

LA INTERNACIONALIZACIÓN DE LA EVALUACIÓN DE LOS APRENDIZAJES EN LA EDUCACIÓN BÁSICA

EMILIA FERREIRO

LA EVALUACIÓN ES INTRÍNSECA AL FUNCIONAMIENTO DE UN SISTEMA educativo, el cual tiene la responsabilidad de garantizar que sus egresados tengan los conocimientos y las habilidades requeridas por el programa de formación. Siempre ha habido evaluación en el contexto educativo, a todos los niveles. Pero lo novedoso es la *internacionalización o globalización de las evaluaciones*. También son novedosas las *agencias evaluadoras*. La Organización para la Cooperación y el Desarrollo Económico (OCDE), cuyos orígenes tenían muy poco que ver con la educación, ha plegado a sus miembros –y a varios países no miembros– a una evaluación internacional, que ha tenido gran impacto, la cual trataré más adelante. Las autoridades locales (las secretarías o ministerios de educación de los diferentes países) hacen evaluaciones que están en gran medida “modeladas”, o cuando menos influenciadas, por los parámetros internacionales. América Latina no es una excepción.

A nivel de la educación superior hace años que resultan visibles los efectos benéficos, tanto como los desfavorables, de los múltiples tipos de evaluación que se aplican (evaluaciones de instituciones, de programas, de individuos, etcétera) y las variadas instituciones o grupos que funcionan como evaluadores. No es el propósito de este artículo hablar de eso. Por lo que toca a la educación básica la evaluación tiene, en México y en toda América Latina, mucho menos tradición que la evaluación universitaria. De eso trataremos en este artículo.

EVALUACIÓN DE APRENDIZAJES

La Conferencia Mundial realizada en Jomtien, Tailandia, en 1990, marcó un parteaguas en varios sentidos. Primero, porque el Banco Mundial firmó junto a UNESCO un compromiso expresado en la declaración de “Educación para Todos” y en metas a ser conseguidas antes del 2000; segundo, porque a partir de allí el enfoque tradicionalmente centrado en la enseñanza cambió hacia un enfoque en los aprendizajes.

Desde la reunión de Jomtien el Banco Mundial impulsa fuertemente reformas, acciones y proyectos educativos concretos destinados a modificar

la gestión de los sistemas educativos y a garantizar “productos medibles”, o sea, aprendizajes sujetos al escrutinio de pruebas cuyos resultados son expresados por números.

En 1990, en la región latinoamericana, “sólo Chile contaba con un sistema público e institucionalizado que recogía en forma recurrente y sistemática mediciones del rendimiento académico. Diez años después, (...) todos los sistemas educativos cuentan con sistemas institucionales sofisticados de este tipo” (Casassus, 2001)¹.

Actualmente, el interés está puesto en la adquisición de *competencias* más que en conocimientos, y en ese sentido las reformas curriculares tienden en todas partes, incluido México, a formularse en términos de competencias, sin que haya una definición clara de qué se entiende por ello, más allá de un conjunto –que se supone articulado– de conocimientos, habilidades y destrezas capaz de ser transferido a contextos diferentes de aquellos en los cuales se produjo el aprendizaje.²

¿Por qué este cambio de énfasis de programas centrados en contenidos a programas centrados en competencias? Esto tiene que ver con la rápida expansión del conocimiento que permiten los medios electrónicos y la acelerada multiplicación de artículos científicos o técnicos en circulación. Se estima que “la riqueza global de conocimiento acumulado se duplica actualmente cada cinco años” (Brunner, 2001)³. Las revistas científicas han pasado de unas 10,000 en 1900 a más de 100,000 al final del siglo XX.⁴ Es evidente que ningún sistema educativo –menos aún a nivel de la educación básica– puede ofrecer a sus alumnos un conocimiento actualizado. “Aprender a aprender” se ha convertido entonces en un *leitmotiv*.

Esta sobreproducción de conocimientos ha llevado a las grandes agencias (particularmente el Banco Mundial y la OCDE) a la convicción de que todos los conocimientos tienen fecha de caducidad, como productos del supermercado. Puede ser cierto del conocimiento tecnológico en ciertas áreas; no es cierto del conocimiento científico.

Darwin, Durkheim, Einstein y tantos otros continúan alimentando las discusiones científicas. Los grandes constructores de esquemas interpretativos de amplio alcance (que llamamos, en sentido estricto, “teorías científicas”)⁵ no son muchos y tienden a perdurar. Incluso pensadores anteriores al calendario cristiano tienen vigencia: Aristóteles es el punto de partida ineludible para hablar de lógica o retórica y Euclides sigue siendo referencia obligada para hablar de geometría. Entonces, no todo tiene fecha de caducidad, a pesar de la tendencia hacia el conocimiento “inmediatamente aplicable y rápidamente perimible” propiciado por las grandes agencias financiadoras.

Curiosamente, no es una reflexión epistemológica sino práctica la que produce una extraña coincidencia: buscan las competencias básicas, transferibles a distintos dominios, quienes apoyan una formación de base sólida, que garantice aprendizajes futuros, y también las buscan quienes están convencidos del conocimiento perimible y quieren garantizar que la gente siga aprendiendo (el *slogan* de “aprendizaje a lo largo de la vida” está incorporado a todos los discursos de las grandes agencias, lo cual no es ajeno a la inestabilidad del mercado laboral). En muchos discursos de estos últimos se trata al aprendizaje como si fuera “chips” que se cargan y descargan a voluntad, y como los seres humanos no venimos equipados con una tecla *delete*, algunos incluso hablan de la necesidad de entrenar *forgetting abilities*.

¿A QUIÉN SE EVALÚA?

Mientras los programas educativos enfatizan con progresivo énfasis las actividades grupales (discusión para la preparación de actividades y la toma de conciencia de los logros obtenidos) y subrayan el interés de obtener productos grupales (como preparar el periódico mural o una página web), la evaluación sigue siendo individual. Se evalúan los logros de aprendizaje de sujetos individuales. No se evalúa la oferta educativa como tal. Hay en esto

una profunda injusticia. Es como evaluar el efecto de cierta ingesta alimenticia, sin garantizar que dicha ingesta haya tenido lugar.

Evaluar la oferta educativa, o sea, evaluar las situaciones efectivas que favorecerían la obtención de una determinada competencia, supone, en principio, emitir dudas sobre la calidad de la intervención educativa, y nadie quiere hincarle el diente a ese problema, que se aborda de manera sumamente tangencial. Resultado: tenemos evaluaciones individuales hechas con toda independencia de la evaluación de las condiciones que permitirían acceder a dichas competencias. Esto es así, sin que ello involucre un pronunciamiento con respecto al contenido mismo de las evaluaciones, sobre cuya pertinencia y calidad es incluso posible coincidir, como veremos enseguida.

El aprendizaje es una relación tripartita entre un cierto contenido “objeto de aprendizaje”, un aprendiz y un adulto encargado de organizar las condiciones didácticas específicas para hacer posible ese aprendizaje. No corresponde decir *adulto encargado de transmitir el conocimiento* porque no hay ninguna “transmisión” directa. A lo sumo, el adulto puede “presentar” el contenido a ser aprendido, pero el aprendizaje es apropiación y la apropiación supone una actividad cognitiva muy diferente de la pura recepción. Los seres humanos no somos “portadores de conocimientos” como podemos ser portadores de determinados virus, sin saberlo y a pesar de nuestra voluntad.

Siendo una relación tripartita, la responsabilidad del aprendizaje no puede ser exclusivamente del alumno. Sin embargo, a partir de los resultados de aprendizaje se pueden emitir dudas acerca de la calidad de la oferta educativa a menos de suponer, por ejemplo, que todos los niños de determinado país, con resultados muy pobres, son igualmente “pobres de espíritu” (como se decía antes), infradotados (como se dijo en algún momento) o con débiles capacidades de aprendizaje, como se diría ahora. Es evidente que ningún país puede permitirse el desliz de afirmar que bajos

resultados de aprendizaje reflejan falta de capacidad de aprendizaje de su población infantil, porque eso supondría hipotecar su propio futuro (además de no ser “políticamente correcto” en época de atención preferencial a las minorías, del tipo que ellas sean).

Los únicos candidatos, pues, a ser señalados como responsables de los bajos resultados son, en primer lugar, la institución escolar en tanto institución encargada de producir aprendizajes y la familia –o los grupos sociales de pertenencia– caracterizados eventualmente como “deficitarios”. La visión empresarial vigente tiende a atribuir responsabilidad a la institución escolar, en lugar de diluirla en agentes mal identificados sobre los cuales no es posible incidir de manera directa. La institución escolar requiere asignación de recursos financieros y éstos pueden variar según los resultados que sus unidades exhiban.

¿QUIÉN EVALÚA?

Por supuesto, los sistemas educativos han practicado siempre la evaluación del aprendizaje de los alumnos. A menudo, con pruebas cambiantes de un periodo presidencial al otro y con resultados conservados bajo llave, para “insumo interno”.

Los cambios actuales son radicales: una agencia internacional organiza evaluaciones de tal manera que una misma prueba es administrada a alumnos de un número importante de países y los resultados se publican en la página web de dichas agencias, involucrando a la prensa escrita en la difusión de los resultados. Los países se ven entonces como participantes en una especie de *ranking* donde aparecen como mejores o peores que sus vecinos en términos de puntajes globales, o de puntajes desglosados: el sector público con respecto al privado, escuelas urbanas con respecto a rurales y muchas otras relaciones de este tipo.

La OCDE se ha convertido en una de estas grandes agencias de evaluación. La razón por la cual dicha agencia –lejana de la educación en sus

finos originales— aparece encabezando estudios en este campo ni siquiera requiere justificación: “Como respuesta a la necesidad de evidencias del desempeño estudiantil comparable internacionalmente, la OCDE introdujo el programa PISA (Programme for International Student Assessment)”.

PISA se aplica a jóvenes de 15 años, al final de la escolaridad obligatoria, cualquiera que sea el tipo de escuela secundaria al que asistan. “PISA se concentra en temas que los jóvenes de 15 años necesitarán en el futuro y busca evaluar lo que pueden hacer con lo que han aprendido”. Los contenidos corresponden a las áreas de lectura, matemáticas y ciencias.

Se trata de una prueba que se aplica en los mismos países cada tres años, con un enfoque dominante en cada aplicación: en 2000 el énfasis estuvo puesto en lectura; en 2003, en matemáticas; en 2006 estará puesto en ciencias; en 2009 nuevamente lectura, y así en adelante. El énfasis en cierta área se expresa en el hecho de que las dos terceras partes del tiempo de evaluación (2 horas) se dedican a esa área.

En la aplicación del año 2000 se involucró a 28 países de la OCDE —entre ellos México— y a otros 4 países, entre ellos Brasil. En 2002 se sumaron otros 13 países a la misma prueba, incluyendo a Chile y Perú.

En 2003 se sumaron otros 3 países, siendo Argentina uno de ellos. Por lo tanto, tenemos un total de 48 países (entre ellos 5 latinoamericanos) que participaron en lo que oficialmente se conoce como “PISA 2000”.

PISA no utiliza como término clave el de “competencias” sino el de “aptitudes”: “El término *aptitudes* se emplea para englobar los conocimientos y las competencias”. “La aptitud se mide conforme a un patrón continuo, no como algo con lo que un individuo cuenta o no”⁴.

A través de un largo proceso de consultas, traducciones y reformulaciones se construyeron situaciones de prueba que apuntan a evaluar la “aptitud” para el dominio en cuestión o las aptitudes que se consideran involucradas. Hay aquí un evidente

desliz en el modo de expresión. Se intenta evaluar o valorar una cierta aptitud a través de determinadas respuestas. Como se les pone un número a las distintas categorías de respuestas, y se realizan operaciones estadísticas con dichos números, se habla entonces de medición⁸.

No están disponibles todos los ítems utilizados para evaluar “la aptitud para lectura” (unos 140), pero los ejemplos disponibles son bastante ilustrativos como para imaginar el resto. En sus propias palabras “la aptitud para lectura se define en PISA como la capacidad de comprender, emplear y reflexionar sobre textos escritos”. Incluye textos en prosa de tipo narrativo, expositivo y argumentativo. Pero también exige interpretar y comparar gráficas y diagramas, lo cual es muy novedoso en este tipo de pruebas.

Los estudiantes deben comprender la información presentada y responder a algunas preguntas o tareas particularmente exigentes. Por ejemplo, se presenta un texto de una página a cuatro columnas, con diagramación propia de revista contemporánea (un texto rodeado por tres recuadros informativos, con diferente color de fondo, todo sin ilustraciones). El título del texto es “Armas científicas de la policía” y trata sobre el uso del análisis genético en la investigación criminológica. Los recuadros agregan información de tipo divulgación sobre ADN, la “credencial de identificación genética” y temas similares. La tarea del estudiante es leer con atención el texto y responder preguntas de comprensión acerca de su propósito y de la información específica presentada. Algunas respuestas requieren elegir entre cuatro opciones, bastante próximas entre sí, cuya diferencia requiere hacer una comparación de la información presentada en el texto principal con la que se encuentra en alguno de los recuadros.

Una lectura comparativa y crítica no es la que se suele practicar en el común de las escuelas. No es extraño que los países de América Latina se hayan ubicado en los últimos lugares de esta evaluación. México ocupó el último lugar en el primer grupo de países que participó en PISA 2000, pero luego Brasil se

agregó, ocupando el último lugar. Finalmente Perú desplazó a Brasil de esta última posición. En los tres países mencionados los resultados de PISA tuvieron repercusiones políticas. En México, el Instituto Nacional de Evaluación de la Educación (INEE) publicó varios documentos para ayudar a leer esos resultados y matizar el impacto de las gráficas comparativas (análisis disponibles en su página electrónica). En Brasil, la comisión de educación de la Cámara de Diputados organizó varias sesiones de debate alrededor del tema del retraso en lectura. Finalmente, Perú fue declarado en estado de “emergencia educativa” al darse a conocer los resultados que le conciernen, como si se tratara de una catástrofe producida por inundaciones o movimientos telúricos.

No es mi propósito discutir esos resultados, sino dar una idea del impacto que tienen estas evaluaciones. La pregunta de difícil respuesta es: ¿A qué se deben esos magros resultados? Vayamos al otro extremo de la escala. El gran ganador de este certamen comparativo fue Finlandia. Y los finlandeses respondieron a la inquietud de los otros países con expresiones muy honestas: “La razón más importante reside en los profesores de educación básica y en su capacitación. En Finlandia, la profesión docente es muy valorada. Son muchos los postulantes, podemos escoger a los mejores, seleccionamos apenas al 10% de los postulantes. En todas las encuestas los profesores de educación básica aparecen como teniendo el mismo prestigio social que los profesores universitarios, los médicos y los abogados”⁹. ¿Cuándo podremos decir algo así en América Latina?

OTRAS EVALUACIONES INTERNACIONALES

PISA evalúa expresamente comportamientos deseables, al nivel de un lector competente del mundo contemporáneo. No se preocupa por saber si esas competencias se han enseñado o no. Más bien, trata de incidir, a través de los resultados de evaluación, sobre los países involucrados, para que esas

competencias se enseñen. Como tal, la evaluación internacional tiene un efecto de rebote muy fuerte sobre el currículo y la organización de la enseñanza.

Otra evaluación con fuerte repercusión fue la que realizó UNESCO-ORELAC, a través del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), en 1997, en once países de la región. Dos diferencias muy importantes con respecto a PISA: se evaluaron a niños de 3° y 4° grado de educación primaria y en este caso las pruebas se basaron en los programas curriculares de cada país, logrando un consenso entre los países después de dos años de negociaciones para construir los ítems de una prueba única¹⁰. Contrariamente al caso de PISA, no se han hecho públicos los ítems tal cual sino categorías genéricas. Por ejemplo, en el caso de lenguaje, se evaluó la identificación de tipos de textos; el reconocimiento de la función de un texto; la distinción entre emisor y destinatario de un texto; la distinción entre una secuencia temporal y causal explícita; y la identificación de concordancias gramaticales y categorías tales como sustantivo, adjetivo, verbo, pero no sabemos a través de cuáles reactivos. La evaluación incluía también una parte de redacción, que finalmente no fue integrada en el informe porque no hubo criterios homogéneos en los distintos países para atribuir puntajes. Por lo tanto, los resultados publicados corresponden a pruebas de lectura.

La sorpresa de estos resultados fue Cuba. Los niños cubanos sobrepasan a todos los otros de América Latina tanto en lectura como en matemáticas y tanto en tercero como en cuarto grado. Las diferencias son tan grandes que no hay manera de formar un grupo con Cuba y otros países. Cuba constituye un grupo por sí mismo. Probablemente estos resultados reflejen más de 30 años de priorización de la educación en los hechos y no solamente a nivel declarativo, así como los resultados a largo plazo de la superación del analfabetismo y el incremento general del nivel educativo de la población.

En todo caso, es de lamentar que Cuba no haya sido incluida en PISA 2000. De todos modos,

el contraste Finlandia-Cuba sirve para mostrar cuán difícil es saber a qué hay que atribuir los buenos resultados en estas pruebas.

OBSERVACIONES FINALES

Los resultados de aprendizaje no tienen un patrón de medida único. Si cambian las pruebas, cambian los resultados. No hay instrumentos consensuados sino instrumentos que se inventan con motivo de cada evaluación, ya sea porque han cambiado los programas de estudio o porque se pone énfasis en nuevas competencias que se estiman necesarias para el futuro (también, por supuesto, porque las pruebas anteriores fueron fuertemente criticadas por algunos sectores sociales con “peso específico”).

En Latinoamérica no hay experiencia acumulada sobre las consecuencias del bilingüismo de buena parte de nuestra población autóctona en los resultados de aprendizaje. Las evaluaciones se hacen en español para todos los niños. El informe 2003 del INEE tiene datos reveladores como el siguiente: al comparar los resultados de pruebas nacionales en lectura (alumnos de 6° grado) se constata que “sólo 2.5 % de los alumnos que egresan de las escuelas de educación indígena logran niveles altos de lectura, mientras el 14% de los egresados de escuelas urbanas públicas lo alcanzan” (señalemos, de paso, que ambos porcentajes son preocupantes)¹¹. En el análisis de los resultados del LLECE también se señala la incidencia de las lenguas indígenas en los resultados de pruebas de lectura en Bolivia y México (Perú y Ecuador no participaron en esta evaluación).

En razón del alto impacto social de la difusión de los resultados de evaluación, “la prueba pasa a convertirse en el foco de orientación de los procesos instruccionales en la escuela”, dice Juan Casassus, un prominente miembro del LLECE, experto en los problemas de evaluación. Este autor señala, en el artículo ya citado, que “en la medida que se crea una cultura evaluativa, la prueba pasa

a tener un peso mayor que el currículo en la cotidianidad del proceso pedagógico en el aula. En ese momento, la prueba tiende a transformarse, en los hechos, en el currículo real”. Si sumamos a esto que las reformas educativas tienden cada vez más a expresar los objetivos en términos de competencias y no de conocimientos, el resultado, según este autor, es que “los currícula, en la práctica, se vacían de contenido”. Esta tendencia se acentúa con la periodicidad de administración de las pruebas. Entre las pruebas internacionales, las regionales y las nacionales es difícil encontrar un año calendario en el cual alguno de los niveles de la educación básica no sea sometido a escrutinio valorativo. Preguntarse a qué intervalos pueden administrarse exámenes de evaluación sin interferir en el desarrollo normal de los procesos de aprendizaje no es, pues, una banalidad ya que la evaluación, como lo estamos viendo, no es algo que se hace “por fuera” del proceso educativo sino que constituye una fuerte intervención en el proceso mismo.

Cuando se constatan resultados preocupantes, no es evidente lo que se debe hacer para cambiar la situación. Es difícil comparar resultados de aprendizaje, en sí mismos, entre países que difieren en todos los parámetros posibles de comparación, como ocurre con México cuando se lo compara con los otros países de la OCDE (Canadá, Estados Unidos, Finlandia, Suecia, Noruega, Alemania, Japón, etcétera.). De nada sirve declarar al país en “emergencia educativa”. ¿Qué es lo que hay que corregir para estar a la altura de los otros? ¿El salario de los docentes y su preparación? ¿El nivel cultural general de la población? ¿Aumentar las horas de asistencia a clase? ¿Poner computadoras esperando que, mágicamente, este recurso técnico suplante al docente mal preparado y peor pagado? ¿Reducir el número de alumnos por clase? ¿Publicitar los casos extremos, o sea, una escuela que funciona a pesar de un entorno desfavorable, esperando que el ejemplo cunda? ¿Lograr que el sindicato docente incorpore la superación académica como

una de sus demandas? ¿Entregar libros y renovarlos periódicamente? ¿Publicitar el *ranking* de escuelas según el rendimiento de sus alumnos? ¿Hacer investigación didáctica para detectar los procedimientos que garantizan mejores experiencias de aprendizaje?

Lo que seguramente no hay que hacer es aplicar ciegamente las recetas de los organismos internacionales, éstos que actúan como si todo lo supieran de antemano, que ignoran las diferencias culturales, que sugieren escuelas mixtas en Afganistán y aseguran que los 300 días efectivos de clase declarados van a coincidir, en México, con un calendario ideal que hace abstracción de las festividades locales y de los “puentes” nacionales. La evaluación incide en la cultura escolar, pero la cultura escolar resiste. Los mecanismos de simulación también forman parte de la resistencia al cambio. Finalmente, de eso se trata: una institución –la escuela– que ayudó a forjar una nación porque tenía una utopía y la supo perseguir, y que ahora ha perdido el prestigio social y ha perdido los ideales que la vieron surgir debe cambiar, es evidente. ¿Serán las evaluaciones internacionales el agente de cambio? ¿Los docentes se sentirán convocados o excluidos? ¿De dónde vendrá la chispa que los encienda nuevamente y que los ponga a funcionar cambiando y siendo agentes de cambio, pero sabiendo cuál es la dirección y el sentido del cambio?

El *para qué* y el *por qué* de la educación están en crisis porque nuestros países latinoamericanos también lo están, sufriendo una globalización opresora y enajenante. La educación es quizás el lugar de manifestación de esta crisis.

NOTAS

¹ Artículo disponible en <http://lece.unesco.cl/documentosdigitales/>

² Por ejemplo, el nuevo Programa de Preescolar de la SEP 2004 señala que: “Una competencia es un conjunto de capacidades que incluye conocimientos, actitudes, habilidades y destrezas que una persona logra mediante procesos de aprendizaje y que se manifiestan en su desempeño en situaciones y contextos diversos”, p.22.

³ Brunner, José Joaquín, “Globalización, educación, revolución tecnológica”. *Perspectivas*, vol. xxxi, n°2, 139-156.

⁴ Este incremento en las revistas científicas no tiene sólo que ver con un incremento del conocimiento publicable, sino también con los

requisitos de evaluación a los que estamos sometidos los científicos.

⁵ La palabra “teoría” ha sido banalizada en la tradición anglófona, al menos en ciencias sociales, donde se aplica a micro-teorizaciones de un campo restringido del conocimiento.

⁶ Todas las citas referidas al programa PISA corresponden a la publicación en español “Conocimientos y aptitudes para la vida – Resultados de PISA 2000”, co-edición OCDE-Santillana, 2002.

⁷ La variabilidad en las denominaciones, y sus fluctuaciones al pasar del inglés a las distintas lenguas, muestra la ausencia de teorización sobre ese “constructo mental” (competencia, aptitud) al cual se hace vagamente referencia.

⁸ “Inteligencia es lo que mide mi test” decía a principios del siglo xx Lewis Terman, el exitoso adaptador de la primera escala de medición de la inteligencia, construida por Binet y Simon. Como puede apreciarse, seguimos en la misma longitud de onda, un siglo después.

⁹ Entrevista con Jouni Ensio Valijarvi publicada en el periódico *Clarín*, Buenos Aires, 28 septiembre 2003, y reproducida en varios sitios Internet.

¹⁰ Ver en la página web de LLECE el informe con el título “Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados en tercero y cuarto grado”.

¹¹ Instituto Nacional para la Evaluación de la Educación. *Panorama educativo de México – Indicadores del sistema educativo nacional 2003*.