

UNA MIRADA TÉCNICO-PEDAGÓGICA ACERCA DE LAS EVALUACIONES DE LA CALIDAD DE LA EDUCACIÓN ¿Qué pretendemos evaluar, qué evaluamos y qué conclusiones podemos extraer de la evaluación?

Alejandro Tiana Ferrer (UNED – IEA)

Los años noventa se caracterizaron por la puesta en marcha y el desarrollo de un buen número de reformas educativas en los países latinoamericanos. Una vez superada la desconfianza hacia la educación que caracterizó a los años setenta y ochenta, los gobiernos de la región se lanzaron a una intensa actividad reformadora, que se dejó sentir en muchos países y que abarcó diversas esferas de la realidad educativa (Gajardo, 1999).

Las reformas emprendidas en esos años persiguieron varios objetivos, entre los que cabe destacar la extensión de la escolarización (especialmente en la enseñanza secundaria), la mejora de la calidad de la educación impartida y el aumento de la equidad en la distribución de los servicios educativos. Aunque no sea éste el lugar para hacer un balance de la década, no cabe duda de que los objetivos previstos se alcanzaron de manera muy desigual, consiguiéndose algunos logros destacables en el primer sentido de los mencionados, pero logrando unos resultados mucho más modestos en los otros dos. La situación es tal que un reciente informe internacional sobre el estado de la educación en los países de América Latina adoptaba como título el lema *Quedándonos atrás*, expresando de ese modo las carencias que aún aquejan a los sistemas educativos de la región (Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe, 2001).

El desarrollo de políticas nacionales de evaluación

Las reformas de los años noventa afectaron a diversos aspectos de la realidad educativa, que conviene detallar. Entre las orientaciones adoptadas por dichos procesos de reforma, y aun sin ánimo de exhaustividad, pueden distinguirse varias líneas de actuación que, si bien no se dieron siempre de manera simultánea, estuvieron presentes de uno u otro modo en las diversas iniciativas nacionales:

- Un primer aspecto en que se avanzó considerablemente, aunque no sin problemas, fue en la redistribución de las competencias y responsabilidades en materia de educación, lo que supuso la puesta en marcha de políticas de descentralización y de autonomía escolar, acompañadas en ocasiones de procesos de privatización, cuyos efectos fueron ambivalentes (Winkler y Gershberg, 2000).
- Una segunda orientación de las reformas, especialmente influyente en muchos países de la región, consistió en la introducción de cambios en la organización curricular, que llevaron generalmente aparejada una revisión de los planteamientos fundamentales en que se inspira la definición y la construcción del currículo.

- Un tercer dominio de acción de las reformas tuvo que ver con los procesos de formación inicial y de capacitación del profesorado, considerado habitualmente pieza clave de las reformas, pero no siempre tan atendido como debiera en su papel de agente transformador.
- Un cuarto ámbito de reforma estuvo concretamente relacionado con la educación secundaria o media, que se ha convertido en la pieza clave y más conflictiva de los sistemas educativos y cuya expansión se ha ido también produciendo en América Latina en las últimas décadas (Braslavsky, 2001).
- Por fin, una última orientación tuvo que ver con el desarrollo de mecanismos y modelos de evaluación, dicho de otro modo, con la consideración de la evaluación como un poderoso instrumento al servicio de la gestión de la educación y de la mejora de su calidad (Tiana, 1996).

Esta última orientación de las reformas de los noventa cobra especial relevancia desde la perspectiva que se adopta en este trabajo, pues es precisamente el impacto de esos mecanismos de evaluación lo que está sometido a crítica y discusión. Por este motivo, merece la pena realizar un par de comentarios adicionales sobre sus rasgos más destacados.

Una primera observación que cabe destacar desde este punto de vista es que la mayoría de los países de la región pusieron en marcha sus propios sistemas nacionales de evaluación a lo largo de los años noventa, generalmente en el marco de procesos más amplios de reforma educativa. Con configuraciones institucionales muy diferentes entre sí y con ámbitos de actuación diversos, fueron varios los gobiernos que se sumaron a la experiencia pionera que otros países como Chile venían desarrollando ya con anterioridad. Argentina, México, Colombia, Bolivia o Brasil, por no citar sino a algunos casos destacados, dieron pasos decididos en esa dirección, de manera que al final de la década eran mayoría los países latinoamericanos que contaban con algún tipo de sistema nacional de evaluación (Ravela, 2001).

Sin embargo, el indudable interés por la evaluación de la educación que se despertó en muchos países no marchó asociado sino mucho más tardíamente con un interés paralelo por participar en estudios comparativos internacionales de evaluación del rendimiento educativo. Aparte de la valiosa experiencia regional del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, en el que participaron un total de trece países de América Latina y el Caribe, los estudios promovidos por organizaciones tan reconocidas como la IEA o la OCDE no han alcanzado todavía en la actualidad cotas altas de participación (Tiana, 2000). Durante los años noventa, fueron más bien los programas de cooperación orientados hacia el reforzamiento de los sistemas nacionales de evaluación, como el impulsado en esa dirección por la OEI, los que lograron una participación más amplia.

Así pues, y hablando siempre en términos generales, la situación de la evaluación educativa en la región latinoamericana se caracteriza hasta el momento por la existencia de un contraste llamativo entre el desarrollo de ambiciosos programas

nacionales de evaluación y una limitada participación en estudios comparativos internacionales. El desarrollo de esos programas nacionales está basado en la puesta en marcha de un buen número de instituciones y agencias de evaluación, cuya experiencia comienza ya a ser importante.

Una mirada a los sistemas nacionales de evaluación

Dando un paso más en el análisis, conviene que nos preguntemos por las principales características que presentan esos sistemas nacionales de evaluación que tanto auge adquirieron durante la última década. Tres son las preguntas clave que debemos hacernos, que están ligadas respectivamente a otras tantas dimensiones de la evaluación: ¿para qué se evalúa?, ¿qué se evalúa? y ¿qué uso se hace de la información?

a) Una multiplicidad de propósitos

La primera pregunta tiene que ver con los propósitos que orientan la evaluación de la educación. Y anticipando la respuesta, hay que reconocer que los sistemas nacionales de evaluación persiguen una multiplicidad de propósitos, al orientarse hacia la consecución de uno o varios de los siguientes objetivos:

- Ayudar a las escuelas y a los profesores a mejorar la calidad de la educación que imparten, proporcionándoles una información relevante y significativa que les permita valorar su situación real y sus logros.
- Conocer y valorar en qué medida se están alcanzando los objetivos educativos o los estándares propuestos para los distintos niveles o parcelas del sistema educativo, contribuyendo al mismo tiempo a establecer metas y niveles de referencia.
- Forzar a las instituciones escolares y a los agentes educativos a generar dinámicas de cambio, por medio de la comparación o la competencia.
- Rendir cuentas a la sociedad o a las autoridades educativas acerca de los logros conseguidos, eliminando así la opacidad del sistema educativo y de las instituciones escolares.
- Establecer sistemas de incentivos para las instituciones y para los profesionales de la educación.
- Certificar, acreditar o seleccionar instituciones, programas de estudio, profesores o estudiantes.
- Conocer y valorar la situación del sistema educativo, así como el impacto producido por las políticas adoptadas.
- Introducir elementos de racionalidad en el debate público y en la construcción de políticas en el ámbito de la educación.

Es cierto que estos propósitos no son en todos los casos compatibles entre sí, pero no lo es menos que las políticas nacionales de evaluación seleccionan algunos frente a otros, les conceden distinto énfasis, establecen prioridades entre ellos o los interpretan de diferente manera. Así, por ejemplo, un propósito compartido por muchos, como es el caso de la incentivación del cambio a través de la comparación, no deja de ser objeto de polémica. Mientras que en algunos países ha llevado, por ejemplo, a la publicación de listas de clasificación de escuelas o universidades, con el ánimo de forzar una competencia abierta por conseguir más estudiantes o más recursos, en otros se ha

rechazado frontalmente esa posibilidad y se han establecido mecanismos de confidencialidad de los datos que impiden ese tipo de uso, aunque permitiendo beneficiarse de la comparación. Como puede apreciarse, un mismo propósito puede inspirar políticas muy diferentes.

b) Una diversidad de actuaciones

La segunda pregunta tiene que ver con la actividad desarrollada por los sistemas nacionales de evaluación. Y también en este caso hay que anticipar la existencia de una considerable diversidad de modalidades de actuación, referidas a su vez a distintos ámbitos educativos. Entre los principales aspectos que son objeto de evaluación en la actualidad, hay que destacar cuatro campos prioritarios:

- Un primer campo que es objeto de considerable atención se refiere a la evaluación de los logros conseguidos por los alumnos. Muchas de las actividades desarrolladas en la actualidad por los sistemas nacionales de evaluación, y sin duda las más importantes de entre ellas, están orientadas a la valoración de los resultados de la educación, medidos a través de los logros que consiguen los estudiantes. En la mayoría de los casos, esas actuaciones se centran en algunas áreas curriculares seleccionadas, especialmente las que se consideran fundamentales para la formación de los jóvenes (lenguaje, matemáticas, ciencias, historia). A veces se complementan con la evaluación de los logros conseguidos en otros ámbitos de la formación escolar, como la educación cívica, el dominio de las tecnologías de la información y la comunicación o las estrategias de aprendizaje. Las evaluaciones suelen estar referidas a etapas o momentos clave del proceso educativo, tales como el tránsito de una etapa a otra o el final de la escolarización obligatoria o de un determinado nivel de estudios. En ocasiones se evalúa a la totalidad de los estudiantes de los grados seleccionados, mientras que otras veces se trabaja solamente con muestras representativas. En cualquier caso, es ésta una línea de actuación prioritaria para los sistemas nacionales de evaluación.
- Un segundo campo que también atrae considerable atención se refiere al rendimiento que logran las instituciones educativas. El interés que ha generado la apertura de la “caja negra” que constituyen las instituciones ha estimulado el desarrollo de diversos programas orientados a su evaluación. Tanto las escuelas primarias y secundarias como las universidades han sido el objetivo prioritario de muchos planes de evaluación y de acreditación puestos en marcha en diversos países latinoamericanos. Esos planes de evaluación tienen características heterogéneas, pero suelen coincidir en conceder una atención especial a los resultados conseguidos por los estudiantes, aunque sin dejar generalmente de lado el análisis de la organización y el funcionamiento de la institución de que se trata.
- Un tercer campo que ha ido desarrollándose recientemente, si bien con mayores dificultades que los dos anteriores, es el relativo al desempeño profesional de los docentes. La convicción de que la actuación de los profesores está en la base de los resultados conseguidos por los estudiantes, aunque no sea su único determinante, ha generado un conjunto de actuaciones de diverso tipo, entre las que también se cuentan algunos programas de evaluación. Se trata de acciones que no han dejado de

generar cierta polémica, tanto por sus características técnicas como por el tipo de uso que se hace de la información obtenida. En muchos casos, quizás en la mayoría, se ha optado por utilizar la evaluación docente en el marco de planes de incentivos profesionales, eludiendo sus aspectos más punitivos, pero en otros casos la evaluación se ha utilizado como instrumento de selección o de promoción profesional, o incluso se ha asociado con mejoras salariales o laborales.

- Un cuarto campo que también ha atraído el interés de los sistemas de evaluación se refiere al impacto de las políticas educativas adoptadas. En un contexto de reformas educativas, como las impulsadas en la región durante los noventa, no es extraño que se haya planteado abiertamente la pregunta acerca de los efectos y el impacto de los procesos de cambio iniciados y de las medidas adoptadas. Desde esta perspectiva, han coexistido dos tendencias que deben diferenciarse, aunque ninguna de ellas haya llegado a atraer tanta atención como las arriba mencionadas. La primera se refiere a la evaluación de los programas más emblemáticos de los procesos de reforma, los denominados “programas estrella”, aquellos en que las autoridades educativas han realizado una mayor inversión y depositado mayores esperanzas. La segunda se refiere a la evaluación de los propios procesos de reforma, con la pretensión de establecer un mecanismo de seguimiento sistemático de sus efectos. Pese a la limitada experiencia desarrollada en ambas direcciones, no cabe duda de que se trata de un campo llamado a experimentar un mayor desarrollo en el futuro próximo.

Aunque los cuatro campos mencionados no agotan la totalidad de los ámbitos abarcados por los sistemas nacionales de evaluación, no cabe duda de que se trata de los más destacados y los que han experimentado un mayor empuje. Nuevamente hay que insistir en que no todos ellos han sido desarrollados en todos los países, ni del mismo modo, ni con la misma energía, pero sí puede afirmarse que son objeto de atención especial por parte de los sistemas nacionales de evaluación.

c) Una pluralidad de usos

La tercera pregunta tiene que ver con el tipo de uso que se hace de la información obtenida por medio de la evaluación. Y al igual que en los dos casos anteriores, también hay que reconocer la existencia de una pluralidad de situaciones. Intentando analizar los usos más habituales de los resultados de la evaluación, podemos distinguir cinco tendencias predominantes:

- Una primera tendencia consiste, por paradójico que pueda resultar, en la renuncia a hacer uso de la información obtenida. Aunque no resulte muy justificable, ni desde el punto de vista de la eficacia administrativa, ni desde la perspectiva de la transparencia, ni menos aún desde las exigencias del comportamiento político democrático, hay que reconocer que existen casos en que los resultados de la evaluación no son publicados, ni difundidos, ni apenas utilizados.
- Una segunda tendencia, que es la más habitual, consiste en la elaboración y difusión de informes de evaluación. Los informes pueden ser de diverso tipo, tener una cobertura variable (nacional, regional, sectorial) y estar dirigidos a audiencias distintas (familias, profesores, autoridades educativas, investigadores y técnicos,

etc.), llegando incluso a adoptar formatos diferentes. Lo importante en este caso no son las características concretas de los informes, que pueden ser de hecho muy diferentes, sino la confianza subyacente en que la difusión pública de los resultados puede satisfacer las demandas de información planteadas por los diversos agentes educativos. En algunos casos se han diseñado y aplicado estrategias bastante completas de difusión, logrando un impacto social y mediático importante.

- Una tercera tendencia consiste en la devolución confidencial de resultados a las instituciones participantes en la evaluación, con el propósito de que puedan utilizar los datos recibidos para elaborar sus propios planes de mejora, sin verse necesariamente expuestas al escrutinio público. Los modos concretos que puede adoptar esa devolución son muy variables, pero lo que caracteriza esa modalidad de actuación es su confianza en el valor intrínseco que encierra la comparación y en la capacidad de reflexión y de mejora que tienen las instituciones y los agentes de la educación cuando se enfrentan con su propia realidad.
- Una cuarta tendencia, no excesivamente habitual y desde luego controvertida, consiste en la publicación de los resultados obtenidos, en forma de tablas de clasificación. En algunos casos la publicación se realiza por los responsables de la evaluación, mientras que en otras ocasiones se permite o incluso se favorece que sean otros agentes (prensa, investigadores) quienes publiquen la información. Este tipo de uso resulta más amenazador para las instituciones educativas que los anteriores, motivo por el cual ha recibido críticas y generado resistencias.
- Una quinta tendencia consiste en la utilización de los resultados de la evaluación para proporcionar orientación pedagógica a instituciones y profesores. Este tipo de uso de la información obtenida a través de la evaluación adopta dos modalidades distintas. En primer lugar, son varios los sistemas nacionales que han elaborado instrumentos tales como cuadernillos de asesoramiento a los profesores, en los que se analizan las respuestas correctas de las pruebas, así como los errores más habituales que suelen producir los estudiantes, y se finaliza ofreciendo sugerencias didácticas más o menos concretas. En segundo lugar, aunque sea de forma más esporádica, también se han desarrollado acciones de orientación para instituciones determinadas, ayudándolas a elaborar sus propios planes de mejora, y se han incluido análisis de los resultados de la evaluación en programas de formación docente. Este tipo de uso, que es menos tradicional que la difusión de informes, está recibiendo una atención creciente por parte de diversas administraciones educativas.

Igual que sucedía en los epígrafes anteriores, también hay que mencionar aquí que los usos mencionados no son necesariamente excluyentes, sino que son cada vez más las administraciones educativas que los combinan en distintos modos. Es cierto que algunos de ellos resultan contrapuestos, como puede ser la alternativa existente entre la publicación de los resultados de las escuelas y la entrega confidencial de los datos, pero no siempre es el caso. De hecho, la combinación de la publicación de informes, la devolución de resultados a las instituciones participantes en la evaluación y la elaboración de instrumentos de orientación didáctica constituye una realidad bastante extendida en los países de la región.

A la vista de la situación que se acaba de exponer cabe extraer una primera conclusión, consistente en que los sistemas nacionales de evaluación de la región latinoamericana no recurren a estrategias de acción uniformes ni adoptan enfoques de evaluación idénticos. Y esta conclusión no es sólo una confirmación de los hechos que se han presentado, sino que se debe incluso considerar lógico que exista tal variedad. En efecto, si aceptamos la existencia de una multiplicidad de propósitos, una diversidad de actuaciones y una pluralidad de usos, debemos aceptar la diversidad interna de los sistemas de evaluación. Serán precisamente los propósitos concretos que se persigan, los estilos de actuación que se adopten y los usos que se prefieran, los que determinarán las características del sistema de evaluación.

Una mirada a las pruebas nacionales de logros

Como se indicaba en el apartado anterior, los sistemas nacionales de evaluación que se han desarrollado en los países latinoamericanos desde los años noventa vienen prestando una atención especial a la medición de los logros conseguidos por los alumnos. Los instrumentos que se aplican con esa finalidad son las genéricamente denominadas *pruebas nacionales*, que merecen una mirada especial por la importancia que han llegado a adquirir.

Las pruebas nacionales presentan bastantes diferencias de unos países a otros, aunque también ofrecen algunas similitudes. Entre sus elementos comunes, quizás el más destacado sea la preocupación prácticamente universal que demuestran por evaluar el rendimiento alcanzado en Lenguaje y Matemáticas. No obstante, aunque se trata de dos áreas especialmente atendidas por los sistemas de evaluación, no hay que pensar que sean los únicos dominios evaluados. En efecto, la medición de los logros en ambas áreas suele ir acompañada por la evaluación de otras áreas curriculares complementarias (entre las que sobresalen las Ciencias, la Historia o las Lenguas extranjeras) o de otras capacidades de tipo transversal, no ligadas necesariamente a áreas específicas (como el autoconcepto, algunas estrategias de aprendizaje o ciertas actitudes).

Las pruebas nacionales suelen aplicarse preferentemente en los grados terminales de las principales etapas educativas (sobre todo, al final de la enseñanza primaria y de la secundaria). No obstante, también se aplican a veces en otros grados que tienen un especial interés desde el punto de vista diagnóstico (por ejemplo, el final de ciertos ciclos), aunque la identificación de tales puntos focales varíe de unos sistemas educativos a otros. El reciente desarrollo de algunos estudios internacionales ligados a la edad de los evaluados y no al grado que cursan, como es el caso del proyecto PISA de la OCDE, aún no ha encontrado traducción directa en los sistemas nacionales de evaluación latinoamericanos, ni es previsible que lo haga a corto plazo. El interés que despierta la valoración de los logros conseguidos al final de determinadas etapas supera hoy en la región al que suscita la valoración del nivel de formación a ciertas edades. La participación en aquellos proyectos internacionales parece satisfacer suficientemente las preguntas que puedan plantearse en este último sentido.

Dependiendo del propósito que guía a la evaluación, las pruebas se aplican en ocasiones a poblaciones o cohortes estudiantiles completas, mientras que en otros casos se limitan a muestras representativas de estudiantes. La primera decisión permite hacer

un uso más singularizado de los datos obtenidos, tal como comparar el rendimiento de diversas escuelas o estudiantes, mientras que la segunda es más adecuada para usos diagnósticos. Por otra parte, hay países en que las pruebas se aplican cada año, aunque es más frecuente que se adopten ciclos plurianuales, de distinta duración.

Las pruebas nacionales parecen satisfacer ampliamente a las autoridades educativas, a la vista del uso tan extenso que se hace de ellas. No obstante, en un análisis más profundo también presentan insuficiencias y problemas, que requieren una consideración más atenta.

Tres son las principales insuficiencias de las pruebas nacionales de logros que han subrayado los especialistas (Ravela, 2000). La primera consiste en la existencia de algunas debilidades técnicas en los procesos de desarrollo y validación de los instrumentos de medición. La segunda se refiere a la insuficiente calidad y capacidad que las pruebas ofrecen para evaluar aprendizajes complejos. La tercera tiene que ver con el insuficiente aprovechamiento que suele hacerse de la información obtenida.

Los problemas a los que se ha hecho alusión son de distinto tipo. Para comenzar, habría que señalar algunos relativos al diseño y confección de los instrumentos de evaluación:

- Un primer problema tiene que ver con el contenido de las pruebas, asunto que plantea dos tipos de dificultades. La primera se refiere a la relación que debe existir entre las pruebas y los objetivos educativos. Lo deseable es que las pruebas estén referidas a los objetivos establecidos en el sistema educativo, de manera que permitan valorar el grado en que se alcanzan las metas previstas. Aunque este planteamiento resulta lógico y plausible, plantea dificultades cuando no existen objetivos explícitos (lo que sucede en muchos casos), o cuando los objetivos de cada curso o etapa se formulan de manera poco concreta. Cuando no se pueden determinar los objetivos de forma precisa, resulta sin duda más difícil elaborar las pruebas. La segunda dificultad se refiere al tratamiento de la diversidad curricular existente entre diferentes regiones, distritos o escuelas. La mayoría de los modelos curriculares aplicados en América Latina conceden cierto grado de autonomía a las autoridades regionales o distritales o a las propias escuelas para definir el currículo. En consecuencia, no resulta sencillo hacer unas pruebas que abarquen todas las enseñanzas efectivamente impartidas en todo el sistema educativo. Las soluciones posibles son dos: elaborar pruebas de mínimos, lo que implica reducir el campo de evaluación, o elaborar pruebas comunes a partir de un consenso profesional o técnico. Ambas soluciones son utilizadas en la región, sin que quepa conceder prioridad a una de ellas sobre la otra.
- Un segundo problema tiene que ver con algunas disyuntivas que se plantean en relación con el diseño y la confección de las pruebas. La primera disyuntiva se refiere al enfoque de las pruebas, que pueden estar centradas en los conocimientos adquiridos por los alumnos o abarcar también el desarrollo de actitudes. Las pruebas que miden unos y otras no son iguales, ni siquiera plantean las mismas exigencias técnicas, lo que obliga a decidir su enfoque. Una segunda disyuntiva se refiere al carácter de las pruebas, pudiéndose optar por la elaboración de pruebas referidas a

criterios de logro o por la confección de pruebas de tipo normativo, adaptadas a una distribución de resultados estadísticamente normal. La tercera disyuntiva se refiere a la delimitación de la población sujeta a evaluación, que permite optar por pruebas de aplicación censal (dirigidas a toda la población estudiantil de un determinado grado o edad) o de tipo muestral (dirigidas a muestras representativas de población).

Otro grupo importante de problemas están relacionados con la explicación de los resultados conseguidos. La mayoría de los sistemas nacionales de evaluación no suelen contentarse con medir los logros alcanzados por los estudiantes, sino que también pretenden explicar por qué se producen tales resultados. Sin embargo, dicho intento de explicación tropieza con varios problemas y dificultades:

- Un primer problema tiene que ver con el grado de coherencia realmente existente entre la evaluación y el currículo. Como se señalaba al comienzo del trabajo, el desarrollo de los sistemas nacionales de evaluación ha ido generalmente asociado a otros procesos de reforma curricular. No obstante, esa simultaneidad no ha asegurado siempre la existencia de una adecuada coherencia entre ambos aspectos. En los casos más extremos, puede incluso decirse que ambos procesos se han desarrollado en direcciones divergentes. Cuanto más importancia se ha concedido a las pruebas, mayor ha sido su impacto sobre el desarrollo curricular.
- Un segundo problema, menos complejo pero no menos influyente, está relacionado con el tipo de interpretación que permiten los diversos modelos de pruebas. Como se señalaba más arriba, dos son los modelos de pruebas utilizados más habitualmente: criteriosales y normativas. Las pruebas de tipo criterial proporcionan una base sólida para valorar la suficiencia de los logros conseguidos, ya que se apoyan en una definición operativa de los objetivos que deben alcanzarse. Por el contrario, las pruebas de tipo normativo exigen alguna operación adicional para poder valorar la suficiencia de los logros alcanzados, dado que se refieren a situaciones de normalidad estadística. La existencia de estos dos modelos alternativos obliga a considerar cuidadosamente qué usos se quieren hacer de las pruebas, antes de decidir el que conviene aplicar.
- Un tercer problema se refiere a las dificultades conceptuales y técnicas que se plantean a la hora de buscar explicaciones. Por una parte, hay que reconocer la debilidad teórica de los modelos habituales de producción educativa, que en buena parte deriva del problema que plantea la causalidad en educación. Esa debilidad influye en la selección de los factores que se consideran asociados al rendimiento, pocos de los cuales cuentan con confirmación empírica sólida (Scheerens, 1996). Además, la explicación de los resultados se ha basado tradicionalmente en la distinción de dos tipos de variables, unas extrínsecas y otras intrínsecas. Mientras que las primeras (nivel socioeconómico y cultural, recursos destinados a la educación) han sido objeto de análisis abundantes, las segundas (organización del sistema educativo, procesos institucionales, procesos de aula) vienen resultando menos concluyentes y aún requieren investigaciones adicionales.
- Un cuarto problema tiene que ver con la voluntad que a veces existe de buscar muchas vías de explicación de los resultados obtenidos, que puede traducirse en un

exceso de datos recogidos. Cuando ese exceso de datos va unido, como ocurre en ocasiones, con una insuficiente delimitación conceptual de las variables seleccionadas y con limitaciones en el análisis estadístico de los datos, la situación puede llegar a resultar inmanejable. La consecuencia suele ser una evidente infrutilización de los datos y una ineficiencia en el uso de los recursos disponibles.

Un último grupo de problemas que plantean las pruebas de logros están relacionados con el uso que se hace de las mismas y con las interpretaciones a que dan pie:

- Un primer problema tiene que ver con los distintos tipos de uso que pueden hacerse de la información. Los resultados de la evaluación pueden utilizarse con una finalidad formativa, orientada al desarrollo institucional y profesional de los agentes implicados, para fomentar la competición, e incluso para impulsar políticas de mercado en el ámbito educativo. Obviamente, las interpretaciones que permiten tales posibilidades son diferentes y persiguen distintos efectos. Es necesario señalar que, si bien esos son los extremos de un continuo de posibilidades de uso, existen otras opciones intermedias en que dichos propósitos pueden combinarse en diferente proporción. La decisión sobre el tipo de uso que se pretende hacer de los datos de la evaluación condiciona en buena medida el diseño de la evaluación.
- Un segundo problema tiene que ver con la existencia de diversas posibilidades de comparación (con una norma o criterio, con otras realidades semejantes, consigo mismo a lo largo del tiempo), cada una de las cuales plantea sus propias exigencias y ofrece diferentes posibilidades. El uso de la comparación plantea un problema complicado, que tiene que ver con la justicia de la comparación. El intento de dar respuesta a esa cuestión ha impulsado el desarrollo de técnicas de cálculo del denominado *valor añadido*.
- Un tercer problema consiste en el excesivo reduccionismo en que a veces se cae cuando se pretende interpretar los resultados de la evaluación. Ese defecto puede venir originado por una excesiva limitación del contenido de las pruebas, que no permite llegar a conclusiones razonables sobre el logro de los objetivos propuestos, o por una interpretación poco rigurosa de los datos, que puede incluso llegar a distorsionar las interpretaciones.

A la vista del análisis que acaba de hacerse de las pruebas nacionales de logros, se llega a la conclusión de que constituyen la práctica de evaluación más habitual en la región, encontrándose bastante asentadas en la actualidad. No obstante, como se ha puesto de manifiesto, presentan algunas deficiencias y problemas que deberán solventarse en el futuro, si se quiere asegurar su credibilidad y se pretende que contribuyan a mejorar la educación.

Algunas reflexiones finales

Si bien es cierto que las pruebas nacionales de logros constituyen la práctica más habitual de evaluación educativa en América Latina, no es la única que se desarrolla en la actualidad. Como se indicaba en las primeras páginas del trabajo, existen otros tres

ámbitos que son también objeto de evaluación. El primero corresponde a los programas y las políticas de educación, que despiertan especial interés en el actual contexto de revisión de las reformas educativas puestas en marcha durante los años noventa. En sus formas más elaboradas ha servido para diseñar mecanismos de seguimiento de los procesos de cambio, aunque se trata de una práctica escasamente implantada. El segundo ámbito corresponde a las instituciones educativas, para cuya evaluación se han diseñado y aplicado algunos modelos, tradicionales o innovadores, aunque sin que pueda decirse que esta tendencia haya alcanzado todavía suficiente desarrollo. El avance más importante que se ha logrado en este sentido consiste en la expansión de una conciencia nueva acerca del valor que tiene la evaluación institucional con vistas a la mejora de la calidad de la educación. El tercer ámbito, y al mismo tiempo el más controvertido, corresponde al desempeño docente. Aunque en algunos países se han desarrollado mecanismos de valoración de la tarea docente que están conectados con la medición del rendimiento de los estudiantes, no existen todavía experiencias suficientes ni de larga tradición en este sentido. Las experiencias que han explorado otras direcciones son aún más escasas.

En conjunto, puede afirmarse que estos otros tres ámbitos abarcados por la evaluación educativa están actualmente infradesarrollados en relación con las pruebas de rendimiento. Hay que subrayar, no obstante, que la evaluación de tales aspectos puede contribuir notablemente a la mejora de la calidad de la educación. Precisamente de esa potencialidad procede el interés que despiertan en muchos países de la región.

Como conclusión final, cabe afirmar que la evaluación es una realidad integrada que cuenta con varios componentes: estudios internacionales de evaluación, estudios diagnósticos nacionales del rendimiento, evaluación de las instituciones educativas y otras evaluaciones complementarias de diversos aspectos de la actividad educativa (preparación y desempeño de los docentes, impacto de las políticas aplicadas, funcionamiento de algunos programas relevantes). Aunque no todos ellos se hayan desarrollado con la misma energía, los sistemas nacionales de evaluación no deberían infravalorar la importancia de concebirlas de forma integrada, de manera que las acciones emprendidas se refuercen mutuamente.

El colofón a estas reflexiones que provoca la mirada a los sistemas nacionales de evaluación educativa consistiría en que la evaluación debe considerarse como un instrumento para la mejora de la calidad de la educación y no como una panacea capaz de solucionar todos los problemas que la educación plantea. Otra consideración diferente de ésta no puede sino conducir a algunos errores de cierta importancia.

Referencias bibliográficas

- Braslavsky, C. (2001): *Educación secundaria: ¿cambio o inmutabilidad?*, Buenos Aires, Santillana.
- Comisión Internacional sobre Educación, Equidad y Competitividad Económica en América Latina y el Caribe (2001): *Quedándonos atrás. Un informe del progreso educativo en América Latina*, Santiago de Chile, PREAL.

- Gajardo, M. (1999): *Reformas educativas en América Latina. Balance de una década*, Santiago de Chile, PREAL (documento de trabajo n° 15).
- Ravela, P., ed. (2000): *Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina?*, Santiago de Chile, PREAL.
- Ravela, P. (2001): *¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?*, Santiago de Chile, PREAL.
- Scheerens, J. (1996): “Can the School Effectiveness Knowledge Base Guide School Management?”, en *II International Conference on School Management. Participative Management and School Evaluation*, Bilbao, Universidad de Deusto, pp. 98-119.
- Tiana, A. (1996): “La evaluación de los sistemas educativos”, *Revista Iberoamericana de Educación*, n° 10, 1996, pp. 37-61.
- Tiana, A. (2000): *Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación*, Washington, BID – Departamento de Desarrollo Sostenible.
- Winkler, D. y Gershberg, A.I. (2000): *Los efectos de la descentralización del sistema educacional sobre la calidad de la educación en América Latina*, Santiago de Chile, PREAL (documento de trabajo, n° 17).