

Evaluación educativa

El tema de la evaluación del sistema educativo se ha instalado como una de las principales cuestiones sobre las que se debate en la sociedad actual. Concomitantemente, en las políticas educativas de la Alianza "Cambiamos" la evaluación se muestra como la solución a todos los problemas y como la posibilidad de alcanzar "la calidad educativa" tan deseada. En este contexto, la centralidad de la evaluación y la aparente preocupación por los resultados de las distintas pruebas estandarizadas no son más que "grandes titulares" presentados por el gobierno actual, como su principal estrategia para justificar las transformaciones que se quieren realizar desde una perspectiva meritocrática, estandarizante, mercantilista y privatizadora, escondiendo la verdadera y última intención que consiste en la imposición definitiva de un proyecto de restauración conservadora en el campo educativo.

Por el contrario, desde nuestra Organización Sindical sostenemos que la perspectiva de la evaluación que el Ministerio intenta imponer como "rendición de cuentas" no significa para nada una solución a los problemas educativos, ya que la misma se reduce a un sistema operativo de control, medición y ranking para justificar los ajustes en educación, deprecia la inversión estatal y comprimir el presupuesto educativo. Todo ello, con la clara intención de abonar el terreno para la privatización educativa y responder así a los intereses de las empresas que pretenden lucrar con la educación.

Los sentidos políticos de la evaluación

Desde el punto de vista conceptual la evaluación es una categoría compleja y polisémica, donde los diversos sentidos e interpretaciones que se le atribuyen dependen de las concepciones teóricas, políticas y epistemológicas desde las que se la aborde. Cada una de sus acepciones remite a distintos significados, finalidades y modalidades de evaluación, ya que -parafraseando a Pierre Bourdieu- el campo de la evaluación es también un campo en disputa.

El interés por la actividad evaluativa en el ámbito educativo data de más de medio siglo. Sin embargo, la evaluación del sistema educativo en su conjunto se instala, en nuestro país, durante la década de los noventa, y se expresa a través de la evaluación de los aprendizajes como expresión de calidad educativa y de los programas focalizados que caracterizaron, durante esa época, a las políticas públicas. La lógica que primaba en ese momento era la del control del gasto y la racionalización, de acuerdo a los requerimientos de los organismos internacionales de crédito. Así, la evaluación se caracterizó por ser un proceso con un abordaje tecnocrático, orientado por el paradigma economicista de la rendición de cuentas, hegemonizado por la ideología neoliberal. En el marco político-ideológico de esa época surgen los Operativos Nacionales de Evaluación (ONE), que se realizaban en todo el país con un criterio totalmente meritocrático, ignorando contextos e historias y produciendo una mayor segmentación del sistema educativo al hacer públicos los resultados obtenidos por las distintas escuelas. Era, entonces, la opinión pública la que juzgaba el valor de la calidad, comparando estos resultados y generando los rankings que penalizan o premian a las instituciones educativas. Esta situación sirvió para desatar una intensa campaña mediática desde algunos sectores políticos, empresariales y medios de comunicación que, respondiendo a los intereses del poder hegemónico, han culpabilizado a las y los docentes y a sus organizaciones sindicales de la "mala calidad" de la educación.

En los últimos años, especialmente a partir de la sanción de la Ley de Educación Nacional 26.206, la evaluación educativa cobró un sentido diferente, ya no fue concebida como control, para penalizar o premiar, sino como herramienta o instrumento -entre muchos otros- que brinda información para orientar las acciones de política pública. El término evaluación del sistema educativo hace referencia a una serie de dimensiones que se relacionan con: las características de las instituciones educativas, los estilos de conducción, las propuestas curriculares, los recursos disponibles, las condiciones de trabajo, las estrategias didácticas, los aprendizajes de los alumnos, las características de puesto y del proceso de trabajo docente, entre las principales. Desde este nuevo marco, entonces, se considera que la evaluación de los aprendizajes a través de pruebas estandarizadas no mide la calidad educativa, sino un aspecto muy reducido de uno de sus múltiples componentes que es "el aprendizaje".

Aún reconociendo la existencia "de hecho" del uso extendido de las pruebas estandarizadas, desde una perspectiva crítica igualmente cabe decir que la información que aportan siempre debe ser complementada con otros indicadores e investigaciones que nos permitan evaluar las distintas dimensiones que integran el concepto de "educación de calidad" como, por ejemplo, la inclusión e integración, las condiciones materiales y de equipamiento, las condiciones institucionales y laborales docentes, entre las principales. Este enfoque integral de la evaluación aporta una mirada más amplia del concepto, que comprende el contexto y las condiciones en que se producen los procesos de enseñanza y aprendizaje¹. De lo contrario, es necesario reconocer y alertar sobre la extrema relatividad y arbitrariedad de dichas pruebas estandarizadas.

En el seno del Consejo Federal de Educación (CFE) -órgano encargado de la concertación, acuerdo y coordinación de la política educativa nacional, cuyas resoluciones son de cumplimiento obligatorio para las jurisdicciones- se aprueba la Resolución 99/2010, que establece un Plan de Acción de la Evaluación de la Calidad Educativa 2010-2020. En dicha resolución se reafirma este enfoque de la evaluación, en tanto expresa que "evaluar los aprendizajes supone focalizar la atención en la comprobación de éstos, como uno de los múltiples componentes de la calidad (...) La calidad de la educación, es una construcción colectiva y una responsabilidad compartida por todos los involucrados en el hecho educativo y su evaluación debe asumir, entonces, un carácter eminentemente formativo".

A partir de este nuevo marco normativo, se logró instalar el debate sobre el concepto de calidad educativa y/o educación de calidad, como así también sobre la legitimidad o no de su medición a través de las pruebas estandarizadas. Sin embargo, aún con este nuevo marco normativo, que dio lugar a otro paradigma donde la evaluación de aprendizajes se reduce a un indicador más -entre otros- de la calidad educativa, el foco y el interés puesto en los resultados de las pruebas estandarizadas continuaron estando presentes en el sentido

¹ Esta mirada del concepto de evaluación se plasma también en la Ley de Educación Nacional. En el artículo 85, inciso d, de la LEN se prevé implementar "una política de evaluación concebida como instrumento de mejora de la calidad de la educación". En el artículo 94, se establece que "el Ministerio de Educación, Ciencia y Tecnología tendrá la responsabilidad principal en el desarrollo e implementación de una política de información y evaluación continua y periódica del sistema educativo para la toma de decisiones tendientes al mejoramiento de la calidad de la educación, la justicia social en la asignación de recursos, la transparencia y la participación social." Es interesante señalar que el Estado Nacional deberá concertar la política de información y evaluación "en el ámbito del Consejo Federal de Educación. Las jurisdicciones participarán en el desarrollo e implementación del sistema de evaluación e información periódica del sistema educativo, verificando la concordancia con las necesidades de su propia comunidad en la búsqueda de la igualdad educativa y la mejora de la calidad" (artículo 96). Por su parte, el artículo 97 establece que si bien se harán públicos los datos, "la política de difusión de la información sobre los resultados de las evaluaciones resguardará la identidad de los/las alumnos/as, docentes e instituciones educativas, a fin de evitar cualquier forma de estigmatización, en el marco de la legislación vigente en la materia."

común de la sociedad y en los medios de comunicación, que, habitualmente y respondiendo al poder hegemónico, hacen uso de esta información para denostar la educación pública y en detrimento de las y los trabajadores de la educación.

Evaluación educativa en tiempos de restauración conservadora

La concepción de evaluación del gobierno actual pudo verse reflejada desde sus primeras acciones. Y el lugar que se le ha dado a la misma es tan central que, actualmente, se impulsa en el Congreso Nacional un proyecto de ley que promueve la creación de un Instituto Nacional de Evaluación, en el cual queda muy clara la intención de definir una política de evaluación aislada de las políticas educativas del Ministerio de Educación nacional y del Consejo Federal de Educación.

1. Las características del operativo de evaluación "Aprender 2016"

Como se ha venido denunciando desde las organizaciones sindicales docentes y desde los sectores académicos comprometidos en la defensa del derecho a la educación pública, el operativo Aprender 2016 da cuenta del cambio en la concepción y el sentido de la evaluación en tiempos de restauración conservadora y del lugar que ella ocupa dentro de las políticas educativas del gobierno actual. En este sentido, la política de evaluación pasa a ser un fin en sí mismo, en lugar de adoptar el rol instrumental que se le había asignado en el proyecto educativo del período anterior. En el llamado "Operativo Aprender" subyace la idea de que el cambio educativo es el resultado de la propia política de evaluación, que actuaría como un incentivo externo. Sin embargo, la evaluación y la información que se desprende de ella no tienen la capacidad de cambiar el desempeño del sistema por sí solo; por el contrario, son las políticas educativas más amplias, en las que la política de evaluación debe insertarse, las que orientan estos procesos.

Respecto de los aspectos técnico-pedagógicos, el operativo Aprender realiza una reducción de los contenidos de las pruebas. Por un lado, prioriza dos áreas en la educación primaria y en 2º y 3º año de la educación secundaria: Lengua y Matemática. Sólo se mantienen cuatro áreas -Lengua, Matemática, Ciencias Sociales y Ciencias Naturales- en 5º y 6º año de la secundaria, donde la prueba es censal (lo que hace suponer que se quiere ir a una prueba de acreditación). Por el otro, reduce la cantidad de preguntas de cada prueba, eliminando las preguntas que requerían el desarrollo de la respuesta por parte de los estudiantes. De esta manera, el operativo Aprender viene a reducir el aprendizaje a una respuesta correcta o incorrecta, limitando considerablemente la posibilidad de indagar acerca de los procesos de aprendizaje.

Asimismo, con este operativo se pasa de una evaluación realizada cada tres años a una evaluación que se aplicará todos los años. El planteo de evaluar todos los años crea expectativas falsas acerca del cambio educativo -la educación no mejora sólo por el hecho de evaluar todos los años o de desarrollar una cultura de la evaluación- e intenta instalar a la evaluación como un mecanismo de control para justificar diversos ajustes.

El operativo Aprender es presentado por el gobierno como un proceso de evaluación formativa que busca lograr cambios educativos a partir de la devolución de los resultados a cada escuela. Esta afirmación encierra dos contradicciones. En primer lugar, las pruebas Aprender fueron confeccionadas utilizando solamente ítems de respuesta cerrada, condición que reduce al mínimo la retroalimentación a las escuelas. Para poder hacer un diagnóstico, la evaluación necesita recoger más información que la que se produce a través de respuestas cerradas de elección múltiple, que nos dicen simplemente si el aprendizaje tuvo lugar o no. Por el

contrario, los ítems de respuestas abiertas, además de señalar los errores cometidos por los alumnos, nos orientan sobre qué es específicamente lo no aprendido.

En segundo lugar, el énfasis en lo formativo se produce a partir de una mirada lineal y simplista, donde la evaluación se asimila a la rendición de cuentas, a la competencia y a la supuesta mejora de la educación. La experiencia de todos estos años nos indica que los sistemas de evaluación a gran escala no pueden ser pensados para producir cambios en las instituciones educativas. Aunque se realice la devolución de los resultados de estas pruebas a las escuelas, es difícil mejorar la enseñanza si no se cambian las condiciones estructurales que impiden a las y los docentes reflexionar en forma colectiva sobre los mismos, sobre lo que saben los alumnos, lo que pueden hacer, en qué contexto se produce la enseñanza, con qué recursos, etc. Este es un proceso de reflexión más propio de la autoevaluación institucional participativa que de la evaluación estandarizada de los aprendizajes.

Por último, es necesario remarcar que hay una reducción tecnocrática y mecanicista del lugar que ocupan los supervisores, directores y docentes, a los que se les pide que participen de una evaluación externa sin haberlos convocado previamente para discutir la concepción de evaluación, sus objetivos, propósitos y cómo se enmarca esta política dentro de las políticas educativas de la nueva gestión. Se niega así la dimensión político-pedagógica de todo el proceso de evaluación.

2. Los resultados del Aprender 2016

En primer lugar, cabe mencionar que los resultados de las pruebas fueron presentados a través de una conferencia de prensa encabezada por el Presidente de la Nación y el Ministro de Educación y Deportes el mismo día que se iniciaba, en todo el país, la Marcha Federal Educativa para exigir la apertura de la Paritaria Nacional Docente. Los resultados fueron primero a la prensa y a la sociedad en general, y nunca se discutió el contenido de todas las dimensiones del operativo con las y los trabajadores de la educación y sus organizaciones sindicales.

Además, el uso que se le dio a estos resultados estuvo focalizado en el desprestigio de la educación pública estatal, sus docentes y sus organizaciones sindicales, que reclamaban -y aún lo hacen- por las condiciones de trabajo y por el derecho a la educación pública.

El propio informe publicado al presentar los resultados de las evaluaciones combinadas de Lengua y Matemática para 5º y 6º año del secundario por sector de gestión concluye: "Los estudiantes que asisten al sector privado alcanzan elevados niveles de desempeño Satisfactorio Avanzado que duplican al porcentaje de los estudiantes de gestión estatal. A su vez, en el sector estatal, el porcentaje de estudiantes que alcanzan desempeños por debajo del Nivel Básico y Básico casi duplican a los estudiantes del sector privado que alcanzan estos niveles de desempeño."

La presentación de resultados focalizados en sector estatal vs sector privado es en sí una definición política y bien está demostrado que esta diferenciación es falaz si no se contempla el nivel socio-económico de las y los alumnos y las condiciones materiales de las propias escuelas. Son muchos los informes nacionales, latinoamericanos e incluso internacionales que afirman que la diferencia estatal/privada deja de existir -es decir, no es estadísticamente significativa- cuando las escuelas se igualan en los indicadores mencionados.

Por otro lado, se ha presentado como "exitoso" un operativo que tuvo altísimo rechazo en la comunidad educativa. Si bien el informe publicado afirma una cobertura total de escuelas participantes -para el sector estatal es del 87,9% y para el privado del 97%- , el mismo informe aclara que en relación con la cantidad de

alumnos, la cobertura es apenas del 66,1% para el sector estatal y del 83,4% para el privado. Es probable que la cobertura sea mucho menor, ya que se considera para su cálculo que un estudiante ha participado de la prueba Aprender cuando ha contestado el 50% o más de alguna evaluación (podría haber respondido el 50% de los ítems de una sola prueba y se lo consideraba como participante)².

Respecto del procesamiento de los datos relevados durante la toma de la prueba en 2016, es importante aclarar que para establecer los puntos de corte entre los 4 niveles del Aprender se utilizó una metodología semejante a la utilizada por el ONE. En ambos casos se empleó el método de *Bookmark*, una de las metodologías más empleadas a nivel internacional para establecer niveles de desempeño en pruebas estandarizadas (UNESCO/TERCE). En este método se acude al juicio de expertos y docentes que trabajan con un cuadernillo que contiene las preguntas de la prueba ordenadas según su dificultad empírica, desde la más fácil a la más difícil. El trabajo de los jueces consiste en revisar las preguntas así ordenadas y seleccionar la primera de ellas, que se estima que el sujeto límite (un estudiante que está en el borde inferior de la categoría de desempeño cuyo puntaje de corte se está definiendo) tendría una mayor probabilidad de responder correctamente. Esta decisión individual es luego discutida en pequeños grupos donde se da a conocer el juicio de los demás participantes, y donde al término de la discusión e intercambio de opiniones cada juez tiene la posibilidad de volver a establecer su punto de corte.

Un problema que ha estado siempre presente es la ya clásica polémica, planteada desde hace varios años, sobre la arbitrariedad de los procedimientos, en tanto el concepto de dificultad es diverso entre las personas y a través del tiempo. En cualquier caso, tiene un componente subjetivo que genera grandes dudas sobre la manipulación de los resultados del Aprender 2016, dada la clara intencionalidad política del gobierno, que los dio a conocer en pleno conflicto con los docentes. Podríamos hablar de un cierto grado de arbitrariedad en cuanto a los puntos de corte que separan a los distintos niveles de desempeño. Esto se encuentra agravado al subdividir el nivel más bajo de desempeño en dos niveles: básico y por debajo del básico, utilizando sólo criterios estadísticos, y con muy pocos ítems (a veces uno solo) en el nivel "por debajo del básico".

3. Los mitos instalados sobre las pruebas estandarizadas

En América Latina los sistemas de evaluación y las pruebas estandarizadas de aprendizaje se han expandido y profundizado en casi todos los países a partir de la década de los '90. Desde la concepción neoliberal asociada a la medición de resultados, esta concepción se impone de la mano de reformas educativas marcadas por la descentralización, estímulos o impulsos a la privatización, mercantilización de la educación y precarización laboral de los trabajadores, entre otras.

En algunos países como Chile, Perú y México, estas políticas avanzaron también hacia la evaluación docente en un contexto de culpabilización y estigmatización gestado desde los sectores de poder, desvalorizando el

² Es importante recordar que en el documento sobre las dificultades del ONE 2013, Elena Duro menciona como aspectos negativos de esta evaluación la no respuesta de los alumnos por inasistencia de estudiantes o pruebas incompletas por agotamiento; y el bajo nivel de participación de los alumnos en las cuatro áreas de conocimiento a causa de un operativo desplegado en cuatro días distintos. Pareciera que el Aprender a pesar de reducir las áreas a dos y de hacerse en un solo día tiene también problemas de cobertura, pese a contar con un enorme aparato publicitario montado sobre el operativo (y a la enorme cantidad de dinero disponible para llevarlo a cabo).

trabajo docente y promoviendo, además, una supuesta pérdida de interés por parte de estudiantes hacia el aprendizaje.

Detrás de las pruebas estandarizadas como medición de la calidad subyace una concepción de la educación como formación de "capital humano", en lugar de entender a la educación como un derecho social y humano que permite la formación integral de personas. Además, siguiendo la línea argumental de Sacks (1999), estas mediciones no son objetivas y dan poca información educativa relevante. Un estudio realizado por Casassus (2007)³, a propósito de la profundización de las políticas de evaluación en Chile, muestra que su impacto fue precisamente el contrario del que se pretendía, ya que no mejoró la calidad educativa ni logró mayor equidad en su distribución. Entre sus argumentos destaca que el "estándar" es un método para uniformar y simplificar lo que es diverso y complejo y trae como consecuencia la reducción de los contenidos de aprendizaje, en tanto se "enseña para la prueba", fragmentando el conocimiento y evitando la reflexión crítica y la profundización de contenidos.

Desde una mirada técnica, en todas las pruebas estandarizadas se seleccionan los contenidos que en las pruebas piloto muestran diferencias estadísticas y se dejan de lado aquellas consignas que, por ejemplo, han podido resolver todos los y los estudiantes. Esto indica que las pruebas estandarizadas no contemplan los contenidos más trabajados por todas las instituciones escolares. "El creador de estas pruebas, Robert Glaser, ha sido muy claro en afirmar que estas pruebas no miden lo que saben los alumnos, ni lo que ellos saben hacer (...) La única competencia que miden estas pruebas referidas a norma es, según Glaser (1963; 1997), la capacidad de recordar procedimientos (para las matemáticas) o reconocer un resultado cuando se le dan alternativas múltiples. Es un error conceptual decir que este tipo de prueba mide lo que saben y/o saben hacer los alumnos de las escuelas. Es un error más grave aún el equiparar el puntaje que se obtiene, con el objetivo de tener una educación de calidad. Puntaje no es calidad." (Casassus, 2007: 74).

Si analizamos los resultados de estas pruebas estandarizadas vemos que, lejos de producirse mejores resultados, en la práctica sus puntajes vienen disminuyendo o bien están estancados en la última década. Por ejemplo, en Colombia, a pesar de implementar un examen anual en educación secundaria, sus resultados entre 2005 y 2009 no varían en términos estadísticamente significativos en ninguna de las áreas evaluadas. En Ecuador, entre 1996 y 2007, los puntajes para 3° grado de educación básica disminuyeron o se estabilizaron. En lo relativo a los resultados para 7° grado de educación básica, se observa que en lenguaje la situación se mantuvo más o menos estable, pero disminuyó significativamente en matemáticas. Las mismas tendencias pueden observarse en la gran mayoría de los países latinoamericanos⁴.

En síntesis...

La CTERA no desconoce la importancia de la evaluación educativa; por el contrario, la entiende como una herramienta muy valiosa, pero no la única, que hay que considerar en todo proceso de enseñanza y de aprendizaje. Esto significa comprender, a su vez, que el mejoramiento de la calidad educativa no se puede reducir a la implementación de operativos estandarizados de evaluación. Desde la concepción de educación que sostenemos, una prueba estandarizada que homogeniza, homologa y disciplina, que espera que todos

³ CASASSUS, J. (2007) "El precio de la evaluación estandarizada: pérdida de calidad y segmentación social". En: Revista Brasileira de Política e Administração da Educação. Enero/Abril, vol. 23, N° 1, pp. 71-79, Porto Alegre.

⁴ Para mayor detalle, véase: IEAL (2015) "La estandarización de la evaluación. Las pruebas nacionales e internacionales ¿medición o evaluación?", Internacional de la Educación para América Latina, abril de 2015.

tengan los mismos estándares de aprendizajes, no podrá nunca evaluar la calidad de un sistema educativo, donde cada escuela y cada contexto constituyen una singularidad.

La educación de calidad es un concepto polisémico, una construcción colectiva y una responsabilidad compartida que no puede pensarse sin la garantía de la inclusión socioeducativa. En este sentido, la evaluación tiene que asumir su carácter eminentemente formativo y de acción político-pedagógica.

Es por ello que desde CTERA se fomenta un proyecto pedagógico cuyo objetivo principal es la definición de un conjunto de políticas educativas que, en el ámbito del Ministerio de Educación y Deportes de la Nación, contribuyan con el mejoramiento de las condiciones de enseñanza y de aprendizaje, a los efectos de garantizar una Educación Pública de la calidad en la Argentina. Y, en ese marco, se plantea la pretensión de avenirse a los fines y objetivos de la política educativa nacional establecidos en el artículo 11º de la Ley 26.206 de Educación Nacional y ajustarse a las disposiciones y principios fijados en los artículos 94º, 95º, 96º, 97º, 98º, 99º de la misma Ley, referidos a la implementación de una política de información y evaluación del sistema educativo para la toma de decisiones tendientes al mejoramiento de la calidad de la educación.

CTERA exige que se respete el Acuerdo Paritario Nacional firmado el 29 de noviembre de 2013, en el que el Estado Nacional se compromete con el financiamiento de un programa de formación docente permanente, gratuito y en ejercicio, garantizando la participación de los sindicatos docentes en la construcción político-pedagógica de los dispositivos de formación. En ese acuerdo se define un componente de autoevaluación institucional formativa, integral y participativa, que concibe a los propios docentes como productores de conocimiento y de los procesos de mejora de la enseñanza y de los aprendizajes. De esta manera, la evaluación es parte de una de las fases de la producción de conocimiento y del proceso de trabajo docente.

En tal sentido, se expresa la necesidad de plantear un sistema o dispositivo de evaluación como dimensión de un proceso más complejo de formación que requiere la participación de las organizaciones sindicales docentes para su definición. Esta manera de pensar la evaluación se contrapone a las concepciones mercantilistas que la colocan como un "fin en sí mismo" y que se han denunciado a lo largo de todo este documento de posición sindical, política y pedagógica.

Documento analítico elaborado por el "Equipo de trabajo sobre evaluación educativa" de la Secretaría de Educación de CTERA, constituido por Liliana Pascual, Luz Albergucci, Miguel Duhalde, María D. Abal Medina.

Ciudad de Buenos Aires, 20 de julio de 2017.