *International Journal of Educational Development*, 31, 319-23.
- Wagner, Daniel A. (2011b). *Smaller, quicker, cheaper: Improving learning assessments in developing countries*. Paris and Washington: UNESCO IIEP and EFA Fast Track Initiative of Global Partnership for Education.
- Wagner, Daniel A. (ed.). (2014). *Learning and education in developing countries: Research and policy for the post-2015 UN Development Goals*. New York: Palgrave Macmillan.
- Wagner, Daniel A.; Castillo, Nathan M. (2014): "Learning at the bottom of the pyramid: Constraints, comparability and policy in developing countries". *Prospects: Quarterly Review of Comparative Education*, 44, 627-638.
- Wagner, Daniel A.; Murphy, Katie M. y de Korne, Haley (2012). *Learning first: A Research Agenda for Improving Learning in Low-Income Countries*. Washington: Brookings Institution.
- Witkin, H. A.; Moore, C. A.; Goodenough, D. R. y Cox, P. W. (1977): "Field-Dependent and Field-Independent Cognitive Styles and their Educational Implications". *Review of Educational Research*, 47, 1-64.

## Notas biográficas

**Daniel A. Wagner** es *UNESCO Chair*, profesor de educación, y director del Instituto Internacional de Alfabetismo y el Programa Internacional de Desarrollo Educativo en la University of Pennsylvania (Estados Unidos). El Dr. Wagner tiene una extensa experiencia en temas educativos nacionales e internacionales y ha servido como asesor a UNESCO, UNICEF, el Banco Mundial, USAID, DFIC, y otras agencias en cuestiones de desarrollo internacional. El Dr. Wagner es Fellow de la Asociación Americana de Psicología, la Asociación Americana de Antropología y la Asociación Americana de Investigación Educativa.

**Nathan M. Castillo** es Doctor en Educación por la University of Pennsylvania (Estados Unidos) y Máster en Política Educativa Internacional de la University of Harvard. Actualmente es investigador asociado en el Instituto Internacional de Alfabetismo de la University of Pennsylvania. Anteriormente, ha trabajado en el Centro para el Desarrollo de la Educación en Washington, DC. Del 2006 al 2008, fue coordinador municipal para la salud escolar mientras servía en el Cuerpo de Paz de los Estados Unidos en Guatemala.

# A tradução da política de avaliação pelos programas de pós-graduação em educação no Brasil

## The Translation of the Evaluation Policy by the Brazil's Postgraduate Programmes in Education

Regina Célia Linhares Hostins<sup>1</sup>

### Resumo

O artigo se insere no âmbito dos estudos sobre políticas educacionais, notadamente sobre as políticas de avaliação da pós-graduação em Educação, intensificadas no Brasil, na década de 1990, com a reforma educacional de matriz neoliberal. Busca-se analisar os processos de interpretação e tradução da política pelos Programas de Pós-Graduação em Educação que, na avaliação do triênio 2010, emergiram do conceito 5 para 6, alcançando o patamar de excelência, equivalente ao dos centros internacionais de pesquisa na área. Trata-se de pesquisa documental, cujas fontes de análise são documentos produzidos: 1) pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -CAPES, vinculada ao Ministério da Educação e responsável pela regulação das políticas de pós-graduação no Brasil; 2) pelos Programas de pós-graduação e postados no sistema de coleta CAPES, nos anos que compuseram a avaliação trienal 2010 (2007-2009). Os documentos estão disponíveis no site da CAPES e são de acesso livre. O estudo tem como base teórico-metodológica, a teoria da interpretação/tradução da política no contexto da prática (Theory of policy enactment) proposta por Ball, Maguire e Braun. Os resultados indicam que a política de avaliação, mesmo que altamente delimitada e configurada por rígidos critérios e padrões internacionais não é simplesmente implementada –linear e acriticamente– no contexto da prática. Nos limites consentidos, nos variados contextos relacionais envolvidos, os atores produzem uma leitura ativa dos textos da política, que se transforma em objeto de recontextualização, reconfiguração e tradução.

### Palavras-chave

Política educacional, política de avaliação, pós-graduação, «policy enactment», performatividade.

### Resumen

El artículo aborda las políticas educativas, especialmente las políticas de evaluación de postgrado implementadas en la reforma educativa neoliberal de Brasil en los años noventa. El propósito es analizar los procesos de interpretación y traducción de la política dirigida a los programas de postgrado en educación que, en la evaluación del trienio 2010, emergieron del concepto 5 a 6, alcanzando el nivel de excelencia, equivalente al de los centros de investigación internacional en el área. Se trata de una investigación documental, cuyas fuentes de análisis son documentos producidos por 1) la Coordinación de Perfeccionamiento de Personal de Nivel Superior –CAPES, vinculada al Ministerio de Educación y responsable de la regulación de las políticas de postgrado en Brasil; 2) los programas de postgrado y aplicados en el sistema CAPES, en los años que comprendieron la evaluación trienal 2010 (2007-2009). Los documentos están disponibles en la plataforma de CAPES y son de acceso libre. El estudio tiene como base teórico-empírica, la teoría de la interpretación-traducción de la política en el contexto de la práctica (*Theory of policy enactment*) propuesta por Stephen Ball, Maguire y Braun. Los resultados indican que la política de evaluación, incluso estando muy delimitada y configurada por criterios severos y modelos internacionales, no es simplemente implementada –de manera lineal y acrítica– en el contexto de la práctica. Dentro de los márgenes permitidos, en los variados contextos relacionales implicados, los actores producen una lectura activa de los textos de la política, que se transforma en un objeto de recontextualización, reconfiguración y traducción.

### Palabras clave

Política educative, política evaluadora, postgrado, «policy enactment», performatividad.

<sup>1</sup> Universidade do Vale do Itajaí, Brasil, Regina Célia Linhares Hostins, [reginalh@univali.br](mailto:reginalh@univali.br)

## Abstract

The study discusses educational policies, especially postgraduate evaluation policies implemented in Brazil's neoliberal education reform in the 1990s. The aim is to analyze the interpretation and translation of Brazil's postgraduate education programmes' evaluation policy which, in the evaluation of the triennium 2010, emerged from the concept 5 to 6, reaching the level of excellence, equivalent to the international research centers in education. It is a documentary research, whose sources of analysis are documents produced: 1) by the Coordination of Improvement of Higher Education - CAPES, linked to the Ministry of Education and responsible for the regulation of postgraduate policies in Brazil; 2) by the postgraduate programmes and applied in the CAPES system, in the years of the triennial evaluation of 2010 (2007-2009). The study is based on the theory of 'policy enactments' proposed by Stephen Ball, Maguire and Brown. The results indicate that the evaluation policy, even if highly delimited by strict criteria and international standards, is not simply implemented – linearly and uncritically – in the context of practice. Within the limits allowed, in the relational contexts involved, the actors produce an active reading of the policy texts, which becomes an object of recontextualization, reconfiguration and translation.

## Key words

Educational policy, evaluation policy, postgraduate education, policy enactment, performativity.

Recibido: 30-03-2017  
Aceptado: 02-05-2017

## 1. Introdução

O ano de 1996 marcou a proposição de mudanças na política de avaliação da pós-graduação no Brasil, momento em que os ajustes neoliberais da reforma educacional atingiram seu ponto máximo. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, vinculada ao Ministério de Educação e responsável pela implantação dessa política no país, a partir dessa data, introduziu um «novo paradigma de referência», fortalecida pela necessidade de introduzir indicadores que expressassem os níveis de concorrência e de competitividade entre os programas. Nas palavras do ex-presidente da agência, Abílio Afonso Baeta Neves (Brasil, 2002:6), as mudanças na sistemática de avaliação fundamentavam-se no «reconhecimento claro de que a avaliação não discriminava mais a excelência», pois a maioria expressiva dos programas obteve conceitos A e B, tanto nas avaliações de 1994, como nas de 1996. Nesse sentido, a CAPES mudou para uma avaliação competitiva entre os programas, a partir de indicadores de excelência de qualidade, que a cada momento poderiam ser movidos para cima.

Sob a concepção dominante dos necessários ajustes neoliberais na «transformação das formas de estado» (Shaw, 1997:497) e na configuração das políticas, os quais envolviam a propagação epidêmica de privatizações, flexibilidade e avaliação, a reforma universitária foi implantada no Brasil e produziu mudanças radicais no ensino superior e na pós-graduação.

A Lei da Reforma Educacional (Lei das Diretrizes e Bases da Educação Nacional – LDBEN N 9394/1996) aprovada em 1996 e as regulamentações dela decorrentes, ampliaram as possibilidades de um arranjo particular de governança para tornar o sistema de ensino superior «mais acessível a uma estratégia de acumulação global» (Robertson *et al.*, 2006:233).

Na pós-graduação, a reforma promoveu mudanças nos financiamentos das pesquisas, nas relações com o setor privado via mestrados de caráter profissional, na internacionalização e no controle da eficiência por meio do sistema de avaliação.

Em face dessa reforma, a CAPES, a partir de 1998, realizou novas adequações na política de avaliação que passou a compreender as atividades de acompanhamento anual dos programas e a Avaliação Trienal propriamente dita, efetuada no ano subsequente ao do fechamento do triênio, momento no qual era atribuída a nota resultante da avaliação. Inicia-se assim o processo de avaliação continuada da pós-graduação no Brasil.

Na década de 2000, definem-se as condições nas quais a política de pós-graduação deveria enquadrar-se e os critérios de avaliação valorizados. «Entre as novidades ou alterações figuram a pós-graduação stricto-sensu de caráter profissional, o incentivo à cooperação internacional [...], ou ainda a vinculação dos resultados da avaliação à homologação do Conselho Técnico-Científico (CTC) da agência» (Brasil, 2002a:26).

Três estratégias foram relevantes na mudança de abordagem da política: 1) a avaliação continuada com o monitoramento anual dos Programas e emissão de conceito no ano após o fechamento do triênio; 2) a adoção de uma nova escala de avaliação com scores de 1 a 7, sendo 1 e 2 para scores abaixo do requerido e 6 e 7 para performances de excelência; 3) a introdução de um conjunto de critérios internacionais para distinguir a excelência: alta qualidade e presença do corpo docente no exterior, pesquisa científica e produção intelectual.

O que se depreende dessas estratégias é que, na reforma de pós-graduação brasileira, na década de 1990, a política de avaliação enfatiza a competitividade, a performance e a performatividade<sup>2</sup> das instituições, dos Programas, dos professores e estudantes, que passam a concorrer entre si em busca das melhores credenciais e do acesso aos financiamentos de suas pesquisas. Nesse caso, a competitividade posiciona os vencedores e os perdedores, os produtivos e os improdutivos, numa «luta em torno da visibilidade» (Ball, 2006:693).

Para acessar às informações e controlar os programas, a CAPES criou um Sistema computadorizado de dados, o qual é continuamente atualizado pelos Programas. O sistema desenvolveu um modo mais preciso de controle e exploração dos indicadores de produtividade e no provimento de informações quantitativas e qualitativas de monitoramento da produção intelectual de professores e alunos.

Paralelamente ao sistema de coleta de dados, a agência constituiu um sistema de avaliação entre pares, com a criação das Comissões de Avaliação por áreas – CA, as quais são compostas por professores/pesquisadores de renome nas respectivas áreas de conhecimento, responsáveis por realizar as avaliações fundamentados nos critérios estabelecidos. As CA avaliam os Programas e submetem seus pareceres à homologação do Conselho Técnico-Científico (CTC) da agência.

Essa complexa infraestrutura de dados e de envolvimento dos pesquisadores como sujeitos e objetos da avaliação, associada às medidas em escalas e aos critérios de comparabilidade, com base em critérios internacionalizados definem o que Ozga e Lingard (2007) denominam de «política de número». Essa se espalha para todos os estados-nação, produzem efeitos nas políticas nacionais e transformam as relações entre pesquisadores, professores, estudantes e instituições.

O artigo discute performance e performatividade na política de avaliação da pós-graduação no Brasil, tendo como foco de análise os discursos produzidos, tanto pela Coordenação da Área de Educação (CA-

---

<sup>2</sup> Stephen Ball emprega o conceito de performatividade como «uma tecnologia, uma cultura e um modo de regulação que emprega julgamentos, comparações e exibições como meios de incentivo, controle, disputas e mudança» (Ball, 2003:216).

ED) na orientação, acompanhamento e avaliação da pós-graduação, como pelos Programas avaliados como de excelência e padrão internacional, nos anos que compuseram a avaliação trienal do ano base 2010 (2007-2009), notadamente os que emergiram do conceito 5 para o 6, no triênio.

A escolha do período em foco –2007 a 2010– não foi aleatória. Neste intervalo de tempo, consolida-se o processo de avaliação continuada da pós-graduação, após um período de reformulações e sucessivas adaptações da sistemática na década de 1990.

Na avaliação do primeiro triênio (1998-2000), após implantação do novo modelo avaliativo, o processo passou por ajustes, discussões, sérios embates e negociações entre as Comissões de Avaliação por Áreas, o Conselho Técnico-Científico - CTC da agência e os programas de Pós-Graduação. Somente no triênio iniciado em 2001 esse processo se estruturou no formato como se apresenta nos dias de hoje, com algumas alterações qualitativas nos últimos três anos.

Tendo como sustentação a abordagem teórico-metodológica de Ball *et al.*, (2012), parte-se do pressuposto de que as políticas são interpretadas e traduzidas de diferentes e variadas formas para contextos específicos. «Assim, o material, o estrutural e o relacional precisam ser incorporados na análise de políticas para fazer o melhor sentido na colocação da política em ação no nível institucional» (Ball *et al.*, 2012: 21). Para os autores, nos limites consentidos, nos variados contextos materiais e relacionais envolvidos, os atores produzem uma leitura ativa dos textos da política, que se transforma em objeto de recontextualização, interpretação e tradução.

A teoria da interpretação/tradução da política no contexto da prática (*Theory of policy enactment*) tem sido discutida por Stephen Ball *et al.* (2012) e empregada como recurso teórico-metodológico na análise de políticas. Para eles a interpretação é uma leitura inicial, com o objetivo de aproximar-se do sentido da política. Trata-se de uma leitura política e substantiva vinculada à linguagem da política. A tradução por sua vez vincula-se à compreensão do texto dentro dos limites da ação no quais ocorre um processo de re-representação, reordenação e refundamentação que se processam por meio de várias práticas materiais e discursivas (Ball *et al.*, 2012).

Tendo por referencial a abordagem teórico-analítica apresentada, no estudo aprofundam-se as discussões e análises no campo preciso da política de avaliação da pós-graduação e, a partir desse *locus* particular, indaga-se: como os programas de Pós-Graduação em Educação (que emergiram da nota 5 para a 6 no triênio avaliado) interpretam e traduzem as múltiplas exigências da política de avaliação da Pós-Graduação dentro dos recursos a eles disponibilizados? Como isso se revela na construção dos seus textos nos relatórios de avaliação?

Essas questões norteiam a análise documental e o debate sobre a tradução da política de avaliação da pós-graduação no Brasil pelos principais atores da política, ou seja, os Programas avaliados.

## 2. Contribuições de Stephen Ball para análise das políticas educacionais

O método teórico-analítico desenvolvido por Stephen Ball e colaboradores destaca a «natureza complexa e controversa da política educacional, enfatiza os processos micropolíticos e a ação dos profissionais que lidam com as políticas no nível local e indica a necessidade de se articularem os processos macro e micro na análise de políticas educacionais» (Mainardes, 2006:49).

A teoria da interpretação/tradução da política no contexto da prática (*Theory Of Policy Enactment*), foi desenvolvida, mais recentemente, por Ball, Maguire e Braun (publicada no livro: *How schools do policy: policy*

*enactments in secondary schools*, 2012), a partir de uma pesquisa realizada em escolas secundárias na Inglaterra e financiada pelo Conselho de Pesquisa Social (ESRC) intitulada: *Policy enactments in the secondary school*, desenvolvida entre outubro de 2008 a abril de 2011.

«*Enactments*» constituem respostas em curso à política, às vezes durável, às vezes frágil, nas diversas redes e cadeias de relações, mas essa resposta não é direta nem reproduz linearmente as diretrizes da política. É na interação e inter-relação entre diversos atores, textos, conversas, tecnologias e objetos (artefatos) que a política é interpretada, traduzida, reconstruída e refeita em diferentes, mas similares caminhos (Ball *et al.*, 2012).

Para os autores, a interpretação é apenas uma leitura inicial, com o objetivo de aproximar-se do sentido da política. A tradução, por sua vez, vincula-se à compreensão do texto dentro dos limites da ação, no quais ocorre um processo de re-representação, reordenação que se processam por meio de várias práticas materiais e discursivas.

O aprofundamento de abordagens dessa natureza assume um papel significativo no campo de pesquisa em políticas educacionais, considerado relativamente novo no Brasil e «ainda não consolidado em termos de referenciais analíticos consistentes, e que, de resto, sofre as indefinições resultantes da crise de paradigmas que afeta as ciências sociais e humanas na contemporaneidade» (Azevedo e Aguiar, 1999:43).

Em uma perspectiva que se distancia de posições binárias, Ball (2011:44) defende uma posição epistemologicamente diferente e declara-se interessado em questões sobre «a ontologia da política», ou sobre a «forma como nós tornamos as políticas encarnadas». O autor busca romper com a ideia de que as políticas são feitas para as pessoas e estas as implementam. As políticas são antes objetos de alguma forma de tradução ou de leitura ativa; um tipo de «ação social criativa». Em seus processos e atos, é necessário capturar não os efeitos sobre coletividades sociais abstratas, mas antes a interação complexa de identidades, interesses, coalizões e conflitos (Ball, 2011).

Braun, Ball e Maguire (2011) exploram as formas com a qual os professores e gestores interpretam, adaptam ou transformam as políticas por meio da lente de seus valores, conhecimentos pré-existentes e práticas. Eles analisam o papel de diferentes tipos de atores políticos (policy actors) e identificam diferentes tipos de políticas, temas da política e as formas com que a política fala aos professores. Os autores consideram que enquanto muita atenção tem sido dada para avaliar quão bem as políticas são implementadas, ou seja, o quão bem elas são realizadas na prática, menos atenção tem sido dada para a compreensão e documentação das formas com as quais as escolas realmente lidam com o múltiplo, e às vezes, com as demandas opacas e contraditórias de diferentes «tipos» de política.

Em face desses pressupostos, no artigo ora apresentado, busca-se evidenciar as interpretações e traduções da política pelos atores políticos envolvidos de forma direta na avaliação da pós-graduação, seja como avaliadores ou como avaliados.

### 3. Procedimentos Metodológicos

A pesquisa, de caráter qualitativo, foi conduzida por meio de pesquisa documental cujas fontes foram coligidas no portal da CAPES ([WWW.capes.gov.br](http://WWW.capes.gov.br)) no menu Avaliação.

#### 3.1. População e delimitação do estudo

Foram objetos de estudo os Programas de Pós-Graduação em Educação que no triênio 2010 da ava-



liação continuada da CAPES emergiram da nota 5 para a 6 e, a partir de então, passaram à condição de Programas internacionais de excelência, com nível de desempenho altamente diferenciado em relação aos demais programas da área.

Considerou-se que os Programas nessa condição poderiam melhor expressar o movimento de interpretação e tradução das políticas no contexto da prática. Afinal, para garantir a passagem da nota 5 para 6 deve haver um grande investimento do Programa e, por consequência, de todos os atores envolvidos, para alcançar os critérios de alto desempenho requeridos como:

1. Conceito Muito Bom nos cinco quesitos avaliados;
2. Média de produção bibliográfica por docente permanente de 21 pontos, ou seja, cerca de 20% acima da necessária para conceito Muito Bom (o equivalente em pontos a 1,5 artigo em periódico internacional A por docente/ano)
3. Número de titulados de doutorado correspondente a, no mínimo, 80% do total de docentes permanentes no decorrer de todo triênio.

Os Programas que atingem esses requisitos mínimos sofrem uma segunda avaliação, realizada por uma subcomissão, de acordo com os critérios definidos CTC sendo eles:

Item 1: Nível de Qualificação, de produção e de desempenho equivalente ao dos centros internacionais de excelência na formação de recursos humanos.

Item 2: Consolidação e liderança nacional do Programa como formador de recursos humanos para a pesquisa e a pós-graduação.

Item 3: Inserção e impacto regional e nacional do Programa considerando indicadores de integração e solidariedade com outros programas com vistas ao aprimoramento do sistema de pós-graduação e de visibilidade de sua atuação.

Considerando os resultados da avaliação dos triênios dos anos base 2007 e 2010 participaram da amostra quatro (4) programas: um pertencente a uma universidade confessional e três pertencentes a universidades públicas (uma estadual e duas federais), a saber: Pontifícia Universidade Católica do Rio Grande Do Sul – PUCRS, Universidade Federal Fluminense – UFF, Universidade de São Paulo – USP e Universidade Federal de São Carlos – UFSCar.

### **3.1.1 Perfil dos Programas estudados**

*Programa de Pós-Graduação em Educação da Pontifícia Universidade Católica do Rio Grande Do Sul – PUCRS:* O Programa faz parte da Faculdade de Educação e é integrado pelo Curso de Mestrado (que teve início em 1972 e foi credenciado em 1976) e o Curso de Doutorado (que foi criado em 1988 e credenciado em 1995). À época, possuía corpo docente formado por 14 professores.

*Programa de Pós-Graduação em Educação da Universidade Federal Fluminense – UFF:* O Programa mantém o Mestrado Acadêmico, desde 1971, e o Doutorado, criado em 1995. O corpo docente do Programa, em 2009, constituiu-se de 33 professores permanentes e 10 professores colaboradores.

*Programa de Pós-Graduação em Educação da Universidade de São Paulo – USP:* Criado em 1971, o Programa contou, no ano de 2009, com 96 docentes, dos quais 77 do grupo de docentes permanentes, 14 colaboradores e 5 colaboradores aposentados que estavam se desligando do programa.

*Programa de Pós-Graduação em Educação Especial da Universidade Federal de São Carlos – UFSCar*: o Programa completou em 2009, 32 anos de funcionamento do curso de mestrado e dez anos do curso de doutorado, e neste período consolidou um centro reconhecido de pesquisas no país e no exterior, com um corpo docente constituídos por 29 docentes.

### 3.2. Procedimentos de coleta e análise de dados

Foram selecionados os documentos: 1) Cadernos de Indicadores: gerados a partir dos dados preenchidos pelos Programas no processo de avaliação continuada e 2) Documentos de Áreas, Relatórios de Avaliação e Resultados da Avaliação de Programas gerados pela CAPES e preenchidos pela Comissão de Avaliação da Área de Educação (CA-Ed).

Os procedimentos de aproximação e análise documental foram assim sistematizados:

1. Leitura prévia dos documentos cadastrados no Coleta CAPES pelos Programas selecionados e divulgados nos Cadernos de Indicadores buscando indícios reveladores das semelhanças e diferenças entre um texto e outro, mediante anotações prévias das regularidades e das diferenças, serviram de elementos para uma primeira classificação ou descrição das evidências mais significativas.
2. Tabulação dos dados em planilha eletrônica Excel.
3. Análise da planilha no sentido de «rastrear a conexão íntima» dos dados. Nessa etapa foram destacadas as palavras-chave dos textos e as categorias de menor e maior complexidade, por meio da Análise de conteúdo.
4. Análise de correlação entre os Programas alternando leituras verticais e horizontais das categorias por meio de tabela dinâmica.
5. Sistematização e aprofundamento das análises estabelecendo correlações com o referencial teórico estudado e outros autores e obras evidenciados na análise da empiria.

#### *Documentos coligidos*

BRASIL. CAPES. Relatório de avaliação 2007-2009. Trienal 2010. Comissão de área da Educação.

BRASIL. CAPES. Relatório de avaliação 2004-2006. Trienal 2007. Comissão de área da Educação.

BRASIL. CAPES. Comissão de Área da Educação. Ficha de Avaliação do Programa de Pós-Graduação PUC/RS - Pontifícia Universidade Católica do Rio Grande do Sul. Ano base 2010 (2007-2009).

BRASIL. CAPES. Caderno de Indicadores: proposta do programa. Pontifícia Universidade Católica do Rio Grande do Sul. Ano base 2010 (2007-2009).

BRASIL. CAPES. Comissão de Área da Educação. Ficha de Avaliação do Programa de Pós-Graduação da UFF - Universidade Federal Fluminense. Ano base 2010 (2007-2009).

BRASIL. CAPES. Caderno de Indicadores: proposta do programa. UFF - Universidade Federal Fluminense. Ano base 2010 (2007-2009).

BRASIL. CAPES. Comissão de Área da Educação. Ficha de Avaliação do Programa de Pós-Graduação da UFSCAR - Universidade Federal de São Carlos. Ano base 2010 (2007-2009).

BRASIL. CAPES. Caderno de Indicadores: proposta do programa. UFSCAR - Universidade Federal de São Carlos. Ano base 2010 (2007-2009).

BRASIL. CAPES. Comissão de Área da Educação. Ficha de Avaliação do Programa de Pós-Graduação da USP - Universidade de São Paulo. Ano base 2010 (2007-2009).

BRASIL. CAPES. Caderno de Indicadores: proposta do programa. USP - Universidade de São Paulo. Ano base 2010 (2007-2009).

WEBCAPES. Disponível em <http://www.capes.gov.br>.

#### 4. A tradução da política pelos Programas do Pós-Graduação: construção dos discursos em torno do critério de organicidade

A análise iniciou com o estudo do documento: Propostas dos Programas (PO), –de cada um dos anos avaliados– estabelecendo correlações com a ficha de avaliação do Programa pela CA-Ed, no sentido de apreender quais mudanças, incorporações, alterações e adaptações foram sendo processadas na trajetória e como os programas foram interpretando e traduzindo os elementos da política e construindo seus discursos em resposta a ela.

A análise, sustentada na teoria de Ball (2011) e Ball *et al.*, (2012), parte do pressuposto de que os Programas se consolidam com base na informação que constroem e transmitem sobre si. Na construção do discurso, estes se articulam num jogo representacional de luta por visibilidade e de intensificação da produção e da qualidade.

Esse exercício evidenciou uma riqueza de dados e um expressivo potencial de discussão da Política de Avaliação, desde a perspectiva dos Programas. Para os propósitos deste artigo e considerando os limites do texto, buscar-se-á analisar como a categoria *organicidade* priorizada no instrumento de avaliação da Comissão de Área (CA-Ed) e central no processo de avaliação, a partir desse período, foi traduzida e se materializou na produção dos discursos dos Programas<sup>3</sup>.

Essa categoria avaliava a coerência, consistência e abrangência entre área(s) de concentração, linhas e projetos de pesquisa, publicações de docentes, estrutura curricular, dissertações e teses dos Programas, de modo delimitar de maneira clara o objeto de especialidade da produção do conhecimento e da formação realizada no Programa e por ele oferecida (Brasil, CA-Ed, 2007).

Buscou-se então identificar o diálogo que se estabeleceu entre, de um lado, o avaliador com os quesitos, as apreciações e as recomendações em torno da organicidade e de outro, o avaliado com as justificativas, as descrições minuciosas e as representações performáticas.

Nesse suposto diálogo há um desenho de expectativas e um movimento intenso de descrições, definições, estratégias e representações na tentativa de correspondê-las. Instala-se assim, o que Ball (2010:39) define como novo modo de regulação em que o que está à frente não é mais a organização do poder dentro de formas de tempo-espaço definitivos. É a base de dados, a reunião de avaliação, a avaliação anual, a elaboração de relatórios e os formulários, as inspeções, a avaliação dos pares que estão à frente. «A questão não é sobre a possível certeza de ser sempre vigiado, como ocorre no panóptico. Se trata, ao contrário, da incerteza e da instabilidade de ser julgado de diferentes maneiras, por diferentes meios e diferentes

3 Utiliza-se a denominação numérica para identificação do Programas: P1, P2, P3 e P4, definida de forma aleatória para evitar o uso nominal.

agentes». O que se produz então é um «estado de permanente e consciente visibilidade» (Ball, 2010:40).

Tão complexa é essa questão no interior dos programas que na avaliação do triênio 2007, a CA-Ed identificou três tendências em relação à organicidade de linhas de pesquisa, nos programas avaliados:

*«1) aqueles que apresentaram linhas de pesquisa em uma abordagem que as aproximou do conceito tradicional (tendência menos evidente dentre as três); 2) programas que indicaram linhas muito abrangentes, aparentemente criadas de maneira indutiva para agregar o conjunto de projetos existentes; e, finalmente, 3) programas que mencionaram linhas abrangentes que, em verdade, estavam mais próximas do conceito de área de concentração, visto que envolviam não apenas projetos de pesquisa, mas também disciplinas e outras atividades específicas. A opção por uma grande abrangência facilita a inclusão de projetos de pesquisa das mais variadas temáticas, mas tende a obscurecer a organicidade do programa».* (Brasil, CA-Ed, 2007:25)<sup>4</sup>.

Vale destacar que, no triênio anterior, o de 2004, a CA-Ed alerta, especificamente no relatório de avaliação de um dos Programas, o P1, para a necessidade de manter uma coerência entre as linhas e projetos, afinal:

*«[...] problemas destacados na avaliação anterior ainda persistem e, em alguns casos, se agravaram. [...] Ainda fica nebuloso o eixo conceitual que articula as 10 linhas de pesquisa do Programa que, no seu conjunto, abordam grande parte da área da Educação. Da mesma forma, não se percebe com clareza a lógica que orientou a distribuição de alguns dos projetos entre as mesmas».* (Brasil, CA-Ed, P1, 2004:1, grifo nosso).

Essa questão parece se constituir em ponto de embate entre o Programa e a Comissão de Área. Nos triênios subsequentes, o Programa explicita seu posicionamento em relação a organicidade:

*«A aparente diversidade característica das linhas de pesquisa, que refletem múltiplas alternativas disciplinares e epistemológicas, ganha maior organicidade ao se compreender que, as mesmas, abrigam uma concepção da educação como fenômeno multireferencial, que exige estratégias de pesquisas advindas de áreas diferentes. Para além disso, o fato é que as disciplinas oferecidas se ancoram, também elas, em perspectivas teórico-metodológicas e nos estudos e pesquisas desenvolvidos pelos professores».* (Brasil, P1, 2009:3, grifo nosso).

Se para a Comissão de avaliação a organicidade do Programa avaliado parecia nebulosa, para P1, esta, se ancorava na concepção da *multireferencialidade* manifesta nas *perspectivas teórico-metodológicas e nos estudos e pesquisas desenvolvidos pelos professores*. Nesse embate, observa-se um Programa que constrói argumentos para responder, a seu modo, aos critérios da política criando um conceito capaz de manter como orgânica a *aparente* diversidade de suas linhas de pesquisa.

Se, do ponto de vista da política esse critério busca assegurar a unidade, no contexto da prática, no interior dos Programas, ele assume outras dimensões, por que diretamente vinculado às experiências e trajetórias acadêmicas do corpo docente e do Programa.

Para reafirmar sua concepção em torno da organicidade como *«multireferencialidade»*, P1 descreve as estratégias adotadas no interior do Programa: os Encontros de Pesquisa das Linhas e a nucleação dos

<sup>4</sup> Daqui para frente, serão apresentados os recortes da empiria em itálico e as referências destacando: fonte do documento (Brasil), quem produziu o documento: CA-Ed, o Programa que está sendo avaliado P1, P2, P3 ou P4, o ano de produção do documento e a página. Quando se tratar de documento produzido pelo Programa a referência será: Fonte (Brasil), Programa (P1, P2, P3 ou P4), ano e página.

professores por grupos de pesquisa:

*«Tais encontros agruparam tanto os alunos e professores a elas filiados, como outros interessados. [...]. A realização dos Encontros oportuniza, assim, significativa mediação concreta entre as disciplinas e o desenvolvimento das pesquisas, pois permitem uma saudável circulação e intercâmbio de questões, opções temático-metodológicas e soluções práticas na condução dos trabalhos dos alunos. Isso também ocorre em atividades específicas de muitos dos 42 grupos de pesquisa cadastrados no diretório de grupos do CNPq (grupos coordenados por um mesmo docente estão sendo unificados, o que poderá representar uma redução desse número nos próximos anos). Ao mesmo tempo, está se desenvolvendo um processo de nucleação de todos os professores em grupos de pesquisa».* (Brasil, P1, 2009:3).

Nas estratégias adotadas, evidencia-se um modo particular de responder às exigências da política e uma forma peculiar de traduzir suas diretrizes. Um misto de capitulação e resistência, como sugere Ball (2010). Significa dizer que o Programa resiste mantendo sua concepção ampliada de organicidade, mas capitula ao intensificar as formas de controle dos grupos existentes com a estruturação colegiada de discussão, o acompanhamento e a autoregulação do funcionamento das Linhas de Pesquisa e disciplinas viabilizadas, por exemplo, pelos Encontros de Pesquisa.

Ball define isso como um jogo particular de regulação ou governamentalidade, as quais têm, simultaneamente, uma dimensão social e interpessoal:

*«Elas são subentendidas dentro de um complexo institucional, de equipe, de grupo e de relações comunitárias - a comunidade acadêmica, a escola, os departamentos, a universidade, por exemplo. Nós nos embasamos na avaliação dos pares, nós redigimos as prestações de conta, nós atribuímos graus aos outros departamentos, nós repreendemos nossos colegas por sua pobre produtividade, nós planejamos, pomos a funcionar e alimentamos os procedimentos departamentais e institucionais para monitorar e melhorar os resultados».* (Ball, 2010:41).

Trata-se, pois, de formas de resposta/resistência e acomodação à performatividade.

O P4 também procura definir um conceito para expressar sua perspectiva de organicidade. Trata-se do conceito de *campo de confluência*:

*«[...] o Programa passou a organizar suas atividades de pesquisa e de docência em Campos de Confluência. A adoção dos Campos de Confluência decorreu de um amplo debate interno e também das discussões nacionais que marcaram a década de 1980, no campo da pós-graduação em educação. Sua concepção entende a educação como um processo e uma prática constituída e constituinte das relações sociais. No plano epistemológico e teórico, a organização do Programa em Campos de Confluência tem como objeto a construção de conhecimentos, buscando orientar a ação política, as práticas educativas institucionalizadas em formas específicas de ação educativa e as práticas que se forjam nos movimentos, lutas sociais e outros espaços-tempos educativos na cidade e no campo».* (Brasil, P4, 2009:3, grifo nosso).

Essa concepção contribuiu para a avaliação do quesito organicidade pela Comissão que afirma ter sido possível, a partir dessa estratégia adotada pelo Programa, construir uma representação discursiva da sua proposta.

*«O Programa organiza-se em seis Campos de Confluência, desdobrados em Linhas de Pesquisa. As linhas de pesquisa e os projetos em andamento são consistentes, abrangentes e atualizados. Há*

*relação dos projetos com as linhas de pesquisa em que se inserem e obteve conceito Muito Bom. A descrição da estrutura curricular foi detalhada, permitindo observar a organicidade entre as linhas, os projetos de pesquisa, a estrutura curricular e as temáticas das dissertações. [...]. Considerando-se tais análises, o Programa foi conceito Muito Bom quanto à coerência, consistência e abrangência da estrutura curricular neste triênio». (Brasil, CA-Ed, P4, 2010:1, grifo nosso).*

O P2, mesmo tendo sido avaliado com o conceito muito bom neste quesito nos triênios avaliados, descreve no seu relatório de 2009, que iniciou um intenso movimento de transformação de suas estruturas e formas de organização ao levar em conta as exigências internas e externas de avaliação:

*«O Programa [...], neste triênio, passou por um complexo processo de modificações em sua estrutura e organização [...]. Isso ocorreu, em parte, devido ao conjunto de exigências institucionais advindas tanto da nossa Universidade quanto das alterações processadas no âmbito do Sistema Nacional de Pós-Graduação e estabelecidas pela CAPES. Por outro lado, devido à própria necessidade de atualização, renovação e mudança, fruto dos sucessivos processos de auto-avaliação que temos levado a efeito. Nesse sentido, trata-se de mudanças que realizamos no currículo, nas Linhas de Pesquisa, no quadro docente, nos Grupos de Pesquisa e em algumas sistemáticas e normativas de funcionamento do Programa. Vale considerar que essas modificações vêm sendo implementadas progressivamente a partir dos últimos anos, conferindo a elas mais ênfase nos anos de 2008 e 2009. [...]. Ao longo deste Relatório, portanto, serão apresentadas e descritas diferentes iniciativas de modificação, buscando-se assegurar uma visão articulada da organicidade do Programa, tanto em seu âmbito mais particular quanto no âmbito institucional local, regional, nacional e internacional». (Brasil, P2, 2009: 2, grifo nosso).*

O uso de táticas discursivas e de estratégias diferenciadas na gestão são algumas das formas de tradução da política pelos Programas que são exigidos a construir uma variedade de descrições textuais de si mesmos. Ball (2010) considera que em dois sentidos, o simbolismo é tão importante quanto a substância. Primeiro, eles simbolizam e representam o consenso incorporado da instituição, evidenciando atividades em torno de um conjunto de prioridades acordadas. Segundo, eles demonstram «o padrão do esforço compartilhado que desloca ou subsume diferenças, desacordos e divergências de valor». (Ball, 2010:48).

No caso de P2, simbolismo e substância se encontram na demonstração do esforço da instituição, em definir *diferentes iniciativas de modificação, buscando-se assegurar uma visão articulada da organicidade do Programa*.

As iniciativas adotadas repercutem positivamente na comissão que, por sua vez, reconhece o trabalho do Programa:

*«A relação dos projetos com as linhas de pesquisa em que se inserem é conceito Muito Bom. A descrição da estrutura curricular foi detalhada, permitindo observar organicidade entre as linhas, os projetos de pesquisa, a estrutura curricular e as temáticas das dissertações e teses. O conjunto das disciplinas e suas respectivas bibliografias é atual e está em consonância com o corpo docente. Considerando-se tais análises, o Programa foi Muito Bom quanto à coerência, consistência e abrangência da estrutura curricular neste triênio». (Brasil, CA-Ed, P2, 2010:1).*

Nesse discurso, uma demonstração do modo como a política pode tornar os Programas mais responsivos. E nesse sentido, os textos escritos têm especial papel por concederem espaço à performatividade e possibilitarem o uso do significado correto, ou do significado esperado pela política.

Nesta mesma perspectiva, o P3 anuncia e reafirma como parte de sua identidade a busca da coerência

na configuração de sua proposta. O documento do Programa destaca que este se consolidou, a partir de dois fatores que contribuíram para tal: a especificidade do Programa e os resultados do sistema de avaliação da CAPES.

O Programa considera que a exigência da Capes em torno da delimitação das áreas de concentração e da organicidade com as linhas de pesquisa contribuiu para a valorização da sua especificidade:

*«Se antes todos os programas eram genericamente denominados como sendo de “educação” na atualidade os programas tendem a ser mais delimitados e, conseqüentemente cresce o número de programas específicos. Neste contexto a especificidade, que já é tradicional do programa desde sua criação se torna um ponto a favor para sua identidade e organicidade, e aliado ao fato dele permanecer sendo o único programa específico [...], 32 anos após sua criação, potencializa sua contribuição para a pós-graduação na área de Educação em nosso país».* (Brasil, P3, 2009:3).

Na avaliação da comissão essa concepção e estrutura adotada pelo Programa é coerente e contribuiu para o alcance do conceito Muito Bom:

*«A descrição da estrutura curricular foi detalhada, permitindo observar organicidade entre as linhas, os projetos de pesquisa, a estrutura curricular e as temáticas das dissertações e teses. O conjunto das disciplinas e suas respectivas bibliografias é atual e está em consonância com o corpo docente. Considerando-se tais análises, o Programa obteve conceito Muito Bom quanto à coerência, consistência e abrangência da estrutura curricular neste triênio».* (Brasil, CA-Ed, P3, 2010:1).

«Multireferencialidade educacional», «campo de convergência», «visão articulada», «especificidade/identidade» são importantes táticas discursivas utilizadas para demonstrar a conexão com a política, mas também para expressar um modo específico de pensar, de agir e de criar conceitos e simbolismos por parte dos Programas. Um paradoxo, um misto de adequação e resistência. «Elas significam, encapsulando ou representando um valor, a qualidade ou a valia de um indivíduo ou de uma organização dentro de um campo de julgamento» (Ball, 2010:38). Elas atuam como medidas de produtividade e de resultados, como formas de apresentação da qualidade, mas também revelam construções próprias, resoluções traduzidas, reconstruídas em diferentes, mas similares caminhos.

Está se tratando aqui do que Ball, Maguire e Braun (2012) chamam de «enactments». Ou seja, está se falando das respostas em curso à política, às vezes durável, às vezes frágil, às vezes resistentes e sempre configuradas nas diversas redes e cadeias de relações.

## 5. Considerações finais

O percurso e o exercício de análise empreendido nesse estudo permitiram discutir importantes indicadores, princípios e tendências que delineiam a política de avaliação da pós-graduação e sua tradução nos Programas de Pós-Graduação em Educação, que na avaliação do triênio 2010 obtiveram avanços nos seus conceitos, isto é, emergiram da nota 5 para a nota 6, alcançando nesse triênio a avaliação de desempenho equivalente ao dos centros internacionais de excelência na área.

A avaliação da pós-graduação brasileira tem sido motivo de debate, o qual ganhou força, no final da década de 1990 quando se estabeleceram critérios e padrões de avaliação em caráter continuado e quando se definiu a pesquisa como atividade central e precípua da pós-graduação brasileira. O momento representou uma mudança de paradigma especialmente porque estabeleceu índices comparativos e níveis

de qualidade traduzidos em conceitos/notas, que implicaram em alterações não somente nos objetivos e processos educativos, como também interferiram fortemente na reengenharia de valores (Ball; Cribb, 2005) das instituições e programas de pós-graduação, com ênfase nos resultados, na definição de novas obrigações e na valorização de iniciativas empreendedoras de alunos, professores e gestores.

A pesquisa se inseriu nesse contexto de mudanças de paradigma, com o propósito de analisar e compreender o complexo processo de interpretação e tradução das políticas, levando em conta as variações de contextos, de recursos e de participação e leitura ativa dos atores do processo.

A análise de dados me permitiu observar que a política de avaliação, mesmo que altamente delimitada e configurada por rígidos critérios e padrões internacionais não é simplesmente implementada –linear e acriticamente– no contexto da prática. Nos limites consentidos, nos variados contextos materiais e relacionais envolvidos, os atores produzem uma leitura ativa dos textos da política, que se transforma em objeto de recontextualização, interpretação e tradução.

Do ponto de vista geral pode-se afirmar que os discursos adotados e as ações efetivadas no interior dos Programas evidenciam tanto um movimento em busca da organicidade e da reafirmação de suas identidades, como de fabricação de discursos para responder à política.

No sentido amplo de pesquisa, as dimensões analisadas –nos Documentos das Comissões de Área, e nos documentos da Proposta dos Programas– revelam respostas e experiências ativamente vivenciadas no contexto da prática, redefinidas por táticas discursivas e traduzidas no interior dos Programas.

Neste sentido, ganhou expressão no estudo a categoria de organicidade (coerência e coesão), dada sua regularidade, relevância e ênfase no teor dos documentos e da política como um todo. Observou-se que essa categoria foi central no processo de avaliação e, particularmente, demonstrou o modo como os atores da política foram moldando seu significado, em um verdadeiro embate de resistência e capitulação aos critérios e diretrizes da política.

Táticas discursivas foram utilizadas para demonstrar a conexão com a política, mas também para expressar um modo específico de pensar, de agir e de criar conceitos e simbolismos por parte dos Programas. Um paradoxo, um misto de adequação e resistência. Essas táticas atuam como medidas de produtividade e de resultados, como formas de apresentação da qualidade, mas também revelam construções próprias, resoluções traduzidas, reconstruídas em diferentes, mas similares caminhos.

## Referências bibliográficas

- Azevedo, J. M. L.; Aguiar, M. A. (1999): “Políticas de educação: concepções e programas” in L. C. Wittmann; R. V. Gracindo (coords.): *O estado da arte em política e gestão da educação no Brasil: 1991 a 1997*. Brasília: INEP.
- Ball, Stephen J. (2003): “The Teacher’s Soul and the Terrors of Performativity”. *Journal of Education Policy*, 18 (2), 215-228.
- Ball, Stephen J. (2006): “Performativities and Fabrications in the Education Economy: Towards the Performative Society” in: Lauder H. *et al.*, (orgs): *Education, Globalization and Social Change*. Oxford: Oxford University Press.
- Ball, Stephen J. (2010): “Performatividades e fabricações na economia educacional: rumo a uma sociedade performativa”. *Educação & Realidade*, 35 (2), 37-55.



- Ball, Stephen J. (2011): “Sociologias das políticas educacionais e pesquisa crítico-social: uma revisão pessoal das políticas educacionais e da pesquisa em política educacional” in Stephen J. Ball; J. Mainardes. *Políticas Educacionais: questões e dilemas*. São Paulo: Cortez.
- Ball Stephen J.; Cribb, A. (2005): “Towards an ethical audit of the privatisation of education”. *British Journal of Educational Studies*, 53, (2), 115-28.
- Ball, Stephen J.; Maguire, Meg; Braun, Annette (2012). *How schools do policy: policy enactments in secondary schools*. London: Routledge.
- Brasil. Capes (2002). *50 anos*. Brasília: CAPES/CPDOC/FGV.
- Brasil. Capes (2002a). *Infocapes: Boletim Informativo da CAPES*. v. 10, N. 4, Brasília, CAPES/MEC, out - dez.
- Braun, Annette; Ball, Stephen J.; Maguire, Meg (2011): “Policy enactments in schools introduction: towards a toolbox for theory and research”. *Discourse: Studies in the Cultural Politics of Education*, 32 (4), 581-583.
- Mainardes, Jefferson (2006): “Abordagem do ciclo de políticas: uma contribuição para a análise de políticas educacionais”. *Educ. Soc.*, 27, (94), 47-69.
- Ozga, Jenny; B. Lingard. (2007): “Globalisation, Education Policy and Politics.” in B. Lingard; Jenny Ozga. (eds.): *The-Routledge Falmer Reader in Education Policy and Politics*. London: Routledge Falmer.
- Robertson, S. L.; Bonal, X; Dale, R. (2006): “GATS and the Education Service Industry: The Politics of Scale and Global Reterritorialization” in H. Lauder *et al.*, (eds.): *Education, Globalization and Social Change*. Oxford: Oxford University Press.
- Shaw, Martin (1997): “The State of Globalization: towards a theory of state transformation”. *Review of International Political Economy*, 4 (3), 497-513.

## Nota biográfica

**Regina Célia Linhares Hostins** es Doctora en Ciencias de la Educación por la Universidade Federal de Santa Catarina (2006), Brasil. Realizó un posdoctorado en 2013, en el Institute of Education, University of London, como colaboradora del Dr. Stephen Ball, en el campo de la Sociología de la Educación, desarrollando estudios sobre aspectos teórico-metodológicos en la investigación de políticas educativas. Es profesora e investigadora del Programa de Pós-Graduação em Educação en la Universidade do Vale do Itajaí y coordina el grupo de investigación «Observatório de Políticas Educacionais». Participa en la red de investigadores del Observatório Nacional de Educação Especial, un consorcio compuesto por 22 universidades brasileñas.

# TESTED: Una reflexión sobre medir en educación

## TESTED: A reflection on measuring in education

Carmen Carmona Rodríguez<sup>1</sup>

### Resumen

Hace décadas el debate educativo estaba centrado, principalmente, en la cobertura y el acceso a la educación básica para todos los niños y niñas, aspecto que en algunos países todavía es un reto a conseguir. Sin embargo, actualmente se presta mayor atención a los resultados de aprendizaje alcanzados por los estudiantes. Resultados que en algunos países provienen de pruebas estandarizadas que están diseñadas para seleccionar y no para incluir. Por ello, es necesario considerar la influencia del contexto como factor de segregación en los centros educativos, hecho que puede de forma indirecta fomentar la desigualdad de oportunidades en el alumnado. El objetivo del presente artículo es reflexionar sobre los factores y consecuencias que tiene la realización de evaluaciones al alumnado, especialmente cuando están estandarizadas. Se analizan aspectos a considerar cuando se interpretan los resultados de pruebas estandarizadas, así como la relevancia de las variables de contexto. En particular, se profundizará a través del documental *TESTED* (2015) en el proceso de preparación, consecución y consecuencias del polémico *Specialized High School Admission Test* (SHSAT) que puede marcar el futuro de los estudiantes de entre 12-13 años de la ciudad de Nueva York. El análisis del documental con el propio director visibiliza una realidad de desigualdad a partir de los resultados de un test, algo que también podría llegar a ocurrir en otros países sino se tiene en cuenta que la educación es un proceso permanente que facilita el aprendizaje, el desarrollo de competencias, la experiencia y la incorporación plena de los valores tanto individuales como sociales.

### Palabras clave

Evaluación, desigualdad, contexto socio-económico, SHSAT.

### Abstract

For decades, the educational debate was mainly focused on the coverage and access to basic education for all children, which in some countries is still a challenge to be achieved. However, nowadays more attention is dedicated to the learning outcomes achieved by students. These results in some countries come from standardized tests that are designed to select and not to include. Therefore, it is necessary to consider the influence of context as a segregation factor in schools, a fact that may indirectly promote inequality of opportunities for students. The aim of this article is to reflect on the factors and consequences of the evaluation of students, especially when they are standardized. We analyze aspects that must be considered when interpreting the results of standardized tests, as well as the relevance of the context variables. In particular, this will be deepened through the documentary *TESTED* (2015) in the process of preparation, achievement and consequences of the controversial *Specialized High School Admission Test* (SHSAT) that can determine the future of students between 12-13 years in New York City. The analysis of the documentary with the director shows a reality of inequality from the results of the test, fact that could also happen in other countries whether we forget education is a permanent process that facilitates learning, development of competencies, experience and full incorporation of individual and social values.

### Key words

Evaluation, inequality, socioeconomic context, SHSAT.

Recibido: 28-03-2017  
Aceptado: 02-05-2017

<sup>1</sup> Universitat de València, [carmen.carmona@uv.es](mailto:carmen.carmona@uv.es)

## 1. Introducción

Medir la educación ha sido un tema clave que ha suscitado diversas opiniones y críticas al respecto (Popkewitz, 2013; Rindermann, 2007). Sin embargo, es primordial que sea un tema a debate dentro y fuera de la comunidad educativa, entre agentes implicados, y sociedad en general.

Entre las acciones que se han puesto en marcha para medir la educación cabe mencionar los programas orientados a analizar el rendimiento del alumnado en los sistemas educativos, recurriendo a diversos instrumentos, enfoques y en base a diversos modelos de sistemas de indicadores, que establecen una perspectiva completa acerca del estado, la situación y la evolución de la educación a escala nacional o internacional (Tiana, 2012).

En España, actualmente el sistema educativo a través de la Ley Orgánica 8/2013, de 9 de diciembre, para la Mejora de la Calidad Educativa (LOMCE) ha establecido para diferentes niveles que se realicen evaluaciones individualizadas al alumnado para medir competencias específicas relacionadas con la comunicación, matemáticas y ciencias.

A partir de las implicaciones que ha supuesto la LOMCE, el presente artículo no trata de reflexionar sobre la necesidad o no de realizar estas evaluaciones, y sus correspondientes mediciones, sino que profundiza más allá intentando analizar qué aspectos se evalúan en general, qué factores están relacionados con los resultados del alumnado en las evaluaciones, y todo ello bajo el análisis y prisma del documental *TESTED* (2015), en el que se realiza un acercamiento más personal y real sobre los tests comparando el proceso de preparación del *Specialized High School Admission Test* (SHSAT) en la ciudad de Nueva York, y el significado que tiene esta prueba para un grupo de adolescentes de entre 13 y 14 años. A través de la reflexión con el propio director del documental, este artículo intenta poner de manifiesto los factores a tener en cuenta en la evaluación y medición de la educación, los posibles factores asociados a los resultados obtenidos y las posibles consecuencias que deberían de tenerse en cuenta para desarrollar políticas educativas que no lleven a una segregación en los centros.

## 2. La evaluación del alumnado

La evaluación del alumnado y en concreto la medición del rendimiento a través de pruebas, que no es lo mismo que la medición del aprendizaje, o evaluación de las competencias del alumnado ha sido, es y será un aspecto relevante que preocupa a las familias, profesorado, y agentes educativos implicados (Carmona *et al.*, 2014; Jornet, 2016).

A nivel internacional, España ha participado en los estudios previos de PISA (Programa para la Evaluación Internacional de Estudiantes), de PIAAC (Programa Internacional para la Evaluación de las competencias de la población adulta) y TALIS (Estudio internacional de la enseñanza y el aprendizaje), de la Organización para la Cooperación y el Desarrollo Económico (OCDE), así como en los estudios TIMSS (Estudio Internacional de Tendencias en Matemáticas y Ciencias) y PIRLS (Estudio Internacional de Progreso en Comprensión Lectora), de la Asociación Internacional para la Evaluación del Rendimiento Educativo (IEA, por sus siglas en inglés), y en el Estudio Europeo de Comunicación Lingüística (EECL), de la Unión Europea.

En términos generales, la evaluación de la educación es un proceso primordial ya que la Educación como pilar fundamental de toda sociedad, es la que marca el rumbo del progreso, igualdad, y equidad desarrollando de manera integral tanto a niños como a jóvenes. Sin embargo, la forma en la que se mide la

educación y se interpretan los resultados, son aspectos clave que inciden en la dirección de las decisiones que se toman en materia de política educativa a nivel nacional e internacional.

En la actualidad, es un hecho que España se encuentra entre los países de la Unión Europea con uno de los índices más elevado de fracaso escolar. En torno al 20% de los jóvenes de entre 18 a 24 años no ha completado ni siquiera el nivel de Educación Secundaria Obligatoria y no sigue ningún tipo de educación-formación (Ministerio de Educación, Cultura y Deporte-MECD, 2016). Este hecho ha fomentado el interés en profundizar en el análisis de los informes nacionales e internacionales que evalúan las competencias del alumnado en primaria y secundaria.

En ese sentido, la OCDE (Organización para la Cooperación y el Desarrollo Económico) organismo que proporciona infraestructura, materiales y especialistas para realización del estudio PISA, declara como objetivo la evaluación para comprobar el nivel de competencias adquiridas por el alumnado de 15 años (al finalizar la educación obligatoria) para su desarrollo en la sociedad (OCDE, 1999). Los resultados de PISA en principio son útiles para realizar un análisis comparativo del nivel de ciertas competencias específicas en más de 80 países, es decir es una herramienta con la que «se pretende colaborar para construir mediciones globales avanzadas para explorar los conocimientos, habilidades y atributos característicos que son relevantes para el éxito del estudiante y que son esenciales para la plena participación en las sociedades modernas» (Schleicher, 2016:3).

Respecto a las dos últimos informes de PISA, se observa que en el del 2012, los resultados indicaban que España se situaba significativamente por debajo del promedio de la OCDE en lectura, matemáticas y ciencias (OCDE, 2014). Sin embargo, los resultados generales son más favorables en el último informe PISA (OCDE, 2016), en el que España consigue una puntuación media en «ciencias» igual a la del promedio de la OCDE, y solo 2 puntos por debajo del total de la Unión Europea (UE), diferencia no significativa (PISA, 2015). En «lectura», España consigue una puntuación media en lectura de tres puntos por encima del promedio de la OCDE y dos por encima del total de la UE, y en «matemáticas», se consigue una puntuación media en matemáticas 4 puntos menos que el promedio de la OCDE y 7 puntos por debajo del total de la UE.

Una vez conocidos los resultados de las dos últimas oleadas de PISA (OCDE, 2012; 2015), el planteamiento que se debería de atender es el de: ¿y para qué sirve conocer esas puntuaciones? ¿cómo se utiliza esa información para la mejora educativa? ¿qué interpretaciones se pueden extraer con esos datos? ¿tan positivo es situarnos en ciencias al nivel del promedio de los países de la OCDE y la UE?

A pesar de todas las evaluaciones que el alumnado tiene que realizar (autonómicas, nacionales, e internacionales), y por ende los centros y profesorado apoyar, en 2013 se reforma la Ley Orgánica 2/2006, de 3 de mayo, de Educación (LOE), y se modifica por la Ley Orgánica 8/2013, de 9 de diciembre en varios aspectos, y uno de ellos es en materia para la evaluación. En cuando a Educación Primaria, actualmente, la LOMCE establece que se realice una evaluación individualizada al finalizar 3º y 6º curso de Educación Primaria. En concreto, el texto de la LOMCE en el artículo 21 queda redactado de la siguiente manera:

*«Artículo 21. Evaluación final de Educación Primaria. 1. Al finalizar el sexto curso de Educación Primaria, se realizará una evaluación individualizada a todos los alumnos y alumnas, en la que se comprobará el grado de adquisición de la competencia en comunicación lingüística, de la competencia matemática y de las competencias básicas en ciencia y tecnología, así como el logro de los objetivos de la etapa. 2. El Gobierno, previa consulta a las Comunidades Autónomas, establecerá los criterios*

*de evaluación y las características generales de las pruebas para todo el Sistema Educativo Español con el fin de asegurar unos criterios y características de evaluación comunes a todo el territorio. 3. El resultado de la evaluación se expresará en niveles. El nivel obtenido por cada alumno o alumna se hará constar en un informe, que será entregado a los padres, madres o tutores legales y que tendrá carácter informativo y orientador para los centros en los que los alumnos y alumnas hayan cursado sexto curso de Educación Primaria y para aquellos en los que cursen el siguiente curso escolar, así como para los equipos docentes, los padres, madres o tutores legales y los alumnos y alumnas. Las Administraciones educativas podrán establecer planes específicos de mejora en aquellos centros públicos cuyos resultados sean inferiores a los valores que, a tal objeto, hayan establecido».*

Desde hace décadas, España y sus comunidades autónomas han participado en diferentes evaluaciones nacionales, aspecto que ha hecho que las nuevas directrices de la LOMCE en cuanto a la referencia de la evaluación individualizada, no resulte ser una novedad, sino una evaluación más a añadir a las que ya se están realizando en los centros de primaria y secundaria, un ranking para los centros y alumnado, pero de los cuales todavía se desconocen las posibles repercusiones a corto y largo plazo.

### 3. Las evaluaciones y sus limitaciones

En cuanto a las evaluaciones, PISA a nivel internacional tiene una importante repercusión en temas de políticas educativas, pero no hay que olvidar que únicamente se centra en tres competencias consideradas troncales: ciencias, lectura y matemáticas. Esencialmente, se valora cómo el alumnado puede extrapolar su conocimiento, sus destrezas cognitivas y sus actitudes a contextos en principio extraños al propio alumno, pero con los que se tendrá que enfrentar a diario en su propia vida. Además en cada ciclo, a los tres dominios troncales se añadió un cuarto innovador, como fue la resolución creativa de problemas en 2012, la resolución colaborativa de problemas en 2015 o la competencia global, ya en preparación para el próximo ciclo de PISA 2018 (PISA, 2015). Es decir, es necesario evaluar otras competencias que no hagan referencia sólo a contenidos, sino también a aspectos sociales y de relación. ¿De qué sirve obtener puntuaciones altas en esos tests si luego ese alumnado no respeta a sus compañeros y compañeras de clase o no sabe trabajar en equipo?

Los resultados del informe PISA a lo largo de los años ha recibido diversas críticas en relación a sus limitaciones. Entre ellas, la escasa fiabilidad de los indicadores asociados con la organización y el clima escolar (basados en las opiniones de los directores de los centros), la ausencia de datos relativos al gasto de los centros y el hecho de que alguna información en relación al contexto solo esté disponible para algunas oleadas del estudio (Cordero *et al.*, 2013; González-Such *et al.*, 2016). Pero principalmente porque para comparar realmente sistemas educativos sería necesario diferenciar qué aportan las puntuaciones PISA en cada país y qué aporta la escuela, así como en qué contribuyen otros factores, tanto naturales como sociales (Carabaña, 2015a) Además, la generalización e interpretación de los resultados de los informes PISA a lo largo de los años, ha impactado a través de los medios de comunicación en la sociedad, en particular de la prensa escrita de carácter informativo-descriptivo, incidiendo en la opinión pública y elaborando un uso inadecuado de los resultados de dichos informes, sin tener en consideración la limitaciones del propio estudio (González-Mayorga *et al.*, 2017).

Al igual que ocurre con PISA, «en las evaluaciones diagnósticas establecidas en la legislación española se enfatiza un enfoque sistémico, si bien en la práctica lo que suele aportarse como evaluación son únicamente medidas de logro acerca del alumnado. Por desgracia, este mismo hecho se observa en la mayor parte de evaluaciones nacionales e internacionales» (Jornet, 2016:2).

La evaluación del sistema educativo por parte de las Administraciones educativas (nacional y autonómica) permite acceder a información en base a indicadores que, por un lado, pretenden ofrecer una visión realista de la situación del proceso de enseñanza-aprendizaje y, por otro, ponen de manifiesto los logros y las carencias del sistema. Esta visión realista ayudará a reorientar las tendencias educativas reforzando las buenas prácticas y proponiendo medidas de mejora. Por tanto, es relevante conocer qué se evalúa, cómo y qué resultados generales se obtienen. Sin embargo, la relevancia de la evaluación debe incidir en averiguar no solo qué contenidos relacionados con competencias específicas el alumnado obtiene, sino que es necesario también evaluar otro tipo de competencias necesarias para el desarrollo integral de la persona, que vive en una sociedad en la que se relaciona con otras personas y con las que es necesario colaborar, compartir y reflexionar si se pretenden realizar cambios profundos en la sociedad actual. Por tanto, las escuelas necesitan dar un empuje hacia el cambio de la «acumulación de conocimiento» y guiar al alumnado hacia un mundo en el que la colaboración y la comunicación es vital para comprender todo lo que sucede a nuestro alrededor (Schleicher, 2015)

Para que la evaluación permita comparaciones y visiones de conjunto debe ser equitativa y atender a la diversidad. Por ello los contextos socioculturales del alumnado y centros, datos que se recogen a través de los cuestionarios de contexto, se consideran imprescindibles para poder explicar debidamente los resultados. En definitiva, las evaluaciones «pretenden» ser formativas para el conjunto del sistema educativo, contribuyendo a su conocimiento, transformación y mejora y a facilitar orientaciones para las políticas educativas.

#### 4. Variables de contexto: aspectos clave para la evaluación y su interpretación

Desde el estudio de Coleman *et al.*, (1966) en el que se le da valor explicativo a la influencia del nivel socioeconómico y cultural de las familias en el rendimiento del alumnado, muchos otros estudios han corroborado esta relación.

En los últimos 30 años, la investigación ha confirmado una asociación entre los indicadores socioeconómicos y el rendimiento académico de los niños como un subproducto del contexto en el que viven (por ejemplo, Bradley y Corwyn, 2002; Danziger, 1995; Duncan y Brooks-Gunn, 1997; Marchesi *et al.*, 2004; McCulloch y Joshi, 2001; Shonkoff y Phillips, 2000). En concreto, los resultados podrían derivarse de variables como el ingreso familiar, la ocupación de los padres y el entorno de la comunidad que influyen directamente en el progreso académico (Van Ewijk y Slegers, 2010). En general, los estudios constatan mejores rendimientos en estudiantes procedentes de entornos familiares y sociales más favorecidos (Gil, 2014). Sin embargo, las relaciones entre los indicadores del estatus socioeconómico de las familias y los resultados académicos proporcionan también una oportunidad para entender cómo las habilidades cognitivas de los niños en las escuelas están en parte influenciadas por sus condiciones de vida, ya que estas relaciones son una representación para estimar el grado de desigualdad y desventaja en nuestras comunidades (Maggi, *et al.*, 2004; Willms, 1999).

Las variables de contexto ayudan a entender la realidad del alumnado, situando la interpretación del rendimiento bajo el prisma del desempleo o la ocupación inestable, así como bajo las condiciones económicas inciertas en los padres jóvenes y solteros que pueden ser variables importantes que afectan al contexto socioeconómico de la familia y al rendimiento escolar del primer ciclo de primaria (D'Angiulli *et al.*, 2004).

En concreto, respecto a las evaluaciones internacionales, algunas de las principales conclusiones extraídas de trabajos que han realizado revisiones sobre PISA y sus resultados son: que los factores socioeconómicos son relevantes en la explicación del rendimiento en las pruebas que han medido las competencias específicas. En particular, las variables asociadas con el nivel educativo de los padres, así como el entorno socioeconómico y cultural en el centro son factores a considerar (Cordero *et al.*, 2013; Scandurra, 2013), del mismo modo hay que reflexionar sobre los resultados inferiores que obtiene el alumnado inmigrante respecto a los nacionales. Se ha observado que existe una relación entre el nivel socioeconómico y cultural de las familias y las influencias culturales sobre el lenguaje (Hoff y Tian, 2005). A nivel europeo, un estudio sobre PISA a lo largo de 12 años en varios países de la Unión Europea (Francia, Italia, Alemania, y España), indica que la pobreza educativa ha disminuido en términos generales a lo largo del periodo 2000-2012 para todas las competencias evaluadas, excepto en el caso de España y Francia que se observa un cierto aumento de la pobreza educativa en algunas de las competencias que se evalúan, en particular matemáticas y ciencias, respectivamente (Díaz Caro *et al.*, 2016).

Los estudios que intentan explicar el rendimiento académico han ido aceptando un modelo complejo de interacción de factores (Tejedor, 2003). Entre los factores analizados, se ha demostrado que existen variables que explican un mayor porcentaje que otras como son la inteligencia, personalidad del sujeto, influencia ambiental, familiar, y social (Garcés, 2012; Mella y Ortiz, 1999; Sirin, 2005; Tejedor, 2003). Además, se constata que conforme se producen cambios en el contexto, éstos pueden afectar de forma directa al nivel educativo de los alumnos pues los estilos cognitivos evolucionan a la par que evoluciona el mundo (García, *et al.*, 2012). Asimismo, en un estudio reciente se demuestra que el nivel ocupacional y educativo de los padres, la obtención de ayuda familiar y la percepción que el alumno tiene sobre la valía que le atribuye su familia como estudiante, son variables fundamentales para analizar el desarrollo del rendimiento académico (Fajardo, *et al.*, 2017), factores socio-afectivos que también son relevantes y no siempre se han considerado relevantes para influenciar el rendimiento académico.

## 5. TESTED: una herramienta para la reflexión

La transición de la Educación primaria a la Educación Secundaria Obligatoria puede ser un momento estresante y crucial en la vida de los estudiantes. Es más, es un momento de cambio no solo de etapa sino también de cambios asociados al desarrollo fisiológico y madurativo propio de la edad de 12-13 años, inicio de la adolescencia.

Tal y como se propone en la LOMCE, en materia de evaluación, en 6º de primaria se llevaría a cabo la segunda evaluación individualizada donde se evaluaría la competencia comunicativa y la matemática. Hecho que coincide también con las pruebas que en Estados Unidos se realizan en el octavo grado (unos 12-13 años) para pasar de la «*middle school*» a la «*high school*». En Estados Unidos a diferencia de España, «de momento», estas pruebas son en general obligatorias para acceder a cualquier instituto. En general se realizan pruebas para medir la competencia en inglés y en matemáticas, además de otras evaluaciones que consisten en una entrevista, portafolio o en algunos centros de enseñanza artística, una audición. Estas pruebas son importantes para los adolescentes porque delimitan en qué instituto estudiarán, y como consecuencia asociada, a qué posible universidad podrían acceder en un futuro.

Además de estas pruebas que se realizan por cada estado, en la ciudad de Nueva York se da una peculiaridad y es la de que para entrar a uno de los ocho mejores institutos (escuelas de élite), no solo de la ciudad sino reconocidos a nivel estatal, es obligatorio realizar un test concreto para acceder. Es el polémico

co *Specialized High School Admission Test* (SHSAT), una prueba anual que consiste en evaluar la competencia en matemáticas e inglés.

Aunque todavía existe poca evidencia de que los estudiantes con mejor rendimiento, obtengan los mejores resultados al haber asistido a una escuela secundaria de élite en Nueva York (Abdulkadiroğlu *et al.*, 2014; Dobbie y Fryer, 2014), la alta demanda para entrar a estos ocho institutos ha disminuido incluso la oportunidad de intentar realizar el test. Se ha argumentado, por ejemplo, que el SHSAT premia la intensa preparación de los exámenes e inhibe la diversidad en las escuelas de élite, que son predominantemente escuelas con alumnado de origen asiático (64%) en las tres escuelas más grandes, más antiguas y con más prestigio, de raza blanca (22%) y mayor porcentaje de chicos que chicas (57%).

Estos ocho institutos de secundaria son considerados los más prestigiosos a nivel nacional. La aceptación en estos centros es del 20% de las solicitudes, y para los estudiantes afro-americanos y latinos, el porcentaje es incluso inferior, se situaría en un 5% y 7% respectivamente. En algunos centros este porcentaje de representación oscila entre un 2% y 3% (Corcoran y Baker-Smith, 2015). Este dato ha alarmado a diferentes organismos, tales como la OCR-Organización de Derechos Cívicos que indica que se están produciendo políticas discriminatorias en relación a la admisión en estos centros especializados de personas afroamericanas y latinas.

El proceso general de admisión a un instituto en la ciudad de Nueva York es el sistema más amplio de elección en el país. Los estudiantes no eligen en función de la zona donde viven sino que deben de solicitar por orden de preferencia de uno a doce centros. Esto hace que exista un número de estudiantes muy superior para las plazas existentes en algunos de los institutos públicos (Tullis, 2014). La forma de seleccionar a los estudiantes varía en función de los criterios que cada centro establece. Algunos centros seleccionan en función de las notas obtenidas, en otros centros es un sorteo como si fuera una lotería, y otros centros aceptan estudiantes que provienen de áreas específicas de la ciudad (Tortoriello, 2016). Dado que el número de estudiantes que se quedaban sin plaza en centros públicos era muy alto, en 2004 se realizó una reforma para reducir el número de estudiantes sin plaza y ese número de estudiantes se redujo de 31.000 a 3.000 en ese mismo año (Tullis, 2014). Sin embargo a pesar de la reducción en el número de estudiantes que se quedan sin una plaza en un centro público, es relevante destacar que el prestigio de los centros no es el mismo, es decir existen grandes diferencias entre un tipo de centro u otro, esto es importante porque si los centros reciben poca financiación, este hecho afecta de forma directa al alumnado, especialmente al que vive en zonas con menos recursos, en familias con pocos ingresos y al alumnado con un rendimiento bajo.

Por tanto, el reclamo que tienen las llamadas «Specialized High Schools» para asegurarse de que el alumnado tendrá más opciones en un futuro es una oportunidad para algunos jóvenes de la ciudad de Nueva York, y la competición es muy alta. La forma de acceso es un proceso selectivo y separado de las otras pruebas que tienen que realizar para el resto de institutos públicos. Existen nueve institutos públicos con la denominación de *Specialized High School*, ocho están destinadas al desarrollo de contenidos más académicos, y todas ellas utilizan el SHSAT como único criterio de admisión, excepto el noveno centro que se centra en un itinerario artístico, es el centro *Fiorello H. La Guardia High School of Music & Art* que tiene un sistema de admisión que conlleva la evaluación de portafolios y audiciones. Los ocho centros denominados como institutos de élite en la ciudad de Nueva York son:

- *Bronx High School of Science*
- *Brooklyn Latin School*



- *Brooklyn Technical High School*
- *High School for Math, Science and Engineering at City College*
- *High School for American Studies at Lehman College*
- *Queens High School for Sciences at York College*
- *Staten Island Technical High School*
- *Stuyvesant High School*

En la actualidad, se están realizando estudios que analizan el impacto de los SHSAT's, y que confirman que estas pruebas estandarizadas han jugado un papel negativo en la segregación de los niños por raza y clase en las escuelas de la ciudad de Nueva York (Knoester y Au, 2017). Este fenómeno conlleva el planteamiento de un objetivo claro de política educativa para hacer disminuir la segregación en los centros, con el propósito no solo de mejorar los resultados globales, sino de reducir las desigualdades existentes en los resultados.

Otro dato asociado con el nivel socioeconómico y los resultados académicos es que en Estados Unidos se han encontrado evidencias que indican que la repetición de curso está relacionada con el grupo étnico, el status social de los padres, y la pobreza de la zona (Hauser *et al.*, 2000), así como también con los ingresos de los padres y los hogares monoparentales (Hauser *et al.*, 2007). En España, otros estudios han encontrado que además «repiten casi únicamente los alumnos de conocimientos bajos o los hijos de madres con pocos estudios» (Carabaña, 2015b:25).

Estas afirmaciones provienen de diversos estudios, y confirman en términos generales lo que hace 50 años Coleman *et al.*, (1966) en su informe indicaron. Sin embargo, también es necesario que desde un acercamiento más personal, y desde una aproximación a la investigación dialógica, sea crucial escuchar las experiencias y narrativas sobre el impacto que los tests tienen en el alumnado de entre 12-13 años, qué sienten, experimentan y piensan sobre cómo un test puede o no condicionar su vida, más allá de los análisis estadísticos derivados de la asociación entre persona, contexto y rendimiento en las competencias específicas que evalúan unas pruebas.

Desde el documental TESTED (<http://www.testedfilm.com>) y a través de dos entrevistas con el propio director Curtis Chin (CH), una en la revista *Futura* (Carmona, 2016) con motivo de la presentación del documental por parte del director en la *Universitat de València* en septiembre de 2016, y otra realizada en marzo de 2017, se reflexiona y dialoga sobre el impacto de la «medición en educación». Curtis Chin es profesor visitante en la Universidad de Nueva York. Ha escrito programas de televisión para ABC, Disney Channel y Nickelodeon, así como proyectos para NBC y Fox. Ha recibido numerosos premios como *The National Endowment for the Arts*, *New York Foundation for the Arts*, y el *San Diego Asian American Film Foundation*. Fue co-fundador del *Asian American Writers Workshop* y *Asian Pacific Americans for Progress*. Chin ha aparecido en MSNBC, CNN, y NPR, y en la revista *Newsweek* y otros medios de comunicación. También es miembro activo del *Writers Guild of America*. El documental ha sido proyectado en más de 150 lugares y 8 países, está siendo un empuje para poner a debate el tema de la medición en educación a través de las reflexiones de los participantes del documental y de las que se generan a partir de él (<http://www.uah.edu/news/campus/uah-hosts-screening-of-acclaimed-documentary-tested-with-filmmaker-curtis-chin>).

El documental aborda la lucha de las familias neoyorquinas para que sus hijos accedan a una de las escuelas públicas de élite de la ciudad. En concreto se graba a una docena de estudiantes de octavo curso de diferentes razas, culturas y niveles socioeconómicos, que aspiran, mediante la superación de la prueba

SHSAT conseguir una plaza en uno de los mejores institutos. En el documental se observan las desigualdades extremas de una ciudad como Nueva York, donde los afroamericanos y los hispanos, que son el 70% de la población en edad escolar, representan menos del 5% a las escuelas secundarias públicas de élite, mientras que «blancos americanos» y asiáticos constituyen hasta el 73%. En el documental también aparecen informantes clave como los investigadores Pedro Noguera y Diane Ravitch. En la actualidad, 1 de cada 2 niños en las escuelas públicas estadounidenses (aproximadamente el 52%) provienen de hogares con bajos recursos (Southern Education Foundation, 2015). Por tanto, no es sorprendente que la creciente pobreza y la desigualdad económica en educación marque las oportunidades disponibles para los niños (Noguera, 2017).

Es relevante mencionar como Ravitch fue durante muchos años una de las fuertes defensoras en realizar pruebas al alumnado y en la rendición de cuentas en educación. Hace algunos años, sin embargo, reconsideró sus creencias y, en un libro influyente, *La muerte y la vida del gran sistema escolar americano*, se desmarcó de sus anteriores ideas y se unió a la oposición. Actualmente sostiene que la reforma en educación ha de combatir el esfuerzo deliberado para reemplazar a las escuelas públicas con un sistema de mercado. El enfoque antinatural en las pruebas ha producido resultados perversos pero predecibles. Ha reducido los currículos a los alumnos probables, con exclusión del arte y la cultura. Y por tanto, ha alentado la manipulación de los resultados en los exámenes estatales, donde enseñar para pasar el test, es considerado poco profesional y poco ético, y común en muchas de las escuelas públicas (Ravitch, 2011).

Algunas de las reflexiones de las que el propio director del documental partió estuvieron relacionadas con el hecho de la desigualdad racial y económica en Estados Unidos, con la explicación de por qué existen tantas diferencias en unos centros u otros de secundaria en la ciudad de Nueva York, y si realmente existiera igualdad, debería de verse reflejada en una mayor diversidad en los centros. En una primera entrevista al director fue interesante conocer qué es lo que le motivó a escribir y realizar el documental sobre esta temática (Carmona, 2016):

*«Había acabado uno sobre la igualdad racial en los Estados Unidos cuando leí un artículo en el New York Times sobre las escuelas neoyorquinas que me inspiró. También mi experiencia personal me había hecho ver que no todo el mundo había tenido como yo unas oportunidades tan maravillosas gracias a la educación pública de calidad que recibí, y me preguntaba por qué». (CH).*

Tras la revisión de los estudios centrados en las evaluaciones internacionales en los anteriores apartados, se ha observado que no se da realmente voz a las familias y adolescentes, actores principales del rendimiento, de las puntuaciones en los tests, del apoyo a sus hijos e hijas, y como parte relevante de esas denominadas «variables de contexto» que se miden:

*«Durante la filmación nos percatamos de cuánto les importaba participar. Creo que muchas veces, cuando vienes de una familia inmigrante o pobre, tienes la sensación de que a nadie le interesa tu historia y no se te pregunta la importancia que tiene para ti la enseñanza. Por ejemplo, todo el mundo que entrevistáramos quería que a sus hijos los fuera bien en la escuela, pero las familias pobres se enfrentaban a retos adicionales. Por ejemplo, estaba el caso de una madre que no dejaba ir su hijo a los talleres gratuitos de preparación de la prueba porque era demasiado peligroso volver en casa solo de noche. En los barrios de clase media no es un tema que les preocupe». (Carmona, 2016:24).*

Otro aspecto a tener en cuenta y que afecta en la decisión de realizar el test e incluso poder tener una buena puntuación es el contexto socioeconómico y familiar: «Lo que se critica es la preparación para el examen

*porque es injusto que algunas comunidades puedan tener acceso a una mayor preparación que otras o que hayan tenido acceso a centros de educación primaria con mejores condiciones.» (CH, 2017). De hecho, en Estados Unidos, los resultados de los tests se utilizan para diversas acciones en educación, se realizan para medir el desempeño del alumnado, y también para evaluar la «calidad» de los centros para tomar decisiones en cuanto a financiación, contratación y remuneración del profesorado o incluso decisiones sobre cerrar o no un centro:*

*«Si vives en una comunidad multinacional donde existen escuelas públicas bien condicionadas de recursos y profesorado, no hay tanta preocupación por la educación de tu hijo, pero en otro tipo de comunidades es un problema [...]. El sistema público americano en realidad se organiza distrito a distrito, donde uno de los retos más allá de la escuela es el desempleo, recursos económicos bajos y la delincuencia». (CH).*

Es decir, los estudiantes de clase media pueden esquivar las dificultades del sistema, pero no los que proceden de familias con pocos recursos y viven en ciertas comunidades. Por tanto no se puede hablar de las desigualdades al comprobar la poca diversidad en los centros de élite sino se trata primero los temas que crean esas diferencias, y en ese sentido se trata de hablar sobre desempleo, discriminación, y falta de recursos en las familias.

Otro factor determinante para conseguir una buena puntuación en el SHSAT está en parte relacionado con el conocimiento de la existencia de ese test, y el apoyo de las familias:

*«Existen claras diferencias entre los grupos culturales y étnicos en relación a la percepción y expectativas de las familias hacia el test. Inmigrantes de segunda o tercera generación que proceden de países donde la evaluación del rendimiento es un componente importante en la educación, suelen puntuar mejor en el test, por ejemplo los estudiantes que su familia proviene de países asiáticos. Sin embargo, hay comunidades, como la afroamericana, en las que no existe una cultura de preparación del test, ya que tampoco ven que pueda ser realmente una oportunidad para solucionar sus problemas». (CH).*

Un ejemplo como anécdota que Curtis Chin indicó fue que: *«Los domingos mientras los adolescentes de otras comunidades van a la iglesia, las familias de origen asiático llevan a sus hijos a escuelas de idiomas o escuelas de preparación del SHSAT».*

El impacto que el documental está teniendo en las diversas comunidades de Estados Unidos, es fundamental para propiciar un cambio, una reflexión en los educadores no solo de ese país sino de la gran mayoría de países en los que se ha proyectado. El documental ha hecho visible que la problemática no está solo en la realización o no del test y sus consecuencias, sino que de partida existen desigualdades económicas y culturales que hacen que el test no sea una prueba tan objetiva para medir las competencias del alumnado.

*«Creo que el hecho de que tantísimos educadores se hayan ofrecido para utilizar el documental en sus programas educativos, y que los sindicatos del profesorado hayan mostrado su interés, [...] indica que está teniendo repercusión. Lo hice pensando en los profesores y los padres, y por eso estoy muy entusiasmado. Me ha sorprendido el hecho de que sociólogos, psicólogos y orientadores estén empleándolo. Lo que queremos con este documental es que la gente piense en el efecto que la pobreza y el racismo tienen en cualquier debate. A veces tenemos esta tendencia a debatir sobre determinadas cuestiones desde un punto de vista estrictamente teórico, sin conexión con las experiencias de la vida real por las que pasan las familias. Se nos ocurren políticas que quizás parecen geniales sobre el papel, pero cuando se ponen en práctica no tienen los resultados deseados y, lo que es peor, hacen más grande el problema» (Carmona, 2016:25).*

El documental hace reflexionar también sobre el aspecto global alrededor de las problemáticas que envuelven las evaluaciones del alumnado. De hecho, por ejemplo según comenta Curtis, en Springfield (Estados Unidos), se volvía a debatir sobre la idea de utilizar una prueba estandarizada como único criterio para acceder a las escuelas públicas de secundaria. Al igual que ocurre en muchos otros países, existe un debate sobre la evaluación, y en España la LOMCE en parte indica que será necesario realizar evaluaciones individualizadas, e informes sobre el nivel del alumnado así como de centros. Pero lo importante es que: «*personas de diferentes países puedan aprender las unas de las otras sobre el efecto que los tests pueden tener en las personas y comunidades*» (CH, 2017).

## 6. Discusión

La evaluación del alumnado es un tema actual a debate que tiene muchas implicaciones para diferentes colectivos, pero especialmente y del que no debemos olvidarnos, el alumnado. Antes de redactar, promulgar, o modificar una ley es necesario plantearse si realmente se están recogiendo las ideas, necesidades, oportunidades de los colectivos implicados directa o indirectamente en dicha ley. Es por ello que es necesario que los responsables del sistema educativo tengan información lo más objetiva posible. Así pues, es necesario que se tenga información sobre cuatro grandes áreas en la evaluación: resultados, contexto de los estudiantes, procesos de enseñanza aprendizaje y políticas y gobierno educativos (González-Such *et al.*, 2016; OCDE, 2016).

En relación al contexto escolar, tal y como sucede en la ciudad de Nueva York, la incorporación de alumnado inmigrante a la ESO en España —que habla un idioma materno diferente, tiene pautas conductuales propias del contexto socio-cultural de origen, un sistema de valores que puede ser opuesto, y además experiencias académicas dispares en comparación con el alumnado autóctono— ha cambiado sustancialmente el contexto de actuación docente.

Sin embargo, a pesar de observar esa realidad, la mayoría del profesorado no alcanza a comprender la importancia o incidencia de estos factores en las posibilidades formativas de sus estudiantes. Es decir, existe una tendencia de los docentes a justificar y asumir que, dado el contexto cultural y familiar mayoritariamente precario de esos chicos y chicas, los resultados desfavorables son difícilmente cambiables (Fernández Sierra, 2017). Estos resultados están en línea con el aspecto que Curtis Chin, director de TESTED, mencionó también en la entrevista: «*Las expectativas del profesorado para que el alumnado sienta interés por solicitar el SHSAT y poder tener una oportunidad para retar al contexto, es primordial*» (CH).

De lo anterior se deriva plantearse la cuestión sobre cuáles serían los factores que estarían más relacionados con la calidad y equidad en educación, ¿todos los niños y niñas están en una situación favorable para recibir ese mínimo al menos de calidad? Es por ello necesario prestar atención al uso de la información de variables de contexto recabada a través de los diferentes informes que se emiten a partir de las evaluaciones autonómicas, nacionales e internacionales. En ese sentido los cuestionarios de contexto deberían de atender a la: descripción de la población de estudiantes, el contexto en el que sucede el aprendizaje; así como el estudio de relaciones entre el contexto y el rendimiento escolar (Jornet, 2016).

A pesar de conocer cuáles son los factores que más pueden incidir en la calidad de la educación, es necesario reconocer que existen desigualdades que los sistemas educativos no deben de reproducir, sino que deben de eliminar la existencia de una alta segregación escolar, socioeconómica, étnico-racial y entre los

estudiantes inmigrantes. Superar este problema es, sin duda, uno de los mayores desafíos que los países tienen para poder llegar a la construcción de sociedades más inclusivas, cohesionadas y justas.

## Referencias bibliográficas

- Abdulkadiroğlu, Atila; Angrist, Joshua y Pathak, Parag (2014): “The Elite Illusion: Achievement Effects at Boston and New York Exam Schools”. *Econometrica*, 82 (1), 137-196.
- Bradley, Robert H. y Corwyn, Robert F. (2002): “Socioeconomic status and child development”. *Annual Review of Psychology*, 53, 371-399.
- Carabaña, Julio. (2015a). *La inutilidad de PISA para las escuelas*. Madrid: Catarata
- Carabaña, Julio. (2015b): “Repetir hasta 4º de Primaria: determinantes cognitivos y sociales según PIRLS”. *RASE: Revista de la Asociación de Sociología de la Educación*, 8 (1), 7-27.
- Carmona, Carmen (2016): “Entrevista a Curtis Chin”. *Futura*, 33, 22-26.
- Carmona, Carmen; García-Bellido, María del Rosario y Bakieva, Margarita (2014): “Reflexiones de un grupo de gestores en Valencia sobre las evaluaciones de sistemas educativos” en Jesús M. Jornet, Mercedes García-García, y José González-Such (eds.): *La evaluación de sistemas educativos: informaciones de interés para los colectivos implicados*. Valencia: Servei de Publicacions.
- Casey, Beth M.; Dearing, Eric; Vasilyeva, Marina; Ganley, Colleen M. y Tine, Michele (2011): “Spatial and Numerical Predictors of Measurement Performance: The Moderating Effects of Community Income and Gender”. *Journal of Educational Psychology*, 103 (2), 296-311.
- Chin, Curtis, Wolman Adam (productores) y Churtis, Chin (director) (2015). TESTED. Estados Unidos. Disponible en <http://www.testedfilm.com>
- Coleman, James S.; Campbell, Ernest Q.; Hobson, Carol J.; McPartland, James; Mood, Alexander M.; Weinfeld, Frederic D. y York, Robert L. (1966). *Equality of Educational Opportunity*. Washington: US Government Printing Office.
- Corcoran, Sean P. y Baker-Smith, Christine (2015): “Pathways to an Elite Education: Application, Admission, and Matriculation to New York City’s Specialized High Schools”. Working Paper. NY: *Research Alliance for New York City Schools and NYU Institute for Education and Social Policy*.
- Cordero, José Manuel; Crespo, Eva y Pedraja, Francisco (2013): “Rendimiento educativo y determinantes según PISA: una revisión de la literatura en España”. *Revista de Educación*, 362, 273-297.
- D’Angiulli, Amedeo, Siegel, Linda S., y Hertzman, Clyde (2004): “Schooling, socioeconomic context and literacy development”. *Educational Psychology*, 24 (6), 867-884.
- Danziger Sandra K. y Danziger Sheldon (ed.) (1995): “Child poverty, public policies, and welfare reform” [Special issue]. *Children and Youth Services Review*, 17, 1- 2.
- Díaz Caro, Carlos, Pérez Mayo, Jesús y Crespo Cebada, Eva (2016): “Evolución de la pobreza educativa en España y en el entorno Europeo: 2000: 2012”. *Investigaciones de Economía de la Educación*, 11, 207-226.
- Dobbie, Will y Roland G. Fryer, Jr. (2014): “The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools”. *American Economic Journal: Applied Economics*, 6 (3), 58-75.

- Duncan, Greg J. y Brooks-Gunn, Jeanne (1997). *Consequences of growing up poor*. New York: Russell Sage Foundation.
- Fajardo Bullón, Fernando; Maestre Campos, María; Felipe Castaño, Elena; León del Barco, Benito y Polo del Río, María Isabel (2017): “Análisis del rendimiento académico de los alumnos de Educación Secundaria Obligatoria según las variables familiares”. *Educación XXI*, 20 (1), 209-232.
- Fernández Sierra, Juan (2017): “Alumnado inmigrante en la ESO: Vulnerabilidad pedagógica del sistema educativo”. *Educación XXI*, 20 (1), 121-140.
- Gil Flores, Javier (2014): “Factores asociados a la brecha regional del rendimiento español en la evaluación PISA”. *Revista de Investigación Educativa*, 32 (2), 393-410.
- González-Nagel, Eugenio J. (2012): “Cuestionarios de Contexto en los estudios de la IEA. Bordón”. *Revista de pedagogía*, 64 (2), 29-40.
- González-Mayorga, Héctor; Vidal, Javier y Vieira, María José (2017): “El impacto del Informe PISA en la sociedad española: el caso de la prensa escrita”. *RELIEVE*, 23 (1), 1-17.
- González-Such, José; Sancho-Álvarez, Carlos y Sánchez-Delgado, Purificación (2016): “Cuestionarios de contexto pisa: Un estudio sobre los indicadores de evaluación”. *RELIEVE*, 22 (1), 1-26.
- Hauser, Robert M.; Pager, Devah I. y Simmons, Solon J. (2000). *Race-ethnicity, social background, and grade retention*. (CDE Working Paper No. 2000-08). Madison: Center for Demography and Ecology, University of Wisconsin.
- Hauser Robert M.; Frederick Carl B. y Andrew, Megan (2007). *Grade Retention in the Age of Accountability*. (CDE Working Paper No. 2007-04). Madison: Center for Demography and Ecology, University of Wisconsin.
- Hoff, Erika y Tian, Chunyan (2005): “Socioeconomic status and cultural influences on language”. *Journal of Communication Disorders*, 38, 271-278.
- Jornet, Jesús M. (2016): “Análisis metodológico del Proyecto PISA como evaluación internacional”. *RELIEVE*, 22 (1), 1-26.
- Knoester, Matthew y Au, Wayne (2017): “Standardized testing and school segregation: like tinder for fire?”. *Race Ethnicity and Education*, 20 (1), 1-14.
- Maggi, Stefania; Hertzman, Clyde; Kohen, Dafina y D’Angiulli, Amedeo (2004): “Effects of neighborhood socioeconomic characteristics and class composition on highly competent children”. *The Journal of Educational Research*, 98 (2), 109-114.
- Marchesi, Álvaro; Martínez, Rosario y Martín, Elena. (2004): “Estudio longitudinal sobre la influencia del nivel sociocultural en el aprendizaje de los alumnos en la Educación Secundaria Obligatoria”. *Infancia y Aprendizaje*, 27 (3), 307-323.
- McCulloch, Andrew y Joshi, Heather E. (2001): “Neighborhood and family influences on the cognitive ability of children in the British National Child Development Study”. *Social Science and Medicine*, 53, 579-591.

- Mella, Orlando y Ortiz, Iván (1999): “Rendimiento escolar. Influencias diferenciales de factores externos e internos”. *Revista latinoamericana de estudios educativos*, 1, 69-92.
- MECD-Ministerio de Educación, Cultura y Deporte (2013). PISA 2012. Programa para la Evaluación Internacional de los Alumnos. Informe español. Volumen I: Resultados y contexto. Madrid: MECD.
- MECD-Ministerio de Educación, Cultura y Deporte (2016). Datos y Cifras curso escolar 2015/2016. Madrid: Secretaría General Técnica. Subdirección General de Documentación y Publicaciones (en línea) <http://www.mecd.gob.es/servicios-al-ciudadano-mecd/dms/mecd/servicios-al-ciudadano-mecd/estadisticas/educacion/indicadores-publicaciones-sintesis/datos-cifras/Datosycifras1617esp.pdf>
- Noguera, Pedro A. (2017): “Introduction to Racial Inequality and Education: Patterns and Prospects for the Future”. *The Educational Forum*, 81 (2), 129-135.
- OCDE (1999). Measuring student knowledge and skills. A new Framework for assessment. Paris: OECD (en línea) <http://www.oecd.org/edu/school/programmeforinternationalstudentassessment-pisa/33693997.pdf>
- OCDE (2014). PISA 2012. Technical Report. Paris: OECD Publishing. (en línea) <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>
- OCDE (2016). PISA 2015. Context Questionnaires Framework. En OECD, PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy. París: OECD Publishing. doi: <http://dx.doi.org/10.1787/9789264255425-7-en>
- Popkewitz, Thomas (2013): “PISA: números, estandarización de la conducta y la alquimia de las materias escolares”. Profesorado. *Revista de curriculum y formación del profesorado*, 17 (2), 47-64. Recuperado de <http://www.redalyc.org/articulo.oa?id=56729526004>
- Ravitch, Diane (2011). *Life and Death of the Great American School*. New York: Basic Books.
- Rindermann, Heiner (2007): “The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations”. *European Journal of Personality*, 21 (5), 667-706.
- Scandurra, Rosario Ivano y Calero, Jorge (2016): “Modelling Adult Skills in OECD Countries (September 7, 2016)”. *IEB Working Paper*, 2016/17. Available at SSRN: <https://ssrn.com/abstract=2836094>
- Schleicher, Andreas (2015). *Schools for 21<sup>st</sup> Century Learners: Strong Leaders, Confident Teachers, Innovative Approaches*, *International Summit on the Teaching Profession*. Paris: OECD Publishing.
- Schleicher, Andreas (2016): “Desafíos para PISA”. *RELIEVE*, 22 (1), 1-8
- Sirin, Selcuk R. (2005): “Socioeconomic status and academic achievement: a metaanalytic review of research”. *Review of Educational Research*, 75, 417-453.
- Southern Education Foundation (2015). A new majority: Low income students now a majority in the nation’s public schools (en línea) <http://www.southerneducation.org/getattachment/4ac62e27-5260-47a5-9d02-14896ec3a531/A-New-Majority-2015-UpdateLow-Income-Students-Now.aspx>
- Tejedor, Francisco J. (2003): “Poder explicativo de algunos determinantes del rendimiento en los estudios universitarios”. *Revista española de pedagogía*, 61, 5-32.

Tiana, Alejandro (2012): “Analizar el contexto para obtener el máximo beneficio de la evaluación”. *Bordón. Revista de pedagogía*, 64 (2), 13-28.

Van Ewijk, Reyn y Slegers, Peter (2010): “The effect of peer socioeconomic status on student achievement: a meta-analysis”. *Educational Research Review*, 5 (2), 134-150.

Willms, J. Douglas (1999). “Quality and inequality in children’s literacy: The effects of families, schools, and communities” en Daniel P. Keating y Clyde Hertzman (eds.): *Developmental health and the wealth of nations: Social, biological, and educational dynamics*. New York: The Guilford Press.

## Nota biográfica

**Carmen Carmona Rodríguez** es Doctora en Ciencias Sociales y de la Conducta (2006) por la Universidad de Groningen, Países Bajos. Actualmente es profesora del Departamento de Métodos de Investigación y Diagnóstico en Educación-MIDE en la Universitat de València. Ha publicado artículos en revistas internacionales sobre los procesos de comparación social y *burnout* entre el profesorado, autoeficacia y metas del alumnado universitario, y en particular sobre valores, aculturación y competencias interculturales. Es miembro de Comenius Association, la Asociación Interuniversitaria y de Investigación Pedagógica (AIDIPE), la European Educational Research Association (EERA) y la International Association for Cross-Cultural Psychology (IACCP). Ha sido profesora invitada en la Universidad de Maryland (EEUU) y en la Universidad de Groningen (Países Bajos), entre otras.



# La medida de la calidad educativa: acerca de los *rankings* universitarios<sup>1</sup>

## Educational Quality Measurement: On University Rankings

Alejandra Montané, José Beltrán y António Teodoro<sup>2</sup>

### Resumen

En el presente texto se revisa desde una perspectiva crítica la construcción de sistemas transnacionales de evaluación de la calidad a través de los rankings internacionales. En la primera parte, reflexionamos sobre el concepto de calidad y presentamos diferentes tipos de rankings: los denominados *League Tables* (ARWU, THE, QS) y los centrados en resultados de investigación (*SCImago Institutions Ranking*). Una vez conocidas las características y posibles utilidades de los mismos, analizamos dialógicamente una serie de riesgos a tener en cuenta para acabar incidiendo de nuevo en el modelo de calidad que proponen y sus posibles efectos.

### Palabras clave

Calidad, evaluación, educación superior, *rankings*.

### Abstract

The present text assumes a critical stance to examine the design of transnational systems of quality assessment through international rankings. First, different categories of ranking procedures are presented, namely, those known as *League Tables* (ARWU, THE, QS); and those centred on research outcomes (*SCImago, Institutions Ranking*). Next, after introducing their potential uses, we dialogically set these types of rankings against a series of risks worth considering. And last, we conclude by offering a reflection on the quality model reflected by each of them, and on their overall effects.

### Key words

Quality, evaluation, higher education, rankings.

Recibido: 28-03-2017  
Aceptado: 02-05-2017

1 Los autores pertenecen al grupo de Clacso *Universidades y políticas de educación superior*. Este artículo se enmarca en el proyecto Erasmus + de la Unión Europea TO-INN. From Tradition to Innovation in Teacher Training Institutions. Referencia: 573685-EPP-1-2016-1-ES-EPPKA2-CBHE-JP.

2 Alejandra Montané López, Universitat de Barcelona, [smontane@ub.edu](mailto:smontane@ub.edu); José Beltrán Llavador, Universitat de València, [jose.beltran@uv.es](mailto:jose.beltran@uv.es); António Neves Duarte Teodoro, Centro de Estudos Interdisciplinares em Educação e Desenvolvimento (CeIED), Universidade Lusófona de Humanidades e Tecnologias, Lisboa, [teodoro.antonio@gmail.com](mailto:teodoro.antonio@gmail.com)

## 1. Introducción

Una determinada concepción de la evaluación de la calidad de la educación superior ampara conceptual e ideológicamente el auge de la aparición de los *rankings* mundiales de universidades que son resultado de la globalización, como refieren Marope, Wells, y Hazelkorn (2013:8), y una señal inequívoca de que las universidades viven realmente «en una época de mediciones y comparaciones». Los *rankings*, cuyo origen es deportivo (Barsky, 2014), se interpretan como un reflejo de la calidad absoluta de las instituciones de educación superior, o por lo menos, un modo creíble de cuantificarla.

Partiremos del supuesto de que la propia noción de calidad es polisémica, compleja, no estática y no se puede entender más que de manera relacional, esto es, poniéndola en juego dialécticamente con otras dimensiones como las relativas a la excelencia, gobernanza o equidad, y a las que hemos prestado atención en ocasiones anteriores (Beltrán y Montané, 2015; Beltrán *et al.*, 2016; Montané *et al.*, 2017). Consideramos que el fundamento de cada instrumento de evaluación está en la definición que se adopte de calidad, incluso Pérez-Esparrells y López-García (2009) argumentan que es erróneo considerar que los *rankings* son un mecanismo idóneo para medir la calidad de la educación superior.

Esta suerte de razón instrumental asocia el fenómeno de los rankings, por una parte, con la *internacionalización de la educación superior* dirigida hacia «economías intensivas en conocimiento» y a la búsqueda global de talentos (Hazelkorn, 2015), por otra parte, remite a las tendencias de *mercantilización de la educación universitaria* alentando una «carrera por el prestigio» –término aún más discutible– destinada a incidir en la elección de universidades por parte de los estudiantes (Hazelkorn, 2015).

Este contexto globalizador y de mercado se hace patente a través del interés de muchos Estados por contar con universidades de élite atractivas para los investigadores de alto nivel y los inversores. Todo ello con el objetivo de convertirse en motores de la generación y difusión de conocimiento a nivel mundial consolidando el modelo de las *world class universities* –las que están en puestos relevantes de los *rankings*– basadas en un discurso ideológico y político comprometido con la adecuación de los sistemas de educación superior de todo el mundo a los requisitos del modelo de producción de los mercados (Santos *et al.*, 2016). En este sentido, los gobiernos de algunos países han fomentado la internacionalización y han financiado la excelencia reformando el modelo de gobernanza de las universidades (Wang, Cheng y Liu, 2012). Así, compartimos la hipótesis de la «gobernanza epistémica», según la cual asistimos a una gobernanza que opera sobre las concepciones acerca de la realidad social de los sujetos (Alasuutari y Qadir, 2016:633-652) y «afecta a las relaciones de poder en los modos de crear, estructurar, coordinar y aplicar el conocimiento» (Campbell y Carayannis, 2013) sostenida a través de una nueva imagen de la geopolítica internacional del conocimiento (Bruner y Uribe, 2007).

Las temáticas relacionadas con la evaluación de la calidad universitaria son múltiples, complejas e interrelacionadas y aunque en el presente texto incidimos en aspectos generales de los *rankings* y sus riesgos desde una perspectiva sociológica, otros interesantes estudios como el de Barsky (2014) trata los procesos por los cuales las revistas académicas internacionales (dominantemente en inglés) determinan el sistema de organización de la información científica, por ejemplo, con la creación del *Institute for Scientific Information* (ISI) que mantiene una íntima relación, y es un aliado indispensable, en la elaboración actual de los *rankings*.

En nuestro texto presentamos, en primer lugar, una aproximación crítica al uso de la noción de calidad, una descripción conceptual y metodológica de los principales *rankings* seguida de una reflexión en torno a los principales riesgos que éstos conllevan.

## 2. Crítica de la calidad pura

Los escenarios recientes de la calidad en Educación Superior forman parte de un proceso complejo y dinámico de gobernanza epistémica que se ha ido configurando en el marco del estado evaluador (Elliot, 2002). No nos detendremos en los desarrollos de las políticas de calidad. Un examen de los mismos en Europa Occidental y con una perspectiva transnacional se ha llevado a cabo en diferentes contribuciones (véase, entre otros, Neave, 1995; Lorenzo Quiles y Cruz de Gracia, 2015). Lo cierto es que el término calidad resulta en la actualidad incuestionable (¿quién osa oponerse a la calidad?) y sin duda es un reflejo muy claro del giro lingüístico en sociología que Richard Harvey Brown sintetiza como: «el proceso de construir conocimiento sociológico –desde cualquier perspectiva teórica– que tiene mucho en común con los procesos con los que la gente común junta realidad y significado en sus vidas cotidianas». (Brown, 1983: 129). En efecto, «la noción de calidad ha sido socializada hasta tal punto entre el común de los ciudadanos-consumidores que se acepta su existencia (en términos ontológicos: se esencializa) y se constituye en criterio universal en la percepción y elección (...) de bienes y servicios». (Aguilar, 2006: 16-17). En el ámbito educativo su uso es tan frecuente como difícil –por no decir imposible– su definición. Todos hablamos de calidad dando por supuesto que sabemos de lo que hablamos aunque no sepamos definir el objeto o la cualidad a la que nos referimos. Forma parte de aquellos términos que son resultado del triunfo de la ambigüedad o cuya textura es tan abierta y maleable que acaban convirtiéndose en etiquetas, mantras o entelequias metafísicas (hurtadas de cualquier contenido político o histórico), al servicio de consensos amplios y fáciles. En palabras del sociólogo de la educación Carlos Lerena, «la cuestión radical no consiste en preguntarse o saber qué es lo que quiere decir la expresión *calidad de la enseñanza*; se trata, no de saber eso, sino precisamente de saber qué es lo que con esa expresión se quiere callar, esto es, qué se quiere evitar decir o qué se quiere ocultar». (Lerena, 1989:95). Estaríamos, entonces, ante una metáfora fundamental («root metaphor»), o lo que es lo mismo, ante una palabra a través de la cual percibimos el mundo social (Alasuutari y Qadir, 2016:637-638).

En nuestro lenguaje ordinario, no decimos que *hacemos* calidad educativa, sino que *tenemos* calidad educativa, como una propiedad que nos pertenece y no como un proceso del que participamos. Y con ello nos referimos a un producto (mensurable, contable) cuyo valor reside en el *quantum* o cantidad que expresa, más que en la experiencia a la que podría remitir. Cuando la calidad, así percibida y cristalizada, se naturaliza y se *sustantiviza* como un objeto de mero cálculo, y no se *adjetiva* como experiencia relevante, se está contribuyendo a la construcción de una suerte de marco cultural, de una gobernanza epistémica sustentada en una supuesta «objetividad tecnocrática» (Innerarity, 2016:11). En este sentido, la propuesta de los *rankings* presenta un primer gran inconveniente objeto de debates y críticas que se basa en pretender que, a través de unos índices, o indicadores, se capte una realidad multidimensional y que, a pesar de la cautela que se debería tener ante ellos, arrastren la capacidad performativa creando una suerte de nueva realidad que empuja a muchas universidades a planificar el modo de mejorar su puntuación en ellos.

Esta inversión de la calidad entendida como cantidad obedece a una confusión epistemológica sobre la que se ha erigido una poderosa industria académica. Es la confusión, que ya denunció el físico Georgescu-Roegen (1971), derivada de no distinguir entre conceptos aritmómorficos y conceptos dialécticos, ambos tan inevitables como necesarios para las ciencias sociales. Los primeros, aquellos que tienen forma aritmética, se pueden hacer corresponder con la serie de los números naturales o reales. Así, si queremos calcular la «longitud» de una mesa, la mediremos utilizando números. Pero hay conceptos que no son susceptibles de medición con números, que no son representables numéricamente, son dialécticos,

y escapan a las leyes de la lógica formal. ¿Cómo podemos representar numéricamente la belleza, la vida o la muerte? Del mismo modo, ¿cómo podemos representar numéricamente la calidad? En realidad, la factoría de los *rankings* padece la enfermedad de la aritmomanía, una nueva versión del clásico positivismo de las ciencias, que invalida como no científico cuanto escapa a la medición aritmomórfica, a aquello que no es susceptible de un registro numérico. Los datos numéricos –inevitables y necesarios– no admiten discusión. Ahora bien, las nociones dialécticas –sin las cuales no existiría el lenguaje en el que se basa el conocimiento y la ciencia– no tendrían ningún sentido sin la discusión ni el debate del que se nutren y que generan. La pretensión de medir o calcular (una actividad aritmomórfica) la calidad educativa (una noción dialéctica) –como si esta tuviera peso, longitud o amplitud– es una contradicción en los términos. Sin embargo, buena parte de la literatura académica y del negocio sobre la calidad de la educación se basa en este oxímoron. No existe la calidad pura, ni la manera de medirla, pues el significado de este término depende del uso que hagamos del mismo en contextos determinados, bajo condiciones específicas y para intereses preestablecidos.

### 3. Sobre los rankings

Una vez establecido el marco discursivo, describimos y analizamos cuatro de los *rankings* más conocidos, tres de ellos de la *League table* –QS, ARWU y THE–, y otro basado en los resultados de investigación, el SCImago *Institutions Ranking*. Este último, determinado por sus autores no como un *ranking* sino como una ordenación de resultados.

Metodológicamente se diferencian en que algunos se sustentan en análisis cuantitativo de la producción científica, otros se basan en informaciones subjetivas sobre el prestigio y reputación y otros combinan ambos métodos. Los *rankings* sintéticos convierten los datos de cada indicador en puntuaciones y a cada una de ellos se les asigna un peso determinado en función de lo que consideran más importante los autores del *ranking* o del concepto de calidad subyacente. La ordenación dependerá de las puntuaciones obtenidas y las mejores universidades serán las que más cerca estén de los máximos valores.

El IREG (*Group International Ranking Expert*) movimiento académico vinculado favorablemente con la elaboración de los *rankings* (Barsky, 2014:64) se estableció en 2004 en el marco de la Reunión de Seguimiento de la Mesa Redonda sobre «Indicadores Estadísticos para Evaluación de la Calidad de Instituciones de Educación Superior: Metodologías de Ranking y Tablas de Posiciones», convocada conjuntamente por el Centro Europeo de la UNESCO para la Enseñanza Superior (CEPES) y el Instituto de Políticas de Educación Superior (IHEP) y elaboraron el documento titulado «Principios de Berlín sobre los *Rankings* de Instituciones de Educación Superior» que quiere servir de guía para productores y consumidores de *rankings*.

En dicho documento se describe la orientación a la filosofía de mercado de los *rankings*, destacando que estos no deben ser el único método de evaluación de los sistemas de Educación Superior. No obstante, tanto los *rankings* como los *benchmarking* pretenden medir la calidad de servicios, programas, proyectos, productos, organizaciones e instituciones forzando la comparación entre diferentes instituciones y sistemas y la competición entre ellos (Krüger y Molas, 2010).

Los usuarios de los *rankings*, según el IREG, pueden esperar y deben exigir que los *rankings*, como los *Principios de Berlín* indican, sean claros acerca de sus objetivos y grupos de destinatarios, reconozcan la diversidad de misiones institucionales, identifiquen las fuentes de información utilizadas y especifiquen los contextos culturales e históricos, teniendo en cuenta que existen diferentes valores y creencias acerca de

lo que constituye la «calidad» en la educación superior. En el informe se afirma que los usuarios deben ser conscientes de que existen *rankings* diferentes con propósitos diferentes, que se dirigen a grupos diversos, seleccionan indicadores distintos y utilizan metodologías diferentes (IREG, 2015). Como señala Barsky (2014:65), parafraseando una irónica cita de Stella y Woodhouse (2006), parece como si una alerta sobre el peligro de fumar fuera convertida en un documento que dijera: «no hay problemas con fumar siempre y cuando usted evite los peligros del cigarrillo».

Retomaremos estos aspectos al final del texto, e iniciamos el recorrido siguiente de presentación de los *rankings* de un modo descriptivo y crítico buscando un debate interdisciplinar en un área copada por la Bibliometría y la Cientimetría que codifican como verdades reveladas (Barsky, 2014) ideas que disimulan las dificultades históricas y limitaciones conceptuales con que se han construido los sistemas de evaluación de la calidad de las universidades y de las publicaciones científicas.

### a) *Academic Ranking of World Universities (ARWU)*

El *Ranking Académico de las Universidades del Mundo (Academic Ranking of World Universities - ARWU)* fue publicado por primera vez en junio de 2003 por el Centro de las Universidades de Clase Mundial (*Center for World-Class Universities - CWCU*) de la Escuela Superior de Educación de la Universidad Jiao Tongde Shanghái de China, y su periodicidad es anual. Aunque el ARWU se creó con el objetivo de conocer el posicionamiento de las universidades de educación superior en China en comparación con las universidades de clase mundial, es el *ranking anual más utilizado por universidades de investigación en todo el mundo*.

Como afirma Barsky (2014) su creación estuvo estrechamente asociada a la decisión del sistema universitario chino, centralizado a nivel estatal y con el objetivo de tener un modelo a seguir para el avance de la formación de científicos en China siguiendo como modelo ideal el vigente en los países capitalistas más avanzados.

Se trata de una lista de universidades en que se clasifican unas 500, que representan el 2 % del total mundial. Siguiendo las aportaciones de Barsky (2014), aparecen 149 universidades norteamericanas, con Harvard como el máximo referente, con 17 sobre las 19 primeras, seguidas de las británicas Oxford y Cambridge. Las universidades chinas aparecen a partir de la posición 151, lo que evidencia la importante diferencia entre el nivel de la calidad atribuido a estas universidades y el modelo ideal deseado.

ARWU publica tres *rankings* de carácter internacional: 1) *ranking* general para el que utiliza seis indicadores, 2) el *ranking* segregado por áreas de conocimiento y 3) el *ranking* por disciplinas específicas para los que se utilizan cinco indicadores, cuatro de ellos iguales que en el ranking general.

Para el *ranking* mundial de 500 universidades (ARWU) se utilizan indicadores relacionados con la *calidad docente*, basados en el número de alumnos y profesores que han ganado premios Nobel y medallas Fields. *La productividad científica* se mide a través de los números de investigadores altamente citados, de artículos publicados en revistas de *Nature* y *Science*, y de artículos indexados en *Science Citation Index-Expanded* (SCIE) y *Social Sciences Citation Index* (SSCI). Se tiene en cuenta también el número de artículos publicados en el *Science Citation Index-Expanded* (SCIE) y el *Social Science Citation Index* (SSCI) durante el año anterior y el rendimiento per cápita respecto al tamaño de una institución. Califica más de 1.200 universidades al año y las 500 mejores se publican en la web, a las 100 primeras se les asigna una posición individualizada y a las restantes una posición global.

Para el *ranking* mundial de 200 universidades desagregadas por áreas de conocimiento (ARWU-FIELD) se asigna un puesto individual para las 50 primeras universidades y el resto se agrupan. Se valoran cinco áreas de conocimiento: 1) Ciencias naturales y matemáticas; 2) Ingeniería, Tecnología e Informática; 3) Ciencias de la vida y agricultura; 4) Medicina clínica y Farmacia y 5) Ciencias Sociales.

Para el *ranking* mundial de 200 universidades desagregadas por disciplinas (ARWU-SUBJECT) igual que el anterior, se asigna un puesto individual para las 50 primeras universidades y el resto se agrupan. Las disciplinas analizadas son: 1) Matemáticas, 2) Física, 3) Química, 4) Informática y 5) Economía y empresa.

ARWU es el único *ranking* de carácter internacional que depende exclusivamente de datos obtenidos de forma independiente de las instituciones objeto de estudio y que no utiliza encuestas de opinión. Su diseño es exclusivamente investigador y es muy valorado por la calidad de sus datos y precisión y confianza de sus herramientas de análisis (Docampo, 2008).

Pero, como afirman Santos, Teodoro y Da Costa (2016) la principal crítica se centra en que todos los indicadores se destinan a una sola de las misiones de la universidad, la investigación. Como afirman estos autores «pensar que el número de premios Nobel o medallas Fields pueden ser un indicador de calidad de una universidad es seguramente una ficción». (2016:38). Continúan afirmando que la utilización exclusiva de bases de datos como Thomas Reuters, básicamente de lengua inglesa, desvalorizan las ciencias sociales y humanas.

**Tabla I. Descriptor del *ranking* - ARWU**

<b>Ranking ARWU</b>	
Institución responsable	Shanghai Jiao Tong University
Web	<a href="http://www.shanghairanking.com/">http://www.shanghairanking.com/</a>
País	China
Ámbito	Internacional
Antigüedad	2003
Periodicidad	Anual
Rankings que ofrece	Tres: uno general, uno desagregado por áreas y uno desagregado por disciplinas.
Tamaño	El general es de 500 universidades y los <i>rankings</i> desagregados de 200 cada uno. Los tres se elaboran sobre un total analizado de más de 1.000 centros.
Procedencia de los datos	Instituciones independientes (como la de los premios Nobel), universidades, bases de datos de Thompson Reuters y fuentes estadísticas (nacionales e internacionales).
Criterios de inclusión	El <i>ranking</i> evalúa a cualquier universidad que haya ganado un Premio Nobel, una medalla Field, que haya publicado en <i>Nature</i> o en <i>Science</i> , que tenga investigadores muy citados o con una cantidad significativa de artículos en SCIE y SSCI. Su inclusión en el <i>ranking</i> depende de la puntuación obtenida. La inclusión en uno de los tres <i>rankings</i> no depende de estar incluido o no en los otros dos. Se puede estar en los tres, en dos o solo en uno
Tipo de indicadores para el <i>ranking</i> general	<p>Relativos (1 =10 %):</p> <p>Rendimiento docente (10 %). Es la suma de los cinco indicadores absolutos entre el total de profesores.</p> <p>Absolutos (5=90 %):</p> <p>Alumnos premiados (10%).</p> <p>Profesores premiados (20%).</p> <p>Profesores más citados (20%).</p> <p>Artículos en <i>Nature</i> y <i>Science</i> (20%).</p> <p>Artículos en SCIE y SSCI (20%)</p>

Tipo de indicadores para el por áreas y disciplinas	Absolutos (5=100 %):
	Alumnos premiados (10 %).
	Profesores premiados (15 %).
	Profesores más citados (25 %).
	Artículos en SCIE y SSCI (25 %).
	Artículos publicados en las revistas mejor valoradas respecto del total de revistas en cada campo (25 %)

Fuente: Adaptado de FICUE (2014:24).

## b) Times Higher Education World University Rankings (THE)

El *Times Higher Education World University Rankings* (realizado por la revista *Times Higher Education*) proporciona información relacionada con la docencia, la internacionalización, la innovación, la investigación y la repercusión científica de las universidades a través de un *ranking* general y otro desagregado por áreas de conocimiento a partir de 13 indicadores. Asimismo, ofrece un ranking de universidades de menos de 50 años de antigüedad.

El *ranking* THE recoge la información a través del análisis bibliométrico de publicaciones y citas en los últimos diez años dentro de cada una de las seis áreas temáticas (Artes y Humanidades, Ciencias clínicas, Ingeniería y tecnología, ciencias biológicas, ciencias físicas y ciencias sociales), y una encuesta *on line* a académicos de todo el mundo realizada por Thomson Reuters.

Como señala la *Guía de buenas prácticas para la participación de las universidades españolas en los rankings internacionales* elaborada por el Ministerio de Educación y Deporte y coordinado por Sanz-Casado (2015:24), THE excluye a las instituciones que no tienen estudiantes de grado, las que enseñan una sola disciplina, o aquellas cuya producción en investigación fue menor 200 artículos al año. El *ranking* generalista de THE tiene dos grandes diferencias con ARWU. La primera es que un tercio de la ponderación del *ranking* se basa en una encuesta realizada a más de 10.000 académicos. El empleo de este tipo de indicadores subjetivos ha sido objeto de fuertes críticas, ya que, entraña deficiencias por su subjetividad y por la composición territorial de los expertos consultados ya que es imposible que un evaluador conozca todas las universidades del mundo para emitir un juicio (Pérez y Esparrells, 2009). Por otra parte, los datos los obtienen de las propias universidades y no de fuentes externas, con el problema de homogeneización entre universidades y países a la hora de su correcto empleo (Sanz-Casado, 2015).

**Tabla II. Descriptor del ranking - THE**

<b>Ranking THE</b>	
Institución responsable	Revista Times Higher Education
Web	<a href="http://www.timeshighereducation.co.uk/worlduniversity-rankings/">http://www.timeshighereducation.co.uk/worlduniversity-rankings/</a>
País	Reino Unido
Ámbito	Internacional
Antigüedad	2010
Periodicidad	Anual
Rankings que ofrece	Dos: uno general y otro desagregado por áreas de conocimiento
Tamaño	El general contiene 400 universidades. El <i>ranking</i> por áreas, 50
Procedencia de los datos	Las propias universidades, Thomson Reuters y la opinión de académicos de todo el mundo

Criterios de inclusión	Todas las universidades son analizadas de manera automática, exceptuando aquellas que no ofrezcan formación de grado, que enseñen en un campo excesivamente especializado o hayan publicado menos de 200 por año. Su inclusión en los <i>rankings</i> depende de su puntuación
Tipo de indicadores para el <i>ranking</i> general	<p>Relativo respecto el tamaño de cada universidad (10= 37%):</p> <p>Alumno/profesor.</p> <p>Doctorados/licenciados o graduados.</p> <p>Doctorados/profesores.</p> <p>Ingresos totales /profesores.</p> <p>Ingresos por investigación/profesores.</p> <p>Artículos publicados por profesor/total profesores.</p> <p>Estudiantes extranjeros/nacionales.</p> <p>Docentes extranjeros/nacionales.</p> <p>Publicaciones con un coautor extranjero/ total de publicaciones de la universidad. - Ingresos de empresas/profesores.</p> <p>Absolutos respecto el tamaño de cada universidad (1=30%):</p> <p>Número de citas que generan los artículos publicados por una universidad.</p> <p>Subjetivos (sondeo de opinión) (2=33%): -</p> <p>Prestigio adjudicado a la docencia del centro.</p> <p>Prestigio adjudicado a la investigación del centro</p>

Fuente: Adaptado de FICUE (2014:42).

### c) QS World University Rankings®

La organización británica *QS World University Rankings* (*ranking* QS) comenzó a publicar su propio comparativo de las mejores universidades a nivel internacional en 2009 a raíz del fin de su sociedad comercial con la revista *Times Higher Education* con la cual elaboró conjuntamente el *ranking* THE entre 2004 y 2009<sup>3</sup>.

Utiliza tres tipos de datos según el informe FICUE (2014): Encuestas *on line* entre académicos referidas al prestigio de las instituciones, las bases de datos *Scopus*<sup>4</sup> para la información sobre la producción científica, ratios de recursos docentes e internacionalización extraída de las propias universidades para tratar la calidad educativa.

Desde su creación, el *QS World University Ranking* identifica más de 52 materias, pero solo se incluyen en el *ranking* aquellas que cumplen tres requisitos: recibir más de 20 respuestas entre académicos en las encuestas *on line*, que lleven más de cinco años con publicaciones en ese campo, y que la disciplina impartida se ofrezca en formación de grado y en programas de posgrado (FICUE, 2014).

Ofrece cuatro tipos de *rankings*:

- 1) El general, construido a partir de seis indicadores (reputación académica, reputación para las empresas, ratio de profesores por alumno, número de citas, internacionalización del profesorado e internacionalización del alumnado).
- 2) El de las mejores universidades de menos de 50 años, elaborado con los resultados del *ranking* general.

3 Este *ranking* se difunde directamente como QS a través de *U.S. News & World Report* de los Estados Unidos. Hay que señalar que la ruptura mencionada con el *Times* no fue por cierto amigable y ambos grupos se critican duramente en sus documentos y compiten fuertemente por sus posiciones en el mercado de evaluación (Barsky, 2014).

4 *Scopus* es del grupo editorial Elsevier, creado en el año 2004, mientras que *The Times* se asoció con Thompson Reuters, lo que muestra la importante batalla económica por el monopolio del suministro de la información que está detrás de estos procesos supuestamente académicos.



- 3) El desagregado por áreas de conocimiento (Artes y Humanidades, Ingeniería y Tecnología, Ciencias biológicas y medicina, Ciencias naturales y Ciencias sociales y gestión empresarial), a partir de cuatro indicadores.
- 4) El desagregado por disciplinas específicas, con los mismos indicadores que el anterior.

**Tabla III. Descriptor del QS World University Rankings**

<b>Ranking QS</b>	
Institución responsable	Quacquarelli Symonds (QS)
Web	<a href="https://www.topuniversities.com/university-rankings">https://www.topuniversities.com/university-rankings</a>
País	Reino Unido
Ámbito	Internacional
Antigüedad	2011
Periodicidad	Anual
Rankings que ofrece	Tres: uno general, uno desagregado por áreas y uno desagregado por disciplinas
Tamaño	El general es de 800 universidades; el desagregado por áreas de 400; y el de disciplinas de 200. Los tres se elaboran sobre un total analizado de unas 2.000 instituciones
Procedencia de los datos	Dos encuestas, base de datos <i>Scopus</i> , universidades y otras instituciones educativas de cada país.
Criterios de inclusión	Las universidades seleccionadas son analizadas de manera automática. Su inclusión en el ranking depende de su puntuación.
Tipo de indicadores para el <i>ranking</i> general	Objetivos: Relativos (4= 50 %): - Profesores contratados/alumno (20 %) - Citas que recibe cada profesor/ total de profesores (20 %) - Docentes extranjeros/nacionales (5 %) - Estudiantes extranjeros/nacionales (5 %) Absolutos (0) Subjetivos (sondeos de opinión) (2=50 %): - Reputación académica (40 %) - Reputación para empresas contratantes (10%)

Fuente: Adaptado de FICUE (2014:8).

#### d) *SCImago-Institutions Rankings (SIR)*

SIR desarrollado por el grupo de investigación *SCImago LAB*<sup>5</sup>, es un recurso para la evaluación de universidades y de instituciones de investigación científica de todo el mundo, que genera información de *rankings* con el fin de analizar los resultados de investigación. SCImago incluye una selección de indicadores bibliométricos, una cobertura inmensa de revistas de impacto y concentra sus intereses en países iberoamericanos (Lorenz y Cruz, 2015:168). SIR incluye y clasifica instituciones académicas que realizan investigación, ordenadas por un indicador compuesto que combina tres grupos de indicadores basados en: (1) investigación, (2) innovación e (3) impacto social medida por la visibilidad de su web. Como señalan sus autores, no se trata propiamente de un *ranking* si no de una ordenación a partir de la producción científica de las instituciones que debe entenderse como un rango predeterminado de distribución, no como una propuesta de *ranking*. El único objetivo de estos de informes es caracterizar los resultados de la investigación de la organización con el fin de proporcionar información útil para

<sup>5</sup> Los miembros del grupo SCImago pertenecen al Consejo Superior de Investigaciones Científicas (CSIC) y a las Universidades de Granada, Alcalá de Henares y Carlos III de Madrid, Extremadura, Oporto (Portugal), Universidad Nacional de la Plata (Argentina) y Pontificia Universidad Católica de Valparaíso (Chile).

las instituciones, los políticos y los gestores de oficinas de investigación, ya que es una herramienta válida capaz de analizar, evaluar y mejorar los resultados de investigación.

Sus parámetros de análisis son:

- Producción total de documentos publicados (tiene en cuenta la base de datos Scopus).
- Citas recibidas por publicación.
- Colaboración internacional (proporción de artículos publicados en cooperación con investigadores y centros extranjeros).
- Indicador SJR (*SCImago Journal Rank*) normalizado (media global de la relación entre el SJR de las revistas en que publica la misma institución y el SJR promedio de las revistas del área temática).
- Índice de citación normalizado (proporción entre el nivel de citas recibidas por la institución y el nivel medio de citas en las áreas temáticas de los artículos)

Las instituciones participantes se seleccionan con el único criterio de que tienen que ser instituciones universitarias y de investigación con más de 100 artículos publicados e incluidos en la base de datos Scopus. A efectos de la clasificación, el cálculo se genera cada año a partir de los resultados obtenidos durante un intervalo de cinco años cuyo último año corresponde a dos años antes de la edición del *ranking*. En la edición 2016 los resultados utilizados fueron los del período de cinco años (2010-2014). En la edición 2016 a nivel mundial se clasificaron en su totalidad a 5.147 instituciones, siendo 2.894 las correspondientes al sector de Educación Superior.

**Tabla IV. Descriptor del Ranking - SIR**

<i>Ranking SIR</i>	
Institución responsable	SCImago Research Group
Web	<a href="http://www.scimagoir.com/">http://www.scimagoir.com/</a>
País	España
Ámbito	Internacional
Antigüedad	2009
Periodicidad	Anual
Rankings que ofrece	Uno
Tamaño	Analiza y publica datos de 3.290 instituciones (no solo universidades)
Procedencia de los datos	Base de datos Scopus
Criterios de inclusión	Haber publicado al menos 100 documentos científicos de cualquier tipo (artículos, revisiones, cartas, conferencias, etc.) durante el último año del periodo tomado en cuenta para realizar cada ranking anual
Tipo de indicadores respecto tamaño de la universidad	Absolutos (7): - Artículos publicados Artículos en colaboración con instituciones extranjeras Artículos publicados en las revistas mejor valoradas Promedio de citas recibidas/total de citas en el mundo Grado de concentración o dispersión temática de la producción científica de la institución Artículos más citados/total de artículos de un área. - Artículos en los que el autor principal pertenece a la institución evaluada

Fuente: Adaptación FICUE (2013).

#### 4. Luces y sombras de los *rankings*

A pesar de las críticas que puedan hacerse en contra de los *rankings* –muchas de ellas fundamentadas, entre otras cosas, por su cuestionable construcción metodológica–, no se puede negar su legitimidad (Altbach, 2006) ni el impacto que tienen en el ámbito universitario. Posiblemente uno de los efectos más destacables sea la poderosa influencia que ejercen «en las decisiones de los estudiantes sobre la universidad a la que deberían asistir, y cuanto estarían dispuestos a pagar» por un título de determinada institución (Ordorika y Lloyd, 2013:210).

De acuerdo con Marmolejo (2010), sin importar lo que pensemos, los *rankings* se han convertido en parte del entorno institucional de la educación superior. Aunque es cierto que estos comparativos «pueden ser una herramienta útil que ayuda a guiar el mejoramiento institucional» en el nivel superior (Marmolejo, 2010:2), también lo es que deben analizarse con escepticismo y manejarse con cuidado, como suscribe Baty (2012), editor del *ranking* THE.

Si bien los *rankings* son un instrumento legítimo y útil en muchos aspectos, uno de los principales problemas estriba en la facilidad que tienen para simplificar la realidad y crear la impresión de que una universidad es mejor que otra, basándose en una serie de indicadores usados arbitrariamente para medir la calidad de las universidades (Cheol Shin *et al.*, 2011; Erkkilä, 2013; Federkeil *et al.*, 2012; Marope *et al.*, 2013; Ordorika y Lloyd, 2013).

Los *rankings* internacionales, principalmente de la *League table*, comparten el fuerte impacto mediático, pero, también, importantes críticas sobre su construcción metodológica, entre ellas, el hecho de que en algunos solo se analicen determinados campos científicos, se consideren solo unos pocos indicadores, se estudie solo un grupo de instituciones del sistema universitario o se abuse de indicadores bibliométricos o asociados a la producción de publicaciones científicas mostrando solamente una dimensión de las actividades universitarias (Sanz-Casado *et al.*, 2013). Como señalan Pérez-Esparrels *et al.*, (2010) existen opiniones que clasifican los *rankings* como un «sin sentido» ya que más allá de las limitaciones que presentan pecan de ingenuidad (Harvey, 2008) al medir exclusivamente lo que es posible medir, no lo que es relevante medir (Stella y Woodhouse, 2006).

Otro aspecto a destacar es la poca diferenciación entre el tamaño de las universidades. Así, una universidad muy grande, pero con un nivel de resultados muy pobre es un problema mucho mayor que una universidad con ese mismo nivel pero más pequeña. Pérez y Aldás (2016) señalan que algunos de los *rankings* más citados son de volumen, puesto que la mayoría de variables con las que se construye proporcionan datos absolutos (número de premios nobel o medallas Fields), pero no relativos al tamaño de la universidad. Otros *rankings*, como el *QS World Universities Ranking*, utilizan variables normalizadas por tamaño.

Los actuales *rankings* no suelen contemplar la responsabilidad social de las universidades, y no contemplan la multidimensionalidad, ni suelen analizar las asociaciones entre los indicadores, así como la implicación de otros factores externos

En este sentido, diversos informes, entre el que destaca la *Guía de buenas prácticas para la participación de las universidades españolas en los rankings internacionales*, (Sanz-Casado, 2015) señala una serie de riesgos a tener presentes en relación a los *rankings*, entre los que hemos seleccionado algunos, propuesto otros y ampliado el contenido a modo de diálogo:

*Riesgo de confundir lo que se puede medir con la que es importante medir.* El planteamiento sistémico y organizado de la información de las publicaciones (ISI, THOMPSON) mantienen una estrecha relación con los principales *rankings* ya que éstos basan sus resultados de investigación basándose en dichas bases de datos, descuidando otros aspectos irrenunciables en la actividad de las universidades como la docencia, la innovación o la responsabilidad social. En la medida en que no se utilicen datos sobre estas otras actividades no será lícito extrapolar los resultados de investigación a las demás actividades (Sanz-Casado, 2015:9). La precisión reclamada para la medición puede en realidad ocultar la importancia de lo que no se mide. Existe, pues, una tensión entre lo que es fácilmente mensurable, pero puede no ser significativo, y lo que es de gran importancia, pero no puede medirse (Unterhalter, 2016).

*Riesgo de utilizar indicadores sintéticos poco robustos.* La principal característica de los *rankings* es que simplifican una realidad multidimensional y justamente esta característica los hace muy sensibles a las variables utilizadas. A ello se une la dificultad de agregar todos estos indicadores en uno solo para construir un índice sintético o la dificultad de contemplar conjuntamente variables relativas a cantidad y calidad. La asignación de pesos a los indicadores, o el empleo de indicadores basados en reputación implican un determinado grado de subjetividad (Sanz-Casado, 2015:9). ARWU recurre al número de Premios Nobel y medallas Fields como indicador para estimar la calidad de la docencia y el profesorado que trabaja en las universidades. Puede cuestionarse, sugieren Federkeil *et al.*, (2012), si este indicador en realidad mide la excelencia en la investigación o mide el prestigio de las universidades para la atracción de talento.

*Rankings* como el THE o el QS utilizan algunos indicadores que los diferencian del ARWU, pues se valen de encuestas de opinión (conocidas como *reputation surveys*). Las encuestas utilizadas como herramientas para estimar la calidad de las universidades únicamente pueden ser relevantes si las personas que participan en la muestra tienen una opinión fundamentada en hechos sobre los que están evaluando, aspecto seriamente cuestionado (Flores y Villaseñor, 2015).

*Riesgo de fijarse solo en la élite.* Los *rankings* son elitistas por naturaleza: más de 16.000 universidades en el mundo nunca podrán clasificarse en alguno de estos comparativos (Rauhvargers, 2011). Los *rankings* internacionales se centran en un limitado número de universidades –las 500 que contempla el *Ranking* de Shanghai representan menos del 3% de las existentes en el mundo– utilizando criterios inaplicables para evaluar a la gran mayoría de las instituciones que conforman los sistemas universitarios de cualquier país, incluidos aquellos con sistemas universitarios más potentes (Sanz-Casado, 2015:10). Por analogía, la utilidad los *rankings* para los sistemas universitarios requiere considerarlos desde una perspectiva más amplia e incluyente que la utilizada por los *rankings* internacionales. Esto sucede, como afirman Sanz-Casado *et al.*, (2013:3), porque muchos *rankings* toman como modelos de universidad las de Estados Unidos (o de los países anglosajones) en los que predomina un tipo de universidad más elitista y orientada fundamentalmente a la investigación. Este modelo impide detectar la diversidad de tipologías que existen en otros contextos y que resultan igualmente valiosas para el desarrollo integral de su entorno.

*Riesgo de comparar universidades heterogéneas.* Las universidades poseen distintas especializaciones por campos científicos o en sus actividades (más o menos orientadas a la docencia, a la investigación, o al desarrollo tecnológico). Las universidades también son heterogéneas en cuanto a su historia, coexistiendo universidades varias veces centenarias con universidades de reciente creación (Sanz-Casado, 2015:10). Marginson (2006) llama la atención sobre que los *rankings* globales usan de esta forma sobreentendida

una distinción discriminatoria entre dos tipos ideales de universidades, la universidad de investigación intensiva y la de educación vocacional, masificada. Por su propia naturaleza institucional compleja, las universidades, que agrupan carreras disímiles que integran disciplinas con tradiciones diferenciadas en relación a los campos del conocimiento y de las aproximaciones epistemológicas a estos, parecen no poder ser sometidas a un tratamiento similar a escala planetaria. No todas las universidades tienen los mismos objetivos ni la misma historia institucional y nacional, por lo que es cuestionable la consideración de las «mejores universidades».

*Riesgo de confundir las partes con el todo: distribución territorial.* Tomar en consideración el número de veces que son citadas las publicaciones de los investigadores como indicador de calidad representa otro problema, ya que los científicos norteamericanos que son los más numerosos tienden a citar a sus compatriotas (Flores y Villaseñor, 2015), dejando de lado especialmente aquellos que no son anglófonos. El inglés es el idioma de las ciencias duras, pero no necesariamente es el idioma principal en campos como las humanidades y las ciencias sociales, donde los investigadores hispanos son más representativos (Altbach, 2006). Esta cuestión representa otra de las características de los *rankings* que perjudica, por una parte, a las universidades de los países en donde el inglés no es la lengua oficial y por otra a determinadas disciplinas como las Ciencias Sociales. Acerca de esta notable deficiencia metodológica, Federkeil *et al.*, (2012: 36) señalan que «la reputación de las universidades como atributo de su calidad [...] se ve fuertemente afectado por la estructura de la muestra en términos de su distribución regional».

*Riesgo de no comprender la dimensión de negocio de los rankings.* Siguiendo a Barsky (2014) podemos afirmar que los mercados globales incluyen también el gran negocio de la evaluación universitaria: los *rankings*. A partir de 2007 los *rankings* fueron compilados por la consultora de temas educacionales Quacquarelli-Symonds (QS), dirigida por Nunzio Quacquarelli. En el año 2009 Times Higher Education cortó sus vínculos con QS y firmó un acuerdo con Thomson Reuters.

Thomson Reuters, nacida en 2008, es una de las grandes compañías que suministra información para empresas y profesionales sobre diversos temas (Barsky, 2014:37) y principalmente es el mayor proveedor de información financiera del mundo. Su plataforma proporciona a los académicos y administradores universitarios acceso a bases de datos del mundo de 12.000 de las revistas de mayor impacto académico y más de 110.000 resúmenes de congresos. Abarca las Ciencias Exactas y Naturales, las Ciencias Sociales, Artes y Humanidades. Tiene su sede en Nueva York, opera en más de 100 países y emplea unas 50.000 personas (Barsky 2014:38). Esta alianza entre una poderosa compañía y la medida de la calidad de las universidades debería hacernos pensar en las reglas del juego con una orientación comercial que acaban asumiéndose como imperativos irrenunciables y forman un entramado que trasciende cualquier concepción de calidad.

## 5. Consideraciones finales

La llamada «cultura de la evaluación» (Elliot, 2002), desplegada en el campo de la educación superior, ha requerido y propiciado el desarrollo de diversos métodos e instrumentos para medir, calificar y dar seguimiento al desempeño y resultados de las funciones académicas y actividades de gestión de las instituciones. Dentro de la gama de modalidades evaluativas (Brooks, 2005), concretamente la evaluación comparativa, ha adquirido relevancia a medida que ofrece referentes para contrastar los logros y avances de instituciones ante los resultados obtenidos por otras unidades del conjunto. En tal contexto, los *rankings* y tablas de posiciones (*league tables*), cuya elaboración se remonta varias décadas atrás en el ámbito anglosajón a nivel nacional, se ha expandido internacionalmente.

Si captamos con cierta distancia crítica el conjunto descrito en el texto, podemos apreciar una serie de características en las que parece no importar tanto el significado como el significante. En la actualidad, en efecto, asistimos a la configuración de una neoescolástica, en la que de algún modo se reedita el debate entre realistas y nominalistas (el nombre de la calidad no es la calidad), y en la que prevalece más el cumplimiento de una serie de requisitos formales que de cuestiones sustantivas. En segundo lugar, en los mecanismos de aseguramiento de la calidad se aprecia un énfasis en los resultados (concebidos como productos), antes que en los procesos educativos. Paradójicamente, las exigencias de calidad están medidas o mediadas por instrumentos y procedimientos hegemónicos principalmente de carácter cuantitativo, que acaban cristalizando en una poderosa industria contable (indicadores, *rankings*, comparaciones...). Esta manera de medir la educación desplaza las lógicas institucionales hacia nuevas lógicas que subordinan el conocimiento a lo que se ha dado en llamar economía del conocimiento. Por otra parte, llama la atención que, además de los instrumentos al servicio del crecimiento productivo, no se hayan desarrollado en igual medida instrumentos vinculados primordialmente al desarrollo social o de ciudadanía.

En el fondo, la consideración de la medida de la calidad educativa a través de los *rankings*, parte de un equívoco, a saber: medir significa registrar, informar, es una actividad descriptiva; evaluar significa emitir juicios de valor fundamentados, es una actividad normativa. Pero medir no es evaluar. No se pueden derivar consecuencias normativas –como de hecho se hace– de cuestiones descriptivas.

Lo que se desprende del marco que hemos presentado es la necesidad de un profundo cambio de perspectiva para recuperar y reconstruir los discursos sociales de los actores que participan y están implicados en las propias instituciones. Al fin y al cabo, el valor de la calidad educativa no depende tanto de la rendición de cuentas (*accountability*) sino de la rendición de sentido (*sense-ability*). Por eso, los instrumentos que ayudan a medir la calidad educativa han de estar al servicio del valor educativo (es decir, de aquello que vale la pena), y no a la inversa: una calidad educativa subordinada a los instrumentos que la miden, ignorando el horizonte de los auténticos fines emancipadores que ha guiado y ha de seguir guiando el desarrollo de la educación superior.

En los *rankings* analizados en el presente texto se percibe el supuesto tácito de que una buena universidad es aquella cuya investigación es altamente competitiva sin medir otros parámetros. El prestigio en ese ámbito eclipsa otras virtudes educativas.

## Referencias bibliográficas

- Aguilar, Luis (2006). *Todo sea por la calidad. Tramar el cambio en educación*. Alzira: Germania.
- Alasuutari, Pertti y Qadir, Ali (2016): “Imageries of the social world in epistemic governance”. *International Sociology*, 31, 16-652.
- Altbach, Philip (2006): “The Dilemmas of Ranking” en Philip Altbach (ed.): *International higher education: Reflections on policy and practice*, 77–80. USA: Chesnut Hill, [www.bc.edu/content/dam/files/research\\_sites/cihe/pubs/Altbach\\_2006\\_Intl\\_HigherEd.pdf](http://www.bc.edu/content/dam/files/research_sites/cihe/pubs/Altbach_2006_Intl_HigherEd.pdf)
- Barsky, Osvaldo. (2014): “La evaluación de la calidad académica en debate: los *rankings* internacionales de las universidades y el rol de las revistas científicas”. Buenos Aires: Ed. Teseo, Universidad Abierta Interamericana.
- Baty, Phil (2009): “Rankings 09: Talking Points”. *The Times Higher Education*, 8.

- Baty, Phil (2012): "Rankings don't tell the whole story - Handle them with care. *University World News*, 227.
- Beltrán, José y Montané, Alejandra (2015): "Higher Education in Spain: Framework for Governance and Autonomy in Spanish Universities" en K. M. Joshi y Saeed Paivandi (eds.): *Global Higher Education: Issues in Governance*. New Delhi: B. R. Publishing House.
- Beltrán, José; Montané, Alejandra; Gabaldón-Estevan, Daniel (2016): "Higher Education in Spain: Framework for Equity" en Saeed Paivandi, Saeed y K. M. Joshi (eds.): *Equity in Higher Education. A Global Perspective*. New Delhi: Studera Press.
- Brown, Richard Harvey (1983): "Theories of rhetoric and the rhetorics of theory: Toward a political phenomenology of sociological truth". *Social Research*, 50, 126-157.
- Brooks, Rachele L. (2005): "Measuring University Quality". *Review of Higher Education*, 29 (1), 1-22.
- Bruner, José Javier; Uribe, Daniel (2007). *Mercados universitarios: el nuevo escenario de la educación superior*. Chile: Ediciones Universidad Diego Portales.
- Campbell, David F. y Carayannis, Elias, G. (2013). *Epistemic Governance in Higher Education*. New York: Springer.
- Docampo, Domingo (2008): "Rankings internacionales y calidad de los sistemas universitarios". *Revista de Educación*, número extraordinario, 149-176.
- Elliot, John (2002): "La reforma educativa en el Estado evaluador". *Perspectivas*, XXXII, 3.
- Erkkilä, Tero (2013). *Global University Rankings: Challenges for European Higher Education*. New York: Palgrave Macmillan.
- FICUE (2014): Fotografía Internacional de la Calidad de las Universidades Españolas (2014) Estudio de la posición de las universidades españolas en algunos rankings internacionales. Fundación para la Proyección Internacional de las Universidades Españolas, Universidad.es [www.universidad.es](http://www.universidad.es)
- FICUE (2013): Análisis de la posición de las universidades españolas en los rankings internacionales. Fundación para la Proyección Internacional de las Universidades Españolas, Universidad.es [www.universidad.es](http://www.universidad.es). [Versión provisional].
- Federkeil, Gero; Van Vught, Frans A. y Westerheijden, Don F. (2012). "An Evaluation and Critique of Current Rankings" en F. A. van Vught and F. Ziegele (eds.): *Multidimensional Ranking. The Design and Development of U-Multirank*. Netherlands: Springer Science+Business Media B.V.
- Flores, Jorge Enrique y Villaseñor, Jorge Ignacio (2015): "Perspectivas actuales sobre los rankings mundiales de universidades". *Revista de Educación Superior*, 44 (175), 44-67.
- Georgescu-Roegen, Nicholas (1971). *The Entropy Law and the Economic Process*. Cambridge: Harvard University Press.
- Gómez, Tomás y Puente, Xavier (2012a): "El valor de los ranking universitarios (segunda parte)". Bitácora de gestión universitaria. Disponible en: <http://finanzasuniversitarias.edunomia.es/2012/02/26/el-valor-de-los-rankings-universitarios-segunda-parte/>
- Gómez, Tomás y Puente, Xavier (2012b): "El valor de los rankings universitarios". Economías Inteligentes.com. Disponible en: <http://economiasinteligentes.com/2012/04/22/el-valor-de-los-rankings-universitarios.html>

- Harvey, Lee (ed.) (2008): "Ranking on Higher Education Institutions: a Critical Review". *Quality in Higher Education*, 14(3), 187-207.
- Hazelkorn, Ellen (2015): *Rankings and the Reshaping of Higher Education: the Battle for World-Class Excellence*. London: Palgrave Macmillan.
- Illich, Ivan (1974). *La convivencialidad*. Barcelona: Barral editores.
- Innerarity, Daniel (2016): "La sociedad de los cálculos". *El País*, 29 de octubre.
- IREG – Observatory on Academic Ranking and Excellence (2015): Pautas para grupos de interés de rankings académicos. Disponible en: [http://iregobservatory.org/en/pdf/IREG-Guidelines\\_Spanish.pdf](http://iregobservatory.org/en/pdf/IREG-Guidelines_Spanish.pdf).
- Krüger, Karsten y Molas, Alba (2010): "Rankings Mundiales de Universidades: Objetivos y Calidad". *Ar@cne revista electrónica de recursos en Internet sobre Geografía y Ciencias Sociales*, 129.
- Lerena, Carlos (1989): "De la calidad de la enseñanza. Valor de conocimiento y valor político de una entelequia". *Política y sociedad*, 3, 91-99.
- Lindblad, Sverker (2008): "Navigating in the field of university positioning: on international rankings lists, quality indicators and higher education governing". *European Educational Research Journal*, 7 (4), 438-450.
- Liu, Nian; Cheng, Ying y Liu, L. (2005): "The Academic ranking of world universities using scientometrics." *Scientometrics*, 64, 101-109.
- López, Ana María; Pérez-Esparrells, Carmen; Montañez, Myriam (2008): "Una aproximación a un indicador sintético para la evaluación del sistema universitario público español". *Investigaciones de economía de la educación*, 3, 105-112.
- Lorenzo Quiles, Oswaldo y Cruz de Gracia, Efraín (2015): "Calidad y evaluación de la Educación Superior. Una perspectiva transnacional a través de los ranking". *DEDiCA. Revista de Educação e Humanidades*, 8, julio, 155- 174.
- Lorenzo Quiles, Oswaldo y Cruz de Gracia, Efraín (2014): "Rankings de evaluación de las instituciones de educación superior. Un análisis comparativo". *Pensamiento y acción*, 19, 62-73.
- Marginson, Simon (2006): Global University Rankings at the end of 2006: Is this the hierarchy we have to have? OECD/IMHE & Hochschulrenkonferenz, Germany: Workshop 4-5, Bonn. Institutional Diversity: Rankings and typologies in higher education. [http://www.cshe.unimelb.edu.au/people/marginson\\_docs/IMHE\\_Bonn\\_rankings041206.pdf](http://www.cshe.unimelb.edu.au/people/marginson_docs/IMHE_Bonn_rankings041206.pdf)
- Marginson, Simon (2010). "National and International Rankings of Higher Education" en P. Peterson; E. Bakery and B. McGaw. *International Encyclopedia of Education*, 3ª (ed.) Amsterdam: Academic Press,
- Marginson, Simon y Van der Wende, Marij (2007): "To rank or to be ranked: The impact of global rankings in Higher Education". *Journal of Studies in International Education*, 11 (3-4), 306-329.
- Marmolejo, Francisco (2010): "The Madness of Rankings". *The Chronicle of Higher Education*. <http://chronicle.com/blogs/worldwise/the-madness-of-rankings/27100>.
- Marope, Priscilla Toka Mmantsetsa; Wells, Peter J. y Hazelkorn, Ellen (eds.) (2013). *Rankings and accountability in Higher Education: Uses and Misuses*. UNESCO.



- Montané, Alejandra; Beltrán, José y Gabaldón-Estevan, Daniel (2017): "Higher Education in Spain. Framework for Quality Assurance" en Stamelos Georgios, K. M. Joshi y Saeed Paivandi (eds.): *Quality Assurance in Higher Education. A Global Perspective*. Delhi, India: Studera Press.
- Neave, Guy (1995): "Las políticas de calidad: desarrollos en Enseñanza Superior en Europa Occidental". *Revista de Educación*, 308, 7-29.
- Oroderika, Imanol y Lloyd, Marion (2013): "A decade of international university rankings: a critical perspective from Latin America" en Priscilla TokaMmantsetsa Marope; Peter J. Wells y Ellen Hazelkorn (eds.): *Rankings and accountability in Higher Education: uses and misuses*. Paris: UNESCO Publishing. <http://unesdoc.unesco.org/images/0022/002207/220789e.pdf>
- Pérez, Francisco; Aldás, Joaquín (dirs.) (2016): "U-Ranking 2016. Indicadores sintéticos de las universidades españolas". *Instituto Valenciano de Investigaciones Económicas*, Fundación BBVA, marzo (cuarta edición).
- Pérez-Esparrells, Carmen y López, Ana María (2009): "Los rankings de las instituciones de Educación Superior: Una revisión del panorama internacional". *Calidad en la educación*. Disponible en: [http://www.cned.cl/public/secciones/seccionpublicaciones/doc/63/cse\\_articulo809.pdf](http://www.cned.cl/public/secciones/seccionpublicaciones/doc/63/cse_articulo809.pdf)
- Pérez, Francisco. (dir.) (2014): Rankings ISSUE 2014. Indicadores sintéticos de las universidades españolas. Fundación BBVA e IVIE. [http://www.fbbva.es/TLFU/dat/Informe\\_URanking\\_FBBVA\\_Ivie\\_2016.pdf](http://www.fbbva.es/TLFU/dat/Informe_URanking_FBBVA_Ivie_2016.pdf)
- Pulido, Antonio (2009). *El futuro de la Universidad. Un tema para debate dentro y fuera de las Universidades*. Madrid: Delta Publicaciones.
- Rauhvargers, Andrejs (2011): "Global University Rankings and Their Impact. Brussels: European University Association". [www.eua.be/pubs/Global\\_University\\_Rankings\\_and\\_Their\\_Impact.pdf](http://www.eua.be/pubs/Global_University_Rankings_and_Their_Impact.pdf)
- Santos, Eduardo; Teodoro, António y Da Costa Junior, Reinaldo (2016): "Rankings Universitários: entre a regulação do mercado e a difusão de modelos organizacionais - o caso brasileiro" en Angelo Del Vecchio y Eduardo Santos (orgs.): *Educação Superior no Brasil; Modelos e Misseoes Institucionais*. Sao Paulo: BT Academica.
- Sanz-Casado, Elías *et al.*, (2013): "Rankings nacionales elaborados a partir de múltiples indicadores frente a los de índice sintético". *Revista española de documentación científica*, 36 (3).
- Sanz-Casado, Elías. (coord.) (2015). *Guía de buenas prácticas para la participación de las universidades españolas en los rankings internacionales*. Madrid: Ministerio de Educación, Cultura y Deporte.
- Shin, Jung Cheol; Toutkoushian, Robert K. y Teichler, Ulrich, (eds.) (2011). *University Rankings. Theoretical Basis, Methodology and Impacts on Global Higher Education*. Dordrecht: Springer.
- Stella, Antony y Woodhouse, David (2006): "Ranking of Higher Education Institutions". *Occasional Publications series*, 6.
- Unterhalter, Elaine (2017): "Negative capability? Measuring the unmeasurable in education". *Comparative Education*, 53 (1), 1-16.
- Wang, Qi; Cheng, Ying; Liu, Nian (2012). *Building World-Class Universities Different Approaches to a Shared Goal*. Rotterdam: Sense Publishers.

## Notas biográficas

**Alejandra Montané López** es Doctora en Pedagogía. Licenciada en Filosofía y Ciencias de la Educación. Profesora e investigadora del Departamento de Didáctica y Organización Educativa de la Universidad de Barcelona. Miembro de RIAIPE: Red Iberoamericana de Investigación en políticas educativas y del grupo de investigación consolidado Formación docente e innovación pedagógica (FODIP), iniciativas en las que ha desarrollado una amplia producción científica. Forma parte del grupo de CLACSO Universidades y Políticas de Educación Superior.

**José Beltrán Llavador** es Doctor en Filosofía. Profesor titular e investigador del Departamento de Sociología y Antropología Social de la Universidad de Valencia. Miembro de RIAIPE: Red Iberoamericana de Investigación en políticas educativas y del comité de investigación (CI 13) de Sociología de la Educación en la Federación Española de Sociología (FES), iniciativas en las que ha desarrollado una amplia producción científica. Forma parte del grupo de CLACSO Universidades y Políticas de Educación Superior.

**António Teodoro** es Doctor en Ciencias de la Educación por la Universidade Nova de Lisboa. Profesor Catedrático de la Universidade Lusófona de Humanidades e Tecnologias (ULHT), en Lisboa. Profesor de sociología de la educación y educación comparada. Director del Centro de Estudos Interdisciplinares em Educação e Desenvolvimento (CeIED), de la Universidade Lusófona. Vice-Presidente del Comité de Investigación de Sociología de la Educación (RC04) de la Asociación Internacional de Sociología (2006-2014). Cofundador del Instituto Paulo Freire de Portugal. Coordinador de la Red Ibero-Americana de Investigación en Políticas Educativas (RIAIPE). Fundador de la Sección de Educación Comparada de la Sociedad Portuguesa de Educación Comparada (SEC-SPCE). Coordinador del grupo CLACSO Universidades y Políticas de Educación Superior.

## Reseñas

Michael James Grenfell, (2014) [2007]. *Pierre Bourdieu*. Londres: Bloomsbury. ISBN: 978-14-4115-469-9, 272 páginas.

Juan Pecourt Gracia<sup>1</sup>

**E**n su trabajo *Pierre Bourdieu*, Michael James Grenfell presenta una visión comprensiva sobre la investigación educativa del sociólogo Pierre Bourdieu que resulta de gran interés. El título del libro puede resultar engañoso, ya que sugiere una presentación general de su trabajo sociológico, cuando en realidad se trata de una indagación sobre su interpretación del mundo educativo. Por ello hay que situar el libro dentro de la colección de la que forma parte (*la Bloomsbury Library of Educational Thought*), dedicada a la publicación de estudios sobre los teóricos más influyentes en el campo educativo, como Rousseau, Dewey, Freire o el propio Bourdieu. Teniendo en cuenta que las ideas de Bourdieu han llegado de forma selectiva, de manera que algunos conceptos han sido más influyentes que otros, este trabajo proporciona la oportunidad de abordar en detalle sus ideas más conocidas, pero también de indagar aquellos trabajos que han pasado más desapercibidos, pero que, pese a todo, siguen teniendo vigencia a la hora de entender qué sucede actualmente con la educación.

El libro de Grenfell está organizado en cuatro apartados. En el primero se contextualiza el trabajo de Bourdieu en relación con el trabajo inaugural de Marx, Weber y Durkheim, y se sitúa en el contexto histórico e intelectual de la Francia de posguerra, donde el sociólogo francés tiene que definir su análisis social frente a figuras tan potentes como Sartre o Lévi-Strauss. En el segundo apartado se presenta una exposición crítica del pensamiento educativo de Bourdieu y se realiza un recorrido por su pensamiento educativo en diferentes momentos históricos. Podemos identificar tres libros capitales que suponen tres hitos en la construcción de su mirada sobre la educación: *La Reproducción* (1970), *Homo Academicus* (1984) y *Nobleza de Estado* (1989). En general, los manuales de sociología de la educación han tendido a resaltar las aportaciones del primero, dejando en segundo plano las innovaciones que introduce en las obras posteriores. El tercer apartado se centra en la recepción de la obra de Bourdieu y su influencia en la nueva sociología de la educación. Y, finalmente, en el cuarto apartado, se realiza una revisión crítica de su trabajo y se reflexiona sobre su actualidad en el nuevo milenio.

Posiblemente, el principal valor del trabajo es, como decíamos, la presentación global que se realiza del análisis educativo de Bourdieu, un recorrido desde los primeros estudios durante la expansión del Estado del Bienestar en Francia hasta las obras posteriores en pleno apogeo del neoliberalismo. A estas alturas, el trabajo *La reproducción*, escrito junto a Jean-Claude Passeron, es bien conocido en los círculos académicos y lo han comentado analistas diversos. Cuando un trabajo tiene éxito y se generaliza (y a veces se cita sin una lectura profunda), es fácil olvidar la dimensión innovadora y de ruptura que tuvo en sus orígenes. Así,

<sup>1</sup> Departament de Sociologia i Antropologia Social. Universitat de València, [juan.pecourt@uv.es](mailto:juan.pecourt@uv.es)

Grenfell recuerda de forma acertada la ruptura que supuso su trabajo frente al análisis marxista preponderante del periodo. Mientras colegas suyos como Touraine, Baudelot y Establet, seguían analizando las instituciones educativas de acuerdo a las nociones marxistas de «clase revolucionaria» o «contradicciones del capitalismo», Bourdieu las abandonó de forma consciente al considerarlas inadecuadas para entender el presente, lanzándose a explorar nuevos territorios. Para ello, introducirá elementos nuevos como el arbitrario cultural, el capital cultural, la violencia simbólica y el *habitus*. La búsqueda que emprende en *La reproducción*, que a pesar de sus innovaciones sigue inscrita en la corriente estructuralista de esos años, y guarda semejanzas con el trabajo de coetáneos como Althusser, continuará en los años posteriores, en los que realizará un intento consciente de desprenderse del determinismo que, en algunos casos, se detecta en su etapa más estructuralista.

En este sentido, el análisis que realiza de *Homo Academicus* y *La Nobleza de Estado* en las décadas posteriores es especialmente relevante. En estos trabajos Bourdieu muestra como su sociología de la educación no es una subdisciplina especializada desconectada del resto, sino que forma parte de un proyecto global que consiste en realizar una sociología de las desigualdades sociales, uno de cuyos mecanismos fundamentales son las instituciones educativas, desde la escuela a las universidades de elite. En *Homo academicus* analiza las dinámicas internas y las transformaciones del campo universitario en el París de los años sesenta, y las relaciona con el levantamiento de mayo del 68. En este trabajo avanza en su noción de campo como herramienta metodológica básica para comprender el mundo educativo. Entiende el campo como un espacio de luchas y conflictos, y relaciona (de forma bastante convincente) las disputas teóricas y metodológicas con las posiciones institucionales en el ámbito universitario. De alguna forma, el origen de clase del individuo (por ejemplo, el del profesor universitario) es refractado por su posición dentro de un campo profesional específico, como es el campo académico. Esto le lleva a desechar el análisis de clases ortodoxo, porque considera que no sirve para comprender lo que está pasando en el terreno educativo. Este análisis de la educación superior como mecanismo reproductor de las desigualdades sociales continuará en años posteriores con *La Nobleza de Estado*, en el que continúa el análisis de los campos educativos. En este caso, abordará el papel de las *grandes écoles* (como la HEC y la ENA) en la formación de las elites francesas y la legitimación de las desigualdades existentes.

Aparte de la visión de conjunto que aporta, un segundo elemento a destacar del trabajo de Grenfell, es la reivindicación de la dimensión crítica propuesta por Bourdieu. Como decíamos, el sociólogo francés eludió el análisis marxista de la educación en un momento en que éste tenía un papel preponderante. Desde los años 80, la reinención del capitalismo impulsada por Thatcher y Reagan ha tenido, como en tantas otras cosas, una influencia importante en las perspectivas que utilizan los nuevos sociólogos de la educación. Se han eclipsado conceptos como resistencia, poder, ideología y control, mientras han ganado peso otros como gestión escolar, organización escolar y rendimiento del alumnado. Las aportaciones más recientes tienden hacia un mayor realismo y pragmatismo. El análisis marxista, defendido en su día por numerosos estudiosos y sustento de una concepción radical de las instituciones educativas, se ha diluido sin que hayan surgido alternativas claras que aborden los espacios abandonados. En este contexto de moderación y retrainamiento de la crítica, el trabajo de Bourdieu proporciona, precisamente, una renovación del aparato crítico que parecía haber envejecido de golpe, como resultado de la desbandada de los marxistas. Su planteamiento, junto a las revisiones críticas realizadas desde la sociología de género o el postcolonialismo, resulta de gran interés para contrarrestar las visiones tecnocráticas que, en ciertas ocasiones, puede transmitir el análisis de la realidad educativa.

Tim Rudd and Ivor Goodson (ed.) (2017). *Negotiating Neoliberalism. Developing Alternative Educational Visions*. Rotterdam: Sense Publishers. ISBN: 978-94-6300-852-5, 219 páginas.

Adriana Marrero<sup>1</sup>

Con frecuencia, la sociología de la educación se mueve con dificultad en el vasto campo que intermedia entre los estudios teóricos, muchas veces de inspiración filosófica, que permiten comprender y dar sentido a las grandes tendencias de los sistemas escolares en las sociedades contemporáneas, y la investigación empírica, por lo general, demasiado preocupada por la exactitud de la medición y la confiabilidad de los datos que pretenden representar aquellas tendencias. No es raro que así suceda. Por un lado, una acentuada conciencia teórica genera una visión más complejizante, más crítica, y menos pragmática frente a las necesidades de evaluación que enfrenta todo sistema educativo y suele, por ello afrontar directamente la tarea de discutir las bases, supuestos y consecuencias de esa evaluación. Por otro, la preocupación por la precisión de las evaluaciones, que ha sido parte importante de las discusiones sobre el estado de la educación en distintas partes del mundo en los últimos tiempos y de múltiples estudios comparativos, lleva implícito un cierto olvido de la «validez» sustantiva de la medición, esto es, un notorio silencio sobre la pregunta de qué es aquello que se está midiendo, cuál es su naturaleza y su historia, en qué tipo de relaciones está inscripto, y hasta qué punto, finalmente, esos indicadores realmente miden aquello que pretenden medir.

Lo teórico y lo empírico, el mundo de las ideas abstractas y de la experiencia, muchas veces cuantificada, suelen funcionar, en la sociología, como poderosos polos de atracción que no permiten ni la equidistancia, ni el diálogo. Una tercera vertiente, de origen etnográfico, elude la tensión entre esos dos polos, simplemente, prescindiendo de ambos. Sin teoría propiamente dicha, escudada en la ingenua pretensión de que es posible comprender la realidad sin supuestos teóricos previos, y evitando, también, pisar el complejo terreno de la recurrencia, que es el terreno de lo cuantitativo, pero también y sobre todo de lo social, se adentra en la perspectiva de unos sujetos que pocas veces son vistos, realmente, como actores y agentes inscriptos en una situación social concreta, de tiempo y de espacio. Las tres tradiciones y los tres modos de estudiar la educación y los sistemas escolares actuales, suelen convivir, o con extrema dificultad, o con recíproca indiferencia, configurando, en los hechos, tres comunidades separadas, con distintos puntos de partida, abordajes y métodos. Esto es particularmente cierto en el mundo hispano de los estudios sobre educación, pero es también un rasgo que caracteriza al trabajo sobre educación en muchas partes del mundo.

<sup>1</sup> Universidad de la República (Uruguay), [adriana.marrero.fernandez@gmail.com](mailto:adriana.marrero.fernandez@gmail.com)

Tal vez sea por todo esto, que el libro que nos ofrecen Tim Rudd y Ivor Goodson, impacta al lector con un sorprendente efecto de frescura, profundidad, agudo interés intelectual y, lo que más extraño, dado el tema sobre el que trata, mucha, muchísima motivación. Me refiero a «*Negotiating Neoliberalism: Developing Alternative Educational Visions*», publicado por Sense Publishers al comienzo en este mismo año 2017.

Las múltiples virtudes de este libro no son fáciles de sistematizar. Destacaré lo central: El neoliberalismo, en especial su impacto a partir de la crisis de 2007, como tema, como factor determinante de importantes cambios sociales y políticos, como factor distorsionante de las prácticas educativas y evaluativas de los docentes y en definitiva, de la vida de educadores y estudiantes afectados en práctica cotidiana y su proyección de futuro.

Cuando ya pensábamos que no quedaba nada nuevo para decir sobre el neoliberalismo y el modo cómo ha afectado las políticas públicas, los sistemas educativos, y las vidas de millones de docentes y estudiantes alrededor del mundo, el libro editado por Tim Rudd y Ivor Goodson, provee una visión más completa y más profunda del fenómeno, tal como se expresa en múltiples dimensiones, instituciones y actores. A diferencia de otros muchos trabajos, sin embargo, los autores logran, en un texto construido con notable maestría, evitar deslizarse hacia cualquiera de los dos extremos tan usuales en la literatura educativa hispanoamericana: o hacia el fatalismo y la impotencia frente a un fenómeno de alcances globales, por un lado, o por otro, hacia vagos y hasta cierto punto inconducentes llamados a una resistencia que queda sin conceptualización, ni análisis.

Desde el punto de vista teórico, el texto es de una riqueza y una densidad notables. Tras un estilo elegante y sofisticado, el lector atento percibirá una sólida estructura argumentativa que abreva de la teoría crítica y de los más recientes desarrollos teóricos en el campo de la sociología y la educación. El discurso vincula con facilidad las propiedades limitantes y recursivas de una estructura matrizada al impulso de la crisis y el empuje neoliberal, con las capacidades de agencia de los sujetos, tal como se expresan a la luz de sus propios modos de representarse su situación, y la relación entre su pasado y su forma de proyectarse hacia el futuro, a partir de sus propias narraciones autobiográficas. Según veremos luego, desde el propósito emancipatorio de su proyecto socioeducativo, los autores proponen herramientas teóricas, conceptuales y metodológicas a fin de generar procesos autorreflexivos de interpretación de la realidad para su transformación. Como se ve, entonces, no se trata solo de un aporte para una mejor interpretación de este mundo sumido en todas las consecuencias de las políticas de «austeridad» –las cuales son cuidadosamente problematizadas– sino de aportar en concreto, a todos los actores vinculados a la educación, un programa para la autocomprensión y modificación de las prácticas propias, con propósitos transformadores.

Estos puntos de partida teóricos y estos métodos de trabajo intelectual, sólidamente articulados a través de ejes claros de análisis, se reflejan a lo largo de los distintos capítulos del libro, que, a pesar de la diversidad de autores y de procedencias –distintos centros académicos de Inglaterra, Irlanda y Noruega– logra una consistencia y una unidad notables.

No es posible, en los límites de esta reseña, dar cuenta de las ideas fuerza de cada uno de los trece magníficos trabajos que componen este volumen. Por ello, me gustaría mostrar, al menos, dos de ellos. Por un lado, el que constituye el segundo capítulo del libro, «Aquí, allí y en todos lados: medición, evaluación y logro», de Michael Hayler, especialista en formación docente e integrante del Centro de Investigación Educativa de la Facultad de Educación de la Universidad de Brighton y parte del grupo de

trabajo que lidera Ivor Goodson. Por otro, el trabajo de los propios editores de este libro, Ivor Goodson y Tim Rudd, en el capítulo de cierre, «Los límites de la educación neoliberal: refracción, reinterpretación y reimaginación».

El capítulo de Hayler, que centra su interés en los múltiples efectos de las mediciones estandarizadas de los aprendizajes en las prácticas y la vida de los educadores y estudiantes, así como en las dificultades para la evaluación tal como es entendida por los docentes, merecería formar parte del programa de lecturas obligatorias de todo investigador en educación. Hayler transita, con una destreza singular, entre políticas económicas y educativas derivadas del nuevo avance neoliberal, antecedentes teóricos y herramientas conceptuales, transcripciones de discursos y reflexiones de docentes a partir de un trabajo sistemático del autor de recolección de información autobiográfica, e interpretaciones del propio autor, que articulan y dan sentido a todos los anteriores. La objetividad de las reformas gubernamentales, la subjetividad de los actores, y el diálogo que entabla entre esas dimensiones el autor, resultan en un texto profundo, interesante, complejo e increíblemente atractivo.

De por sí, la evaluación, como sostiene Michael Hayler, es parte intrínseca del proceso educativo y no tiene por qué tener efectos negativos. En la medida en que se practique una evaluación formativa apoyada en la investigación, puede conducir, por el contrario, al empoderamiento de educadores y estudiantes, y motivarlos a la resistencia respecto a la nueva cultura de la rendición de cuentas. Y ello porque, a pesar de su carácter ideológico, esa cultura y la simple medición de resultados a la que va asociada, solo puede surtir efecto sobre los actores si llega a naturalizarse, y no va a naturalizarse mientras la evaluación siga estando asociada al mundo de lo valorativo, y no meramente, al de lo fáctico e instrumental.

Siguiendo el eje de análisis propuesto por Rudd y Goodson, Hayler da cuenta de las cinco «erres», que articulan los distintos trabajos: «Recuerdo», «Regresión», «Reconceptualización», «Refracción» y «Renovación», (*Remembering, Regression, Reconceptualisation, Refraction and Renewal*). Estos conceptos son entendidos como modos de resignificación de la situación, a efectos de repensar las posibilidades políticas de la educación, y ofrecer un «antídoto», al menos parcial, contra las reformas de la política educativa de los últimos treinta años, su retórica, su promoción y su normalización. La «Resistencia», que el autor propone como la sexta «R», está lejos de convertirse en una consigna vacía. Por el contrario, acá, la «resistencia» tiene como propósito reafirmar la escuela como un lugar desarrollo personal y progreso social. Concebidos como actores situados en la intersección entre la biografía, la sociedad y la historia, los docentes podrían constituirse como actores centrales de este proceso, en la medida en que se comprometan, interpreten y respondan a nivel micropolítico del aula, a través de interacciones significativas con niños y jóvenes. A través de un texto rico en reflexiones de docentes sobre su biografía y circunstancias, el autor fundamenta e ilustra la idea central de que, al otorgar sentido a sus propias vidas y a su trabajo, los actores individuales disponen de libertad para interpretar y cambiar su situación, al mismo tiempo que son constreñidos a ajustarse a ella. Y ello, porque el lugar de la resistencia se encuentra en la subjetividad misma, y constituye una lucha que se produce al mismo tiempo *sobre* aquello en lo que nos hemos convertido y *en contra* de aquello en lo que nos hemos convertido.

El libro, como decíamos, se cierra con un capítulo elaborado por los mismos editores, Ivor Goodson y Tim Rudd, «Los límites de la educación neoliberal: refracción, reinterpretación y reimaginación».

El texto desarrolla, de modo magistral, un andamiaje conceptual que se asienta en importantes antecedentes en la teoría sociológica más avanzada, a la que se incorporan nociones novedosas, profun-

damente reveladoras y heurísticamente rendidoras para la interpretación, cuestionamiento y transformación de la realidad social y educativa. Por esto, el texto tiene, sobre todo, una clara intención de convertirse en un programa para la visualización y aprovechamiento de las oportunidades de agencia, en una realidad que podría aparecer, bajo otra óptica, como irresistible.

Goodson y Rudd centran su análisis en el concepto de «refracción», entendido como un instrumento teórico, y a la vez, metodológico, que permite entender el vínculo particular que se establece entre la influencia de la globalización neoliberal y las múltiples formas en que ésta es asimilada o rechazada por las diferentes realidades concretas. Desde el nivel «supra» o global, el neoliberalismo, se expresa de modos distintos a nivel macro o nacional, meso o institucional, y a nivel micro o de los actores individuales. Al igual que la luz se descompone y cambia su trayectoria al impactar contra distintas superficies, también, en el campo de lo social y educativo, la refracción abre posibilidades de emancipación, esto es, abre espacios que no quedan cooptados por las lógicas neoliberales. A los autores no solo le interesa describir el modo cómo esas influencias son refractadas, sino, además, conocer las causas y las motivaciones que generan esa refracción.

La construcción del concepto mismo de refracción, a partir de una estructura molar ordenada en un espacio de propiedades generado por dos ejes cartesianos, señala las tensiones y complementariedades entre la estructura y la acción, el pasado y los proyectos futuros de los actores, tal como aparecen objetivas en las narrativas biográficas de los sujetos. Cuanta mayor es la refracción, sostienen los autores, mejores son los resultados educativos observados en los distintos contextos nacionales. Forma también, parte del análisis el modo cómo se ha expresado, en diversos períodos, la relevante contradicción entre la independización de la lógica instrumental proveniente de los intereses del mercado, que afecta no solo a los procesos educativos, sino también a su evaluación, por un lado, y los procesos y propósitos de una educación pública basada en la solidaridad y en los valores colectivos, por otro.

Si la refracción significa oportunidades de agencia, la producción misma de la agencia pertenece al sujeto, en aquellos aspectos y lugares que no han sido colonizados y cooptados por la lógica instrumental del neoliberalismo. A partir de la puesta en valor de la capacidad reflexiva del ser humano frente a estos problemas, y del cuestionamiento del lugar de nuestra propia acción en ellos, es como puede generarse una política de vida que logre alinear, de modo cotidiano, cada una de las acciones con nuestros propósitos e ideales. Como se lee en una cita mencionada por los autores: «Supón que lograras esa revolución sobre la que hablas y con la que sueñas. Supón que tu bando haya ganado y tuvieras el tipo de sociedad que querías. ¿Cómo vivirías, personalmente, en esa sociedad? ¡Empieza ahora mismo a vivir así!»

Goodson y Rudd se sitúan, en esto, muy lejos de las perspectivas teóricas que ven en los procesos modernos de exacerbación de la racionalidad instrumental y de pluralismo, el origen de la pérdida de sentido de la existencia. Bien al contrario, en una postura que, al mismo tiempo que deriva de una sólida conceptualización, muestra el lado más luminoso de ella, los autores ubican en el sentido vital, aquello que está incontaminado y es incorruptible por la lógica racional del mercado: no siempre pueden conquistar nuestras mentes, menos aún, nuestros corazones.

Tan riguroso como inspirador, el libro que ponen en nuestras manos Tim Rudd y Ivor Goodson es una oportunidad invaluable de volver a reflexionar, con cuidado, sobre todo aquello que hemos querido alguna vez como educadores, estudiosos y sociólogos de la educación; pero también como personas sensibles y reflexivas, en busca de un mundo mejor. Solo que esta vez, estamos invitados a lograrlo con



la mejor guía posible: la de estos dos grandes intelectuales y colegas, que a través de estas páginas todos quienes se acerquen a ellas, los podrán sentir tan amigos, tan humanos y tan cercanos.

Tal vez por ello, han logrado también, reunir a un notable conjunto de distinguidos colegas y pensadores, también personas sensibles, que han contribuido con sus trabajos a este proyecto. Me limitaré a transcribir los nombres de estos autores y los capítulos con los que colaboran, en su orden y, por fidelidad, en idioma inglés, aunque sus aportes merecen la máxima atención como perspectivas y propuestas alternativas en educación, en el desafío al orden neoliberal que ha venido colonizando la educación. Esos capítulos, y sus autores, son:

1. *Negotiating Neoliberal Education: A Positional Issue and an Oposicional Stance* (Tim Rudd and Ivor Goodson);
2. *Here, There and Everywhere: Measurement, Assessment and Attainment* (Mike Hayler);
3. *Against Academic Labour and the Dehumanisation of Educational Possibility* (Richard Hall);
4. *Neoliberal Schooling, Dehumanisation and an Education* (Peter Humphreys);
5. *Neoliberalism and the Value of Higher Education* (Yvonne Downs);
6. *Beyond 'Entrepreneurialism of the Self': What It Means to be a Student in the Neoliberal University* (Nadia Edmond);
7. *Refractions of the Global Educational Agenda: Educational Possibilities in an Ambiguous Policy Terrain* (Ingunn Elisabeth Stray and Helen Eikeland Voreland);
8. *Education Free for All: Outsourcing, Contracts and Conflicts of Interest in the UK* (Deborah Philips);
9. *Co-operative Education, Neoliberalism and Historical Perspectives: The Dilemmas of Building Alternatives* (Tom Woodin);
10. *Co-operative Education Against Neoliberalism* (John Schostak);
11. *Resisting Neoliberal Education: For Freedom's Sake* (Stephen O'Brien);
12. *There Is An-(No)-Other Way: Surfacing the Hidden Injuries of 'Austerity' – Resistance, Resilience and Professional Responsibility: The Irish Case* (Ciaran Sugrue);
13. *The Limits of Neoliberal Education: Refraction, Reinterpretation and Reimagination* (Ivor F. Goodson and Tim Rudd).

En definitiva, el libro editado por Tim Rudd y Ivor Goodson, constituye una magnífica obra, un trabajo epistemológico, teórico y metodológico riguroso y consistente, y un aporte indispensable para repensar la educación, pero sobre todo, para comenzar a operar ya, de modo inmediato, sin más dilación, una nueva postura ante el mundo y la vida que, en la práctica cotidiana, nos permita aprovechar las múltiples oportunidades de agencia, que nos abre, como un crisol, la refracción.

## Anuncios y próximos números

La RASE participará en la II Conferencia Ibérica de Sociología de la Educación que tendrá lugar los días 13 y 14 de julio en Córdoba y que reunirá a un buen número de personas dedicadas a la sociología de la educación. Una selección de las mejores contribuciones presentadas en Córdoba podrán ser publicadas en el número misceláneo de enero de 2018 para dar difusión de la actualidad de la investigación en sociología de la educación tanto en España, como en Portugal.

Estamos preparando el número monográfico de septiembre de 2017 dedicado a **EL PROFESORADO FRENTE A LAS POLÍTICAS DE RENDICIÓN DE CUENTAS: ¿QUÉ REDEFINICIÓN(ES) DE SU PROFESIONALISMO Y DE SU RESPONSABILIDAD SOCIAL?** Coordinado por Christian Maroy, de la Universidad de Montreal (Quebec, Canadá), Yves Dutercq, de la Universidad de Nantes (Francia) y Mar Venegas, Universidad de Granada (España). El plazo de aceptación de contribuciones se cerró el pasado 28 de febrero por lo que actualmente se encuentran en proceso de revisión. Se trata de un monográfico en el que se pondrá el foco en la rendición de la responsabilidad social del profesorado y de su profesionalismo, en el marco de las políticas de rendición de cuentas. El objetivo es estudiar las formas en que el profesorado de los distintos niveles educativos de la enseñanza obligatoria reaccionan o no ante las políticas de eficacia de los resultados y de responsabilización. Se han recibido un número importante de contribuciones y tendrá una mirada claramente internacional.