

Reflexiones sobre las funciones de los sistemas de evaluación educativa de gran escala. Entrevista a Silvina Larripa

Reflections on the functions of large- scale systems of educational assessment. Interview with Silvina Larripa

E: *La problemática de la calidad de la educación está instalada hace tiempo así como el desarrollo de programas y operativos para su evaluación. ¿Qué aspectos te parecen significativos señalar en cuanto a su surgimiento y desarrollo?*

S: Bueno, la problemática de la calidad de la educación se instala en la agenda de los países de América Latina a fines de la década del '80, comienzos de la década del '90. Obedece a una tendencia internacional. Algunos autores señalan que, resueltos los problemas de cobertura, la pregunta ¿por qué aprenden los estudiantes? pasa a ser un problema fundamental. Este es un costado de la cuestión. En un marco más complejo -digamos- en los '80, en algunos de los países desarrollados, las preguntas por ¿en qué medida el gasto público en educación rinde sus frutos?, ¿en qué medida las inversiones en educación rinden sus frutos?, también se convierten en un problema fundamental en el marco de políticas de ajuste fiscal y bueno, muchos autores hablan del Estado Evaluador y en este sentido me parece que es otro costado que hay que entender para comprender el surgimiento de los sistemas de evaluación de gran escala. En América Latina, podríamos decir que esta preocupación se manifiesta, por ejemplo, en el surgimiento de sistemas de evaluación nacionales de la calidad educativa. La instalación de estos sistemas es una de las condiciones para acceder a los préstamos para las Reformas Educativas que acompañan la instalación de estos dispositivos. En este sentido, la implantación de sistemas de evaluación de la calidad es un requisito, forma parte de la propuesta de reforma educativa con apoyo financiero externo. Esto me parece importante porque si bien no determina las características últimas de los sistemas de evaluación de la calidad, sin la aparición de estos organismos de crédito y de este movimiento de Reforma, no podríamos entender la instalación de estos grandes sistemas que demandan muchos recursos humanos, financieros. Así que ahí hay una primera respuesta a esta pregunta por cómo surgen, cómo surge la preocupación por la calidad.

E: *¿Se han operado cambios en las últimas décadas? Si es así, ¿cuáles?*

S: Yo diría que ahora muy pocos hablan de que estos sistemas evalúan calidad

educativa sino que señalan que estos sistemas evalúan logros académicos o rendimiento académico de los alumnos. De alguna manera se problematizó un poco más este salto que sí, al principio, parecían pegar estos sistemas al entender que tomando pruebas de lápiz y papel, que es el instrumento que suelen usar, se estaba evaluando la calidad del sistema educativo. Porque hay otros indicadores del rendimiento de los sistemas educativos, cuantitativos, que siguen siendo relevantes para conocer qué pasa con la educación en un país o en un conjunto de países. Después bueno, esto desde una perspectiva más personal creo que ahora ha aparecido en el debate el problema de la inclusión; quizás estoy muy teñida de este asunto porque estamos trabajando en educación secundaria, en qué pasa con la educación secundaria hoy, pero creo que esto se puede extender y decir que ahora se habla de cómo incluir a todos los alumnos, cómo garantizar que permanezcan en el sistema educativo y -al mismo tiempo- con aprendizajes significativos ¿no? O sea, con esto quiero decir que los indicadores de rendimiento cuantitativo: deserción, abandono, repitencia, siguen siendo fundamentales para comprender qué pasa con los sistemas educativos y poder generar políticas centradas en la mejora.

E: *¿Dirías entonces que se ha desplazado la preocupación por la calidad?*

S: O por lo menos aparece esta idea de inclusión con calidad ... Lo digo en forma de lema.

E: *¿Cuáles serían, a tu juicio, las principales funciones de estos operativos y sus posibles impactos en diferentes esferas y ámbitos del funcionamiento de los sistemas educativos?*

S: Bueno, voy a dar un punto de vista personal: los sistemas de evaluación de gran escala tendrían que servir para tomar decisiones a esa escala. Es decir, creo que uno de los problemas que vi, por ejemplo, en uno de los casos que analicé (el de un sistema de evaluación jurisdiccional en Argentina) es que se monta un sistema de gran escala con foco en la toma de decisiones institucionales y parece que no hay un uso fuerte de los resultados en el ámbito macro ¿no? O sea, por parte de otras direcciones. No digo el mismo equipo o la propia dirección a cargo de la evaluación, sino de otras direcciones: planeamiento, capacitación docente, currículum, que deberían verse nutridas por este tipo de resultados. Justamente, porque quizás para ofrecer resultados que se puedan usar a nivel institucional, hay otras estrategias, incluso, -digo, no soy economista- pero menos costosas también ¿no?

E: *¿Por ejemplo?*

S: Por ejemplo, conceptualmente la evaluación es una práctica atada a la toma de decisiones: o sea, se diferencia por ahí de la investigación en esto ¿no? Se construyen instrumentos, se releva información, pero el problema del uso de esa información para la toma de decisiones es el que hace a la validez del propio sistema, o programa, o propuesta de evaluación. Es como clave. Entonces desde mi punto de vista, justificar

un programa de evaluación de gran escala cuando los resultados se van a emplear solamente en las escuelas, por los directivos o los docentes, es problemático. Porque no todos comparten esta visión. Desde el punto de vista institucional podríamos pensar que hay tanto por evaluar en el micro nivel de la institución que podemos pensar en un plan institucional de evaluación; en donde los profesores y los coordinadores del ciclo se pongan de acuerdo acerca de cuáles van a ser los criterios para evaluar a los alumnos. En el primer ciclo o en el segundo ciclo, para resolver el problema de la promoción, de los criterios para promocionar o no a los alumnos dentro de cada uno de los ciclos, y obviamente también al final ¿no? Qué instrumentos y qué estrategias son más adecuadas, incluso para evaluar distintos aspectos del currículum. Porque necesitamos distintas estrategias justamente por eso ¿no? Los objetos son distintos, los conocimientos, las habilidades, exigen diversificar los modos de acercarnos a conocer qué saben los alumnos, qué pueden hacer. Entonces en ese sentido hay tanto para coordinar en el espacio de la Institución y para compartir y para socializar en un cuerpo de profesores, que pensar que los resultados de un sistema de evaluación de esta magnitud sólo van a ser empleados por un docente o un grupo de docentes, es reducir el alcance y la potencialidad que -desde mi punto de vista- pueden tener estos sistemas. Entonces, últimamente por lo que estoy más preocupada, a partir del análisis quizás de estos casos -que son sistemas o programas de evaluación en dos de las jurisdicciones del país- es ver en qué medida esos resultados son empleados por quienes están a cargo del planeamiento, el desarrollo de programas y la evaluación de programas que impactan en las instituciones.

E: *Y entonces pensando en este nivel, a escala nacional o jurisdiccional, ¿qué tipo de “usos” podrían pensarse como pertinentes a esa escala?*

S: Bueno, hay un ejemplo concreto en uno de los casos que estoy analizando como parte de la investigación doctoral: a partir de resultados de una evaluación de escala jurisdiccional en el nivel primario, se desarrolló un programa de fortalecimiento de la enseñanza que involucraba distintas áreas: planeamiento, especialistas en desarrollo curricular y por supuesto, destinado a directivos y a docentes. Los resultados de la evaluación fueron un elemento más del diagnóstico para que esas mismas autoridades pensarán en este proyecto que abarcaba todas las escuelas primarias y en este sentido, por eso también era una decisión de escala jurisdiccional. Tenía muchos matices, incluyendo producción de materiales para los docentes y para los alumnos, capacitación de los directivos; bueno, esto se dio en el marco de un cambio en la propuesta curricular de esta jurisdicción...

E: *Teniendo en cuenta que se trata de cuestiones públicas de impacto social, en las que cabe tanto el análisis técnico como la reflexión ética y política, ¿cuáles podrían ser los rasgos de una buena práctica en esta materia y cuáles serían las condiciones para que ellas puedan tener lugar?*

S: Una de las condiciones tiene que ver con sostener en el tiempo estos dispositivos, o estas políticas de evaluación de gran escala, porque es la única manera de que brinden resultados que permitan tomar decisiones no basadas en muy poquita evidencia o en evidencia recolectada en un momento determinado del tiempo. Bueno y acá hay un problema que tiene que ver con los sistemas de evaluación como políticas, en general de un plazo que muchas veces no es el plazo en el que se sostienen las autoridades. Y eso que parece tan obvio, genera muchas dificultades porque se cambian los sentidos con los cuales esa evaluación fue pensada e implementada. Bueno, y ahí se rompe esta idea de que, en realidad, la evaluación a gran escala siempre es la evaluación de una política o para una política ¿no? En este sentido más amplio ¿no? Es la evaluación de una política educativa o para una toma de decisiones en materia de política educativa. Por lo tanto, si fue concebida y después sus resultados fueron empleados desconociendo cuál fue el marco y los propósitos, en el mejor de los casos, quizás se encuentran otros usos; en el peor de los casos, los resultados no son empleados directamente por las nuevas gestiones de gobierno y entonces, se discontinúa y se pierde el sentido ¿no? Desde el punto de vista de los actores escolares esto es importante también porque estos sistemas generan movimientos en las instituciones y en los actores escolares. El no poder justamente vincular el momento en que ellos son convocados para participar -generalmente ya en el momento de la aplicación, a veces en el diseño- pierden también el sentido por el cual esa evaluación es importante y puede contribuir. Ahora lo digo incluso desde el punto de vista de pensar la evaluación para mejorar prácticas a nivel de la institución ¿no?

E: *Esta es una condición, ¿habría alguna otra que te parezca relevante?*

S: Sí, sí... Otra fuerte, digamos, se ubica en el plano técnico. O sea, la evaluación tiene un fuerte componente de conocimiento técnico que quizás desde algunas perspectivas no es tan reconocido ¿no? Yo por el contrario creo que es fundamental. En este mismo caso que estaba comentando se toman, por ejemplo, evaluaciones a alumnos de tercer año y a los dos años se toman evaluaciones a alumnos de quinto año esperando hacer algún tipo de reflexión acerca de si los alumnos continúan aprendiendo. Son los mismos alumnos, entonces, bueno, hay que cuidar cuestiones, por ejemplo, relativas a la muestra. Pero también cuestiones relativas a cómo se construyeron los instrumentos porque depende de los ítems, de las cuestiones, de las preguntas, de las tareas que se incluyan, de la dificultad de las mismas, de los contenidos... si voy a poder ¡o no! decir si los alumnos continúan aprendiendo lo esperado o por el contrario avanzan y obtienen resultados que no son tan buenos o tan deseables como quizás eran los que habían obtenido en el primer ciclo. Y, ¡esto es importante! Lo mismo pasa con los operativos nacionales. Bueno, mi tesis de Maestría fue sobre este punto. Los datos analizados, quiero aclarar, son desde 1995 a 2000. El trabajo muestra problemas técnicos en la construcción de los instrumentos que hacen que no sea posible establecer comparaciones válidas,

longitudinales. Entonces tanto para desarrollar un sistema de evaluación de escala nacional, como en el caso de los sistemas de evaluación jurisdiccionales, se requiere un conocimiento técnico, específico.

E: *Por ejemplo, un aspecto que vos habías analizado en relación con la comparabilidad es la existencia de preguntas de anclaje, ¿no?*

S: Sí, sí... Hay diferentes maneras de lograr que las pruebas sean comparables a lo largo del tiempo. Obviamente si se aplicara el mismo instrumento año a año, estaría resuelto el problema porque son los mismos ítems, tareas, cuestiones que se plantean de forma tal que lo que varía es, solamente, el rendimiento de los alumnos. Esto, entendiendo que las condiciones de aplicación son las mismas, en la misma altura del año, etc. ¿no? Los ítems de anclaje son ítems que son comunes, iguales, en los instrumentos que se aplican en distintos años para el mismo grado escolar. Hay procedimientos que lo que permiten a través de estos ítems comunes es justamente estimar la habilidad que hubieran tenido los alumnos para resolver cualquiera de los instrumentos administrados en cada uno de los años. Ahora bien, no solamente hay que incluir esos ítems de anclaje sino que los mismos, tienen que ser suficientes en función de lo que se llama el marco conceptual, que es el conjunto o el resto de los ítems o cuestiones que componen las evaluaciones. Y allí es donde, haciendo un análisis de qué evaluaba cada uno de los ítems, lo que se observó, fueron fuertes diferencias en cuanto a qué contenidos se evaluaban en cada uno de los instrumentos. Por ejemplo, yo puedo entender una prueba de matemáticas con un gran foco centrado en resolución de problemas o puedo entenderla con un foco centrado en el manejo del sistema de numeración. Si yo construyo una prueba con foco en la resolución de problemas, con muy pocos ítems de numeración y al año siguiente invierto esta decisión de manera fuerte, los resultados dejan de ser comparables porque se refieren a otro objeto. Hay dos cuestiones ahí: dificultad y contenido evaluado. La dificultad con que presento las tareas y el contenido efectivamente evaluado en cada una de esas tareas. Esos son dos aspectos que hay que cuidar para que aquello que se evalúa sea lo mismo y, por lo tanto, los resultados puedan ser comparables. Ese es un ejemplo de aspectos técnicos a considerar en estos sistemas de gran escala que también tienen esta función de monitoreo.

E: *Te estás refiriendo a algunas cuestiones difíciles de resolver tanto en lo político como en lo técnico y que, al mismo tiempo, son decisiones críticas para el desarrollo de estos programas. ¿Hay algún otro aspecto que te parezca importante señalar?*

S: A mí me parece que uno de los aspectos de resolución más compleja, es que cada vez es más difícil pensar las políticas de evaluación de manera aislada. O sea, uno de los peligros que representa la instalación de una política de evaluación aislada de otras es que pasa a cumplir otras funciones, y esto es un problema de difícil resolu-

ción porque la cuestión es quién toma las decisiones, quién usa los resultados. Desde mi perspectiva personal, muchos de estos sistemas o programas de evaluación parecen haberse “autonomizado”. Estoy usando un concepto que empleó una de las personas que yo entrevisté y bueno, me quedé pensando en esta “autonomización” ¿no?

E: *Es interesante, ¿podés avanzar con esta idea? ¿Qué significa que un sistema de evaluación se ha “autonomizado”?*

S: No sé si la palabra es “autonomización”. Estoy trabajando un poco sobre eso, porque son más categorías de política, de política pública, o sea, lo que me está sucediendo es que estos problemas me están llevando a otro campo más amplio, que es -en general- ¿qué pasa con las políticas educativas?, ¿cómo se articulan? ¿Cómo se articula el trabajo de diferentes áreas, al interior de un Ministerio o de un organismo estatal para que el trabajo de una Dirección sea aprovechado por otra? Es muy simple la idea, pero los problemas que estoy encontrando, que encontré y que más me preocupan, aluden a esto. A que las funciones de una evaluación de este tipo se ven seriamente restringidas o directamente dejadas de lado, si el propio programa de evaluación quizás -y con esto no quiero hablar de responsabilidad del propio programa o de quienes trabajan allí- no goza de un espacio en el marco de la política educativa y de cómo se piensa y se planifica esa política educativa, en términos más amplios. Con buenas intenciones, se empiezan a generar otros objetivos... Por ejemplo, lograr acercarse más a las escuelas -que no digo que no sea válido, pero no es el único o quizás el más potente objetivo de este tipo de sistemas-. Las escuelas, los directores, los maestros, siempre están ahí, son interlocutores. Muchas veces, en el caso de jurisdicciones con un gran territorio, son casi como audiencias muy anónimas. Pero la posibilidad de dirigirse a esas audiencias, de preparar cuadernillos para esas audiencias siempre está ¿no? En cambio, hay otros interlocutores, por ejemplo quienes están a cargo del diseño de políticas educativas, que varían generalmente cuando cambia la gestión, las autoridades políticas.

Se pone mucho foco en la escasa utilización de los resultados por parte de los docentes, de los directivos. Digamos, aparte de que muchas veces la información tarda en llegar a las escuelas, etc. No quiero tampoco entrar en esta discusión, la información oportuna es obviamente una característica que hace a un buen dispositivo ¿no? Esto para la toma de decisiones políticas y a nivel institucional, pero hay poca preocupación por el uso de los resultados por parte de quienes están a cargo del diseño de políticas educativas.

E: *Más allá de estos impactos esperados, hay algunos estudios que dan cuenta de un impacto real que tiene el establecimiento de sistemas de evaluación a gran escala, sea nacional o jurisdiccional, en las prácticas de enseñanza y en el currículum ¿no? ¿Cuál es tu opinión al respecto?*

S: Bueno, yo creo que ahí hay dos categorías: sistemas de Evaluación con consecuencias fuertes o consecuencias débiles. En el caso de la Argentina, los sistemas que se implantan son sistemas con consecuencias débiles. Débiles para las Instituciones, para los actores escolares, para los docentes, para los alumnos. No es el caso de Chile, entonces ahí sí podemos hablar de múltiples relaciones entre el Sistema de Evaluación, la Unidad de Currículum, lo que sucede en las escuelas. Los Sistemas de Evaluación con consecuencias fuertes implican que los resultados del proceso de evaluación van a incidir de manera importante en la trayectoria de un sujeto, en la historia de una Institución. Por ejemplo, porque los resultados que obtengan los alumnos de una institución van a ser considerados para tomar decisiones acerca de en qué medida se financia o no algún proyecto de esa Institución. En el caso de la evaluación docente, porque el resultado de una evaluación de gran escala puede llegar a significar, separar al docente del ejercicio de su profesión... En el caso de los alumnos, consecuencias fuertes porque implican o no el ingreso a los estudios superiores. Es el caso de los exámenes de ingreso. Y también hay otras consecuencias fuertes que quizás no están tan formalizadas. Por ejemplo tuve oportunidad de visitar Chile y de conversar con la gente que tiene a su cargo el Sistema de Evaluación; y comentaba que los directores tomaban decisiones acerca de qué docente era asignado a los grados en donde era administrada la prueba SIMCE, que es la prueba nacional que se administra en cuarto y octavo básico y segundo año del nivel medio. Lo que habían visto es que muchos directores colocaban a los mejores docentes en cuarto básico, y que -al mismo tiempo

muchas veces había docentes que se sentían responsables por los resultados y se deprimían profundamente si los alumnos no rendían lo esperado. Bueno, el nombre de la institución se pone en juego porque cuando los resultados están listos se prepara un inserto en la prensa donde no se arma un ranking pero sí aparece el nombre de las instituciones.

En la Argentina esto no es así. Se habla de consecuencias débiles porque justamente los resultados de la evaluación no van a determinar el despido de un docente, por ejemplo, o la posibilidad de no continuar estudios en “x” institución para un alumno... Y se habla -por el contrario, en la bibliografía especializada- de que estas evaluaciones tienen foco en lo que se conoce como función formativa o de reorientación de los procesos de enseñanza, de las propuestas de enseñanza, con vistas a mejorar los aprendizajes de los alumnos. Esto en el plano micro ¿no? En el plano macro podríamos pensar en políticas para mejorar... Son consecuencias débiles pero muy importantes. Quienes abogan por la introducción de Sistemas de Evaluación con consecuencias fuertes, conciben el cambio educativo en un sentido diferente que quienes abogan por Sistemas de Evaluación con consecuencias débiles. Esto me parece importante ¿no? Pedro Ravela en una conferencia, creo que es del año 2003, señala -a mí me gusta mucho esta idea- que detrás de un Sistema de Evaluación hay una teoría del cambio. Bueno, la pregunta sería entonces: ¿realmente el cambio educativo se produce porque las familias van a demandar mejor educación porque tienen información gracias a estos resultados? Y los profesores ¿van a capacitarse porque está en riesgo la posibilidad de continuar su carrera? Hay una idea e incentivos por detrás ¿no? Desde esta perspectiva, las personas cambiamos si nos incluyen en un sistema de incentivos positivos-negativos que hace que nuestras conductas se modifiquen.

Creo que los sistemas de evaluación con consecuencias débiles están apelando a otra idea del cambio, en la que los resultados de un proceso de evaluación son un elemento más para promover transformaciones, que no se producen de manera automática o como resultado de un incentivo externo, sino como resultado de un proceso de apropiación que debe ser reconocido y apoyado por las mismas políticas de evaluación y las políticas educativas más amplias en las que éstas deberían insertarse. El uso de los resultados de las evaluaciones implica un proceso que tiene su tiempo y supone fundamentalmente una profunda interpretación de esos resultados en función de cómo se concibió el proceso de evaluación, sus objetivos, sus propósitos por parte de todos los involucrados (políticos, gestores, supervisores, directores, etc.) que entonces podrán vincular esos hallazgos con lo que se viene haciendo. De este modo, el cambio surge de la convicción de que lo que se puede hacer a partir de esa nueva interpretación de la realidad, es mejor de lo que se hacía hasta entonces.

E: *En relación con el contexto local y pensando en perspectivas a futuro, ¿hay*

algún desafío importante u obstáculos a franquear, tanto a nivel del diseño de un sistema como de su uso para la toma de decisiones?

S: Un tema que no tocamos y que es fundamental, es la definición del referente, de qué es lo que se va a evaluar en términos de cuál es el recorte que se va a hacer. Por ejemplo, si se decide evaluar en función del Diseño Curricular: qué recorte se va a hacer del Diseño y -en este sentido- el Sistema de Evaluación Nacional sigue teniendo quizás algunas dificultades en este aspecto. Tenemos descentralizada la definición del currículum y por lo tanto llegar a un acuerdo acerca de qué se considera fundamental, es muy complejo. Están los núcleos de aprendizaje prioritarios, pero esto no parece ser suficiente. Se generan ruidos sobre todo en las instituciones. Por ejemplo, hay directores que preguntan: “Va a haber una evaluación de Ciencias Naturales en quinto año, nosotros no tenemos Ciencias Naturales en quinto año. ¿Se está considerando que dentro de Ciencias Naturales los contenidos a evaluar van a incluir Biología, van a incluir Física, van a incluir Geología, van a incluir...?”. Me parece que -y esto para todos los Sistemas- abordar una revisión de qué es lo que se considera significativo, relevante de ser evaluado, y que eso sea de dominio público. Previamente, también de discusión pública ¿no? O sea, poder volver. Quizás es muy utópico ¿no? Pero dar de vuelta una discusión acerca de qué es lo que se va a evaluar y para que se avance en una mejor especificación de lo que técnicamente llamamos el referente, pondría en otra situación al Sistema de Evaluación, al modo en que se recortan y luego se interpretan los resultados. Porque los mismos, de alguna manera, son evidencia sobre qué sucede con ese recorte, con esos contenidos. Y acá entramos en otro problema que es que desde la Didáctica hay muchas perspectivas acerca de cómo se entienden los contenidos. A nivel nacional esto es realmente muy complejo porque cada jurisdicción tiene su propio Diseño Curricular, sus modos de aproximarse y de comprender los objetos de enseñanza. Y creo que ésta también es una de las razones por las cuales existen dispositivos jurisdiccionales ¿no?

En el caso de los dispositivos jurisdiccionales, es fundamental que realmente el referente sea de dominio público. Muchas veces los involucrados reciben información sobre el día de la aplicación, el tipo de instrumento, pero hay un punto clave que es qué espacio se les da, por ejemplo, a los supervisores, a los directores de escuela, para participar en el proceso de definición de qué es lo que va a ser evaluado. Esto se hizo en algunos momentos de la historia de los dispositivos jurisdiccionales de evaluación, se problematizó. O sea que aquí también hay una decisión a tomar: ¿el referente va a estar centrado en el Diseño Curricular prescripto? o ¿va a implicar una indagación previa acerca de qué es lo que los alumnos tienen oportunidades de aprender de ese Diseño Curricular prescripto y centrarse en ello? En el caso de una evaluación de nivel medio, esto se complejiza mucho más porque hay múltiples programas, hay algunos que están desactualizados, otros que son demasiado nuevos. Bueno, hay un montón de decisiones importantes

a tomar. Creo que antes hablamos mucho de los propósitos y de los usos, o sea *para qué* una evaluación de esta escala. Y en este momento estamos prestándole atención al *qué*. Yo diría que son los dos aspectos centrales para pensar cualquier propuesta de evaluación: para qué y qué se va a evaluar, porque los resultados van a ser interpretados y utilizados en función de las respuestas que se den a estas dos cuestiones.