
Reflexiones en progreso N° 1 sobre *Cuestiones fundamentales y actuales del currículo y el aprendizaje:*

Evaluación del aprendizaje del estudiante y el currículo: Problemas y consecuencias para la política, el diseño y la aplicación



Título	Evaluación del aprendizaje del estudiante y el currículo: Problemas y consecuencias para la política, el diseño y la aplicación
Serie	Cuestiones fundamentales y actuales del currículo y el aprendizaje
Reflexiones en curso	Octubre 2015 N° 1 IBE/2015/WP/CD/01
Directora OIE	Dr. Mmantsetsa Marope
Equipo de coordinación y producción de la OIE-UNESCO	Innovación y liderazgo en currículo y aprendizaje Renato Operti, Lili Ji, Émeline Brylinski
Autor	Dr. Joshua A. Muskin Consultor OIE-UNESCO Investigador principal no residente, The Brookings Institution
Palabras clave	Educación; Evaluación del aprendizaje; Currículo; Competencias; Competencias del siglo XXI; Enseñanza; Planificación educativa; Formación docente

Nota del equipo de la OIE:

La OIE ha creado la serie *Reflexiones en progreso sobre cuestiones fundamentales y actuales del currículo y el aprendizaje* para abrir un espacio común para una conversación a nivel mundial, la producción colectiva y el debate sobre las cuestiones que preocupan en gran medida a los Estados Miembros de la UNESCO. Su intención es la de apoyar la labor de los países en la incorporación de temas desafiantes en los procesos de renovación y desarrollo de los currículos en diferentes niveles, ámbitos y ofertas del sistema educativo.

Inicialmente, la serie *Reflexiones en progreso* se concentra, entre otros, sobre los siguientes temas centrales: a) la atención y educación de la primera infancia (AEPI) como base del desarrollo y aprendizaje infantil integral; b) la lectura y escritura en los primeros grados de primaria para apoyar el desarrollo de competencias esenciales; c) la cultura y las competencias de los jóvenes en el siglo XXI (incluyendo la educación formal, no formal e informal); d) los currículos de las TICs y la pedagogía inclusiva para contribuir a la obtención de unos resultados del aprendizaje pertinentes y eficaces; e) currículos de Ciencias, Tecnología, Ingeniería y Matemáticas (STEM por sus siglas en inglés) para fomentar el desarrollo sostenible; f) el currículo y la educación para la ciudadanía mundial (la paz, los derechos humanos, el desarrollo sostenible, los valores, la ética, el multiculturalismo, etc.); g) la evaluación para mejorar y apoyar las oportunidades de aprendizaje; y h) la educación inclusiva como principio rector de los sistemas educativos.

La serie de reflexiones cubre una amplia gama de productos de conocimiento, entre ellos: documentos de debate, notas sobre políticas, marcos, directrices, prototipos, cajas de recursos, herramientas de aprendizaje y recursos multimedia. Se debate sobre estos materiales, se

perfeccionan, se utilizan y se difunden, con la participación de agencias e institutos educativos y curriculares y, en particular, con encargados del desarrollo de currículos y especialistas en currículo, expertos en desarrollo, encargados de diseño de políticas, formadores de docentes, supervisores/inspectores, directores/rectores, docentes, investigadores y otros agentes educativos. Además, sirven como material de referencia para las iniciativas de la OIE en materia de formación de desarrollo de capacidades en currículo, el aprendizaje y la educación de calidad – principalmente, maestrías, diplomas, certificados y talleres –, para entablar un diálogo político y técnico que involucre una diversidad de partes interesadas y apoyar un trabajo sobre el terreno sostenible en los países.

A través de blogs y foros electrónicos, animamos al público a interactuar activamente y aportar diversas perspectivas. En efecto, el espacio en línea para la reflexión nos permite estar conectados, facilita el intercambio entre expertos de diferentes regiones del mundo y fomenta la reflexión continua sobre los temas en cuestión. El blog está estructurado para reunir recursos diversos, que incluyen herramientas y documentos (como se mencionó anteriormente) sobre temas específicos a fin de proporcionar un conjunto complejo y nutrido de materiales dirigidos a las necesidades específicas de los Estados Miembros. Las *Reflexiones en progreso* capturarán visiones pertinentes, opiniones y comentarios compartidos por el público, y servirán como recurso clave para apoyar los esfuerzos de los Estados Miembros en la incorporación de las conclusiones pertinentes y prácticas eficaces en las políticas nacionales, los marcos curriculares, el desarrollo del currículo y el quehacer profesional.

Evaluación del aprendizaje del estudiante y el currículo: Problemas y consecuencias para la política, el diseño y la aplicación

Resumen: El rol de la evaluación en la educación ha crecido significativamente en las últimas décadas, una tendencia que tiene dos grandes manifestaciones. Una de ellas es el rápido aumento del número de países y jurisdicciones ya sea participando en encuestas (pruebas) internacionales de aprendizaje o comenzando sus propias evaluaciones de todo el sistema, o ambas. La otra es la siempre creciente importancia de la evaluación para hacer que los sistemas y sus principales actores (especialmente los docentes) sean responsables de los resultados educativos. El reciente compromiso de las naciones del mundo y las principales organizaciones internacionales en Incheon (República de Corea) para una educación "de calidad" para todos en 2030 y el próximo acuerdo mundial sobre los nuevos Objetivos de Desarrollo Sostenible van a exigir más de la educación en términos tanto de equidad como en la forma de percibir la "calidad", por lo cual será necesario adoptar una óptica mucho más pertinente. Aunque la evaluación será vital para este proceso, hay un grave riesgo doble de que los sistemas y sus asociados sigan dependiendo excesivamente de pruebas para conducir sus reformas. En primer lugar, la mayoría de las pruebas más importantes no llegan a todos los estudiantes y se concentran solo en unos pocos temas — principalmente matemáticas y lectura y, a veces, ciencias —, con el resultado común de una reducción del currículo y de otros elementos en el proceso educativo. Del mismo modo, con raras excepciones, dichas pruebas descuidan el rango más amplio de competencias personales, tales como la adquisición de nuevos conocimientos usando una variedad de métodos, y la aplicación práctica de los conocimientos y técnicas básicas que los estudiantes aprenden en la escuela. El segundo riesgo es la falta continua de coordinación en la evaluación con las demás principales funciones del sistema educativo — tal vez sobre todo el currículo, operando en cambio en relativo aislamiento (sino total). Para que la evaluación sea de alta calidad y pertinente y para que dé lugar a mejoras reales en el sistema educativo general y sus resultados, debe estar en armonía completa y funcional con el currículo, la formación y el apoyo de docentes, los textos y materiales, la planificación, el presupuesto y todos los demás componentes. El presente documento explora las formas en que la evaluación es de vital importancia para la educación y postula medios por los que se puede conectar de manera efectiva a las otras funciones clave de la educación para impulsar un sistema nacional hacia 2030.

I. Visión general de la evaluación del aprendizaje de los estudiantes en el contexto de la EPT después de 2015¹ y los ODS²

Como está ampliamente acreditado, Peter Drucker, el gurú de la gestión estadounidense nacido en Austria³, afirmó que lo que se mide se hace⁴. Si bien la intención es explicar lo que sucede en el campo de la gestión, la observación de Drucker se refiere ampliamente a todos los sectores, incluida la educación y, más concretamente, al currículo. El currículo incorpora los conocimientos y capacidades con que una nación u otra jurisdicción o institución *prevé* dotar a sus niños y jóvenes para que estén preparados para asumir la edad adulta de una manera comprometida, productiva y satisfactoria. Para que el currículo oficial (prescrito) genere este resultado, es importante que corresponda al currículo que las escuelas (y otras instituciones educativas) ofrecen en realidad y al currículo que los estudiantes aprenden. Sin embargo, lamentablemente en muchos casos se ve comprometido el currículo oficial, y el contenido real de la enseñanza en clase se asemeja solamente a un facsímil tenue o restringido de lo que el sistema educativo, y de hecho toda la sociedad, espera y requiere que los estudiantes aprendan.

Muchas razones pueden explicar esta diferencia, que en muchos lugares es a veces más un abismo. Según lo explicado por Benavot⁵, una causa frecuente es que los libros de texto que un sistema proporciona a sus docentes no se corresponden con el currículo. Esto puede darse debido a problemas de contenido, o conocimientos y técnicas, pero también puede ser una cuestión de filosofía de la educación, o de enfoque. Por ejemplo, un currículo puede presentar métodos basados en la competencia; pero si los libros de texto siguen teniendo demasiada información, presentando hechos para consumir en lugar de la solución de problemas y otras tareas analíticas y operativas para practicar y afianzarse, la instrucción y el aprendizaje probablemente se retendrán en gran medida de memoria. Un currículo puede sabotearse en un sentido similar. Por ejemplo, aunque propugne una amplia gama de competencias y conocimientos para los estudiantes, puede estar tan sobrecargado de contenidos — tanto en el número de asignaturas como la cantidad de material en cada una — que los docentes sienten que deben virtualmente correr a toda velocidad en las clases para cubrir todo el material⁶. El desarrollo de competencias lleva mucho más tiempo y es más complejo que la distribución de información. Una tercera explicación se encuentra en la capacidad de los docentes,

¹ - Educación para Todos, la iniciativa mundial para lograr la educación básica universal iniciada en Jomtien (Tailandia), en 1990, renovada en Dakar (Senegal), en 2000, y adoptada nuevamente en Incheon (República de Corea), en 2015.

² - Objetivos de Desarrollo Sostenible: la iniciativa para un programa de desarrollo integral, equitativo y sostenible a nivel mundial, está programada para ser finalizada y formalizada en septiembre de 2015.

³ - Véase más sobre Drucker en <http://www.druckerinstitute.com/peter-druckers-life-and-legacy/> (26 de mayo de 2015).

⁴ - Una búsqueda en Internet revela que, en realidad, esta afirmación no aparece en ninguno de los principales escritos de Drucker, aunque la atribución es frecuente y el sentido básico de la frase es un reflejo de la mayor parte de lo que ha escrito; y, tal vez, es algo que en realidad pronunció en una forma u otra en una o en muchas ocasiones. Véase <http://athinkingperson.com/2012/12/02/who-said-what-gets-measured-gets-managed/> (26 de mayo de 2015).

⁵ - Benavot (2012)

⁶ - Aquellos estudiantes que pueden mantenerse al día, una "vieja guardia" de educadores puede luchar, son "brillantes" y por lo tanto están efectivamente determinados para avanzar en la educación formal. La existencia de muchos abandonos, la misma lógica afirma, sirve simplemente para demostrar el intenso rigor y la alta calidad del programa académico. Estas son las conclusiones que pocos sistemas y sociedades pueden permitirse el lujo de seguir creyendo. Las exigencias de la economía y la sociedad en general demandan cada vez mayores niveles de alfabetización, aritmética elemental y otros conocimientos y capacidades cognitivas de todos los ciudadanos; por lo que una selección brutal es rechazada. Además, se entiende ampliamente que hay muchos factores más allá de las capacidades intelectuales de los estudiantes que pueden socavar su rendimiento y progreso académico en los diferentes grados.

que a su vez se manifiesta en unas cuantas dimensiones diferentes. La más obvia, ¿poseen las capacidades y conocimientos necesarios para impartir el currículo? Esto se aplica igualmente a la perspicacia pedagógica y su dominio del contenido de la asignatura. La capacidad también se refiere, sin embargo, a factores tales como unas instalaciones adecuadas, el acceso a materiales de aprendizaje adecuados, el tiempo adecuado y el estímulo y apoyo. Cuando estos otros factores no se corresponden con el contenido y requisitos de un currículo, incluso el docente más capaz y bien informado estará en desventaja a la hora de impartir el currículo.

El "obstáculo" potencial para la aplicación plena y fructuosa de un currículo, que es el foco de este trabajo, es la evaluación del aprendizaje del estudiante. Cuando estas dos funciones no están sincronizadas — un fenómeno demasiado común —, la eficacia de ambas está en peligro. Los riesgos del currículo están secuestrados por normas y estrategias de evaluación "traicioneras" o contradictorias, y la evaluación corre el riesgo de ser irrelevante o, por desgracia, más comúnmente, de brindar señales engañosas y perjudiciales referentes al logro y la rendición de cuentas en el sistema global y sus diversos grupos (estudiantes, padres de familia y sociedad) y asociados. Volviendo nuevamente a Drucker, parecería solamente una hipérbole menor, sino una exageración, decir que los exámenes y evaluaciones de nivel del sistema *retienen a todas las demás iniciativas de educación como rehenes*. Independientemente de las declaraciones ministeriales de política relativas a los métodos pedagógicos, independientemente del currículo e independientemente de lo que los docentes aprendan en cursos de desarrollo profesional, si el contenido y las capacidades sobre las que se examinarán los estudiantes y la manera en la que se espera que expongan el conocimiento no se ajustan al currículo y la pedagogía oficial, la mayoría de los docentes normalmente "enseñará para la prueba". De hecho, si un docente no hiciera esto, los colegas, los padres, los estudiantes e incluso muchas autoridades locales lo juzgarían por ser irresponsable, ya que es la prueba lo que determina la oportunidad de los estudiantes de progresar en su educación formal y, por lo tanto, de lograr los beneficios sociales, económicos y personales que acompañan a la educación. Mientras que muchos argumentarán que enseñar para la prueba no es necesariamente la mejor estrategia para enseñar para *tener éxito* en una prueba⁷, esta es, por desgracia, la solución que muchos defienden, tanto si provienen del Sur como del Norte. En un país como los Estados Unidos de América, donde, según Kamenetz, un estudiante promedio realiza durante los 12 años de su vida académica en la escuela primaria y secundaria un total de 113 pruebas estandarizadas⁸, las consecuencias de una estrategia de enseñanza basada en las pruebas sobre el aprendizaje y las cualidades de los graduados de un sistema deben ser vistas como prodigiosas.

Se espera que los procesos en curso en el marco de la Educación para Todos de 2015 y los Objetivos de Desarrollo Sostenible (ODS) repercutan significativamente en lo que los países enseñan y cómo lo enseñan durante los próximos quince años, ya que debería influir mucho lo que los donantes internacionales y otras instituciones promueven, financian, apoyan técnicamente e investigan. Siguiendo la lógica anterior, estos procesos a escala mundial también incluyen la elaboración y la priorización de los principales indicadores para reflejar el aprendizaje que subyace a estos objetivos, junto con las herramientas y los métodos con los cuales las naciones y sus asociados pueden medir el grado de consecución de estos indicadores de aprendizaje (y de educación). El período previo a 2015 (el final del mandato de la EPT de Dakar) marcó un énfasis mundial sobre la evaluación de las

⁷ - Véase Muskin (próxima publicación, 2015).

⁸ - Kamenetz (2015) y <http://www.americanradioworks.org/the-test/> (5 de junio de 2015).

competencias básicas de lectura y matemáticas y, en algunos casos, de ciencias⁹. Este ha sido el caso de los países de ingresos bajos, medios y altos por igual. Después de 2015, sin embargo, los ODS requerirán a los sistemas educativos que hagan hincapié también en otros aspectos del currículo y, sobre todo, en la gama de competencias cognitivas y no cognitivas que preparan a los estudiantes y les dan confianza para utilizar estas capacidades para un compromiso pleno, productivo a nivel económico, social, ambiental y personal¹⁰.

El presente documento tiene como objetivo explorar lo que significa precisamente este pronóstico para la evaluación. ¿Cuáles de estos nuevos retos y objetivos curriculares pueden y deben evaluar los sistemas y sus muchos interesados y asociados? ¿Cómo pueden los resultados de estas evaluaciones orientar de manera constructiva a los encargados del desarrollo curricular y las muchas personas que son responsables de los diferentes elementos necesarios para impartir el currículo con éxito?

II. Visión general de los fundamentos políticos y operativos de la evaluación del aprendizaje

Antes de responder a estas preguntas, es importante primero recordar lo que la evaluación significa operacional y políticamente dentro de un marco de educación. Definida de manera más simple, la evaluación del aprendizaje es el medio por el cual una institución o persona determina el grado en que una persona ha adquirido un conjunto predeterminado de conocimientos y capacidades. En la práctica, sin embargo, la evaluación no es en absoluto sencilla. Más bien, la evaluación comprende una variedad de dimensiones, todas las cuales tienen un efecto significativo en lo que se enseña y cómo, aunque en diferentes grados. Una forma de analizar la evaluación es a través de los siguientes aspectos:

- El **propósito** de la evaluación;
- el **nivel** de evaluación;
- el **ámbito sustantivo de aplicación** de la evaluación;
- la **naturaleza** de la evaluación; y
- los **efectos** de la evaluación.

El **propósito de la evaluación** se divide más básicamente a través de dos vectores; uno es servir a los fines sumativos y formativos, y el otro, distinguir entre estas dos funciones pedagógicas y la rendición de cuentas. Comenzando con la pedagógica, la *evaluación sumativa* es esencialmente un proceso para determinar lo que los estudiantes han aprendido de un conjunto específico de

⁹ - Esta tendencia está documentada en el Informe de Seguimiento de la EPT en el Mundo de 2015, lo que muestra que durante "los últimos 25 años, el número de evaluaciones nacionales ha crecido considerablemente, pasando de 12 en 1990 a 101 en 2013 (UNESCO, 2015b)."

¹⁰ - Los dos principales regímenes internacionales de evaluación, la Asociación Internacional de Evaluación del Rendimiento Escolar (IEA) y el Programa para la Evaluación Internacional de los Alumnos (PISA), operado por la Organización de Cooperación y Desarrollo Económicos (OCDE), que han sido pioneros de evaluaciones a gran escala en esta dirección. La IEA ha realizado desde 2009 el Estudio internacional de educación cívica y para la ciudadanía (ICCS, por sus siglas en inglés) y ha llevado a cabo el primer Estudio internacional de alfabetización en computación e informática (ICLS, por sus siglas en inglés) en 2013. La OCDE ha operado PISA desde 2000, cuya prueba se centra en las capacidades de los estudiantes para resolver problemas y la aplicación práctica de su lectura, matemáticas y conocimiento de la ciencia.

conocimientos y capacidades al final de un período de instrucción. Esto puede suceder después de completar una sección, un capítulo, un semestre, un año o cualquier otro bloque de tiempo o contenido. Puede servir sobre todo para generar una puntuación o a menudo apenas una parte de una puntuación acumulada o alguna otra señal de un grado de cumplimiento frente al currículo, o puede permitir certificar o acreditar a un estudiante, ya sea para confirmar que ha finalizado con éxito un curso (por ejemplo, un diploma) o para calificarlo y, en muchos casos, seleccionarlo para alguna nueva oportunidad de aprendizaje (por ejemplo, el acceso a la universidad) o una función social, económica o de otro tipo. Por el contrario, la *evaluación formativa* típicamente se refiere a métodos para percibir qué aspectos de un conjunto específico de contenidos de un estudiante o grupo de estudiantes está dominando o no con el objetivo de tomar una acción deliberada para remediar las lagunas. Estos 'controles' pueden producirse en el curso de la instrucción, planteando preguntas, revisando ejercicios, respondiendo a preguntas y similares. También podrían ocurrir en las pausas previstas en el proceso de aprendizaje creadas específicamente para identificar lo que los estudiantes están comprendiendo o no, como cuestionarios, deberes u otras actividades diseñadas para indicar a un docente y sus alumnos qué aspectos requieren mayor esfuerzo, ya sea para toda la clase o para algunos estudiantes de la clase.

Por lo general, las evaluaciones sumativas y formativas se mantienen separadas. Sin embargo, las evaluaciones sumativas pueden tener importantes aplicaciones formativas, incluso aunque los sistemas no las aprovechen (o incluso rara vez lo hagan). Dentro de un curso, un docente que observa en una evaluación sumativa que la mayoría de sus alumnos no ha dominado el contenido que es vital para el aprendizaje en el siguiente bloque puede fructíferamente optar por revisar la lección anterior para consolidar el aprendizaje necesario antes de continuar. Del mismo modo, algunos sistemas, como los de Palestina, Bután y Australia¹¹, han introducido mecanismos para utilizar los resultados de los exámenes u otras grandes evaluaciones para proporcionar a los docentes la orientación y los materiales con los que reforzar su instrucción para la nueva promoción de estudiantes y alertar a los docentes que reciban a los estudiantes previamente evaluados qué contenido han dominado y cuál no totalmente.

La evaluación para la *rendición de cuentas* tiene como objetivo general determinar si las inversiones que un sistema, la sociedad o incluso los asociados hacen en la educación producen los resultados esperados y necesarios en cuanto al aprendizaje. Basándose normalmente en evaluaciones sumativas, la idea es hacer responsables a diferentes actores de todo el sistema de los resultados, tal vez recompensándolos cuando sean positivos y, más a menudo, criticándolos o incluso sancionándolos cuando no lo sean. De esta manera, no alcanzar al menos los estándares mínimos en una evaluación puede dar lugar a la dimisión de un ministro, si el desempeño de todo un sistema es insatisfactorio, o al despido o la reasignación de docentes y directores o el cierre de una escuela o su transferencia a una nueva gestión¹² si se localizan las fallas. Es la perspectiva de estas medidas draconianas a nivel local lo que a menudo obliga a los docentes a enseñar para la prueba en lugar de enseñar a aprender (una distinción que se examina más adelante).

¹¹ - Véase Muskin (próxima publicación, 2015).

¹² - Véase, por ejemplo, Rogers, W. T. (2014: 4).

El *nivel de evaluación* se refiere al mismo tiempo al lugar de la gestión de un proceso y los datos que produce y a la cobertura del contenido de la prueba y su "proximidad" al currículo oficial. La gestión de la evaluación reside en tres niveles básicos. El primero es el de todo el sistema, que utiliza evaluaciones de diversas naturalezas tanto para: a) verificar o certificar que los estudiantes han alcanzado satisfactoriamente los estándares curriculares; y b) determinar el grado en que las diferentes políticas educativas, estrategias e insumos están dando los resultados deseados. En el primer caso, el tipo de evaluación es por lo general un examen que todos los estudiantes toman al final de un ciclo; o por lo menos aquellos que deseen certificarse oficialmente y ser aptos para la selección de oportunidades disponibles para los beneficiarios de un título oficial. En este último caso, la evaluación suele ocurrir en los puntos que no son de transición en la vida académica de los estudiantes y puede basarse en muestras o censos. Estas pruebas a menudo cubren también sólo una pequeña gama de temas; generalmente lectura y matemáticas, y en ocasiones ciencias.

El segundo nivel se encuentra por encima de estas evaluaciones de todo el sistema. Se trata de encuestas regionales e internacionales, o pruebas, de conocimiento y capacidades. En el nivel internacional, son las dos principales pruebas IEA¹³: Progreso en el estudio internacional de alfabetización (PIRLS, por sus siglas en inglés) y Estudio internacional sobre las tendencias de enseñanza en matemáticas y ciencia (TIMSS, por sus siglas en inglés), y el Programa para la Evaluación Internacional de los Estudiantes (PISA, por sus siglas en inglés) de la OCDE. PIRLS y TIMSS han operado desde 1960 y ahora llegan a los estudiantes de los grados cuatro (dos pruebas), ocho y doce (TIMSS solamente) en 57 países (en 2011) y 48 (para el grado ocho, en 2007). PISA comenzó en 2000 y se aplicó a los estudiantes de 15 años de edad (con independencia del grado) en 65 países en 2009. También hay unas cuantas evaluaciones regionales, y todas se centran también en los "dos grandes" temas de matemáticas y lectura. Dos de estos dos funcionan en África — Consorcio de África Meridional y Oriental para la Supervisión de la Calidad de la Educación (SAQMEC, por sus siglas en inglés) y Programa para el análisis de los sistemas educativos de CONFEMEN (PASEC, por sus siglas en francés¹⁴) — y el tercero en América Latina — Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE). Las evaluaciones regionales de la IEA tienen como objeto poner a prueba a los estudiantes en una serie de conocimientos y capacidades que se pretende que reflejen los currículos de la amplia gama de países que participan (la tarea de capturar muchos currículos dentro de una evaluación es probablemente más difícil para los instrumentos de la IEA que para las pruebas regionales, ya que PIRLS y TIMSS llegan a través de muchos más idiomas y contextos). PISA, por el contrario, pone a prueba los conocimientos y capacidades que cree que superan los estándares curriculares para, en cambio, determinar cómo de bien está educando un país a sus jóvenes para la economía mundial. En particular en África, las evaluaciones regionales han representado para muchos países la única evaluación verdadera fuera de los exámenes de fin de ciclo que llevan a cabo los sistemas. Ambos conjuntos de evaluaciones extra-sistémicas permiten a los países comparar los niveles de sus alumnos con los de otros países en un conjunto común de contenidos y capacidades. Aunque tales comparaciones están destinadas a proporcionar inspiración a los sistemas y, en especial, un análisis de fondo para ayudar a guiarlos hacia mejoras estratégicas, y

¹³ - IEA se refiere a la Asociación Internacional de Evaluación del Rendimiento Escolar.

¹⁴ - CONFEMEN es la *Conferencia de Ministros de Educación de los Países de Habla Francesa*; véase www.confemen.org (22 de mayo de 2015).

a veces lo consiguen, la clasificación resultante a menudo es categóricamente inútil e incluso tiene consecuencias perniciosas a veces en muchos países y otras jurisdicciones¹⁵.

El tercer nivel de evaluación es el de las aulas. Mientras que los otros dos presentan propósitos sumativos (aunque, como se ha explicado anteriormente, sumativo también puede ser formativo), las evaluaciones a nivel del aula comprenden las estrategias y objetivos tanto sumativos como formativos. El diseño de estas evaluaciones puede implicar a los docentes individualmente, pero también puede involucrar a grupos de docentes en la creación de instrumentos para evaluar un tema común o elementos de aprendizaje que combinan diferentes temas dentro del currículo, tales como la escritura expositiva, aplicaciones integradas y competencias personales, como la colaboración. Los docentes también pueden encontrar materiales de evaluación — elementos de pruebas individuales, instrumentos y protocolos más amplios — en la esfera pública, ya sean proporcionados por la autoridad educativa central, por el sector privado o por alguna entidad de la sociedad civil. En la mayoría de ámbitos escolares, la evaluación a este nivel la impulsa el docente, pero muchos sistemas también promueven fuertemente la autoevaluación y la evaluación por pares. Cediendo a la finalidad de la evaluación formativa que está básicamente reservada a los docentes, vale la pena señalar el aspecto vital de la "retroinformación", con la cual los docentes (o los pares) utilizan los resultados de la evaluación para ayudar a los estudiantes a percibir con precisión las lagunas de su comprensión o rendimiento con el fin de aclarar las percepciones erróneas o tomar medidas concretas para las clases de refuerzo. Según Hill¹⁶, "hay pruebas convincentes de los meta-análisis de cientos de estudios que indican que la evaluación formativa, cuando se utiliza para proporcionar retroinformación diariamente al docente y los estudiantes, es una de las intervenciones más poderosas jamás registradas en literatura de investigación educativa. Pero rara vez se pone en práctica".

El *ámbito sustantivo de aplicación de la evaluación* se refiere aquí a qué conocimiento y capacidades mide un instrumento o enfoque en particular. Como se indicó anteriormente, las evaluaciones a gran escala, tanto sistémicas como extra-sistémicas, tienden a centrarse en sólo dos o a veces tres temas: lectura, matemáticas y ciencias. Los exámenes, por el contrario, llegarán a través del currículo para evaluar todas las asignaturas, o al menos las académicas y las que se pueden medir utilizando métodos de lápiz y papel. Las evaluaciones realizadas en las aulas también abarcan todo el currículo, cubriendo todos los temas, y permiten la evaluación y la *retroalimentación* con mayor profundidad, precisión y puntualidad. También hay diversidad en las formas en que las evaluaciones requieren a los estudiantes que expongan lo que han aprendido. En un extremo está la presentación de información o procedimientos de memoria: ¿qué fecha?, ¿qué resultado del cálculo?, ¿qué definición?, ¿qué rama del gobierno?, etc. Estas preguntas están adaptadas a las pruebas de opción múltiple y las pruebas de verdadero y falso puntuadas mecánicamente.

En el otro extremo están las preguntas abiertas que miden las respuestas de los estudiantes basadas no sólo, o incluso no necesariamente, en la 'corrección' de hecho de sus respuestas, sino también, y quizás más aún, en el proceso que los hace llegar allí. Tales evaluaciones miden el dominio de los estudiantes de los conocimientos y capacidades del currículo, así como su pensamiento crítico, el

¹⁵ - Véase, por ejemplo, la carta abierta al director de PISA publicada en *TheGuardian* el 6 de mayo de 2014 y firmada por más de 80 académicos y funcionarios de educación; <http://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics> (23 de mayo de 2015).

¹⁶ - Hill, Peter, en Barber y Rizvi (2013: 65).

razonamiento, la comunicación, las competencias inferenciales y muchas otras competencias que necesitan con el fin de llevar el currículo a la práctica significativa y eficaz durante sus estudios y más allá. Estas últimas competencias desafían los formatos de pruebas en las que hay que marcar con un círculo la respuesta correcta, pero también desafían la simple recopilación como contenido que un sistema puede agregar al currículo y los libros de texto. Más bien, se puede argumentar que requieren espacio dentro del currículo en el que los docentes (con formación y orientación) pueden guiar a sus estudiantes para lograr un "aprendizaje más profundo"¹⁷ de los contenidos curriculares proporcionando situaciones a través de las cuales cultivar estas capacidades.

Entre los dos extremos se encuentran las evaluaciones de habilidades de los estudiantes para utilizar la solución de problemas, el razonamiento y la aplicación como se relacionan con el currículo. Estas incluyen cosas tales como problemas para matemáticas, inferir y describir emociones o intenciones de una historia escrita y proponer una secuencia de pasos para la investigación de una hipótesis científica; por ejemplo, respectivamente: a) en lugar de simplemente calcular el área de un círculo por la longitud del radio, tales estrategias pueden pedir a un estudiante que determine el tamaño de una pizza; b) en lugar de identificar los hechos a partir de un texto, se puede dirigir a un estudiante para que seleccione entre un conjunto de posibles motivaciones para la decisión de un héroe; y c) en lugar de calcular la longitud del lado de un triángulo en función del tamaño de un ángulo y otro lado, una prueba puede requerir a un estudiante que describa una estrategia trigonométrica para medir la altura de una torre de telefonía celular. Operativamente, este espacio puede ser igualmente adecuado para pruebas en las que hay que marcar con un círculo la respuesta correcta y preguntas abiertas.

En el fondo, el enfoque es un reflejo de la alta vulnerabilidad de la evaluación a las pautas culturales que podrían ser más familiares para ciertos grupos de estudiantes que para otros, incluso dentro del mismo país. Usando los ejemplos anteriores para ilustrar este riesgo, la pizza no es un alimento universalmente popular o incluso reconocido; lo que se considera heroico puede variar de un contexto a otro; y las torres de telefonía celular no son omnipresentes en todos los contextos. Los encargados de crear las pruebas son sin duda conscientes de este reto y tienen estrategias para gestionarlo. Sin embargo, una simple revisión informal de muestras de preguntas que los principales sistemas de evaluación revelan, sugiere que pueden pasar por alto matices importantes que podrían ser significativos para muchos estudiantes. Los examinados podrían suponer la intención de una pregunta a pesar de una pista desconocida, pero si esto requiere incluso un poco más de tiempo en una prueba cronometrada estrictamente, lo cual es muy posible, hay una desventaja añadida; y ¿por qué molestarse con la pista contextual si es tan probable que confunda como que ilumine una pregunta¹⁸?

Lo que falta en casi todas las pruebas estandarizadas son las competencias de "traslación" mencionadas anteriormente. Como se ha indicado, se trata de los comportamientos, talentos y rasgos que necesitan los estudiantes, y para los que los empleadores, líderes cívicos y otros alrededor del mundo claman, con el fin de utilizar su "conocimiento de los libros" del currículo para ser participantes efectivamente productivos, comprometidos y satisfechos en la "vida real". Sólo para arañar la superficie de la larga lista de tales atributos, estos incluyen la confianza, la perseverancia, la ambición, la curiosidad, el trabajo en equipo, la independencia, el aprender, la comunicación, la

¹⁷ - Véase Pellegrino (2012).

¹⁸ - Véase Muskin (próxima publicación, 2015).

empatía, escuchar, el liderazgo, atraer a seguidores, el juicio, la paciencia y la creatividad. También para repetir, estas son precisamente las competencias que tantos gobiernos han codificado en sus planes y reformas educativas nacionales, *pero que tan pocos han incorporado estratégica, prominente y coherentemente* en sus currículos, así como en la formación docente (inicial y en servicio), los libros de texto, la inspección y, quizás más esencialmente (recordemos la situación de "rehén"), en las normas, los instrumentos y las prácticas de evaluación.

Esta laguna se puede explicar por una razón muy simple: todo esto es difícil de hacer, y tal vez más especialmente la evaluación de dichas competencias. Como dice Levin,

... los atributos no cognitivos o de personalidad específicos necesarios para una vida adulta exitosa son más difusos y más controvertidos y no han cedido a los métodos de medición sencillos utilizados para las pruebas estandarizadas. Simplemente, no hay acuerdo global sobre lo que es de importancia más allá de los logros del estudiante y cómo deben medirse. Por estas razones, y tal vez otras, los debates sobre la educación de clase mundial y los sistemas educativos se han limitado a los logros del estudiante¹⁹.

Sin embargo, ¿es realmente necesario juzgar estas competencias en todos los estudiantes de manera uniforme, estandarizada e incluso totalmente objetiva? Ciertamente, la normalización es importante para la medición de los conocimientos y capacidades que se aprenden, utilizan, ejecutan y representan básicamente de manera similar en todos los contextos. Pero esto no es necesariamente o siempre el caso de las competencias personales, que en cambio a menudo se aprenden, utilizan, ejecutan y representan en formas y combinaciones que son en gran medida únicas para cada persona. En referencia a las diferentes naturalezas, según Gardner, de la inteligencia²⁰, también se puede pensar en cómo la confianza podría manifestarse de manera diferente en los estudiantes (y adultos) cuyos puntos fuertes se encuentran, respectivamente, en la curiosidad verbal lingüística, la creatividad lógico-matemática, la comunicación visual espacial y la cenestesia corporal. Tales diferencias podrían servir como indicio de toda la gama de otras permutaciones que podrían resultar de la combinación de inteligencias y competencias. ¿Cómo estandariza esto un sistema?

- A. Identificando una gama media estadística de indicadores para cada uno.
- B. Creando diferentes pruebas para los diferentes "tipos" de estudiantes.
- C. Estableciendo estándares ideales basados en la investigación de expertos.
- D. No lo estandariza.

La respuesta "correcta" es "D", o por lo menos este es el argumento de este autor. Dada la inmensa variedad en la naturaleza y manifestaciones de estas competencias, parece que su evaluación es la más adecuada para las aulas. Por un lado, una evaluación verdadera y rica requiere una variedad de métodos para ser precisa, válida y completa y, en parte, como consecuencia de lo primero, exige un tiempo y una variedad considerables en las circunstancias o tareas, mediante las que se juzga a los estudiantes. Por otro lado, parece que sólo el conocimiento de primera mano del ejercicio de un estudiante de sus competencias sería la información que verdaderamente permitiría a un docente proporcionar una reflexión, una orientación y unas oportunidades significativas para ayudar a cada estudiante a mejorar.

¹⁹ - Levin, Hank, en People for Education (junio de 2013: 10).

²⁰ - Gardner, Howard (2006). Véase también:

http://www.niu.edu/facdev/resources/guide/learning/howard_gardner_theory_multiple_intelligences.pdf (24 de mayo de 2015).

En esta misma línea, esta información tan valiosa, aunque subjetiva, debe ser información muy útil para que un sistema guíe su labor para fortalecer la enseñanza de estas competencias. Con los resultados agregados en las aulas de un sistema, los encargados de adoptar decisiones, diseñadores curriculares, las autoridades de la formación de docentes, quienes elaboran los materiales y otras personas pueden obtener ideas sobre cómo están evolucionando los estudiantes en estas capacidades y lo que los docentes están haciendo y no están haciendo para fomentarlas. Así, pueden crear y difundir políticas, contenidos, estrategias, documentos, formación técnica y apoyo e incluso herramientas y estrategias de evaluación para apoyar una mejor enseñanza y aprendizaje en estos ámbitos²¹.

Estos puntos de vista no pretenden descartar por completo la oportunidad de evaluar las competencias personales, utilizando indicadores y métodos estandarizados y "objetivos"; ni tampoco descartan por completo la utilidad de tales medidas. De hecho, los docentes pueden beneficiarse enormemente de los indicadores, medidas y herramientas estandarizados; pero probablemente no deberían ser herramientas que limiten la evaluación del desempeño a tareas bien delimitadas con plazos determinados. Una guía con rúbricas y criterios claros y definidos con precisión para la evaluación de la creatividad a través del desempeño de un estudiante, por ejemplo, puede sondear amplia y profundamente sus atributos relacionados de manera que una prueba en la que hay que marcar con un círculo la respuesta correcta o incluso una tarea precisa de tiempo limitado simplemente no puede²². En cambio, el punto principal es poner de relieve las limitaciones de los enfoques estandarizados de las pruebas para capturar información suficientemente rica y fiable y para poder utilizar esta información y tomar decisiones verdaderamente significativas. En su lugar, hay un papel de vital importancia aquí para la evaluación en el aula. Esto puede ser pedir mucho de los docentes, y mucho más de lo que muchos expertos y críticos son propensos a creer que la mayoría de los docentes puede manejar. La visión aquí, sin embargo, es que la mayoría de los docentes puede manejarlo *si el sistema funciona lo suficiente como para crear, ofrecer y apoyar las condiciones, los insumos, el contexto y las competencias de los docentes para que lo consigan*.

²¹ - Este enfoque refleja el plan propuesto por el estado de Massachusetts, EE.UU., para desarrollar un índice de educación creativa e innovadora (Departamento de Educación Primaria y Secundaria de Massachusetts, 2012).

²² - Robert Sternberg (2003: xvii), investigador principal en educación en desarrollo y evaluación de la creatividad, ha creado pruebas estandarizadas para tareas como escribir "historias con diversos títulos como *Los zapatos del pulpo*" y crear "arte para temas como *La Tierra desde el punto de vista de un insecto*". Si bien estos retos sin duda ofrecen a los estudiantes la oportunidad de "dejar volar su creatividad", no existe ninguna garantía de que un estudiante encuentre la inspiración necesaria para exhibir plenamente, ni siquiera parcialmente, su creatividad en uno, o incluso una selección, de estas tareas. Tal vez la presión de un entorno de prueba evoca un grave "bloqueo del escritor" o "bloqueo del artista", o algún otro factor lo reprime en ese instante en ese entorno. ¿O, tal vez, este tipo de tareas simplemente no se corresponde con los intereses de un estudiante o los aspectos en los que es más creativo? Un estudiante de matemáticas muy ingenioso no es necesariamente un artista o visualmente adepto. El hecho de ser un guitarrista y compositor talentoso no se traduce necesariamente en una escritura inspirada e inventiva. ¿Significa que un estudiante es poco creativo si no es capaz de completar una de estas tareas? Y tal vez más significativo, ¿qué decisiones toma un sistema sobre la promoción de la creatividad según la información sobre las pruebas estandarizadas? ¿Aumenta los recursos de creación literaria y las artes, o equipa a los docentes con conocimiento, herramientas y oportunidades para evaluar y cultivar la creatividad en todos los ámbitos que corresponden más estrechamente a las pasiones y talentos de cada estudiante? El trabajo de Sternberg ofrece dos grandes conclusiones que sugieren la última de las dos opciones, la opción del aula; aunque es posible que no estaré de acuerdo necesariamente con esta conclusión: "En primer lugar, la creatividad tiende a ser bastante específica de un ámbito, pero no completamente. En segundo lugar, [la creatividad] tiende a ser bastante pero no totalmente distinta de la inteligencia medida psicométricamente (ibíd.)".

La *naturaleza de la evaluación* se relaciona estrechamente con lo que se está evaluando, los fines previstos de los resultados y quién está realizando la evaluación, como se indicó anteriormente. La evaluación clásica para la rendición de cuentas, la certificación, la selección u otras razones sumativas requiere que todos los estudiantes respondan a la misma gama de preguntas por escrito para situarlos objetivamente frente a un conjunto común de estándares u objetivos de aprendizaje concretos (referida a criterios). Especialmente para los propósitos de selección, el objetivo puede ser situar o clasificar el rendimiento de cada estudiante frente a sus compañeros (referida a normas). Si se realizan como encuestas de aprendizaje a gran escala, las evaluaciones comúnmente favorecen las preguntas de opción múltiple y de verdadero y falso, que los estudiantes responden en hojas en las que hay que marcar con un círculo la respuesta correcta. Ciertos exámenes a gran escala sí incluyen ensayos u otras respuestas de texto completo, como para el bachillerato internacional. Si bien la calidad de las respuestas y la evaluación es a menudo mayor en las respuestas basadas en un texto, también el costo es significativamente mayor²³. Además, para muchas evaluaciones a gran escala, especialmente las que se basan en muestras, un análisis detallado de las respuestas es de poca importancia, con sistemas que buscan su lugar en vez de obtener una visión más amplia de los conocimientos y capacidades de su población estudiantil. De hecho, como explica Green, estas pruebas de forma rutinaria están destinadas *no* a medir lo que cada estudiante sabe, sino, más bien, a diferenciar entre los estudiantes con el fin de obtener un margen suficiente de rendimiento²⁴, para aspirar a la famosa (o infame) curva de campana²⁵.

Si bien dicha información podría proporcionar una señal en general aceptable de la calidad del aprendizaje de los estudiantes y la eficacia del sistema educativo — por ejemplo, los estudiantes de cuarto grado están leyendo a un nivel de segundo grado o, con un poco más de exactitud, a los estudiantes del séptimo grado se les dan bien las divisiones pero no las fracciones —, fracasa lamentablemente en al menos dos formas vitales. Una, como se indicó anteriormente, las evaluaciones a gran escala (que no sean exámenes), olvidan el currículo, marginalizando o excluyendo ciertos temas por completo e incluso reduciendo el contenido curricular de las materias que se evalúan (no se puede incluir mucho en una prueba de un par de horas). Stecher y Chun proporcionan una vívida ilustración de cómo el énfasis en la evaluación de la lectura y las matemáticas en el estado de Washington, EE.UU., dio lugar a una reasignación significativa y considerable del tiempo asignado por los docentes a las demás materias del currículo: más de la mitad de los docentes del cuarto grado encuestados indicaron el aumento de la cantidad de tiempo de instrucción para la lectura, la escritura y las matemáticas, mientras que sacrificaron tiempo para estudiar estudios sociales, ciencias, artes y salud y educación física²⁶. En Botsuana, los agentes de la

²³ - Green (2015). Extrapolando a Green, consideremos un caso en el que todos los alumnos han llegado a dominar perfectamente el currículo, obteniendo el 100% en una evaluación referida a criterios. Una evaluación estandarizada todavía debe apuntar a crear una propagación a través de los niveles de rendimiento; por lo que en tal caso se debe "elevar el listón" presumiblemente más allá del currículo para ser capaz de crear un conjunto claro de "ganadores" y "perdedores". Green pondera dicho enfoque como ético. Uno puede preguntarse igualmente simplemente cómo de eficaz es este enfoque y el resultado para los docentes, un sistema escolar y toda la sociedad. ¿Cómo reaccionan los diseñadores curriculares, quienes elaboran los textos y materiales, las instituciones y servicios de formación de docentes, los planificadores o incluso la sociedad más amplia a esta información, que, básicamente, desafía el currículo, y por lo tanto la visión oficial de una sociedad científica, con el fin de generar una gama aceptablemente amplia de desempeño? ¿Cómo sería la rendición de cuentas en dicha situación para las instituciones y sus agentes?

²⁴ - Para más información sobre la curva de campana, véase Fendler y Muzzafar (2008).

²⁵ - Kamenetz (2015) señala en su entrevista en American Radioworks una diferencia de 30\$ y 75\$ por estudiante para, respectivamente, las pruebas puntuadas por máquinas y las pruebas puntuadas por personas en los Estados Unidos.

²⁶ - Stecher y Chun (2001: 13).

Unidad del Currículo del Ministerio de Educación apuntan que los docentes reducen el currículo oficial previsto en muchas aulas porque optan por concentrarse únicamente o especialmente en el "programa de evaluación" proporcionado a comienzos del año escolar por el Consejo de Exámenes de Botsuana²⁷. Como se mencionó anteriormente, una separación estructural de estos dos ámbitos que, en realidad, deberían estar unidos, los deja debilitados y en riesgo de disminuir todo el esfuerzo educativo.

El segundo fracaso se refiere a la generación de información precisa, deliberada y estratégica que ayude a los sistemas, las escuelas y los educadores a mejorar el nivel de aprendizaje de los estudiantes y, por tanto, su rendimiento en las evaluaciones futuras. Tal resultado es claramente una prioridad para los sistemas. Sin embargo, por alguna razón, tienden a tratar a las evaluaciones principalmente como una alarma, lo que da lugar a que los titulares de los medios señalen el éxito o el fracaso general, y no se vea como una herramienta de diagnóstico perfeccionada que puede señalar las esferas precisas con lagunas de aprendizaje y, por tanto, soluciones precisas.

Un par de enfoques amplios podría servir para tratar las evaluaciones de una manera más de diagnóstico. Uno consiste en proporcionar a las escuelas, los docentes y los agentes educativos que los forman y apoyan un análisis exhaustivo de las respuestas de los estudiantes en ítems específicos de una prueba, lo que les permitiría ajustar el diseño de sus clases y la manera de impartirlas y el de sus propias evaluaciones para concentrarse en cualquier punto débil, usando evaluaciones sumativas con fines formativos. Esto es lo que Bhután ha intentado hacer con la ayuda de Educational Initiatives, una firma de India, al realizar un estado anual del aprendizaje de los estudiantes que ofrece a cada escuela un informe completo de las preguntas de la prueba que sus estudiantes respondieron bien y mal y que ofrece explicaciones para cada una de por qué la mayoría de los estudiantes se equivocaron en las preguntas²⁸. En Sudáfrica, un programa piloto está tratando de hacer lo mismo, para permitir a los docentes de los estudiantes examinados que refuercen su instrucción para la próxima promoción informando a los docentes que tendrán en sus clases a los estudiantes examinados de los puntos fuertes y débiles de sus nuevos estudiantes para que también ellos puedan ajustar su enseñanza²⁹.

Por supuesto, para que los docentes sean capaces de utilizar dicha información de manera eficaz para mejorar su instrucción, se deben cumplir algunas condiciones clave. Para empezar, el análisis debe ser claro y comprensible, debe ser lo suficientemente detallado y pertinente y también oportuno. Por el lado de la "demanda" de la ecuación, los docentes no sólo deben estar suficientemente capacitados y ser capaces de entender los datos y utilizarlos para influir en su enseñanza, sino que deben tener el tiempo y la autoridad necesarios para hacer ajustes a su instrucción, incluyendo aspectos tales como la estimulación, la secuenciación, los ejemplos y ejercicios que utilizan, y la suplementación, entre otros. También deben tener acceso a recursos adecuados y facilidades de instrucción para ajustar su enseñanza.

Tal diagnóstico detallado ofrece sistemas completos con información que pueden utilizar para proporcionar a los docentes recursos didácticos complementarios también. Esto es lo que Palestina

²⁷ - Comunicación personal, 23 de abril de 2015; y Muskin (próxima publicación, 2015).

²⁸ - EducationalInitiatives (2010); y Muskin (próxima publicación, 2015).

²⁹ - Kanjee y Sayed (2013); Kanjee, comunicación personal, 19 de febrero de 2015; y Muskin (próxima publicación, 2015).

ha hecho desde sus "impactantes" resultados del TIMSS en 2007³⁰. Además, los sistemas pueden optar por desarrollar y poner a disposición de los docentes ejemplo de pruebas e instrumentos de evaluación centrados sobre todo en los aspectos del aprendizaje en los que los estudiantes tuvieron más problemas en una evaluación formal previa, como en Australia³¹ y Nueva Zelanda³².

Como se indicó anteriormente, la naturaleza de la evaluación no se limita a pruebas de lápiz y papel ni a objetivos de aprendizaje estandarizados. Especialmente en lo que se refiere a la evaluación del razonamiento, el pensamiento crítico y las capacidades de aplicación de los estudiantes y, quizá sobre todo, de sus competencias personales, o cognitivas y no cognitivas, son posibles muchos otros tipos de evaluación y, se argumenta aquí, incluso se recomiendan. Como ha señalado el Ministerio de Educación de la provincia de Ontario (Canadá) en su guía oficial para la evaluación y presentación de informes en las escuelas de Ontario de los grados 1 a 12, tales alternativas podrían incluir "[...] observaciones formales e informales, debates, conversaciones de aprendizaje, cuestionamiento, conferencias, deberes, tareas en grupo, manifestaciones, proyectos, portafolios, procesos continuos de desarrollo, representaciones, autoevaluaciones y evaluaciones por compañeros, reflexiones, ensayos y pruebas³³". Si bien puede ser importante que haya normas claras incluso para este tipo de evaluaciones, un sistema (y sus asociados) puede desear considerar seriamente si es realmente importante tener datos científicamente puros y comparables sobre todas las dimensiones del aprendizaje. Extrapolando el argumento anterior, cuanto más comparables sean los datos sobre aspectos del aprendizaje que tienden a ser más idiosincrásicos, como cómo un estudiante muestra curiosidad, creatividad o confianza (o incluso cómo un docente nutre estos atributos en sus alumnos), menos precisa será la medida probablemente y menos útiles serán los resultados para el docente en el fortalecimiento del desarrollo de estas capacidades³⁴.

Por último, también es importante mencionar el uso incipiente pero creciente de las TIC (tecnologías de la información y la comunicación) para llevar a cabo la evaluación. Los desarrolladores y educadores defienden las ventajas del uso de la tecnología. Entre ellas destacan la capacidad de mejorar la logística, reducir el fraude y acelerar el análisis y la presentación de informes de los resultados de evaluaciones a gran escala. A nivel del aula, las TIC permiten a los docentes y estudiantes adaptar las pruebas a las necesidades, las capacidades y el ritmo particulares de una clase e, incluso, de cada estudiante (evaluaciones adaptativas) y para generar una retroinformación casi instantánea, no sólo indicando lo que un estudiante ha hecho bien o mal, sino proporcionando también orientación y materiales específicos para remediar los puntos débiles³⁵. Cada vez más, este tipo de soluciones no requiere de Internet, pero puede usar conectividad y dispositivos móviles³⁶.

³⁰ - Matar, Mohammed, correspondencia personal, 17 de enero de 2015; Muskin (próxima publicación, 2015).

³¹ - Cornell-Farrow, Sarah (2014); Muskin (próxima publicación, 2015).

³² - Brown, Gavin T. L. (2014).

³³ - Ministerio de Educación, Ontario (2010: 28).

³⁴ - Esta observación se refiere también a la observación sobre el trabajo de Sternberg, en la nota a pie de página 21.

³⁵ - Redecker (2013: 11) caracteriza la evaluación de adaptación basada en las TIC como "Generación 2,... adaptar la dificultad o el contenido de la siguiente [prueba] presentada o un aspecto de la temporización del siguiente orden del día sobre la base de las respuestas [individuales] de los examinados", cuyo método se remonta a comienzos del decenio de 2000. Véase también las Herramientas de evaluación de Nueva Zelanda para la enseñanza y el aprendizaje (asTTle), <http://e-asttle.tki.org.nz/> (4 de junio de 2015).

³⁶ - Véase Hill y Barber (2014); Brown (s.f.); Muskin (próxima publicación, 2015).

Los principales *efectos de la evaluación* recaen en tres grandes ámbitos de decisión. Uno se refiere a la política, el programa y la planificación de las que el *sistema educativo* es responsable. En este nivel, los resultados de la evaluación pueden desencadenar reacciones que van desde "arreglos" en todo el sistema o ajustes o reformas más sustanciales — por ejemplo, relacionadas con el currículo, el desarrollo profesional, el presupuesto y la asignación equitativa de los recursos — hasta iniciativas más específicas como pueden pertenecer a instituciones particulares, áreas temáticas o estrategias de evaluación, entre otros. El segundo ámbito se refiere a *decisiones sobre todos los estudiantes individuales*. Los resultados de una evaluación o combinación de evaluaciones bajo este pretexto, generalmente, determinan la trayectoria educativa de un estudiante, incluyendo aspectos tales como el nivel de estudios que alcanzará, las áreas de estudio que puede que siga si es capaz de continuar a un nivel de especialización, si será en una institución académica o de formación, la calidad de renombre de la institución en la que continuará su educación o formación, y otros aspectos (ciertamente, otros factores también son importantes en muchos ámbitos, como la situación socioeconómica, la etnia y la geografía). El tercer ámbito es el *aula*, proporcionando información que los docentes (y otros educadores) puedan utilizar (como se describe más arriba) para ajustar y fortalecer su diseño, entrega, evaluación y retroinformación de las clases y que los estudiantes (y sus padres) puedan utilizar para revisar y fortalecer sus estrategias, las condiciones y los resultados de aprendizaje; o tal vez abandonar la escuela y buscar otro tipo de educación.

El argumento central de este trabajo es que es el último de estos propósitos mejorar *el proceso y los resultados* del aprendizaje del estudiante, que deberían guiar en última instancia *todas las evaluaciones*, ya sea directa o indirectamente. Desafortunadamente, no suele ser el caso. En cambio, varios expertos internacionales de evaluación afirmaron claramente, aunque extraoficialmente, durante la investigación para este trabajo que muchas evaluaciones (por lo menos las sumativas a gran escala) suceden a menudo ya sea como una cuestión de prestigio, por razones políticas internas, bajo presión (explícita o implícita) de un asociado financiero, o para medir el logro de educación relativa de una jurisdicción frente a otras jurisdicciones. Incluso cuando un sistema lleva a cabo la evaluación con el claro propósito de "tomar la temperatura" de sus programas de educación, existen escasos ejemplos en todo el mundo de un seguimiento deliberado de diagnósticos y esfuerzos profundos para utilizar los resultados de manera deliberada y estratégica y así fortalecer la calidad de la impartición y los resultados de la educación del sistema³⁷ (la reacción de Palestina ante sus impactantes resultados del TIMSS en 2007 antes mencionada ofrece un ejemplar prometedor).

En muchos casos, incluso si un país quisiera utilizar los resultados de una evaluación a gran escala para tomar decisiones precisas, deliberadas y estratégicas para fortalecer la calidad de sus programas, la impartición y los resultados, el diseño, y quizá sobre todo, la gestión de muchos

³⁷ - Alemania es ampliamente conocida por ser el niño de cartel por una reacción a una evaluación a gran escala que llevó al cambio cualitativo importante de un sistema educativo, a raíz de sus "desastrosos" resultados de PISA en 2000 (véase Figazzolo [s.f.: 14]; y Neumann, Fischer y Kauertz [2010]). Sin embargo, la parte de esta historia que no aparece en la literatura parece ser el diagnóstico que en realidad Alemania llevó a cabo para identificar cómo debe reaccionar precisamente el sistema a la "fiebre" señalada por PISA para mejorar sus resultados en la siguiente ronda de la prueba. Lo que también parece digno de mención, sin embargo, es cómo, 15 años después, Alemania todavía aparece como el ejemplo preferido de una verdadera historia de "cambio de tendencia" de PISA; ¿dónde están los otros? En su lugar, lugares como Singapur, un habitual participante "de altos vuelos" en las principales clasificaciones internacionales de evaluación, según los informes, deciden reorientar sus programas nacionales de educación lejos de un enfoque miope en el éxito en los exámenes internacionales hacia otras prioridades, ya que "... el precio del alto rendimiento académico puede ser asombroso". Según Tan (2010: 53), no es, sin embargo, el costo financiero lo que preocupa al país, sino más bien los costos en cuanto a la pérdida de "creatividad y capacidades de pensamiento entre los estudiantes y miembros de la mano de obra".

programas de evaluación, o lo hacen difícil o excluyen la posibilidad de una reacción tan completa. El diseño de los elementos e instrumentos puede ser de un grado de sofisticación científica tan alto que puede ser difícil utilizar los elementos individuales a efectos de información y remediación³⁸. Además, como presupone Green, los elementos de pruebas reales pueden al final tener poco que ver con los objetivos de aprendizaje clave del currículo.

De un carácter más prosaico, las instituciones de evaluación son a menudo reacias, o se niegan de plano, a compartir sus instrumentos y elementos para fines de diagnóstico (o cualquier otro fin) para poder utilizarlos en versiones posteriores de sus pruebas. En los Estados Unidos, los docentes que supervisan la realización de la prueba deben incluso a veces firmar un documento jurídicamente vinculante de que se comprometen a no divulgar cualquier elemento del instrumento que administran bajo riesgo de enjuiciamiento³⁹. La creciente presencia de empresas privadas en el ámbito de la evaluación internacional y nacional podría parecer agravar esta situación, lo que preocupa no sólo por las futuras versiones de una prueba, sino también por los servicios y productos de efecto indirecto relacionados que venden y que están vinculados a una evaluación. Por último, lo que también perjudica a los sistemas en el uso de los resultados de evaluación para mejorar la calidad de la educación es el factor del tiempo. Por un lado, se encuentra el simple hecho de que los resultados y análisis de la prueba pueden llegar pasado el momento de aplicarla evaluación. Incluso aunque las condiciones de la época de la evaluación apenas hayan cambiado, no es cierto que aún prevalezcan los problemas indicados por los resultados; especialmente cuando los resultados no se comunican hasta tres, cuatro o incluso más años después de una prueba, como se ha atribuido al SACMEQ. Por otro lado, para aquellas evaluaciones a gran escala que se producen una vez cada varios años, un sistema también puede tener que esperar mucho tiempo, incluso para generar resultados que analizar. Esto puede ser especialmente problemático en el caso de que el sistema desee valorar la eficacia de alguna innovación o de otro tipo de intervención entre pruebas.

En definitiva, como se afirma más arriba, toda evaluación debe conducir *deliberada y estratégicamente* a una mejor enseñanza y aprendizaje. En cuanto a las políticas, los resultados y *su diagnóstico profundo* deberían conducir a decisiones claras y coherentes sobre qué enseñar, la mejor manera de enseñar y cómo movilizar y asignar recursos (en todos los sectores) para lograr una óptima calidad y la equidad en todo el sistema. Con respecto a la planificación, los resultados deben guiar a los sistemas en cualquier revisión del currículo y, sobre todo, de los insumos que los sistemas proporcionan directamente a los docentes y otros educadores de primera línea con el fin de traducir el currículo y la visión de educación más amplia en instrucción y aprendizaje en el aula. Esto afecta aspectos como el diseño y la distribución de libros de texto y otros materiales pedagógicos, las instalaciones de la escuela, la formación de docentes, la supervisión y el apoyo, los programas extracurriculares, el apoyo social a los estudiantes y sus familias y comunidades, y las normas, estrategias e instrumentos de evaluación. Por último, en la práctica, los resultados de evaluación generados por el sistema deben hacer básicamente lo que los propios docentes hacen con sus esfuerzos de evaluación formativa. Se trata de proporcionar información detallada pertinente sobre lo que sus estudiantes — ya sea de forma individual o en grupo — entienden y no entienden para que los docentes puedan tomar decisiones sobre los mejores métodos, materiales, contenidos, ritmo, equilibrio, revisión, calendario, gestión, evaluación, y así sucesivamente para incorporar en sus

³⁸ - Guadalupe, Cesar (comunicación personal, 30 de enero de 2015).

³⁹ - Green (2015).

clases y satisfacer las necesidades de aprendizaje de su clase y de los estudiantes individualmente. La evaluación cuesta mucho dinero y, quizás especialmente, tiempo y esfuerzo para despilfarrar en razones distintas que mejorar la calidad de la enseñanza y el aprendizaje.

III. Consecuencias para el currículo de las prioridades y las prácticas actuales en la evaluación del aprendizaje de los estudiantes

Como se ha apuntado, las consecuencias de la evaluación (y de la amonestación icónica de Drucker) para el currículo deben ser evidentes: si un sistema espera verdaderamente que se enseñe todo su currículo, entonces se debe probar todo el currículo, y esto debe ocurrir en formas que importan. Esta conclusión se refiere por igual a los contenidos del currículo, a la forma en que se enseña y a la capacidad de emplear el contenido para resolver problemas y desempeñar con eficacia y satisfacción en todas las esferas de la vida social y personal. También se refiere a la larga lista de competencias personales que la sociedad espera y necesita que el sistema educativo ayude a los niños y jóvenes a cultivar, también mencionado anteriormente.

Como corolario, si un sistema, o sociedad, solo valora verdaderamente y, por lo tanto, en realidad sólo examina partes del currículo, tal vez podría considerar eliminar del currículo aquellas partes que no considera dignas de evaluación. Por un lado, un funcionario de la educación podría percibir esta sugerencia como algo jocoso; las pruebas no deberían guiar el currículo. Por otro lado, dado que en la práctica las pruebas sí que guían gran parte, si no la mayoría, del currículo que se enseña y cómo, parecer que corresponde a los desarrolladores curriculares y las unidades de evaluación trabajar en estrecha colaboración para garantizar que los objetivos se alcanzan. Por desgracia, este tipo de colaboración es a menudo no sólo inexistente, sino que las estructuras de educación pueden suponer barreras funcionales. En un esfuerzo por mantener la integridad metodológica y política de las evaluaciones, la unidad responsable puede ser totalmente independiente del resto de departamentos ministeriales, incluso operar con autonomía parcial o total. Desde este punto de vista, las unidades de evaluación o exámenes operan para cumplir una especie de función de auditoría que no debe estar influenciada por el objeto de su escrutinio. La interpretación de esta autonomía como aislamiento institucional puede dar lugar a una evaluación sin sincronizar, o incluso contradictoria, con el resto del ministerio, con el departamento del currículo y cuyos objetivos sean quizás más vulnerables. Esto puede, a su vez, provocar confusión o conflicto dentro de los departamentos de educación de apoyo — por ejemplo, la formación de docentes, la inspección, la planificación y el desarrollo de materiales — cuando se enfrentan a contradicciones entre lo que está probado y lo que el sistema profesa como su programa oficial de aprendizaje, es decir, el currículo.

Por lo tanto, merece la pena tener en cuenta qué vías de comunicación y cooperación podrían ser valiosas e, incluso, de vital importancia establecer entre una unidad de evaluación y los departamentos curriculares (y otros) de un ministerio de educación. Además de ayudar a asegurar que lo que un sistema establece como metas de aprendizaje para los estudiantes sea lo que sus programas de evaluación miden realmente, la comunicación de los resultados de la evaluación de nuevo y en el sistema promoverá los muchos vínculos necesarios para mejorar la calidad de la enseñanza y el aprendizaje, como se ha desarrollado en la sección anterior. Además, ir más allá de cualquier inferencia de jocosidad del corolario anterior, un análisis de lo que desde el currículo

verdaderamente garantiza la evaluación de hecho puede dar lugar a preguntas acerca de qué contenidos y objetivos son verdaderamente esenciales o incluso sólo interesantes y útiles y cuáles han superado su "tiempo de conservación" y ya no son pertinente, o tal vez nunca lo fueron.

En la conversación con otras autoridades y asociados educativos y del currículo, así como con docentes de muchos países, especialmente de las regiones de bajos ingresos del mundo, se está de acuerdo en que los currículos a menudo están inflados y son difíciles de manejar. Esto, muchos también están de acuerdo, a menudo resulta de un enfoque aditivo para el desarrollo del currículo, simplemente introduciendo nuevos objetivos de aprendizaje, contenidos y estrategias (por ejemplo, métodos basados en competencias o la ciudadanía como asignatura) en un programa académico sin eliminar o, al menos, racionalizando, otros elementos. Esta adición puede ser útil si se completa como parte planeada de una reforma curricular, o puede ser *de facto*, con un nuevo contenido agregado sin dejar de utilizar los antiguos libros de texto y criterios de evaluación (para estudiantes o docentes) y no dar a los docentes una orientación significativa con la que abordar contradicciones inherentes. Sobre todo cuando se combina con el fenómeno global del aumento del tiempo dedicado a la evaluación y el mayor impacto asociado a los resultados, en especial para las escuelas, el conflicto entre el contenido y la impartición de un currículo y la evaluación de su cumplimiento parecerían una fórmula para empeorar la calidad de la educación, en vez de mejorarla.

Tal conclusión provocativa busca justificación en al menos dos argumentos. Uno de ellas es que un currículo con "esteroides" — es decir, excesivamente voluminoso — deja poco tiempo para la evaluación formativa y la retroinformación, las cuales, como se muestra más arriba, constituyen dos elementos vitales del aprendizaje. Cuando los docentes hace un sprint en el programa para cubrir todo el material que el sistema requiere que los estudiantes aprendan en un año, hay muy pocas oportunidades de hacer un balance de lo que están reteniendo y lo que no. Y cuando, inevitablemente, surgen lagunas los docentes a menudo sienten que no pueden hacer una pausa para revisar el contenido y las técnicas que se requiere a sus estudiantes para pasar con éxito a las lecciones posteriores. El requisito de que las escuelas preparen a los estudiantes para las evaluaciones estandarizadas de gran impacto directo, ya sean nacionales o internacionales, roba a los docentes y estudiantes aún más tiempo para dedicar al currículo y a una evaluación formativa y una retroinformación valiosas, así como a veces da lugar a un secuestro virtual del contenido y el proceso de instrucción.

Este último peligro reside en la práctica de enseñar para la prueba. Un docente, y de hecho un sistema, pueden creer realmente que al preparar a los estudiantes para tener éxito en la prueba, los estudiantes están aprendiendo mejor. Sin embargo, esta lógica no es cierta y, de hecho, lo contrario sí puede serlo. Por un lado, enseñar para la prueba, especialmente en contextos de alta responsabilidad, como los Estados Unidos, a menudo implica apuntar al suelo del aprendizaje, o los requisitos mínimos básicos para el cumplimiento de los estándares oficiales, en vez de apuntar al techo, el tejado o incluso el cielo. Cuando un sistema establece su listón para el éxito en la mecánica de la lectura — reconocimiento de la letra, conciencia del fonema, sintaxis, vocabulario, prosa expositiva, etc. —, que es donde muchos de los docentes, si no la mayoría, se detendrán en su instrucción. De hecho, aquí es donde un sistema puede requerir a un docente que se detenga, según han comentado anecdóticamente algunos docentes de los Estados Unidos⁴⁰.

⁴⁰ - Véase Muskin (próxima publicación, 2015).

Algunos líderes de la evaluación sugieren que enseñar para la prueba es aceptable si la prueba es buena. Tal vez sea una posición legítima en el mundo ideal, pero las limitaciones de tiempo, la capacidad de procesar, difundir y utilizar los resultados y otros factores sugieren que tal ideal es manifiestamente difícil de alcanzar⁴¹. Además de reducir el contenido que cubre un docente, ya sea debido a los limitados niveles de evaluación o al tiempo dedicado a la preparación e impartición de las clases, un enfoque sobre el "programa evaluado" puede tanto desmotivar a los estudiantes como socavar su aprendizaje. Centrándose en la gramática y el vocabulario en vez de en la lectura y la escritura de información, el estímulo y placer intelectual o en ecuaciones y fórmulas matemáticas en lugar de en el uso de las matemáticas para resolver problemas apremiantes y contestar preguntas estimulantes compromete el aprendizaje en al menos dos niveles. Por un lado, reduce el entusiasmo del estudiante para estudiar y aprender, convirtiendo la curiosidad y la ambición en rutina y memoria. A un nivel más cognitivo, priva a los estudiantes de la oportunidad de profundizar y consolidar su comprensión de un concepto, contenido o técnica en particular mediante la exclusión de oportunidades para explorar el tema desde una variedad de ángulos y utilizarlo de manera pertinente⁴².

Este mismo razonamiento se refiere al segundo argumento referente a cómo el conflicto entre la evaluación y el currículo disminuye el aprendizaje. Esto es que un currículo abultado y un énfasis en la evaluación para la rendición de cuentas deja poco tiempo y genera poco impulso para el desarrollo de las competencias personales vitales de los estudiantes, como se describió y debatió anteriormente, independientemente de la prioridad otorgada a estas en la visión nacional de educación. Aprender a leer y escribir, rendir en matemáticas o demostrar el dominio de cualquier otro tema para una evaluación estandarizada deja poco tiempo o energía para fomentar las competencias cognitivas y no cognitivas más amplias de los estudiantes, ya sea en actividades extracurriculares, como es más común, o en esfuerzos académicos. Una lección de historia sobre la Segunda Guerra Mundial, que requiere que los estudiantes lean y retengan hechos relacionados y capten interpretaciones basadas en textos es mucho más efectivo para cubrir el contenido de lo que sería un proyecto de grupo que involucra a los estudiantes en la planificación y la realización de una exploración del tema consultando documentos y realizando entrevistas sobre la experiencia y los efectos de la guerra en su propia comunidad y elaborando y compartiendo una presentación multimedia sobre el tema. Además de desaprovechar la oportunidad de ayudar a los estudiantes a descubrir y desarrollar sus talentos personales y otras capacidades del siglo XXI, desaprovechan la oportunidad de aprender el contenido mejor en al menos dos formas. En primer lugar, como se concluyó en el párrafo anterior, rechazan posibilidades de profundizar y consolidar su aprendizaje al observar la materia en gran medida de memoria, en una sola dimensión. En segundo lugar, las mismas competencias que un enfoque de la educación más sólido promovería y perfeccionaría para un rendimiento más productivo en la vida "real" son aquellos atributos que preparan y motivan a los estudiantes para mejorar en sus esfuerzos académicos: la confianza, la curiosidad, la perseverancia, la ambición, la planificación, la autoevaluación, y así sucesivamente. Así, mientras que el aprendizaje basado en textos puede ser más eficiente, podría parecer que está lejos de ser más eficaz.

Con una vista de pájaro de la dinámica entre el currículo y la evaluación, la pregunta esencial parece ser qué combinación de medidas y estrategias de medición permitirían a un sistema utilizar la

⁴¹ - Véase Kamenetz (2015), *The Test: Why Our Schools are Obsessed with Standardized Testing—But You Don't Have to Be*; and listen to a podcast interview with the author at <http://www.americanradioworks.org/the-test/> (25 de mayo de 2015).

⁴² - Véase Pellegrino y Hilton (2012); Bloom, Engelhart, Furst, Hill y Krathwohl (1956); Gardner (2006).

evaluación igualmente para la rendición de cuentas, la certificación o selección y los fines pedagógicos para optimizar la impartición y resultados de la educación. En el análisis anterior, el objetivo ha sido destacar al menos dos direcciones estrechamente relacionadas para encontrar una respuesta adecuada. Una consiste en establecer como una prioridad estratégica concreta el uso de evaluaciones estandarizadas y sus resultados para orientar las decisiones y acciones en todos los niveles del sistema educativo — políticas, planificación y práctica. La evaluación para la rendición de cuentas y la certificación y selección son objetivos ciertamente valiosos; pero estos propósitos no deberían sustituir o abrumar la aplicación de la evaluación para generar resultados y análisis que ayudarán a los actores e instituciones en todo el sistema educativo y, *lo más importante, a los docentes en el aula*, para mejorar la instrucción y el aprendizaje en el aula.

La segunda dirección también apunta a priorizar mejores resultados pedagógicos mientras se busca en el camino un equilibrio constructivo entre los tres objetivos principales de la evaluación y entre las diversas naturalezas, o métodos, para medir el aprendizaje. Como se ha argumentado, la rendición de cuentas y la certificación y selección no están en desacuerdo con la mejora pedagógica. Más bien, tienen mucho valor que ofrecer y, a su vez, con la actuación pedagógica fortalecida, los resultados las evaluaciones de la rendición de cuentas y de la certificación y selección deberían mejorar, por la aclamación del sistema educativo y el beneficio de la sociedad y, por supuesto, de sus niños y jóvenes. Parcialmente implícita en el análisis precedente, pero no menos importante, es la idea de que los sistemas educativos pueden beneficiarse más si colocan un mayor énfasis en la importancia y tiempo en la evaluación realizada en el aula. Por un lado, las evaluaciones que los docentes crean y dirigen pueden ser más precisas, matizadas, completas y multidimensionales y más fiables que una sola prueba estandarizada o una serie de pruebas. Esto se debe a que los docentes pueden: a) vincular su evaluación directamente a donde se encuentre una clase en el currículo en momento dado; b) utilizar un conjunto diverso de métodos de evaluación para percibir y medir las diversas formas en que los estudiantes pueden demostrar su comprensión de un tema o técnica; c) crear condiciones más propicias para tener la confianza de los estudiantes y, por lo tanto, poder mostrar lo que realmente saben y lo que no; d) evaluar las respuestas de las pruebas de cada estudiante con la capacidad de interpretar e identificar dónde la comprensión se puede encontrar tras una respuesta equivocada; e) dar forma a una evaluación al nivel y necesidades de cada estudiante, y muchas otras razones en particular. Técnicamente, las pruebas estandarizadas pueden imitar muchas de estas mismas ventajas, pero hacerlo probablemente implicaría costos exorbitantes para el presupuesto y la logística, aunque la llegada de las estrategias de las TIC para la evaluación podría mitigar estos obstáculos.

Es importante añadir aquí, sin embargo, que esta recomendación no implica un abandono total de las evaluaciones estandarizadas convencionales. Por el contrario, estas son y seguirán siendo muy importantes. Más bien, la proposición es que haya un mayor equilibrio entre los instrumentos estandarizados externos y esas evaluaciones, tanto sumativas como formativas, que los docentes llevan a cabo. Además, como se argumentó anteriormente, existe un gran potencial y un número creciente de ejemplos de evaluaciones estandarizadas que sirven para proporcionar a los docentes elementos, instrumentos y estrategias mejorados y flexibles que pueden utilizar con sus alumnos para evaluar y reforzar su aprendizaje.

La otra ventaja clave para el aprendizaje que la aplicación de la evaluación en el aula bajo el control del docente de los métodos de medición y los resultados puede producir es la capacidad de analizar y

utilizar los resultados de una manera precisa, oportuna, específica y variada. Los docentes pueden tomar decisiones acerca de sus clases a mitad de la impartición, preparándose para las clases del día siguiente o durante un período más largo en reacción a las indicaciones instantáneas de las evaluaciones en el momento, a partir de tareas calificadas o de la información que reciben sobre los resultados de sus estudiantes o de una muestra de otros estudiantes en pruebas externas.

Por supuesto, ser capaz de formular y llevar a cabo las evaluaciones y analizar y utilizar sus resultados en estas formas altamente constructivas requieren una gran capacidad por parte de los docentes y otros educadores de primera línea. Dicha capacidad está apenas garantizada en muchos entornos, si no en la mayoría, y sobre todo en los entornos de países de ingresos bajos y medios, como muchos expertos en el campo de la evaluación se apresuran a afirmar. Más bien, se encuentran con que las capacidades requeridas para un enfoque robusto e informativo de la evaluación superen en gran medida a las de la gran mayoría de docentes. Aun suponiendo que se trate de una afirmación legítima, ¿debería justificar la marginación continua de los docentes en la aplicación de la evaluación y su privación rutinaria de información oportuna y significativa sobre los resultados de las pruebas que los docentes pueden utilizar para reforzar su enseñanza y el aprendizaje de sus alumnos? O más bien, ¿debería el reconocimiento de la baja capacidad en la evaluación conducir a una investigación exhaustiva de las causas de esta situación con el fin de elaborar y aplicar un conjunto de medidas para elevar esta capacidad y, con ello, elevar la calidad de la enseñanza y el aprendizaje? Tal actitud podría ser particularmente afín dado que la cuestión de la "capacidad", como se ha explicado en la sección final del informe, no se refiere solo a los docentes, sino también a muchos otros factores que son controlados por el sistema.

IV. Marco para la armonización de la evaluación del aprendizaje del estudiante con el currículo dentro de la agenda mundial de educación hacia 2030

Una revisión de lo que un sistema necesita evaluar y, por lo tanto, qué capacidad se necesita para llevar a cabo la evaluación y la gestión de sus resultados puede comenzar apropiadamente con una declaración clara del currículo que un sistema verdaderamente quiere y que una sociedad realmente necesita que sus niños y jóvenes aprendan *y, finalmente, al que sean capaces de dar un buen uso*. Canalizando el mensaje central del informe a la UNESCO de la Comisión presidida por Delors de 1996, ¿qué requiere la sociedad que sus graduados de la escuela aprendan a conocer, hacer, estar juntos y ser? Haciendo referencia a la declaración conjunta del Foro Mundial sobre la Educación que se acaba de terminar en Incheon (República de Corea), las expectativas para el aprendizaje están aún más empapadas que antes en las competencias que los estudiantes deben poseer para conseguir el éxito y la satisfacción en todas las esferas de la vida y los medios de subsistencia:

"La educación de calidad fomenta la creatividad y el conocimiento, garantiza la adquisición de las competencias básicas de lectura, escritura y cálculo, así como de aptitudes analíticas, de solución de problemas y otras habilidades cognitivas, interpersonales y sociales de alto nivel. Además, la educación de calidad propicia el desarrollo de las competencias, los valores y las actitudes que permiten a los ciudadanos llevar vidas saludables y plenas, tomar decisiones con conocimiento de causa y responder a los desafíos locales y mundiales mediante la educación para el desarrollo

sostenible (ESD) y la educación para la ciudadanía mundial (ECM)" (párrafo 9⁴³; la cursiva es nuestra).

Ciñendo todos los compromisos a una "perspectiva de aprendizaje a lo largo de toda la vida que abarque a todo el sector y aborde las cuestiones de acceso, resultados, equidad y calidad de la enseñanza para todos — niños, jóvenes y adultos — desde la atención y educación de la primera infancia hasta la enseñanza superior y el aprendizaje de adultos, en contextos educativos formales, no formales e informales⁴⁴", es una obligación fundamental proporcionar a todos una educación que importa. Lo que más importa en el aprendizaje de cara al futuro también será sin duda uno de los temas acuciantes de las discusiones, debates, negociaciones y acuerdos que se producirán en la Cumbre Mundial sobre el Desarrollo Sostenible de las Naciones Unidas en la ciudad de Nueva York del 25 al 27 de septiembre de 2015. Allí, las naciones del mundo se reunirán para adoptar y comprometerse a una serie de objetivos, metas y acciones generales que tienen por objeto la eliminación de la pobreza mundial, el conflicto y la desigualdad y la consecución de poblaciones educadas, saludables, seguras y productivas y de un planeta sostenible y saludable.

La primera parte de esta formulación es el tema central de un documento que la Oficina Internacional de Educación de la UNESCO preparó para la Conferencia de Incheon, cuyo propósito se captura bien en el título: Reubicar y reconceptualizar el currículo para el eficaz cumplimiento del Objetivo de Desarrollo Sostenible N° 4 para un desarrollo integral y unos medios de vida sostenibles⁴⁵. Al abordar en particular el aspecto de la evaluación formal de esta ecuación en el presente informe, podría parecer más crítico considerar tres objetivos operativos:

- i. confirmar que los conocimientos y capacidades deseados y requeridos (idealmente, el currículo) se han aprendido no sólo como información para la retención a corto plazo y la representación en una prueba, sino como un conjunto coherente de conocimientos y competencias que el estudiante puede invocar y utilizar para navegar con eficacia y satisfacción su camino en la vida al tiempo que contribuye, tanto activa como pasivamente, a la consecución de los ODS (evaluación del aprendizaje, o rendición de cuentas, pero por mejorar la enseñanza y el aprendizaje y contribuir a los objetivos de la sociedad, no por vergüenza o aclamación);
- ii. contribuir de manera constructiva al proceso de aprendizaje, proporcionando retroinformación vital para los docentes y los estudiantes (evaluación para el aprendizaje) y sirviendo como herramienta adicional de instrucción (evaluación como aprendizaje⁴⁶);
- iii. ofrecer las estrategias, condiciones e incentivos "correctos" para motivar, preparar y guiar a los docentes y otros educadores para enseñar el currículo completo y toda la gama de objetivos y competencias de aprendizaje, motivar a los estudiantes a aprender a conocer, hacer, estar juntos y ser en un mundo dinámico de ODS (objetivos de desarrollo sostenible) en lugar de reducir el currículo principalmente para "enseñar para la prueba" y, por lo tanto, crear alumnos que aprenden de memoria y consumidores de información con otras competencias limitadas.

⁴³ - Texto completo de la Declaración de Incheon, "Educación 2030: Hacia una educación inclusiva y equitativa de calidad y un aprendizaje a lo largo de la vida para todos", está disponible en <http://es.unesco.org/world-education-forum-2015/about-forum/declaracion-de-incheon> (4 de junio de 2015).

⁴⁴ - De la página web del Foro Mundial sobre la Educación 2015 de la UNESCO: <http://www.unesco.org/new/es/education/themes/leading-the-international-agenda/education-for-all/world-education-forum-2015/> (26 de mayo de 2015).

⁴⁵ - El borrador está disponible en: http://www.ibe.unesco.org/fileadmin/user_upload/Publications/drafts/WEF_IBE_Position_Paper_eng.pdf (8 de junio de 2015).

⁴⁶ - Véase Ministerio de Educación, Ontario (2010) y Black y Wiliam (2001).

La consecución de estos objetivos requerirá una reflexión, investigación y negociación significativas y una reforma de la evaluación en la mayoría de los sistemas y regímenes educativos, si no prácticamente en todos. Además, la tendencia en muchos lugares parece estar desgraciadamente lejos de los tipos de decisiones que los sistemas deben tener en cuenta para evaluar para un mundo de ODS. La historia tras el último libro de Anya Kamenetz sobre pruebas en Estados Unidos ilustra este punto sin rodeos: aunque se dispuso a investigar y escribir sobre algunos de los más emocionantes modelos pedagógicos del país — por ejemplo, el aprendizaje basado en proyectos, el aprendizaje semipresencial y *makerspaces*—, tuvo que cambiar de rumbo, porque lo que se encontró en la realidad fue que estas innovaciones, en realidad, son raramente utilizadas y exprimidas, afirma, por las presiones de las pruebas estandarizadas. Tal vez los movimientos de base incipientes pero aparentemente crecientes que se oponen a la tiranía de las pruebas marcan una reversión de esta tendencia. También se podría esperar que surjan nuevos puntos de referencia y compromisos mundiales del Foro Mundial sobre la Educación en Incheon y que los Objetivos de Desarrollo Sostenible estimulen un intenso escrutinio y reformas significativas de los currículos y de las normas y estrategias de evaluación por los países y las organizaciones internacionales por igual.

Anticipar ese cambio está en la raíz de lo que la iniciativa que la Comisión especial sobre métricas de los aprendizajes se comprometió a hacer desde 2012⁴⁷. Gestionado por el Instituto de Estadísticas de la UNESCO y la Brookings Institution, un centro de estudios con sede en los EE.UU. y asesorado y apoyado por muchas de las principales instituciones bilaterales y multilaterales internacionales de educación y organizaciones prominentes de la sociedad civil e investigadores, la Comisión especial sobre métricas de los aprendizajes primero llevó a cabo una consulta global masiva⁴⁸ para identificar un conjunto básico de ámbitos de aprendizaje y esferas de medición⁴⁹ hacia el que todos los sistemas educativos pueden educar a sus niños y jóvenes. Con estas listas, el reto se ha orientado hacia la identificación de los indicadores y estrategias para cada ámbito y esfera por la que medir el grado de cumplimiento, evaluando por igual a los estudiantes, instituciones y sistemas educativos. Para lograrlo, la Comisión especial sobre métricas de los aprendizajes se ha asociado con 15 países y otras jurisdicciones en cuatro continentes⁵⁰ — apodados los "campeones del aprendizaje" — para desarrollar, experimentar y validar una serie de métodos de evaluación y protocolos que podrían

⁴⁷ - Otras iniciativas internacionales para centrarse en la promoción de una mejor y más útil la evaluación del aprendizaje y la calidad de la educación incluye el nuevo catálogo de evaluaciones del aprendizaje del UNESCO Institute for Statistics (véase

http://www.uis.unesco.org/nada/en/index.php/catalogue/learning_assessments), el enfoque de los sistemas para unos mejores resultados educativos del Banco Mundial (SABER - véase: <http://saber.worldbank.org/index.cfm>) y el inminente programa de evaluación para el aprendizaje de la Alianza Mundial para la Educación.

⁴⁸ - En la consulta para la primera fase de la Comisión especial sobre métricas de los aprendizajes participaron más de 500 personas, entre ellas docentes, ministros y funcionarios ministeriales, representantes de la sociedad civil, académicos e investigadores, procedentes de 57 países en seis continentes.

⁴⁹ - Los siete ámbitos de aprendizaje son: 1) el bienestar físico; 2) las competencias sociales y emocionales; 3) la cultura y las artes; 4) la alfabetización y la comunicación; 5) los enfoques del aprendizaje y la cognición; 6) la aritmética y las matemáticas; y 7) la ciencia y la tecnología. Las siete esferas de medición son: 1) el acceso y terminación de las oportunidades de aprendizaje; 2) la exposición a una amplitud de oportunidades de aprendizaje a través de los siete ámbitos; 3) las experiencias de la primera infancia que promueven el desarrollo y el aprendizaje en varios dominios; 4) la capacidad de leer; 5) la capacidad de utilizar los números y aplicar este conocimiento a situaciones de la vida real; 6) una capacidad adaptable y flexible establecida para satisfacer las exigencias del siglo XXI; y 7) el indicador del "aprendizaje para todos", que mide el porcentaje de niños que han completado un ciclo de aprendizaje y han logrado resultados de aprendizaje adecuados. Véase <http://www.brookings.edu/about/centers/universal-education/learning-metrics-task-force-2> (5 de junio de 2015).

⁵⁰ - La lista de los campeones del aprendizaje comprende a Bogotá (Colombia), Botswana, Buenos Aires (Argentina), Etiopía, Kenya, Kirguistán, Nepal, Ontario (Canadá), Pakistán, Palestina, Rwanda, Senegal, Sudán, Túnez y Zambia.

servir de modelo o, al menos, de inspiración para que otros países los tomen prestados, los adapten o los emulen.

Una de las primeras lecciones de la iniciativa de los campeones del aprendizaje es que el logro de este objetivo requerirá prestar atención a algunos aspectos clave, una conclusión que refleja la experiencia relacionada de muchos otros sistemas a través de múltiples contextos en los últimos años. Fundamentalmente, estos aspectos implican en el primer momento un compromiso oficial y de la sociedad con una visión consensuada de un "graduado escolar educado" y un objetivo claramente articulado y un conjunto de objetivos estratégicos por los que pueda hacerse efectiva la visión. La esperanza ahora en 2015 es que cada país y jurisdicción haga referencia a la Declaración de Incheon y los ODS para examinar y bien confirmar, revisar o descartar y empezar de nuevo en el compromiso con una visión, metas y objetivos del siglo XXI para su sistema educativo, que está ajustado a sus prioridades culturales, sociales, económicas y ambientales particulares. El segundo aspecto clave es movilizar y coordinar todos los elementos de un sistema educativo para revisar y ajustar sus políticas, programas y acciones respectivas que promuevan visión renovada o nueva, metas y objetivos con una combinación de esfuerzos distintos y conjuntos. Esto se aplica igualmente a las dimensiones pedagógicas, materiales, políticas, administrativas, técnicas y contextuales de un sistema e incluye más precisamente elementos tales como la enseñanza en clase, el bienestar del estudiante, la planificación y gestión a nivel del sistema y de la escuela, estrategias y materiales pedagógicos y curriculares y la formación, la supervisión y el apoyo a los docentes.

El tercer y último aspecto clave es el de la evaluación, recordando de nuevo la advertencia de Drucker. En este sentido, los países y los actores internacionales desearán reconsiderar precisamente por qué están realizando las evaluaciones, si están evaluando lo que es realmente importante, para el aprendizaje y para la vida, y si la forma en que están evaluando está contribuyendo verdaderamente a la mejora de la enseñanza y el aprendizaje que fomenta los objetivos más amplios del currículo y la sociedad en general. Las dos primeras reflexiones se relacionan estrechamente con el primero de los dos elementos clave identificados anteriormente: la visión, las metas y los objetivos. ¿Está evaluando el sistema lo que la sociedad ha identificado como las competencias prioritarias para llevar a la sociedad hacia sus aspiraciones, ya sean económicas, sociales, éticas, ambientales o relacionadas con la dignidad y la realización personal? Como se ha comentado anteriormente, estas competencias deben abarcar por igual y juntos: a) la información y técnicas básicas, combinadas para constituir el conocimiento; b) el discernimiento y la capacidad técnica para aplicar estos (y otros) conocimientos, y *adquirir nuevos conocimientos*, resolver problemas y funcionar eficazmente en los papeles sociales, personales y económicos; y c) la amplia gama de competencias personales —cognitivas y no cognitivas— que proporcionan a un individuo con la ambición, las actitudes y comportamientos para reunir de manera efectiva todos sus atributos intelectuales, afectivos, sociales, morales e incluso físicos para participar en el mundo, producir y seguir aprendiendo de forma que se contribuya a los ODS y los objetivos más parroquiales de una comunidad de cualquier manera que se adapte a su condiciones, el contexto y el sentido propio de propósito.

V. Avanzar

Volviendo al tema central de este informe, qué preguntas deberían plantearse los sistemas educativos acerca de cómo examinar y revisar sus prioridades, programas y estrategias de evaluación cuando miran hacia el año 2030, el final de los nuevos ODS y la Educación para Todos. Este reto se refiere especialmente al tercer objetivo antes mencionado:

ofrecer las estrategias, condiciones e incentivos "correctos" para motivar, preparar y guiar a los docentes y otros educadores para enseñar el currículo completo y toda la gama de objetivos y competencias de aprendizaje, motivar a los estudiantes a aprender a conocer, hacer, estar juntos y ser en un mundo dinámico de ODS (objetivos de desarrollo sostenible) en lugar de reducir el currículo principalmente para "enseñar para la prueba" y, por lo tanto, crear alumnos que aprenden de memoria y consumidores de información con otras competencias limitadas.

También evoca a la referencia anterior a la capacidad para diseñar y realizar evaluaciones y analizar y utilizar los resultados de manera que no sólo señalen problemas, sino que realmente contribuyan a su diagnóstico y remedio; cuando un termómetro muestra que un paciente está febril, el médico no procede directamente al tratamiento sin completar primero un diagnóstico.

En los debates tanto de la literatura como de las conferencias, la preocupación acerca de la capacidad de evaluación por lo general se refiere a dos esferas. Una es el conocimiento, la capacidad y las herramientas de la unidad de evaluación, evaluación o exámenes de un ministerio. ¿Tiene el equipo de evaluación la capacidad técnica y la tecnología (hardware y software) para poder elaborar pruebas fiables, válidas y apropiadas y archivar, analizar y difundir los resultados de una manera coherente, comprensible y verdaderamente útil? Las otras se refieren a los conocimientos y la capacidad de los docentes y otros educadores locales para llevar a cabo una evaluación efectiva de los estudiantes y utilizar los resultados y los de las pruebas a gran escala para mejorar su enseñanza y el aprendizaje de sus alumnos. Por último, ¿qué pueden y tal vez deberían hacer los sistemas y sus asociados nacionales e internacionales para fortalecer la capacidad en estas dos esferas?

Estas dos esferas son ciertamente buenos puntos de partida para hacer frente a la capacidad de un sistema para evaluar el aprendizaje de los estudiantes y utilizar los resultados para mejorar la enseñanza y el rendimiento. Sin embargo, a pesar de ser una costumbre común, son un lugar imperfecto en que pararse cuando se aspira a garantizar el diseño más eficaz y efectos de evaluación productivos. La capacidad no se limita a las personas y departamentos directamente responsables de la evaluación y sus resultados; principalmente, la unidad de exámenes y las escuelas y los docentes. Más bien, la evaluación eficaz — es decir, la evaluación que genera información precisa, fiable y completa sobre el desempeño de los estudiantes y, por tanto, de todo el sistema *y que proporciona datos y análisis que contribuyen a mejorar directamente la enseñanza y el aprendizaje en el aula y en todas las escuelas* — depende de una capacidad sistémica sólida, articular, fortalecer y coordinarlos intereses y los efectos de todas las unidades ministeriales (e incluso de otros sectores) que se refieran a medir y mejorar el aprendizaje del estudiante. A nivel del sistema, una reflexión fructífera sobre este desafío podría hacer que los encargados de adoptar decisiones participen en la exploración de los siguientes tipos de preguntas:

- ¿Cómo de precisa y completa es la manera en que los exámenes a nivel del sistema y otras evaluaciones reflejan el currículo y otros resultados del aprendizaje importantes buscados por el

sistema y la sociedad en general? ¿Hasta qué punto deben éstos, de hecho, reflejar el currículo? ¿Cuáles son los efectos de las lagunas o las diferencias entre lo que se evalúa y lo que se prevé que se aprenda sobre qué y cómo enseñan los docentes y qué y cómo aprenden los estudiantes? ¿Son estos efectos un problema, y si es así, qué podría hacerse para mejorar la situación?

- ¿Qué contacto, cooperación y colaboración existe entre los exámenes del sistema o la unidad de evaluación y otros departamentos de educación con un impacto en el contenido, la planificación, la administración o los resultados de las pruebas a gran escala? ¿Cómo interactúa la unidad de evaluación con los expertos curriculares para identificar qué contenido cubrir y qué capacidades ofrecer al exigir a los estudiantes que demuestren sus conocimientos y habilidades relacionadas con el contenido? ¿Qué datos y análisis presenta la unidad de evaluación a los departamentos del currículo, la formación de docentes y la elaboración de materiales acerca de las esferas y la naturaleza de los puntos fuertes y las lagunas del aprendizaje de los estudiantes y a estos y el departamento de planificación sobre la ubicación de deficiencias graves? ¿Cómo consultan y coordinan los departamentos la elaboración de estrategias convergentes para mejorar la enseñanza y el aprendizaje que se basan en los resultados de las pruebas?
- ¿Qué datos vitales sobre la enseñanza y el aprendizaje y todo el conjunto de conocimientos y competencias a los que contribuyen estos procesos fundamentalmente NO ofrecen las evaluaciones dirigidas por el sistema? ¿Cuál de estas dimensiones del aprendizaje sería posible y apropiado integrar en estas evaluaciones a gran escala? A la inversa, ¿cuáles no? ¿Cómo pueden contribuir los diferentes departamentos de educación y actores en la primera línea de impartición de la educación — es decir, docentes, directores, agentes locales de educación e incluso padres y estudiantes — a la revisión de las evaluaciones a gran escala para incorporar estas otras dimensiones del aprendizaje? ¿Qué responsabilidades de la evaluación (y oportunidades —véase más adelante—) podría asignar un sistema a estas unidades y actores de la educación en primera línea, y tal vez sobre todo a las escuelas y los docentes, para proporcionar información precisa y útil sobre las competencias de los estudiantes que están mal evaluados o para los que hay muy poco tiempo para medirlos dentro de los instrumentos de gran escala? ¿Cómo podría un sistema combinar y equilibrar sus resultados de evaluación centrales y de nivel escolar para tomar decisiones sobre los estudiantes (certificación, selección, progresión), sobre el rendimiento del sistema (todas las dimensiones) y sobre las acciones que hay que tomar para fortalecer los resultados del aula?

A nivel de la escuela y el aula, la reflexión, decisiones y acciones productivas podrían surgir mientras las autoridades del sistema y los agentes técnicos se unen a los docentes, directores de escuela, agentes locales de educación e incluso padres y estudiantes para explorar otro conjunto de preguntas relacionadas, tales como:

- ¿Tienen los docentes y otros actores de la educación de primera línea los conocimientos y habilidades necesarios para formular y aplicar herramientas y estrategias de evaluación eficaces y pertinentes para utilizar con sus alumnos en el aula, para fines tanto formativos como sumativos? ¿Están suficientemente capacitados para utilizar los resultados de estas evaluaciones, así como los datos y análisis que puedan recibir de las pruebas del nivel del sistema para tomar decisiones estratégicas adecuadas sobre la planificación e impartición de las clases y orientar mejor el aprendizaje de los estudiantes individualmente? ¿Y los padres (e incluso los

estudiantes)? ¿Cuáles son los conocimientos y habilidades que necesitan para llevar a cabo estas funciones específicas y cuál es la mejor manera posible en que el sistema puede formarlos y apoyarlos para obtenerlos y utilizarlos?

- ¿Se entregan los datos y los análisis generados por la evaluación, especialmente de evaluaciones a gran escala, en una forma y un formato que sea comprensible y útil para los docentes, otros educadores locales y los padres y estudiantes? ¿Es la información lo suficientemente precisa, y apunta claramente a los contenidos y habilidades específicas, o representa medias amplias con poca o ninguna indicación de dónde han respondido correctamente los estudiantes y en qué aspectos de un tema han fallado? ¿Los datos y análisis presentan información que pertenece al currículo de una manera obvia? ¿Están distribuidos en forma oportuna, permitiendo a los docentes, otros educadores, padres y estudiantes reaccionar de una manera que permita remediar adecuadamente, asegurando que los estudiantes dominen al menos los conocimientos y capacidades fundamentales que se requieren a medida que avanzan en sus estudios y después en la vida "real"? ¿Reciben los docentes y las escuelas la información a tiempo para tomar decisiones y mejorar la instrucción de la siguiente promoción y para trabajar eficazmente con el grupo sometido a la prueba mientras continúa al siguiente grado?
- Suponiendo que un docente y otras personas no reciben información sobre los resultados de la evaluación del estudiante en una forma clara, exhaustiva y detallada, pertinente y a tiempo para usarla de manera significativa, y suponiendo que poseen el conocimiento y las capacidades necesarias para tomar decisiones estratégicas basadas en los datos, ¿se dan las condiciones que realmente les permiten hacer esto? ¿Hay suficiente espacio en el currículo o el tiempo en el día y el calendario escolares para ajustar el ritmo de la instrucción a la remediación? ¿Puede un docente dedicar una segunda clase sobre un tema para el que el currículo oficial solo prevé una hora si la evaluación muestra que los estudiantes no han comprendido el contenido relacionado? ¿Qué diría un inspector encontrase un docente que está sustituyendo el contenido del libro de texto con otros materiales para ayudar a los estudiantes a aprender mejor una lección que los resultados del examen del año pasado mostraron que sus alumnos no habían comprendido? ¿Pueden los docentes resumir el plan de estudios y así encontrar tiempo para llevar a cabo evaluaciones adecuadas de sus alumnos, ya sean del aprendizaje, para el aprendizaje o como aprendizaje?
- En el mismo sentido, ¿tienen los docentes el tiempo, los recursos y la autorización necesarios para llevar a cabo sus propias evaluaciones, tanto formativas como sumativas? ¿Caben la evaluación y la retroinformación en el plan de estudios y la jornada escolar? ¿El sistema proporciona modelos o elementos, herramientas, rúbricas, directrices y estrategias de evaluación reales que los docentes pueden consultar o utilizar, incluidos los elementos e instrumentos estandarizados? ¿Hay un suministro adecuado para llevar a cabo la evaluación? ¿El protocolo para determinar el aprendizaje acumulado de un estudiante para basar las decisiones acerca de la progresión, la certificación y la selección tiene en cuenta los resultados de las evaluaciones realizadas por los docentes? ¿Permite técnicas de evaluación más profundas y alternativas, como la revisión del portafolio, la evaluación de proyectos y las medidas de comportamiento y otras competencias personales?

La capacidad de crear y llevar a cabo las evaluaciones del aprendizaje y de analizar y utilizar los resultados para mejorar la enseñanza y el aprendizaje es fundamental para el éxito de un sistema y asegurarse de que los estudiantes salgan de la escuela con el conocimiento y las competencias que la sociedad les requiere. A medida que el mundo avanza más hacia el siglo XXI, la naturaleza de estos activos es cada vez más diversa, compleja e importante, como se refleja en los nuevos ODS y los compromisos de la EPT. El logro de esta capacidad y el uso de la evaluación no sólo como un termómetro, sino para contribuir a un diagnóstico rico, preciso y pertinente de lo que los estudiantes están y no están aprendiendo y cómo los sistemas pueden funcionar mejor para mejorar estos resultados requiere reflexión y acción por parte de todos los actores educativos. Estos esfuerzos conciernen a la evaluación, pero también a las otras dimensiones que afectan la calidad y los efectos de la educación, incluyendo aquellos controlados directamente por el sistema educativo y los que son responsabilidad de la sociedad en general. El objetivo de las preguntas anteriores es ayudar a los líderes del sistema educativo y otros actores a unirse a los representantes de diversos asociados del sistema para evaluar la evaluación (de todo tipo), sobre todo por su pertinencia, la minuciosidad de su cobertura y su verdadera utilidad en la toma de decisiones relativas a la mejora de la calidad en todos los niveles del sistema. Este diálogo debería afectar no sólo a qué, cómo y por qué las evaluaciones ocurren y cómo se combinan para contar una historia precisa y estratégica acerca de un sistema y cada una de sus unidades y el rendimiento de los alumnos, sino también a cómo todas estas dimensiones actúan, solas y juntas, para mejorar los resultados educativos en el futuro para acercarse a la visión que una sociedad establece para sí misma.

Bibliografía

- Barber, M. and Rizvi, S. 2013. *Asking More: The path to efficacy*. Pearson.
<http://efficacy.pearson.com/wp-content/uploads/2013/11/Asking-More-The-Path-to-Efficacy-high-res1.pdf> (Accessed 5 June 2015.)
- Benavot, A. 2012. *Primary School Curricula on Reading and Mathematics in Developing Countries*. Technical Paper No. 8, Montreal, UNESCO Institute for Statistics.
- Black, P. and Wiliam, D. 1998. Inside the Black Box: Raising Standards Through Classroom Assessment. <http://weaeducation.typepad.co.uk/files/blackbox-1.pdf> (Accessed 5 June 2015.)
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. 1956. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York, David McKay Co Inc.
- Brown, G. T. L. 2014. AsTTle: a national testing system for formative assessment: how the national testing policy ended up helping schools and teachers. in Lai, M. and Kushner, S. (ed.) *A Developmental and Negotiated Approach to School Self-Evaluation (Advances in Program Evaluation, Volume 14)* Emerald Group Publishing Limited, pp.39 – 56. <http://www.emeraldinsight.com/doi/abs/10.1108/S1474-7863%282013%290000014003> (Accessed 5 June 2015.)
- Cornell-Farrow, S. 17 November 2014. *Concerning Learning: a Review of My School and its Potential Applications to Improving Educational Standards in Developing Countries*. Report prepared for the Office of The Honourable Julia Gillard, former Prime Minister of Australia.
- Educational Initiatives. January 2010. *Student Progress Tracking System*. EI Working Paper Series – Issue 4. <http://www.ei-india.com/wp-content/uploads/2012/07/Student-Progress-Tracking-System-Issue-4.pdf> (Accessed 5 June 2015.)
- Fendler, L. and Muzaffar, I. 2008. The history of the bell curve: sorting and the idea of normal. *Educational Theory*, vol. 58, no. 1; pp. 63-82.
<http://fendler.wiki.educ.msu.edu/file/view/2008+Bell+Curve.pdf> (Accessed 5 June 2015.)
- Gardner, H. 2006. *Multiple Intelligences: New horizons in theory and practice*. Basic Books.
- Green, D. 31 May 2015. *How to Cheat on State Standardized Tests and Not Get Caught*. Education Week: Teacher.
http://blogs.edweek.org/teachers/work_in_progress/2015/05/how_to_cheat_on_state_standard.html (Accessed 5 June 2015.)
- Hill, P. and Barber, M. November, 2014. Assessment Renaissance. Seminar Series Paper No. 239, Centre for Strategic Education. <http://www.cse.edu.au/content/assessment-renaissance> (Accessed 5 June 2015.)
- Kamenetz, A. 2015. *The Test: Why Our Schools Are Obsessed With Standardized Testing – But You Don't Have To Be*. Philadelphia, PublicAffairs Books. (Listen also to the podcast, <http://www.americanradioworks.org/the-test/>) (Accessed 5 June 2015.)

- Kanjee, A. and Sayed, Y. 2013. Assessment policy in post-apartheid South Africa: challenges for improving education quality and learning. *Assessment in Education: Principles, Policy & Practice*, vol. 20, no. 4.
- Massachusetts Department of Elementary and Secondary Education. 2012. Massachusetts Commission to Develop an Index of Creative and Innovative Education in the Public Schools. <http://www.doe.mass.edu/research/reports/>(Accessed 9 January 2015.)
- Ministry of Education, Ontario. 2010. Growing Success: Assessment, evaluation and reporting in Ontario Schools – for Grades 1 – 12. <http://www.edu.gov.on.ca/eng/policyfunding/growsuccess.pdf> (Accessed 5 June 2015.)
- Muskin, J. Forthcoming. Using Student Learning Assessment Data to Improve Teaching and Learning (working title). Washington, DC, The Brookings Institution.
- Pellegrino, James W. and Hilton, Margaret L., Editors. 2012. Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. National Academy of Sciences. <http://www.nap.edu/catalog/13398/education-for-life-and-work-developing-transferable-knowledge-and-skills> (Accessed 5 June 2015.)
- People for Education. June, 2013. Broader Measures of Success: Measuring what matters in education. <http://www.peopleforeducation.ca/measuring-what-matters/wp-content/uploads/2013/10/P4E-MWM-full-report-2013.pdf> (Accessed 5 June 2015.)
- Redecker, C. 2013. The Use of ICT for the Assessment of Key Competences. A Joint Research Centre Science and Policy Report, European Commission, Luxembourg <http://ftp.jrc.es/EURdoc/JRC76971.pdf> (Accessed 4 June 2015.)
- Rogers, W. T. 2014. Improving the utility of large-scale assessments in Canada. *Canadian Journal of Education / Revue canadienne de l'éducation* 37:3.
- Stecher, B. and Chun, T. 2001. School and Classroom Practices During Two Years of Education Reform in Washington State. CSE Technical Report 550, National Center for Research of Evaluation, Standards and Student Testing. <https://www.cse.ucla.edu/products/reports/TR550.pdf>(Accessed 5 June 2015.)
- Sternberg, R. 2003. *Wisdom, Intelligence, and Creativity Synthesized*. Cambridge University Press.
- UNESCO. 2015a. Education 2030: Towards Inclusive and Equitable Quality Education and Lifelong Learning for All. <https://en.unesco.org/world-education-forum-2015/incheon-declaration>.
- UNESCO. 2015b. *EFAGlobal Monitoring Report: Education for All 2000-2015: Achievements and Challenges*. Paris, UNESCO. <http://unesdoc.unesco.org/images/0023/002322/232205e.pdf> (Accessed 25 June 2015.)