

EVALUACIÓN DEL APRENDIZAJE DE LOS ESTUDIANTES INDÍGENAS EN AMÉRICA LATINA

Desafíos de medición e interpretación en contextos de diversidad cultural y desigualdad social¹

ERNESTO TREVIÑO VILLARREAL

Resumen:

Este artículo analiza la problemática de la evaluación del aprendizaje, usando pruebas estandarizadas, en poblaciones indígenas desde el punto de vista de las oportunidades educativas. Combinando conceptos de la psicología transcultural, la antropología y la psicometría, se presentan los principales factores que afectarían las inferencias realizadas a partir de los resultados de las pruebas aplicadas a estudiantes indígenas. Finalmente, se resumen las limitaciones de los exámenes estandarizados y se presentan algunas recomendaciones para circunscribir las inferencias hechas a partir de los resultados en estas pruebas.

Abstract:

This article analyzes the problems of using standardized tests to evaluate the learning of indigenous populations, from the point of view of educational opportunities. By combining concepts of transcultural psychology, anthropology and psychometrics, the article presents the principal factors that would affect inferences made from the results of tests given to indigenous students. Lastly, the limitations of standardized tests are summarized, and some recommendations are provided for circumscribing the inferences based on the results of such tests.

Palabras clave: evaluación del aprendizaje, educación intercultural, igualdad de oportunidades, pruebas de aprovechamiento, psicometría.

Key words: evaluation of learning, intercultural education, equal opportunities, achievement tests, psychometrics.

Introducción

Las poblaciones indígenas han sido históricamente las más desatendidas por los sistemas educativos de América Latina. Asimismo, en esta región del mundo escasea la investigación sobre la relación entre estos gru-

Ernesto Treviño es subdirector de Evaluación de Factores Escolares de la Dirección de Evaluación de Escuelas del Instituto Nacional para la Evaluación de la Educación, Av. Adolfo López Mateos 437 Ote, col. Centro, CP 20000, Aguascalientes, Ags., CE: ernesto.trevino@ineemexico.org

pos indígenas y los sistemas educativos. Las poblaciones indígenas de América Latina, en general, viven condiciones de pobreza más agudas en comparación con la población no indígena. Cifras del año 2002 indican que en Bolivia tres cuartas partes de los indígenas son pobres, mientras que la mitad de los no indígenas viven tal situación. Asimismo, la tasa de extrema pobreza es de 52% entre la población indígena boliviana y de 27% entre los no indígenas (Hall y Patrinos, en prensa). Las tasas de pobreza en Ecuador son de 84% para los indígenas y 46% para los no indígenas y el ingreso mensual de los hombres indígenas representa solamente 55% de los no indígenas (Winkler, 2004). Los indicadores de pobreza varían entre los distintos grupos étnicos mayas de Guatemala. Por ejemplo, el índice de pobreza entre distintos grupos indígenas se ubica en un rango que va desde 63% para los kaqchikel hasta 90% para los mam.

Por otro lado, 42% de la población no indígena está clasificada como pobre. La situación es aún más desigual al comparar la extrema pobreza, pues el rango de este indicador abarca desde 14% para los kaqchikel hasta 38% de los q'eqchi, y 8% para la población no indígena (Winkler, 2004). Durante el año 2002, en México, 90% de los indígenas y 47% de los no indígenas vivían en condiciones de pobreza. En ese mismo año, la incidencia de extrema pobreza fue de 69% para los indígenas y 15% para los no indígenas (Hall y Patrinos, en prensa). Estas cifras revelan que las poblaciones indígenas, además de tener una alta tasa de pobreza, viven claramente en desventaja en comparación con la población no indígena. Esta situación determina, en gran medida, las escasas oportunidades de movilidad educativa y social de los grupos indígenas.

Cualquiera de los indicadores educativos que se analice muestra que los grupos indígenas se encuentran en desventaja. Las cifras siguientes pretenden ejemplificar la situación de marginación educativa que afecta a los estudiantes indígenas, aun antes de someterse a cualquier tipo de evaluación estandarizada. En términos de asistencia a la escuela, en México 15% de los niños y niñas entre 6 y 14 años de edad no asiste, en comunidades donde la población indígena representa 70% o más del total de habitantes. Esto es aproximadamente el doble de la tasa de inasistencia nacional que se ubica en 7.7% (INEGI, 2000). En Perú, cifras del año 2000 indican que 11% de la población quechua —grupo indígena mayoritario en ese país— en el rango de edad de 6 a 17 años no asiste a la escuela. En contraste, 9.3% de los hispanohablantes en ese rango de edad no asiste a la escuela (Kudo, 2002).

La inasistencia escolar se traduce, en el largo plazo, en desigualdades con respecto a los años de escolaridad entre indígenas y no indígenas. En 1999 en Ecuador la brecha de la población adulta era de 3.5 años más para la no indígena. Por su parte, en Guatemala, el promedio de los no indígenas es de 5.5 años, más del doble de los 2.5 años de la población indígena (Winkler, 2004). En México se ha encontrado que, al comparar ambos grupos en áreas rurales en el año 2000, hablar una lengua indígena está negativamente relacionado con la probabilidad de terminar la primaria, aun después de controlar las características socioeconómicas de los alumnos (Mier y Terán y Rabell, 2003). Es necesario precisar que las desigualdades entre indígenas y no indígenas en cuanto a años de escolaridad han disminuido recientemente, en parte gracias a la expansión de los sistemas educativos de América Latina (Mier y Terán y Rabell, 2003; Winkler, 2004). Los indicadores aquí expuestos ponen de manifiesto las desventajas educativas que pesan sobre las poblaciones indígenas en América Latina, quienes tienen también los logros más bajos en los exámenes estandarizados, como se verá enseguida.

El aprovechamiento académico en exámenes estandarizados de la población indígena es significativamente menor que el de la no indígena. En Ecuador, los estudiantes del primer grupo de 5º grado obtienen puntajes inferiores en 20% a los de sus compañeros no indígenas. Los resultados en Perú son similares, puesto que los indígenas en 4º grado tienen puntajes 15% más bajos en relación con la población no indígena (Winkler, 2004). Los estudiantes indígenas de México que cursaron el 6º grado de primaria en el ciclo escolar 2003-2004 obtuvieron puntajes con una desviación estándar por debajo de sus compañeros no indígenas, según las pruebas de estándares nacionales aplicadas por el Instituto Nacional para la Evaluación de la Educación.

Es precisamente el punto de la interpretación de los resultados de estudiantes indígenas en pruebas estandarizadas el tema central de este artículo. El presente estudio busca llenar un vacío en la literatura educativa latinoamericana al combinar el análisis de la equidad y la diversidad en torno a la evaluación de los aprendizajes de la población indígena. La mayor parte de la literatura en educación indígena toca la equidad o la diversidad como dos conceptos mutuamente excluyentes. Por tanto, estos dos temas no han sido explorados de manera conjunta, considerando las tensiones y las coincidencias que existen entre ambos. Un enfoque dialéctico entre

diversidad y equidad al analizar el tema de la evaluación de aprendizajes de grupos indígenas permite un mejor entendimiento de los dilemas que se enfrentan para proveer oportunidades educativas en un contexto de diversidad cultural.

El presente trabajo es un estudio teórico acerca de la medición del aprendizaje de los estudiantes indígenas de América Latina mediante exámenes estandarizados. El estudio tiene por objetivo sentar las bases para profundizar la investigación sobre la medición del aprendizaje de los estudiantes en contextos indígenas. En este documento se propone que las inferencias que se pueden hacer a partir de los resultados de las poblaciones indígenas en exámenes estandarizados son distintas a las relativas a las poblaciones no indígenas. Las diferencias culturales entre ambos grupos influyen en los resultados de los exámenes; de esta manera, para los integrantes de las culturas indígenas las pruebas no solamente miden los constructos que se proponen, sino que sus resultados reflejan una mezcla de las características específicas de cada cultura y el constructo que se pretendía medir. Por lo tanto, los resultados de los indígenas en estos exámenes deberían conllevar interpretaciones distintas a las que se darían para otras poblaciones. Es necesario, entonces, calificar cuidadosamente los resultados de los indígenas en estos exámenes, y queda a la investigación la tarea de desentrañar los efectos de los factores culturales y los constructos que mide el examen. Las pruebas estandarizadas podrían proporcionar información acerca de los aspectos de la cultura dominante que los estudiantes indígenas han aprendido a través de la escuela, pero las limitaciones de los exámenes hacen difícil que éstos brinden información sobre distintas formas de aprendizaje y conocimientos de las culturas indígenas. Finalmente, las inferencias acerca de los resultados en los exámenes podrían tener consecuencias sobre la equidad en la educación, por lo que la calificación cuidadosa de tales inferencias podría ser un valioso instrumento para promover las oportunidades educativas de los niños y niñas indígenas.

Si bien en distintos países se han realizado estudios en cuanto a la evaluación educativa mediante exámenes estandarizados en contextos de diversidad, las contrastantes diferencias culturales, sociales y económicas entre los grupos indígenas y el resto de la población en América Latina requieren de una especificación particular de este problema. En general, los estudios de este estilo se han realizado en países desarrollados, con fuertes

sistemas de apoyo social a las poblaciones más desfavorecidas –como pueden ser los casos de Estados Unidos, Canadá, España, Bélgica y Suiza–. Sin embargo, en el contexto latinoamericano, el más desigual del mundo de acuerdo con cifras de Naciones Unidas, la situación de los pueblos indígenas es de marginación extrema. Además, los contrastes culturales entre los grupos indígenas y la población dominante en América Latina son especialmente agudos, pues ellas se fundan en paradigmas de aprendizaje completamente distintos, como se presentará enseguida. Por otro lado, la relación histórica entre los pueblos indígenas y las culturas dominantes data de más de 500 años de marginación, exclusión y destrucción de las culturas indígenas (Falcón, 2002). Dicha relación es cualitativamente distinta a aquella que se encuentra entre las distintas poblaciones de los países desarrollados. Finalmente, este estudio mantiene como premisa que es necesaria una reflexión que considere tanto las condiciones específicas de las poblaciones indígenas en América Latina como la situación que enfrentan las áreas de evaluación educativa en los países de la región que, en su mayoría, son de reciente creación o consolidación.

La experiencia acumulada en el campo de la psicología transcultural puede colaborar para esclarecer algunos de los tópicos más controversiales en el debate de la medición del logro educativo entre culturas distintas a la occidental. Como se verá más adelante, la psicología transcultural se ha enriquecido al incorporar elementos de la antropología que han permitido una comprensión más profunda de los rasgos particulares de cada cultura. Esta experiencia puede ayudar a evitar los errores que se cometieron en los primeros trabajos transculturales en el campo de la psicología, especialmente con respecto a las inferencias a que se llegaba usando instrumentos estandarizados de medición.

Una de las premisas de este documento es el reconocimiento de la tensión entre interpretaciones modernas y posmodernas del concepto de equidad en relación con los exámenes estandarizados en contextos de diversidad cultural. Por un lado, los enfoques posmodernos consideran que estos exámenes discriminan a aquellas personas que no pertenecen a la cultura dominante, ya que sirven como instrumentos de exclusión de las culturas marginales que, simultáneamente, legitiman la posición de los grupos sociales dominantes. Dentro de este marco lógico, la equidad se alcanzaría cuando la diversidad de culturas que componen una sociedad sean igualmente valoradas. De darse una situación de igualdad entre culturas, la

aplicación de exámenes estandarizados perdería sentido, dado que entra en contradicción con la valoración de la diversidad. Por otro lado, concibiendo la equidad desde el enfoque modernizador, los exámenes estandarizados tienen un papel importante, dado que miden las habilidades y conocimientos que los estudiantes necesitarían para aumentar sus posibilidades de movilidad educativa y social. Este marco teórico supone que las estructuras sociales que impiden la valoración igualitaria de las distintas culturas no cambiarán en el corto plazo, incluso se podría pensar que la valoración igualitaria de las culturas no es un objetivo en este enfoque. Como consecuencia, los estudiantes necesitan adquirir los conocimientos y valores de la cultura dominante, que en este caso es modernizadora.

En la exposición del argumento de este estudio se presenta primero una clasificación de estas pruebas en función de distintos criterios. En segundo lugar se enfatiza la importancia de los exámenes estandarizados en los sistemas educativos contemporáneos, las comparaciones internacionales y las tensiones entre la evaluación estandarizada y el creciente reconocimiento legal de la diversidad cultural. Enseguida se sitúa el tema de la evaluación estandarizada en contextos indígenas dentro del campo de la psicología transcultural. En cuarto lugar se comparan las prácticas educativas de las comunidades indígenas con aquellas que se desarrollan en la educación formal. A continuación se analizan los distintos factores que afectan la validez de las inferencias basadas en la evaluación estandarizada en contextos indígenas, así como sus implicaciones en la equidad. Dentro de este apartado se centrará la atención en el rol que toman las diferencias culturales en la validez de ciertas inferencias. En sexto lugar se presentan algunas preguntas y propuestas en torno a la equidad y la evaluación en contextos indígenas. Al final del documento se incluyen conclusiones generales.

Si bien a lo largo de este documento se usan los términos de cultura indígena, grupos indígenas, comunidades indígenas y poblaciones indígenas, no se pretende sugerir que estas culturas son iguales entre sí. Estos términos se usan para simplificar la exposición, por lo que acudo a la amabilidad de los lectores para que consideren la inmensa gama de culturas indígenas presentes en América Latina. Si bien hay desacuerdos en la forma de contabilizar a los indígenas, se estima que en América Latina existen alrededor de 400 grupos étnicos y lingüísticos, lo que habla de la heterogeneidad cultural en la región (IADB, 2004).

Tipos de exámenes

Los exámenes pueden clasificarse en dos grandes tipos de acuerdo con las consecuencias que dichas pruebas tienen en la vida de los estudiantes u otros actores educativos. Por un lado, existen sistemas de evaluación donde los resultados en los exámenes estandarizados tienen escasas o nulas consecuencias (*low-stakes*) para los actores educativos (estudiantes, docentes o escuelas). En estos casos, la evaluación cumple propósitos informativos. Por otro lado, existen sistemas de evaluación en donde los resultados de las pruebas se utilizan para castigar o incentivar a los actores educativos (*high-stakes*) (Cizek, 1998; Ravela, 2001). Las consecuencias de los exámenes dependen directamente del uso que se le da a los resultados de las pruebas. Por ejemplo, los exámenes tienen altas consecuencias en instancias donde los resultados de las pruebas se usan como instrumentos de selección para admitir estudiantes a alguna institución o en los casos donde se usan para acreditar la consecución exitosa de un nivel educativo. Los exámenes, aun en sistemas de evaluación de consecuencias nulas, tienen impacto sobre la apreciación social y/o las oportunidades educativas de los alumnos.

En Estados Unidos, donde existe una tradición de examinación de estudiantes con propósitos de rendición de cuentas que data de mediados del siglo XIX (Hamilton y Koretz, 2002; Kirby y Stecher, 2004), las políticas de aplicación de exámenes de alto impacto –revitalizadas a finales del siglo XX– están ampliando las desigualdades educativas. La efervescencia de los exámenes para la rendición de cuentas tuvo su corolario en el paquete de reformas denominado *No Child Left Behind* (que ningún niño se quede atrás) que inició en diciembre de 2001 e impuso a los estados de la Unión Americana el requisito de aplicar pruebas de alto impacto a estudiantes de educación básica y media. Esto bajo la lógica de que la evaluación con consecuencias motivará a los docentes, mejorará la gestión de las escuelas y el aprendizaje de los alumnos (Fusarelli, 2004; Kirby y Stecher, 2004).

No obstante, los trabajos sobre las consecuencias de los exámenes estandarizados de alto impacto indican que existen efectos perniciosos del modelo de rendición de cuentas en detrimento de los estudiantes pobres, inmigrantes y/o pertenecientes a minorías étnicas. Específicamente, los estudios han encontrado un efecto diferencial de la política de rendición de cuentas en el cual los estudiantes de sectores socioeconómicos medio y

alto aumentan sus niveles de aprendizaje. Simultáneamente, los docentes y directivos de las escuelas que concentran a estos estudiantes responden al modelo de rendición de cuentas conforme a lo esperado, es decir, se ven motivados y orientan sus acciones hacia la mejora. Por otro lado, los estudiantes de estratos socioeconómicos bajos, inmigrantes y/o pertenecientes a minorías étnicas, han mostrado rendimientos a la baja en los exámenes, así como una mayor tasa de deserción y menores niveles de escolaridad. Asimismo, las escuelas que sirven a estos alumnos tienden a poner en marcha políticas que castigan a los estudiantes más desaventajados y enfatizan la práctica repetitiva de ejercicios similares a los que se incluyen en los exámenes estandarizados por sobre el desarrollo de las habilidades y competencias que se supone miden los exámenes (Apple, 2001; Diamond y Spillane, 2004; McNeil, 2000; Muller y Schiller, 2000; Roderick *et al.*, 1999). A juzgar por esta evidencia, en un entorno hipotético con un sistema de exámenes de alto impacto es indudable que los estudiantes indígenas serían los grandes afectados por esta medida.

Dependiendo de las consecuencias que tienen los resultados de los exámenes, las poblaciones indígenas pueden encontrarse en, al menos, dos casos. En primer lugar, en un ámbito de medición de nulas consecuencias, los alumnos indígenas podrían ser estereotipados como incapaces de transitar exitosamente en el sistema educativo; debido a que exhiben rendimientos más bajos que la población no indígena, tal como se expuso en la introducción. En segundo lugar, en ambientes donde los exámenes tienen altas consecuencias, sus resultados podrían limitar las oportunidades educativas de los alumnos indígenas.

Las pruebas estandarizadas también pueden clasificarse en relación con su referencia a normas o a criterios. Los exámenes referidos a normas intentan comparar el rendimiento de un estudiante en relación con otros que han tomado la prueba previamente y cuyos resultados se utilizaron para crear una escala que sirve para categorizar el logro en percentiles. Los exámenes referidos a criterios son los que tratan de medir el avance de los alumnos con criterios funcionales de desempeño, relacionados con conocimientos o habilidades que el estudiante debería manejar y que se reflejan en puntajes específicos en la prueba (AERA, APA Y NCME, 1999).

Finalmente, las pruebas también pueden dividirse en dos grupos según el contenido que cubran. El primer grupo está constituido por aquellas con referencia a currículo, cuyo objetivo es medir el aprendizaje que han

logrado los alumnos en cuanto un currículo predeterminado. El segundo grupo corresponde a las pruebas sin referencia a currículo, que comúnmente están diseñadas en torno a criterios establecidos por expertos en distintos campos, quienes definen las habilidades, competencias y/o conocimientos que los estudiantes deberían de poseer.

Tensión entre la creciente importancia de los exámenes estandarizados y el reconocimiento de la diversidad

Actualmente, la medición del aprendizaje con exámenes estandarizados es una práctica común que influye directamente en las oportunidades educativas de los estudiantes. En algunos contextos evaluativos los exámenes tienen altas consecuencias, dado que sus resultados se usan para la toma de decisiones tanto en la política educativa como en los recursos financieros que se asignan a las escuelas, los salarios de los docentes y el ingreso o graduación de los estudiantes en los distintos niveles educativos. En Estados Unidos, por ejemplo, el paquete de reforma educativa denominado “Que ningún niño se quede atrás”, exige a los gobiernos estatales que examinen anualmente a todos sus estudiantes desde el primer año de educación básica hasta el último de educación media. Esta reforma propuso el uso de exámenes de altas consecuencias (*high-stakes*) donde las escuelas –y en algunos casos los estudiantes– reciben incentivos o castigos de acuerdo con los resultados en el examen (DOE, 2003). El uso de los resultados de las evaluaciones con propósitos de rendición de cuentas da a los exámenes estandarizados un papel protagónico en la definición de las oportunidades educativas para los estudiantes; dado que influyen en el futuro educativo de niños, niñas y jóvenes por la forma en que las decisiones basadas en los exámenes podrían afectar tanto a las escuelas como al repertorio de opciones educativas disponibles para ellos.

Los países asiáticos son un clásico ejemplo donde los exámenes estandarizados tienen implicaciones para las oportunidades futuras de la población. En esta región del mundo los de ingreso a las instituciones de educación superior constituyen un hito en la vida de los estudiantes. Aquellos que aprueban exitosamente este examen tendrán abundantes oportunidades educativas y mayores probabilidades de seguir una carrera profesional exitosa (Cheng, 1996). El uso de los exámenes estandarizados en el contexto asiático muestra la importancia que los resultados en dichas pruebas tienen en la educación y las posibilidades de trabajo de la población.

En América Latina, los exámenes de ingreso también inciden en las oportunidades educativas se aplican, en muchos países, para el ingreso a la educación media y superior, y se constituyen en filtros que dejan a un número importante de estudiantes sin oportunidad de acceder a estos niveles educativos. Por ejemplo, las instituciones mexicanas de educación media y superior han usado tradicionalmente exámenes de ingreso para seleccionar a los estudiantes. Asimismo, desde la creación del Centro Nacional de Evaluación para la Educación Superior (CENEVAL), distintas universidades públicas en los estados utilizan los mismos exámenes de ingreso (Tirado *et al.*, 1997). En Chile los que buscan ingresar a la educación superior deben tomar la Prueba de Selección Universitaria, que sirve como tamiz para las instituciones de este nivel educativo. Brasil y Colombia, por su parte, también tienen pruebas de admisión a las instituciones de educación superior. En resumen, los exámenes estandarizados parecen afectar las oportunidades educativas de los estudiantes principalmente cuando son usados para seleccionar a aquellos que pueden acceder a los niveles de educación media y superior.

Los exámenes de ingreso podrían entenderse como un filtro piramidal de desigualdades educativas, en el cual los miembros de la cultura dominante alcanzan con facilidad mayores oportunidades, mientras que los de culturas minoritarias son marginados desde las etapas iniciales. Si se conciben las desigualdades educativas como una serie de puntos de transición entre los distintos niveles (Benadusi, 2001), es claro que los exámenes de ingreso restringen la movilidad educativa o la transición hacia los niveles educativos más altos, que son los más valorados dentro de la sociedad. Por lo tanto, los exámenes reducen las oportunidades educativas para aquellos estudiantes con bajos resultados en el de ingreso. Hasta ahora, la investigación en América Latina no se ha enfocado en investigar los efectos de los exámenes estandarizados en las oportunidades educativas para los distintos grupos sociales y culturales.

En América Latina la mayoría de los países cuenta actualmente con áreas de evaluación educativa para monitorear el logro de los estudiantes mediante exámenes sin consecuencias (Ravela, 2001; UNESCO, 2000; Wolff, 1998). Con la excepción de los de ingreso a la educación media y superior, los exámenes que se usan para seguir el desempeño de los planteles no tienen consecuencias sobre los estudiantes, los docentes o las escuelas mismas (Treviño, en prensa).² Sin embargo, los estandarizados han incrementado

su protagonismo desde el inicio de la década de los noventa, y las condiciones actuales sugieren que tomarán un papel aún más importante en el futuro.

Los estudios internacionales de evaluación a través de exámenes estandarizados dan cuenta de la relevancia de la evaluación en contextos multiculturales, en este caso en el ámbito internacional. El Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de la UNESCO, el *Programme for International Student Assessment* (PISA) de la OCDE y el *Trends in International Mathematics and Science Study* (TIMSS) de la IEA son tres ejemplos sobresalientes de proyectos internacionales de evaluación que han despertado el interés de la comunidad educativa mundial. Asimismo, estos proyectos conllevan un fuerte componente de comparabilidad intercultural.

Las experiencias de LLECE, PISA y TIMSS ofrecen luces importantes en el campo de la evaluación en ambientes de diversidad cultural, las cuales trataré de resumir a continuación. En el caso del LLECE los exámenes fueron sometidos a un proceso de consenso entre los países participantes en el que se identificaron contenidos curriculares coincidentes en los 13 países involucrados en el proyecto. Una vez concluido el proceso de consenso, las pruebas fueron certificadas por expertos externos (LLECE, 1998). Estas pruebas fueron traducidas al portugués para su aplicación en Brasil, que también participó del proceso de consenso. El estudio de PISA tradujo y adaptó los exámenes para asegurar que éstos midieran los constructos que pretendían evaluar en cada uno de los países participantes. Análisis estadísticos y consultas con expertos fueron las dos fuentes principales de información para la adaptación de los exámenes. El estudio de PISA consideró, entre otras cosas, la equivalencia de constructos entre las diferentes culturas para la adaptación cultural de las pruebas (Grisay, 2003; McQueen y Mendelovits, 2003). El TIMSS llevó a cabo procedimientos de ajuste de las pruebas para maximizar su comparabilidad a través de los distintos países participantes; para ello se contrataron traductores calificados y se llevaron a cabo procesos de revisión de este trabajo en cada país (Maxwell, 1996).

Los mecanismos para asegurar la comparabilidad de los resultados entre las poblaciones de los distintos países no están exentos de críticas. En el caso del LLECE escasea la documentación que dé cuenta detallada de los procedimientos que se utilizaron, por ejemplo, en los procesos de tra-

ducción. El proyecto PISA, por otro lado, realizó una lista exhaustiva de tareas para asegurar la comparabilidad cultural de los exámenes. El estudio PISA definió los constructos a medir, desarrolló los instrumentos y finalmente realizó estudios piloto para verificar la adecuada medición de dichos constructos en los países participantes. Sin embargo, la selección de constructos se basó primordialmente en habilidades de lectura y matemática, lo que podría haber dejado fuera constructos que son importantes para algunas de las culturas participantes en el estudio. Los resultados del TIMSS 1995 muestran cómo países de culturas afines se aglutinan formando grupos de rendimiento similar, lo que parece asociarse con factores lingüísticos, geográficos e histórico-políticos. Existen cuatro grupos de países claramente definidos: angloparlantes, germanoparlantes, nórdicos, del este de Europa y del este de Asia. Estos resultados llevan a pensar que estos países comparten algunos conceptos en torno a la matemática y las ciencias, así como al aprendizaje y enseñanza de las mismas (Gronmo, Kjaernsli y Lie, 2004).

Finalmente, los tres estudios internacionales citados asumen, implícitamente, que las culturas son unitarias al interior de los países, supuesto que no es correcto para la gran mayoría de las naciones. Por ejemplo, las investigaciones de evaluación nacionales e internacionales comúnmente construyen los exámenes para la medición del aprendizaje en el idioma oficial del país en cuestión, sin considerar que los estudiantes indígenas tienen una lengua materna distinta a la oficial. Por lo tanto, la diferencia en el dominio de la lengua oficial es un factor más que aumenta la disparidad en la evaluación en detrimento de los alumnos indígenas.

En suma, las evaluaciones internacionales han desarrollado procedimientos para la adaptación y/o traducción de los exámenes estandarizados. Esto ha llevado a considerar aspectos que van desde la validez de los constructos hasta análisis estadísticos que permitan determinar la validez de la medición en cierta cultura. La ausencia de especificaciones no-occidentales de distintos constructos en los exámenes estandarizados es, probablemente, la falencia más importante en cuanto a la adaptación de exámenes realizada en estos estudios internacionales.

Bajo las condiciones actuales es fácil prever que los resultados en los exámenes estandarizados tendrán, paulatinamente, mayores consecuencias en la trayectoria educativa de los alumnos, influyendo de manera creciente en las oportunidades educativas de los grupos minoritarios, especial-

mente de las poblaciones indígenas. Las repercusiones que los exámenes estandarizados tienen en las oportunidades educativas hasta el momento no se consideran como un fenómeno prioritario en la investigación educativa en América Latina. Asimismo, la aplicación de exámenes estandarizados se considera como una variable exógena, como un hecho incuestionable que sucede en la vida de los alumnos; esto a pesar del peso que podrían tener en la trayectoria educativa de los niños y niñas.

Los exámenes estandarizados se tornan más problemáticos en el contexto presente donde, de manera progresiva, los países han reconocido la diversidad cultural de sus poblaciones. Por ejemplo, doce países de América Latina han firmado y ratificado la Convención sobre pueblos indígenas y tribales de la Organización Internacional del Trabajo –Argentina, Bolivia, Brasil, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, México, Paraguay, Perú y Venezuela– (ILO, 2003). Dicha convención va más allá de reconocer legalmente a los pueblos indígenas, estableciendo una serie de derechos en cuanto a cultura, acceso a la tierra y políticas sociales, entre otros.

Situando el fenómeno dentro del campo de la psicología transcultural

El uso de exámenes estandarizados para medir el aprendizaje en comunidades indígenas da una perspectiva adicional de la tensión entre los paradigmas *etic* (visión desde afuera) y *emic* (visión desde dentro), que ha estado presente en el seno del campo de la psicología transcultural (Werner, 1979; Suárez-Orozco, 1994). El paradigma *etic* –relacionado con la psicología– busca medir y entender cierto fenómeno usando estándares externos al contexto en el cual se desarrolla dicho fenómeno. Al contrario, la posición *emic* –más relacionada a la antropología– trata de entender cómo cierto fenómeno opera en relación con el contexto y busca comprender los patrones, ritos y costumbres que constituyen los estándares de la cultura local (Werner, 1979). El uso de exámenes estandarizados en poblaciones indígenas es un intento del paradigma *etic* por entrar en el terreno del *emic*, suponiendo la existencia de conceptos universales relacionados con el conocimiento, el aprendizaje, las habilidades y el desempeño (Van de Vijver, 1997). Sin embargo, al suponer que existen constructos universales en las distintas culturas y que hay instrumentos que pueden medir dichos constructos, se incurre en una contradicción fundamental donde se usa la misma vara para entender la diversidad. Bien podría darse el caso en que al aplicar el

mismo instrumento en distintas culturas se estén midiendo fenómenos diferentes al constructo que se pretendía capturar. En el caso del uso de los exámenes estandarizados para entender fenómenos educativos en poblaciones culturalmente diversas queda en suspenso una pregunta: ¿qué nos dicen los resultados de los exámenes estandarizados acerca del aprovechamiento en grupos culturalmente distintos? Y, más específicamente relacionado a esta investigación, ¿cuál es la interpretación adecuada de los resultados de los estudiantes indígenas en los exámenes estandarizados? En estos momentos, la respuesta a ambas preguntas sería “no sabemos con seguridad”. La información sobre aprendizaje recolectada mediante exámenes estandarizados podría estar reflejando el contraste entre los supuestos detrás de las pruebas y la cultura local, tal como pueden ser diferencias en los métodos de aprendizaje o en el manejo del lenguaje usado en el examen. Hasta ahora la investigación no ha desarrollado respuestas a estas interrogantes, en parte, porque aún no se ha llegado al punto de unir los campos de la medición, la psicología transcultural y la antropología en torno al tema de la educación de las poblaciones indígenas. Este texto, al articular estos tres cuerpos de literatura, busca establecer las bases para avanzar en el análisis teórico y empírico de la medición del logro educativo en comunidades indígenas.

Uno de los principales desafíos de la psicología transcultural es el problema de las categorías, que también aparece en la medición del aprendizaje en contextos indígenas. Esta cuestión consiste en usar una categoría psicológica que no es adecuada para la sociedad en la que se aplica (Price-Williams, 1975). Las categorías pueden ser generales o específicas en relación con la medición del aprendizaje; las generales son aquellas relacionadas con la cosmovisión de las sociedades estudiadas, que pueden incluir conceptos como abstracto y concreto, y la diferenciación entre intelecto y emociones (Price-Williams, 1975). Estas categorías generales moldean la manera en que distintos grupos culturales abordan y conciben el aprendizaje, por lo que son relevantes para la medición del mismo en comunidades indígenas. Por otro lado, existen categorías psicológicas específicas estrechamente relacionadas con los procesos cognitivos y, por lo tanto, directamente vinculadas con la medición del aprendizaje mediante exámenes estandarizados. Algunos ejemplos de estas categorías son: la existencia de conceptos lingüísticos universales, la presencia de conceptos matemáticos y científicos, la forma en que se crean los conceptos y la distinción

entre desempeño y competencias (Werner, 1979). El uso de categorías psicológicas permite un mejor entendimiento de las limitaciones e implicaciones de la medición del logro educativo en comunidades indígenas mediante instrumentos que no necesariamente reflejan las categorías culturales de dichos grupos. Los exámenes estandarizados, frecuentemente, asumen la existencia universal de cierto tipo de creencias, formas de interpretar el mundo y de procesos de enseñanza y aprendizaje, sin considerar que estas categorías cambian de una cultura a otra (Padilla, 2001). Bajo estas circunstancias, las inferencias que pueden hacerse a partir de los resultados en un examen estandarizado aplicado a dos grupos culturalmente distintos no necesariamente serán iguales para ambos grupos. De hecho, los supuestos sobre la universalidad de los conceptos medidos en los exámenes limitan las inferencias y el entendimiento que la medición provee de grupos poblacionales cuya cultura no está adecuadamente representada en los constructos que el instrumento mide.

El relativismo cultural de la inteligencia es otro tópico de la psicología transcultural que se relaciona con la medición de los aprendizajes en contextos indígenas. Detrás de este concepto está la idea de que diferentes sociedades desarrollan distintos “tipos de inteligencia” y usan diversos enfoques para aprender y resolver problemas (Price-Williams, 1975). Al medir los aprendizajes educativos en contextos indígenas, es posible encontrar una contradicción entre el tipo de inteligencia que se promueve y cultiva a nivel local y el tipo de inteligencia que se mide a través de las pruebas estandarizadas. Fácilmente puede darse el caso en donde el tipo de inteligencia valorada localmente y usada como marco para la solución de problemas, sea distinta al tipo de inteligencia que mide el examen —puede darse el caso, por cierto, de que ambas estén en directa oposición.

El papel que toman los juicios de valor en el trabajo de psicología transcultural es otro tema relacionado con la medición del logro educativo en contextos indígenas. Price-Williams (1975), propone que los juicios de valor tienen un rol fundamental en la interpretación de los hallazgos de la psicología transcultural y dichos juicios pueden llevar a conclusiones que retraten a la comunidad bajo estudio de manera patológica, distorsionando la realidad. La interacción entre juicios de valor y el uso de un instrumento inapropiado para una población específica puede llevar a interpretaciones y conclusiones erróneas. En cuanto a la medición educativa, al interpretar los resultados en los exámenes de estudiantes de diversos grupos cultura-

les, los juicios de valor pueden tomar un papel negativo al definir a estos grupos como incapaces cuando, en realidad, simplemente no están familiarizados con las tareas incluidas en los exámenes.

Tal como se ha mostrado, la medición del logro educativo de las poblaciones indígenas es un tópico que puede situarse dentro del marco de la psicología transcultural, y al conectar este campo con la medición educativa se evitarían muchos de los errores cometidos en las etapas tempranas de la investigación transcultural.

Conocimiento, aprendizaje y enseñanza entre grupos indígenas

El propósito de esta sección es mostrar las contradicciones entre los enfoques para el aprendizaje en zonas indígenas y aquellos promovidos en la educación formal. El análisis de tales contradicciones muestra las discrepancias entre las prácticas educativas autóctonas de las comunidades indígenas y las que comúnmente suceden en la escuela. Al realizar este ejercicio, es importante tomar en cuenta que los exámenes estandarizados, en general, intentan medir constructos relacionados con la educación formal.

Los sistemas educativos promueven un tipo de aprendizaje impregnado por nociones occidentales.³ La educación formal impulsa el trabajo individual y el pensamiento analítico siguiendo una estrategia lingüística (Tharpe, 1994) en un marco estricto de tiempo para aprender los contenidos que se imparten exclusivamente en la escuela y donde se evalúa a los estudiantes individualmente.

Dentro de esta definición de escolarización, un individuo tiene que trabajar por sí mismo, embarcándose en tareas analíticas y será sometido a una evaluación individual. De hecho, el mecanismo clave de legitimación de la educación formal consiste en el otorgamiento de diplomas o certificados, los cuales se conceden solamente de manera individual. La educación formal busca también desarrollar habilidades analíticas, donde la comprensión se da mediante la separación del objeto de estudio en diferentes partes. Los métodos de enseñanza y aprendizaje son principalmente lingüísticos y, frecuentemente, el estudiante es un agente pasivo en el aprendizaje. El lenguaje se constituye en el principal vehículo para producir y crear aprendizajes. Los estudiantes aprenden mediante explicaciones verbales, teniendo pocas oportunidades de hacerlo por la práctica. Finalmente, en las escuelas, los alumnos tienen un tiempo limitado para asimilar el contenido curricular. Si un estudiante no aprende el contenido bajo esas

condiciones, entonces es penalizado usando diferentes mecanismos, entre ellos la reprobación.

Las comunidades indígenas en América no necesariamente comparten la visión eurocéntrica de la educación, el aprendizaje y el conocimiento. La evidencia que se presenta a continuación se basa en estudios de distintas culturas indígenas. Tharpe (1994) investigó a los navajos y a comunidades hawaianas, De Haan (1999) analizó una comunidad mazahua en México, y Joe (1994) también se ocupó de los navajos. Los estudios citados muestran que las comunidades indígenas difieren de la cultura occidental dominante en, al menos, cinco aspectos que son relevantes al tema de la medición del logro educativo. En primer lugar, las culturas indígenas tienen una visión más colectivista de la organización social (De Haan, 1999; Tharpe, 1994), una característica que también afecta a la organización de la enseñanza y el aprendizaje, especialmente en aspectos como las estrategias pedagógicas y la medición del desempeño. Además, las variaciones en el grado de colectivismo afectan la cognición y el pensamiento infantil, dado que la intensidad del colectivismo en la comunidad es un punto de referencia que define las formas culturalmente apropiadas de participar en el aprendizaje y en las prácticas educativas.

En segundo lugar, las comunidades indígenas y la cultura dominante tienen distintas concepciones del conocimiento. Como se mencionó anteriormente, las sociedades occidentalizadas promueven el pensamiento analítico, donde el aprendizaje ocurre cuando “el todo es revelado a través de desdoblamiento de sus secciones”, mientras que los nativos americanos tienen una percepción holística del pensamiento en la cual “las piezas derivan su significado de las tendencias del todo” (Tharpe, 1994:90). Las formas de pensamiento prevalcientes en cada comunidad están estrechamente vinculadas con la educación local, afectando la “percepción, la resolución de problemas, la interpretación y la acción” (Tharpe, 1994:90). Como consecuencia, cualquier medición del logro educativo, para ser adecuada, debería considerar la forma de pensamiento que prevalece en cada comunidad.

En tercer lugar, las comunidades indígenas tienen estrategias de enseñanza y aprendizaje diferentes a las utilizadas en las sociedades occidentalizadas. Por ejemplo, se usan extensivamente las estrategias observacionales de aprendizaje, que difieren de las lingüísticas de la cultura dominante y la educación formal (De Haan, 1999; Tharpe, 1994). En comunidades indígenas, la competencia se adquiere mediante la participación práctica en actividades

económicas usando un sistema en el que los estudiantes se consideran aprendices y se les da tiempo suficiente para aprender (De Haan, 1999; Joe, 1994; Tharpe, 1994). Dentro de este sistema el marco de tiempo para el aprendizaje está dado por las necesidades del alumno, quien es respetado por quien actúa como profesor. Asimismo, el aprendizaje tiene conexiones simbólicas y concretas con la vida diaria aprenden al participar en actividades económicas específicas relacionadas con su contexto inmediato y que podrían transformarse en una ocupación en el futuro (Joe, 1994).

En cuarto lugar, los métodos de enseñanza en contextos indígenas difieren de aquellos utilizados en la educación formal. En las comunidades el aprendizaje es un proceso que se da a lo largo de toda la vida, en el que se asume que el conocimiento y las habilidades no pueden enseñarse, sino que los niños y niñas las adquieren mediante su participación en la práctica (De Haan, 1999; Joe, 1994). En este mismo sentido, los métodos de enseñanza-aprendizaje son generalmente informales y provienen de distintas fuentes (Joe, 1994). En el proceso de enseñanza, los adultos asumen una actitud no intervencionista, dejando espacio para errores y promoviendo la toma de decisiones colectiva (De Haan, 1999). Las escuelas, en su sentido tradicional, no son culturalmente relevantes para los pueblos indígenas como el navajo, quienes no están de acuerdo con la idea de forzar a los niños a asistir a la escuela (Joe, 1994). En resumen, los métodos de enseñanza en las comunidades indígenas se basan en el aprendizaje mediante la práctica, con una limitada intervención de los adultos y utilizando los errores como oportunidades de aprendizaje. Estas características contrastan con los principios bajo los que se rige la educación formal.

Finalmente, las comunidades indígenas evalúan el desempeño educativo basándose en su propio conocimiento y prácticas educativas, que son distintas a aquellas presentes en la educación formal, aquí el aprendizaje se desarrolla sin que exista la posibilidad de fracaso público. En estos contextos, los docentes comparten la responsabilidad tanto por los errores como por el nivel de competencia del estudiante (De Haan, 1999; Tharpe, 1994). Estas prácticas de evaluación y corresponsabilidad por el aprendizaje son contrarias a las intenciones de los exámenes estandarizados creados centralmente para evaluar a los estudiantes de forma individual.

En resumen, las características culturales de las comunidades indígenas, especialmente aquellas relacionadas con el aprendizaje, son poderosos factores que pueden comprometer la validez de las inferencias hechas a

partir de los resultados en exámenes estandarizados. Esto se debe a que dichas pruebas no necesariamente toman en consideración las formas locales de aprendizaje y de medición del mismo. En esencia, los exámenes intentan medir constructos relativos al aprendizaje, conocimiento y habilidades que podrían llamarse igual en distintas culturas; sin embargo, aunque estos constructos existen en diferentes grupos culturales, sus cualidades varían en concordancia con la cultura local, dado que distintas culturas han creado concepciones culturalmente relevantes de conocimiento, aprendizaje y enseñanza.

Validez, cultura y equidad al medir el aprendizaje en poblaciones indígenas

En esta sección se analizarán las áreas de validez, cultura y equidad que deberían considerarse al analizar el logro educativo en culturas indígenas. Antes de embarcarse en este análisis, es importante destacar algunos puntos que ayudarán a enmarcar la discusión.

Un primer aspecto a tomar en cuenta es que los exámenes estandarizados comúnmente miden habilidades y competencias relacionadas con la educación formal, aun cuando no midan los contenidos del currículo oficial. Dada esta característica, el contenido de los exámenes puede considerarse como un atributo determinado externamente –en este caso por las autoridades educativas– sobre el cual los encargados de desarrollar los instrumentos no tienen control. En este sentido, el contenido no necesariamente es un problema de la prueba. Por otro lado, existen características técnicas intrínsecas a los exámenes sobre las cuales los encargados de desarrollarlos tienen mayor control, tales como el formato, la discriminación de los reactivos y la fiabilidad, entre otros.

El segundo punto que enmarca esta discusión es el uso indebido de los resultados de los exámenes; es posible argumentar que esto existe como práctica generalizada haciendo inferencias incorrectas, especialmente en relación con poblaciones culturalmente diversas. Desafortunadamente, en reiteradas ocasiones se llega a inferencias que no necesariamente están apoyadas por las características técnicas y los propósitos de las pruebas; así es frecuente encontrarse con que los exámenes se usan como un producto genérico que mide el logro educativo, sin considerar las diferencias técnicas entre distintas pruebas. En este contexto, los conceptos como validez, sesgo e impacto adverso rara vez son considerados al hacer inferencias.

Siguiendo la idea previa, un tercer punto importante es la distinción entre validez, sesgo del examen e impacto adverso. Validez es una propiedad de la inferencia hecha a partir de los resultados de los exámenes. Una inferencia es válida en tanto esté alineada con los propósitos del examen, asumiendo que la prueba mide correctamente el o los constructos de interés. En segundo lugar, el sesgo es una propiedad estadística de la prueba que podría afectar la validez. Sin embargo, la evidencia estadística de la existencia de sesgo debe acompañarse de evidencia sustantiva sobre la forma y el contenido del examen. Al analizar el sesgo en las pruebas, es importante considerar que la diferencia en resultados para distintos grupos culturales es evidencia necesaria mas no suficiente para asegurar que el examen está sesgado. El tercer concepto, regularmente confundido en las discusiones de validez y sesgo es el de impacto adverso. De acuerdo con este concepto, un examen podría no mostrar sesgo pero afectar consistentemente a los estudiantes que pertenecen a grupos de bajo rendimiento (Camilli y Shepard, 1994). Por lo tanto, el uso de los exámenes en la toma de decisiones podría afectar desproporcionadamente a estos grupos, en muchas ocasiones, debido a que los resultados de las pruebas se usan erróneamente para tomar decisiones.

Factores que afectan la validez

De acuerdo con el reporte de Estándares para los Exámenes Educativos y Psicológicos –*Standards for Educational and Psychological Testing*– la “validez se refiere al grado en el cual la evidencia y la teoría apoyan la interpretación de los resultados en los exámenes según los usos propuestos para estos últimos” (AERA, APA y NCME, 1999:9; Messick, 1993). Las inferencias derivadas de los resultados de los exámenes deben apoyarse en evidencia sobre su validez. Existen distintos tipos de evidencia que apoyan la validez de las inferencias, entre ellos se cuentan: el contenido del examen, el proceso de respuesta, la estructura interna de la prueba, la relación entre resultados del examen y otras variables, la validez del constructo y las consecuencias de las pruebas (AERA, APA y ncme, 1999). Estos tipos de evidencia para la validez de las inferencias son fundamentales al analizar el logro educativo de los estudiantes indígenas. Primero, como se mostró anteriormente, el contenido de los exámenes estandarizados no necesariamente refleja el tipo de conocimientos, aprendizaje y métodos de enseñanza prevalecientes en sus comunidades. Esta falta de coincidencia puede ser caracterizada como un problema de categorización típico del campo de

la psicología transcultural que consiste en usar una categoría psicológica inadecuada para el grupo bajo estudio (Price-Williams, 1975). En este caso, los exámenes estandarizados podrían ser inapropiados para medir constructos relacionados con la educación entre los grupos indígenas, ya que éstos difieren epistemológicamente de los constructos que promueve la educación formal. Tal como se mostró en los párrafos anteriores, los tipos de conocimiento, el aprendizaje, la enseñanza y los métodos de evaluación usados por las comunidades indígenas están habitualmente en oposición con las concepciones presentes en la educación formal y en la evaluación estandarizada. Por ejemplo, el contenido de una prueba podría estar basado en enfoques analíticos que no necesariamente son compartidos por estas comunidades; por otro lado, los exámenes podrían intentar medir habilidades que no son valoradas ni impulsadas en los medios indígenas. Estas diferencias entre culturas indígenas y no indígenas se convierten en factores que afectan la validez de las inferencias hechas a partir de los resultados que obtienen los estudiantes indígenas en los exámenes estandarizados.

El sesgo es una característica de la estructura interna de los exámenes que también puede invalidar las inferencias hechas a partir de las mediciones en contextos culturalmente diversos. El sesgo en los exámenes puede concebirse como una “diferencia sistemática en la validez de la inferencia, de tal modo que es menos válida para un grupo en comparación a otro” (Koretz, 2003, comunicación personal). Camilli y Shepard (1994) lo definen como “invalides o error sistemático en la medición que un examen hace de miembros de un grupo particular” (p. 8). Siguiendo estas definiciones, el sesgo es una propiedad estadística de los exámenes donde el error sistemático en la medición afecta a los miembros de un grupo. Sin embargo, al analizarlo es necesario poner atención a las propiedades estadísticas y a las características sustantivas de la prueba así como a las inferencias que puedan hacerse a partir de los resultados de las mismas (Camilli y Shepard, 1994; Popham, 2000). Cabe recalcar que las diferencias en puntajes en las pruebas son evidencia necesaria mas no suficiente de sesgo en los exámenes. Por otro lado, el sesgo en las pruebas estandarizadas podría afectar a las poblaciones indígenas, dado que este tipo de exámenes mide constructos educativos cualitativamente distintos a aquellos valorados en los grupos indígenas. Los exámenes también podrían incluir ejemplos y situaciones comunes en la cultura dominante que pueden aparecer como descontextualizados para las culturas indígenas.

Los proyectos internacionales de evaluación mediante exámenes estandarizados como TIMSS y PISA desarrollaron procesos para disminuir el sesgo de los exámenes para su aplicación en contextos nacionales distintos. Aunque, como ya se dijo, estas iniciativas consideran a los países como culturas homogéneas, los procesos llevados a cabo para minimizar el sesgo pueden iluminar el campo de la evaluación en contextos de diversidad étnica y lingüística. En TIMSS y PISA se realizó una selección cautelosa de los reactivos, considerando los sesgos culturales y descartando preguntas que requirieran de conocimientos culturales básicos para su correcta interpretación. Por ejemplo, en el caso de TIMSS se descartaron reactivos que implícitamente demandaran de los alumnos conocimientos de beisbol, dado que muchas de las culturas participantes no están familiarizadas con este deporte (Maxwell, 1996).

En el caso de los exámenes de lectura de PISA se tomaron medidas para minimizar el efecto de las diferencias culturales en los resultados. Durante el desarrollo de las pruebas se prestó especial atención a la equivalencia de los constructos, al formato del examen y a la velocidad para responder. Para asegurar su equivalencia, los constructos se definieron en consulta directa con los expertos representantes de cada país. Además, el proyecto PISA consideró distintos formatos de texto y diferentes contextos que podrían ser familiares a los jóvenes de 15 años en una diversidad de países. También hubo preocupación por usar una gama variada de estímulos en los materiales de las pruebas (McQueen y Mendelovits, 2003). En cuanto a la traducción, PISA operó una estrategia donde en cada país se tradujeron al idioma oficial las dos versiones originales de los exámenes que fueron redactadas en inglés y francés. A partir de ambas traducciones se creó la prueba en la lengua oficial de cada país. Finalmente, un equipo de verificadores seleccionados por PISA compararon las versiones originales con las producidas por cada país (Grisay, 2003). Adicionalmente a la traducción, los exámenes se adaptaron mediante la sustitución de algún término del texto, la modificación de la sintaxis y el cambio de nombres de personas, lugares, organizaciones, etcétera, solamente en los casos donde dicha adaptación mejorara la equivalencia lingüística y cultural y, simultáneamente, no afectara la dificultad de los estímulos de los textos. Una vez que las versiones para cada país estuvieron listas, se pilotearon y se analizó el comportamiento de cada uno de los reactivos en las distintas culturas (McQueen y Mendelovits, 2003).

Como se puede ver a partir de las experiencias internacionales, es posible disminuir el sesgo de los exámenes estandarizados, pero esto requiere

de un esfuerzo amplio, deliberado y técnicamente adecuado. Aún en circunstancias donde los desarrolladores de exámenes tomen pasos para minimizar el sesgo, éste no necesariamente desaparece por lo que es importante tenerlo en cuenta al momento de realizar inferencias a partir de los resultados de las pruebas estandarizadas en contextos diversos.

La evidencia sobre la validez de las inferencias puede provenir de la relación entre los resultados en los exámenes estandarizados y otras variables (AERA, APA y NCME, 1999). En el caso de las comunidades indígenas es importante analizar la relación entre la pertenencia a un grupo y el puntaje en los exámenes, dado que esta relación podría ayudar a clarificar ya sea el distinto significado de los resultados para cada grupo cultural o la diferencia en la medición de un constructo en poblaciones indígenas y no indígenas.

La gráfica 1 muestra el promedio de puntajes en la prueba de lenguaje del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación para aulas con distintos porcentajes de estudiantes indígenas en Bolivia, México y Perú. Los puntajes de este examen han sido estandarizados en una escala con media de 250 y desviación estándar de 50. Como se observa en la gráfica 1, a medida que aumenta el porcentaje de niños indígenas en el aula los puntajes promedio disminuyen, habiendo una diferencia de $3/4$ de desviación estándar entre los puntajes en salones sin niños indígenas y aquéllos donde 75% de los estudiantes son de este origen.

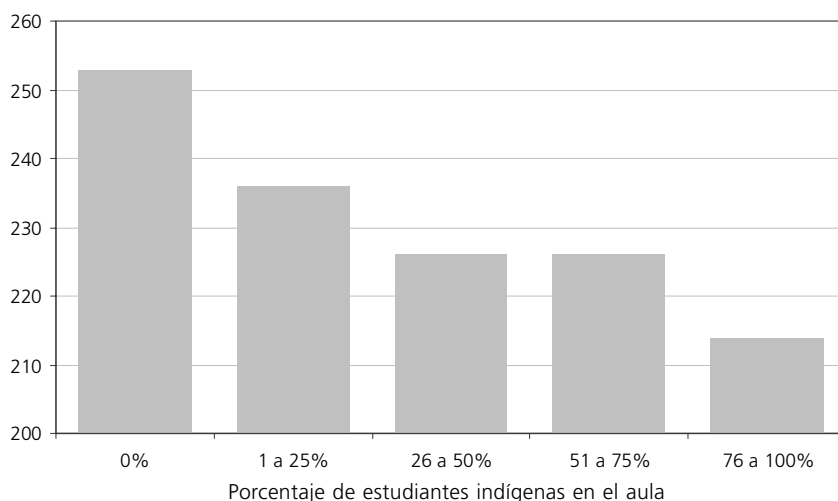
Los resultados de la gráfica 1 muestran que los puntajes en la prueba de lenguaje son distintos en salones cuya composición racial es diferente; las disparidades en puntajes podrían indicar que “el significado de los puntajes es distinto para los miembros de los diferentes grupos” culturales debido a la subrepresentación o irrelevancia de los constructos en los distintos grupos poblacionales (AERA, APA y NCME, 1999:15). Esto sugiere que es plausible plantear la hipótesis de que los constructos que las pruebas están intentando medir están inadecuadamente especificados para su aplicación en contextos indígenas. De comprobarse esta hipótesis, el primer cuestionamiento obvio que viene a la mente, en términos de la representación de constructos, es la incompatibilidad lingüística entre el idioma usado en el examen y la lengua materna de los estudiantes indígenas. Al sustentar la prueba, los estudiantes indígenas estarían en desventaja en comparación con aquellos cuya lengua materna es el castellano. Por lo tanto, los puntajes tendrían diferente interpretación para cada grupo; para los hablantes

de castellano, reflejarían las habilidades lingüísticas medidas por el examen, como pueden ser las habilidades de comprensión lectora y de expresión de ideas mediante la escritura; para los estudiantes indígenas, los puntajes en las pruebas representarían su nivel de manejo en su segunda lengua, que en este caso es el castellano. Por consecuencia, las inferencias que se podrían hacer de los resultados en los exámenes serían diferentes para ambos grupos estudiados.

La discusión anterior pone de manifiesto la importancia de la noción de validez del constructo, que se refiere al grado en que los resultados en un examen reflejan el desempeño de los estudiantes en el concepto que la prueba pretende medir. El ejemplo presentado arriba sugiere que los resultados en las pruebas deben cuestionarse desde la óptica de la especificidad contextual. Esto quiere decir que es necesario preguntarse si la forma en que se mide el constructo es específica a un contexto o si, por el contrario, se puede generalizar en distintos contextos (Messick, 1993). En suma, la generalización de los constructos subyacentes en un examen debe ser materia de análisis cuando las pruebas se usen en contextos de diversidad.

GRÁFICA 1

Rendimiento promedio de los estudiantes en lenguaje por porcentaje de indígenas en el aula para Bolivia, México y Perú, en tercer y cuarto grados



Fuente: LLECE, 2000. Cálculos propios usando la base de datos del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación de UNESCO-OREALC.

La validez de las consecuencias de los exámenes es otra fuente de evidencia para calificar las inferencias. Aquí es importante distinguir entre las consecuencias de los exámenes como evidencia para juzgar la validez de las inferencias y las causadas por decisiones de política que se toman sobre la base de los resultados en las pruebas (AERA, APA y NCME, 1999). En el caso de la medición educativa en poblaciones indígenas, las interpretaciones discriminatorias y los sesgos en la atribución de la responsabilidad por los resultados en los exámenes podrían tener dos tipos de consecuencias. En primer lugar, el uso inapropiado de las pruebas podría llevar a interpretaciones del rendimiento académico que retraten a las poblaciones indígenas como incapaces de tener éxito en la escuela, esto basándose en resultados de exámenes que ni siquiera están escritos en la lengua materna de los estudiantes. En este terreno, los puntajes en los exámenes podrían usarse para apoyar tesis racistas, para explicar el desempeño diferencial de distintos grupos culturales. El segundo factor que afectaría a la validez sería la atribución causal por los resultados. En este sentido, es necesario desarrollar análisis multivariados con diseños de investigación complejos para desentrañar las causas detrás de los resultados de los estudiantes, ya que los exámenes por sí solos tienen como fin describir una situación y su objetivo no es explicarla.

Por otra parte, si las inferencias se limitan correctamente usando los métodos analíticos apropiados, los resultados de tales análisis permitirían establecer relaciones causales entre distintos factores escolares y extra-escolares asociados con el rendimiento de los alumnos. Estas consecuencias no siempre se consideran en los análisis de validez de los instrumentos aplicados a poblaciones culturalmente diversas. Los puntajes podrían, por lo tanto, interpretarse usando un marco patológico, que retrate a los estudiantes indígenas como carentes de capacidades que son valoradas por la cultura dominante pero se manifiestan de forma distinta en su propia cultura. Si bien los teóricos tratan de distinguir entre sesgo y consecuencias, resaltando que un examen podría tener impacto adverso en los grupos de bajo rendimiento a pesar de no ser sesgado, lo cierto es que el bajo rendimiento de un grupo —independientemente del sesgo en el examen— podría tener impacto en la validez referida a las consecuencias.

El papel de la cultura en la validez y la interpretación de los exámenes

La validez cultural de los exámenes estandarizados es un tipo de evidencia a la que no se hace referencia comúnmente en el campo de la medición

educativa. La validez cultural es crucial para entender los resultados que los estudiantes indígenas obtienen en los exámenes estandarizados. Las diferencias culturales pueden ser un factor que afecte las inferencias que pueden hacerse de los resultados en los exámenes, porque es posible que éstos no midan adecuadamente los tipos de conocimiento, aprendizajes y habilidades de ese grupo particular de estudiantes. Solano-Flores y Nelson-Barber (2001) proponen el concepto de validez cultural específicamente en los exámenes de ciencias, sugiriendo que las características socioculturales de los alumnos influyen en su forma de pensar y entender la ciencia. Estas “influencias socioculturales incluyen los valores, creencias, experiencias, patrones de comunicación, estilos de enseñanza y aprendizaje, y epistemologías inherentes al origen cultural de los estudiantes y a las condiciones socioeconómicas prevalecientes en su grupo” (Solano-Flores y Nelson-Barber, 2001:555). Tal como ya se mostró, las poblaciones indígenas tienen formas diferentes de entender el mundo a la cosmovisión que promueve la educación formal. Estas diferencias deben tomarse en cuenta al evaluar a los estudiantes pertenecientes a distintos grupos culturales y al hacer inferencias sobre los resultados obtenidos por ellos en los exámenes. El concepto de validez cultural sugiere que al diseñar los exámenes y analizar los resultados, será necesario limitar la afirmación que sostiene la existencia de conceptos o constructos universales a través de distintas culturas.

En ocasiones, los factores que afectan la validez de los exámenes son confundidos con el sesgo. Por ejemplo, Padilla (2001) afirma que los tres factores principales que sesgan los exámenes estandarizados son:

- 1) el contenido puede estar sesgado, dando una ventaja injusta al grupo que está mejor representado en el contenido del examen;
- 2) las características de la evaluación tales como el formato de la prueba y los procedimientos de su administración podrían favorecer a un grupo en detrimento de otros; y
- 3) el uso inapropiado de los exámenes como predictores del rendimiento académico futuro.

De los tres factores mencionados por Padilla, solamente el contenido del examen puede considerarse como una fuente de sesgo. No obstante, para llegar a la conclusión de que el instrumento está sesgado se requiere, además de evidencia estadística, de un análisis sustantivo del contenido.

En lo que se refiere a las características de la evaluación, son factores que podrían afectar la validez al medir el rendimiento académico en poblaciones culturalmente diversas, sin embargo, no necesariamente pueden considerarse como sesgo del examen. Finalmente, el uso erróneo de las pruebas no representa una fuente de sesgo, debido a que “el uso” no es una propiedad del examen. A pesar de esta confusión, las tres características mencionadas por Padilla son factores que podrían afectar la validez cultural de los exámenes. Por ejemplo, algunos grupos indígenas mantienen un sentido de responsabilidad compartida por el desempeño, mismo que está en contradicción con la forma individualizada de medir los resultados educativos utilizando exámenes estandarizados. Asimismo, el uso inapropiado de las pruebas es otro factor que podría afectar la validez del examen debido a las consecuencias que éste tendría sobre la carrera escolar de los estudiantes, especialmente cuando los resultados se usan para tomar decisiones.

La validez cultural de los exámenes es especialmente importante en las sociedades multiculturales de hoy, dado que este concepto es fundamental para entender las limitaciones de las pruebas estandarizadas al aplicarlas a estudiantes de distintos grupos culturales, así como para establecer las fronteras de las inferencias que pueden hacerse con base en la información proveniente de los exámenes (Solano-Flores y Nelson-Barber, 2001). Asimismo, los estudios de validez cultural ayudarán a calificar los hallazgos relacionados con los puntajes en los exámenes, ofreciendo una comprensión más amplia de los resultados educativos en relación con el contexto.

Las definiciones técnicas de validez, sesgo e impacto adverso no resuelven otros problemas de los exámenes, tales como la suposición de la universalidad de los constructos válidos para distintas culturas y el uso inapropiado de los resultados al hacer inferencias (Hambleton, 1993; Messick, 1993; Van de Vijver, 1997). Es claro que los constructos no necesariamente son universales, dado que diferentes culturas han desarrollado distintos tipos de conocimiento, de enseñanza y de aprendizaje. Asimismo, la orientación epistemológica de las diversas culturas podría incidir en que un mismo constructo se operacionalice de forma diferente entre varios grupos. Por ejemplo, en la cultura aymara si un docente hace la seña que se presenta en la figura 1 y pregunta a sus estudiantes, “¿cuántos dedos tengo aquí?” Los estudiantes contestarán, si saben contar, “dos dedos”. Si el profesor

cambia el gesto y les presenta a los estudiantes la seña que se muestra abajo (figura 2) y pregunta, “¿y ahora, cuántos dedos tengo? Los niños aymara le contestarán “te falta uno”, en vez de decir, “tienes un dedo”.



FIGURA 1



FIGURA 2

¿Esto significa que los estudiantes aymara carecen del conocimiento para hacer sustracciones matemáticas? Ciertamente que no, porque fueron capaces de identificar la esencia de la resta, pero usando un marco de referencia holístico. En los exámenes estandarizados, la respuesta dada por los niños aymara probablemente sería considerada incorrecta, porque esta alternativa no necesariamente estaría incluida como una de las opciones múltiples de respuesta. Este ejemplo muestra la distancia entre las epistemologías de distintos grupos culturales. Si bien los niños aymara son capaces de llevar a cabo el proceso de resta subyacente en la pregunta, la respuesta que ofrecen desde su marco interpretativo sería inadecuada desde una perspectiva analítica. De incluirse esta pregunta en un examen aplicado a estudiantes aymara, donde solamente acepte como respuesta legítima “tienes un dedo”, se podría afirmar que la prueba adolece de subrepresentación del constructo que pretende medir. Esto se debe a que excluye dimensiones importantes con respecto al constructo (Messick, 1993) que son específicas al contexto indígena. Por lo tanto, sería ineficaz para medir el dominio de la resta entre los niños y niñas aymaras. En suma, las inferencias a partir de los resultados en esta prueba hipotética no serían generalizables al contexto indígena (Messick, 1993).

Los exámenes estandarizados podrían ser incapaces de medir ciertas habilidades en diferentes culturas. Como se mostró arriba, las culturas indígenas fomentan habilidades y conocimientos diferentes a los que se aprenden en la educación formal. Por ejemplo, Cole (1996) encontró estrategias

diferenciales de categorización de objetos al comparar a los kpelle de Liberia con estadounidenses y canadienses. Dicha investigación descubrió que las variaciones en el método usado para medir la clasificación de objetos impactó de forma importante el desempeño de los kpelle. Por otro lado, los estadounidenses y canadienses se enfocaron principalmente en buscar categorías pareadas, aunque éstas no necesariamente existieran en este experimento. En este mismo sentido, investigaciones en Yucatán, México, encontraron que personas tanto indígenas como mestizas con más años de escolaridad tuvieron un desempeño más alto en desarrollar actividades teóricas como resolver silogismos simples (Cole, 1996). Esto sugiere que las personas con mayor escolaridad tienden a desempeñarse mejor en actividades de tipo escolar, aunque podrían tener competencias para aplicar la misma habilidad en la vida diaria de una manera distinta a la que se enseña en las escuelas.

La validez de las inferencias de los exámenes estandarizados puede también verse afectada por los supuestos de comparabilidad cultural que subyacen a las pruebas; estos exámenes encierran tres tipos de falacias respecto de la comparabilidad cultural: *a)* equivalencia contextual: los niños y niñas que sustentan el examen están familiarizados con las normas del discurso utilizadas en este último; *b)* equivalencia conceptual: los estudiantes están familiarizados con los atributos de las tareas incluidas en el examen; y *c)* equivalencia lingüística: el lenguaje usado en la evaluación tiene el mismo significado para los niños de distinto origen cultural (Gopaul-McNicol y Armour-Thomas, 2002). Las equivalencias contextual, conceptual y lingüística están directamente relacionadas con el supuesto de la universalidad de constructos. Los exámenes estandarizados pueden ser discriminatorios porque están sesgados hacia la cultura dominante (García y Pearson, 1994, citado por NCR, 1997). Este tipo de sesgo acarrea implicaciones políticas que van más allá de los aspectos técnicos de los exámenes, dado que se cuestiona el proceso mediante el cual ciertos tipos de conocimiento se consideran más importantes que otros. En general, los exámenes estandarizados no incorporan las experiencias de grupos minoritarios (Padilla, 2001). Estos cuestionamientos entran en el campo de la equidad y las tensiones entre la diversidad, los exámenes estandarizados y el uso de sus resultados.

Al medir el logro de los estudiantes y hacer inferencias a partir de los resultados en los exámenes, es clave considerar el papel del lenguaje, las experiencias previas y la falta de oportunidades de aprendizaje. Estudiantes de distintos grupos culturales y sociales experimentan la escuela de

modo diferente, enfrentando complejas relaciones con factores tales como los valores hegemónicos de la cultura dominante, la resistencia de los sistemas de creencias de las culturas marginales, la incompatibilidad entre las culturas de la escuela y del hogar y las dificultades de la aculturación (Gopaul-McNicol y Armour-Thomas, 2002). Estas diferencias, junto con disparidades entre las escuelas, afectan el logro educativo de los estudiantes de diversos grupos culturales.

Perspectivas actuales

en la adaptación cultural de exámenes estandarizados

Actualmente, desde distintas instancias se desarrollan adaptaciones culturales de exámenes estandarizados. Los proyectos internacionales de evaluación educativa son el ejemplo más sobresaliente de esta adaptación, aunque al interior de muchos países se llevan a cabo adaptaciones de distintos exámenes estandarizados. En el ámbito internacional los aportes de Hambleton son los de mayor influencia, pues en ellos se basaron los exámenes del TIMSS y de PISA. En un trabajo seminal sobre comparación internacional, Hambleton (1993) propone métodos para asegurar la adecuada traducción de los exámenes a distintos idiomas y subraya la importancia de hacer adaptaciones culturales adecuadas, de acuerdo con este autor, existen aspectos claves a considerar en la adaptación cultural de pruebas estandarizadas. Los niveles de colectivismo-individualismo de las culturas, como ya se mencionó, afectan la forma en que personas de distintas culturas enfrentan las evaluaciones que pretenden ser individuales. La familiaridad con el formato del examen, la motivación y la ansiedad ante las pruebas son también variables que podrían afectar la validez de las inferencias hechas a partir de los resultados de pruebas estandarizadas en distintas culturas (Hambleton, 1993). En un trabajo posterior, Hambleton y Pastula (2000) proponen que existen tres fuentes de error en la adaptación cultural de los exámenes: diferencias culturales y lingüísticas, técnicas de adaptación, e interpretación de resultados.

La adaptación cultural de exámenes estandarizados ha sido analizada desde distintos paradigmas usando métodos estadísticos que tratan de desentrañar el efecto de diversas variables en los resultados de los exámenes. Un meta-análisis de los primeros 25 volúmenes del *Journal of Cross-Cultural Psychology* encontró que la hipótesis del sesgo cultural en los exámenes era la que mejor predecía el desempeño de distintos grupos culturales. Esto confirma que el

contenido de los exámenes está sesgado hacia la cultura occidental, y que las pruebas tienen poca consideración para medir conocimientos y habilidades relacionadas con las culturas no occidentales (Van de Vijver, 1997). Por otro lado, se ha encontrado también que los aspectos verbales-culturales tienen mayor influencia en los resultados en exámenes estandarizados en comparación con la complejidad cognitiva de las pruebas, esto al contrastar los resultados de inmigrantes y no inmigrantes en los Países Bajos (Helms-Lorenz, Van de Vijver y Poortinga, 2003). En suma, existe evidencia sustancial para apoyar la hipótesis de que los exámenes tienen sesgos culturales que plantean desafíos al momento de realizar inferencias a partir del desempeño de los alumnos pertenecientes a distintos grupos culturales.

Los resultados de estos trabajos son iluminadores para el campo de la medición, pero adolecen de algunos problemas. En primer lugar, los análisis estadísticos podrían sufrir de sesgo de selección, pues la mayoría de los resultados incluyen, primordialmente, tareas típicas de la cultura occidental relacionadas con la educación formal. Por lo tanto, la muestra del tipo de tareas incluidas en estas pruebas no es representativa del universo de tareas existentes en los distintos grupos culturales evaluados (Van de Vijver, 1997). En segundo lugar, la falta de tareas relevantes para las culturas no occidentales en las pruebas estandarizadas plantea dos problemas: por un lado, desconocemos si los estudiantes de culturas no occidentales realmente manejan los conocimientos o habilidades que los exámenes pretenden medir, pues sería necesario incluir en las pruebas preguntas que reflejen los modos de aprendizaje de estas culturas; por otro lado, la evaluación a través de exámenes estandarizados podría estar reñida con las teorías del aprendizaje que postulan que existen distintos modos de aprender un mismo concepto (Bransford, Brown y Cocking, 2000). En conclusión, hace falta flexibilizar las evaluaciones para incluir en ellas tareas relativas a las culturas no occidentales y, así, estudiar con mayor detalle las formas en que se puede evaluar el desempeño de los estudiantes en contextos como el indígena.

Equidad y diversidad

Las evaluaciones mediante exámenes estandarizados son una instancia de continua tensión entre la diversidad y la equidad. Por una parte, al interior de los países ha crecido el reconocimiento de la diversidad cultural, lo que ha tenido efectos en la evaluación del lenguaje en algunos contextos (Kalantzis *et al.*, 1989). Sin embargo, el reconocimiento de la diversidad

no parece haber tenido efectos en la operacionalización de los constructos que subyacen a los exámenes. En el contexto de América Latina, solamente Bolivia ha aplicado pruebas usando las lenguas indígenas y constructos culturalmente relevantes (Treviño, en prensa). De todas maneras, el reconocimiento de la diversidad dentro de las fronteras nacionales está lejos de influir en la operación de los sistemas nacionales de evaluación, los que están principalmente orientados a medir el aprendizaje de los contenidos curriculares y no necesariamente incorporan conocimientos de las distintas culturas que viven en cada país.

Desde el punto de vista de la evaluación educativa, el reconocimiento de la diversidad podría contraponerse con algunos aspectos de la equidad. Aun cuando los sistemas de evaluación se adaptaran a la diversidad lingüística y cultural de los países, las estructuras sociales prevalecientes impondrían sobre las poblaciones indígenas la necesidad de aprender la lengua dominante para que éstas puedan incrementar sus posibilidades de avanzar en el sistema educativo, acceder a servicios sociales y fortalecer la equidad social (Kalantzis *et al.*, 1989). Esta tensión es típica de los sistemas de América Latina, donde el multiculturalismo se ha reconocido, pero los indígenas tienen que manejar la cultura dominante para lograr la movilidad social y económica. En estas condiciones, los sistemas de evaluación enfrentan poderosos incentivos para medir aquellos aspectos que las estructuras sociales y económicas esperan, que corresponden precisamente con los atributos promovidos por la educación formal y moldeados por la cultura dominante.

No existe una solución definitiva a esta paradoja, sino que es posible alcanzar un acuerdo intermedio donde se respete la diversidad mientras se monitorea la equidad. Este acuerdo requiere que cualquier inferencia hecha a partir de los resultados en exámenes estandarizados sea adecuadamente calificada tomando en cuenta las características culturales, lingüísticas y sociales de la población evaluada. Por otro lado, mientras las estructuras sociales requieran del manejo de los códigos de la cultura dominante para la movilidad social, es necesario monitorear el rendimiento de los estudiantes como una forma de evaluar el desempeño del sistema educativo en términos de equidad.

Las consecuencias de la evaluación son un factor crucial en la tensión entre equidad y diversidad. El uso de exámenes estandarizados en la toma de decisiones podría tener impactos significativos en las oportunidades

educativas de los niños, especialmente de aquellos pertenecientes a culturas indígenas. Por lo tanto, los exámenes estandarizados no deberían ser el único instrumento para juzgar el rendimiento académico de un individuo, así como tampoco sus resultados deberían tenerse como un pronóstico confiable de rendimiento académico futuro por dos razones: primera, los exámenes estandarizados son incapaces de medir todos los contenidos que se aprenden en la escuela; y segunda, la confiabilidad de sus resultados disminuye al acotar la muestra de población evaluada, dándose mayor probabilidad de error en la estimación del aprendizaje de individuos. Los exámenes estandarizados deben usarse de manera adecuada, respetando sus limitaciones y, especialmente, no deberían utilizarse para restringir las oportunidades educativas a niños y niñas.

Equidad y medición en contextos indígenas

Las inmensas desigualdades educativas de América Latina y las condiciones de marginación en que se encuentran los grupos indígenas no se resolverán solamente mediante una evaluación adecuada. Para ello se requieren de acciones y políticas que escapen del tema de este documento.

Desde la perspectiva de la equidad es imprescindible cuestionarse si la evaluación mediante exámenes estandarizados, con todas sus limitaciones y consecuencias, es una medida que vale la pena tomar en contextos de diversidad cultural y de enormes desigualdades educativas, sociales y económicas. En primer lugar, el uso de exámenes estandarizados ha sido campo fértil para debates de tipo epistemológico, pedagógico e ideológico en torno a lo que constituye el conocimiento legítimo y las desigualdades educativas; los exámenes estandarizados implícitamente legitiman una serie de conocimientos como válidos y valiosos. En este sentido, las diversas instancias que usan pruebas estandarizadas sancionan el conocimiento valioso y descartan otros tipos de conocimiento que son excluidos en el diseño de los exámenes. Los grupos sociales que tienen conocimientos distintos a los legitimados son entonces marginados por ese hecho, y las consecuencias pueden ir desde la falta de visibilidad de las tradiciones epistemológicas de tales grupos hasta la pérdida de oportunidades educativas y de movilidad social por el hecho de no haber aprobado un examen. En términos pedagógicos, el uso de exámenes de altas consecuencias limita la visión de la educación a la adquisición de aquellos conocimientos y habilidades medidas por los exámenes. Entonces, se emiten juicios evaluativos de la educación

basándose, principalmente, en los resultados en los exámenes y dejando de lado aspectos tan importantes en los que inciden las escuelas y docentes tales como la educación en valores, el aprendizaje para la convivencia, el respeto a los demás y muchas otras áreas donde la educación toma un papel fundamental en la socialización de las nuevas generaciones. En cuanto a lo ideológico, los exámenes enfrentan a los paradigmas de la modernidad y la post-modernidad. La modernidad, tomada en su versión extrema, tiene una visión lineal del progreso y desarrollo económicos en la que las sociedades transitarán distintas etapas que desembocarán en la industrialización y post-industrialización. Bajo este enfoque el uso de exámenes estandarizados es legítimo. Suponiendo que los exámenes tienden a enfocarse en capacidades analíticas, el punto de vista de la modernidad no percibe como problemático el hecho de medir solamente conocimientos y habilidades de tipo analítico, que son apreciados como valiosos para el tránsito de las sociedades a través de las distintas etapas del desarrollo. Por otro lado, la post-modernidad, también en su visión más radical, considera que todos los tipos de conocimiento son legítimos y que ninguno debe prevalecer sobre otro. En este enfoque se desestima por completo el uso de los exámenes por considerarlos elementos de legitimación del conocimiento y poder de un grupo sobre otros. En ambas posturas se encuentran también críticos que, desde su enfoque particular, consideran a los exámenes estandarizados como un mecanismo institucional para marginar a los estudiantes más desfavorecidos y legitimar la posición favorecida de los grupos dominantes. La evidencia mostrada en este documento parece apoyar la tesis que indica que los exámenes de alto impacto son mecanismos perversos que ensanchan las brechas aprendizaje entre los grupos socioeconómicos y culturales.

Las evaluaciones estandarizadas adolecen de limitaciones metodológicas graves para su aplicación en contextos de diversidad. Aun en países con largas tradiciones de evaluación estandarizada, el avance de la ciencia en torno a la adaptación de exámenes para la diversidad está recién en una etapa inicial de concepción teórica y metodológica. Los análisis estadísticos para evaluar la presencia de sesgo e impacto adverso no logran resolver el problema de fondo que proviene de la disparidad epistemológica, más que lingüística, entre culturas diversas.

Las posibilidades de realizar adaptaciones e interpretaciones adecuadas de los resultados de los exámenes en América Latina también son limitadas

dada la carencia de especialistas en este tema y la baja prioridad que éste tiene en las agendas de investigación.

La discusión en relación con la equidad sugiere que la evaluación en contextos de diversidad puede acarrear graves consecuencias para los grupos marginados –en este caso para los indígenas– por lo que debe ser sujeta a un severo juicio. Mayor marginación cultural y social, así como menores oportunidades educativas son consecuencias asociadas con el uso de exámenes estandarizados en grupos culturalmente diversos. Por otro lado, desde el punto de vista de la ciencia es difícil cualificar las inferencias que pueden hacerse a partir de los resultados en exámenes estandarizados.

La mayoría de las evaluaciones de estudiantes indígenas en América Latina ha descartado por completo la relación entre diversidad cultural y resultados en exámenes estandarizados. Si bien es cierto que gran parte de los sistemas de evaluación en la región son relativamente jóvenes y están en proceso de consolidación, esto no justifica que se instalen visiones simplistas de la evaluación estandarizada de poblaciones indígenas que desacreditan sus conocimientos y que, en casos de pruebas de altas consecuencias, podrían conllevar efectos negativos directos sobre las oportunidades educativas de los estudiantes.

Un hecho fehaciente es que la evaluación mediante pruebas estandarizadas está ya instalada en la agenda de las autoridades educativas de América Latina. Ante esta realidad se han ya señalado las limitaciones de este tipo de evaluación en contextos indígenas y sus consecuencias en la equidad. En la siguiente sección, y ante la inminente permanencia de la evaluación estandarizada en la región, se ofrecen algunas recomendaciones simples que podrían avanzar nuestra comprensión sobre la evaluación en contextos de diversidad a la vez que reducirían los impactos adversos e injustos de tales exámenes cuando son aplicados a estudiantes indígenas.

Recomendaciones para una evaluación equitativa de los estudiantes indígenas

Una vez indicadas las dificultades y efectos perversos que podría llevar consigo la evaluación estandarizada en contextos indígenas, ahora se mencionan algunas recomendaciones para minimizar el impacto adverso y avanzar en este campo del conocimiento. En esta sección se discutirán algunos de los retos de la evaluación educativa en contextos indígenas en América Latina; se abordarán, además, formas en que los resultados de los exámenes de

estudiantes indígenas podrían usarse apropiadamente. También se presentarán estrategias que ayudarían a mejorar los exámenes estandarizados; y finalmente, se tocará el tema de la situación de la investigación sobre la medición educativa en contextos multiculturales de América Latina.

Al parecer, existe capacidad técnica en América Latina para realizar evaluaciones apropiadas a las culturas indígenas, pero las posibilidades de sustentabilidad en el tiempo de este tipo de evaluación se muestran débiles. La evidencia disponible indica que Bolivia ha realizado evaluaciones usando exámenes adaptados a las lenguas y culturas indígenas (Barrera, 1998), lo que quiere decir que la evaluación educativa intercultural es técnicamente factible, aunque sería necesario corroborar la calidad de tales evaluaciones. Aun cuando las áreas de evaluación de los países de Latinoamérica cuenten con la capacidad humana para el desarrollo técnico de evaluaciones interculturales, la experiencia boliviana también muestra que este tipo de medición ha sido insostenible en el tiempo. Por lo tanto, es probable que el apoyo político y financiero para este tipo de evaluaciones sea débil.

En las condiciones actuales desconocemos el significado de los puntajes que obtienen los estudiantes indígenas en las distintas pruebas. Los niños y niñas indígenas participan en evaluaciones estandarizadas escritas en castellano (o portugués en el caso de Brasil) y sus resultados comúnmente se interpretan como una prueba irrefutable de sus niveles de aprendizaje. Sin embargo, carecemos de análisis que permitan valorar el funcionamiento de los exámenes en distintas culturas y calificar adecuadamente las inferencias que se pueden hacer en relación con los resultados de los estudiantes indígenas.

También desconocemos el grado de validez de los constructos de las pruebas en contextos indígenas. Por ello, cabe la posibilidad de que las tradiciones epistemológicas indígenas generen respuestas válidas en los distintos dominios que los exámenes pretenden medir, pero dichas respuestas no necesariamente serán consideradas como correctas en las pruebas que se hacen a partir de una corriente epistemológica distinta a las de los grupos indígenas.

La combinación de factores como las desigualdades sociales, el impulso a la evaluación de altas consecuencias y el desconocimiento del significado de los resultados de los indígenas en las pruebas estandarizadas podría traer secuelas graves para las oportunidades de niños y niñas de grupos indígenas. En un contexto donde cada día se habla más de rendición de cuentas, sería riesgoso olvidar que vivimos en la región más desigual del

mundo, donde los exámenes de altas consecuencias perjudicarían aún más a las poblaciones históricamente desfavorecidas, como son los indígenas. De darse este escenario, las evaluaciones no solamente ratificarían la condición de desventaja de los estudiantes indígenas con respecto al manejo de los símbolos de la educación formal, sino que además se constituirían en una barrera injusta para los alumnos indígenas que provienen de culturas distintas y viven en condiciones de marginación.

Las evaluaciones estandarizadas tienen una serie de limitaciones que es necesario considerar al aplicarlas a estudiantes indígenas. Un paso sencillo en esta dirección se refiere a la interpretación correcta de los resultados, provista de una clara delimitación de las inferencias. Factores tales como los propósitos del examen, el contenido, el lenguaje y el formato, entre otros, son características importantes que podrían afectar la validez de las inferencias, por lo que se requiere establecer el alcance de las mismas. Si se evalúan cuidadosamente el sesgo y otros factores que afectan la validez de las inferencias, podría ser posible llegar a conclusiones razonables y sustentables que ayuden a describir el desempeño de los estudiantes indígenas en los distintos constructos que el examen intenta medir. En este ejercicio es fundamental considerar que los puntajes en los exámenes describen una situación y no explican las razones detrás de los distintos niveles de aprendizaje.

Tomando en cuenta la equidad de oportunidades educativas y suponiendo que los exámenes miden adecuadamente los constructos que les subyacen, sería entonces posible mantener los factores que afectan la validez dentro de márgenes aceptables y llegar a inferencias valiosas y adecuadas a partir de los resultados en los exámenes. Por ejemplo, ya que la mayoría de los sistemas educativos de América Latina no ofrecen opciones de educación bilingüe o multicultural más allá de la primaria, los puntajes en las pruebas donde se usa el lenguaje de la cultura dominante mostrarían los niveles de logro de los estudiantes indígenas sobre las tareas relacionadas con la educación formal. Utilizando esta información se podría inferir que el bajo rendimiento de los estudiantes indígenas podría limitar sus oportunidades educacionales futuras, especialmente en la transición a niveles educativos más avanzados. De esta forma, los resultados en las pruebas dan información importante acerca de las desigualdades en el presente, las que influyen en las oportunidades educativas que los indígenas tendrán en el futuro. Esta valiosa información puede obtenerse usando los datos actualmente disponibles sobre el rendimiento de los estudiantes. El único

requisito, no exento de problemas, es que se identifique en cada examen la filiación étnica de los alumnos.

Otra medida relativamente sencilla para mejorar los exámenes sería el muestreo de distintos tipos de tareas orientadas a medir el mismo constructo, pero basadas en las tradiciones epistemológicas de diferentes grupos culturales. De esta manera, se podrían tener algunas preguntas que resulten más apropiadas para las culturas indígenas en cada país y aplicarlas tanto a los estudiantes indígenas como a los no indígenas. Estas preguntas irían acompañadas de otras que midan el mismo constructo usando el enfoque occidental de la educación formal. De esta forma podría compararse el rendimiento de los estudiantes de distintas culturas en un mismo constructo adaptado culturalmente, lo que podría arrojar luces acerca de la magnitud del sesgo en los exámenes estandarizados y formas de adaptarlos en contextos de diversidad cultural.

Los actuales exámenes estandarizados podrían mejorarse si se analiza su compatibilidad en contextos indígenas. Una idea que podría dar resultados inmediatos es la inclusión de un reducido número de preguntas abiertas en un examen que se aplique a una muestra de estudiantes indígenas y otra de no indígenas. Las respuestas a los reactivos abiertos permitirán identificar los procesos cognitivos y clarificar el significado de los constructos en ambos grupos. Asimismo, se pueden aplicar preguntas similares de opción múltiple a ambas muestras. La comparación entre las respuestas a los reactivos abiertos y los de opción múltiple en los dos grupos de estudiantes permitirá una mejor comprensión de los procesos cognitivos que gatilla cada reactivo tanto en indígenas como en no indígenas. Esto proporcionaría información acerca de las diferencias en las respuestas que pueden atribuirse a factores culturales. Además, este ejercicio proveería evidencia acerca del efecto que tiene el formato del examen en el resultado para ambas poblaciones. Esta simple idea puede ayudar a mejorar los exámenes y calificar mejor las inferencias hechas a partir de la medición del logro de los alumnos.

En el contexto de América Latina se requiere más investigación para explorar el tema del sesgo en los exámenes. En países como Estados Unidos, los análisis de sesgo son una práctica común tanto en el campo académico como en el de la política educativa (Freedle, 2003). A lo largo de esta investigación no se encontraron estudios sobre sesgo en exámenes estandarizados en el contexto de América Latina. Esto es indicativo del limitado espacio que recibe este tema en la agenda de investigación de la región.

Conclusión

La práctica de la evaluación educativa en América Latina ha adquirido una importancia creciente en los últimos trece años. El uso de exámenes estandarizados para la toma de decisiones es también un fenómeno progresivo. La evaluación educativa ha florecido en un contexto de reconocimiento legal de la diversidad cultural al interior de los países, causando una tensión entre la medición del conocimiento “estándar” y la aceptación de la diversidad de conocimientos en el discurso educativo. Esta tensión aún no se resuelve.

La combinación de factores como las desigualdades sociales, el impulso a la evaluación de altas consecuencias y el desconocimiento del significado de los resultados de los indígenas en las pruebas estandarizadas podría limitar aún más las oportunidades de los estudiantes indígenas, que son el grupo más desfavorecido en la región más desigual del mundo.

La evaluación estandarizada en contextos indígenas es problemática y lleva necesariamente a la pregunta de cuán adecuado es realizarla. La contraposición entre la epistemología que respalda los exámenes y la de muchas culturas indígenas parece insalvable y, por lo tanto, la evaluación de los indígenas mediante exámenes estandarizados parecería un esfuerzo fútil. Las culturas indígenas fomentan tipos de conocimiento, estilos de enseñanza y aprendizaje, así como métodos de evaluación que, con frecuencia, se contradicen con las prácticas de la educación formal. Por lo tanto, el desencuentro epistemológico podría llevar a invalidar los resultados de las evaluaciones.

En vista de los desafíos que enfrenta el tópico de la medición del logro educativo en comunidades indígenas, se propone que este tema encaja perfectamente en el campo de la psicología transcultural. El problema de las categorías, los tipos de inteligencias y el supuesto de universalidad de constructos son tres temas de la psicología transcultural estrechamente ligados con la medición del rendimiento educativo en contextos indígenas. El campo científico de la medición educativa se beneficiaría enormemente al aprender de los errores cometidos en las épocas tempranas de la psicología transcultural.

Los resultados de los estudiantes indígenas en los exámenes estandarizados deben considerarse con suma precaución. Las diferencias entre el lenguaje y la cultura indígenas y aquellas usadas en el examen podrían imponer limitaciones a las inferencias que se pueden hacer a partir de los resultados

de los exámenes; y éstos no deben usarse para juzgar individualmente. Este tipo de juicios no están garantizados por las características técnicas de las pruebas. Por lo tanto, los exámenes podrían usarse para identificar áreas de oportunidad para mejorar la enseñanza; además, si son acompañados por encuestas de factores asociados, el análisis conjunto de ambos permitiría entender mejor la relación entre desempeño de los estudiantes y algunas variables explicativas de la población y las escuelas. De lo contrario, los resultados en los exámenes en sí mismos no deben utilizarse para atribuir causalidades, pues éstos solamente describen una situación.

Dadas las estructuras de los sistemas educativos de América Latina, los resultados en los exámenes estandarizados podrían ser una variable que anticipe las oportunidades educativas futuras de los estudiantes indígenas. Ya que la educación bilingüe o multicultural no va más allá de la primaria en la mayoría de los sistemas educativos, los alumnos indígenas requieren manejar el currículo y la lengua dominante para avanzar a niveles más elevados del sistema educativo. En este contexto, las pruebas donde se usa el lenguaje de la cultura dominante ayudan a identificar las brechas de desigualdad que afectan a los niños y niñas indígenas.

La evaluación educativa en contextos interculturales aún es débil en la región, pues las áreas de evaluación en muchos de los países están en proceso de consolidación. Aunque existan las capacidades técnicas para la valoración intercultural, el apoyo político y financiero para ésta todavía es, aparentemente, débil.

Aun cuando la evaluación estandarizada pueda ayudar a documentar las desigualdades educativas entre indígenas y no-indígenas queda irresuelta la pregunta del grado de validez de este tipo de evaluación en contextos de diversidad. Los cuestionamientos de la validez de los exámenes estandarizados van de la mano con el uso y las posibles consecuencias que los resultados de los exámenes puedan tener en las vidas de los estudiantes indígenas. Si bien existen medios para graduar la validez de las inferencias hechas a partir de los resultados en las pruebas, esto no soluciona los problemas de fondo que se refieren al uso de los exámenes como mecanismo para limitar injustamente las oportunidades educativas de los indígenas, así como el choque de perspectivas sobre el aprendizaje y conocimiento entre las culturas indígenas y la cultura representada en los exámenes.

En suma, este documento ha buscado sentar las bases para la discusión sobre evaluación estandarizada en contextos de diversidad cultural

y su relación con las desigualdades educativas. En este momento hay más preguntas que respuestas en este campo de investigación, pues a estas fechas aún desconocemos datos básicos como los niveles de sesgo en los exámenes. En América Latina nos queda un largo camino por recorrer en la investigación sobre la medición del rendimiento, la diversidad cultural y la equidad. Será necesario producir más investigaciones empíricas que guíen el entendimiento que tenemos sobre este tema en la región y den luces sobre las alternativas más adecuadas de acción en el campo de la evaluación.

Notas

¹ Agradezco enormemente a Guillermo Solano-Flores, Luz María Moreno, Paula Tápia y a dos dictaminadores anónimos, cuyos comentarios a versiones anteriores de este documento fueron muy valiosos para mejorarlo. Cualquier imprecisión o error en este estudio son totalmente responsabilidad mía.

² Chile es una excepción en América Latina, al menos, en tres aspectos. En primer lugar, el Sistema de Medición de la Calidad de la Educación (SIMCE) publica anualmente información sobre el desempeño de las escuelas de niveles básico y medio, con la intención de

poner presión pública para que mejoren. Segundo, el Ministerio de Educación de Chile usa los datos producidos por el SIMCE para distribuir, en cada región del país, recursos adicionales a las escuelas de más bajo desempeño, buscando tener consecuencias compensatorias. Finalmente, los datos del SIMCE, junto con otra información, han sido usados para otorgar premios al mérito a los docentes de las escuelas (McMeekin, 2000).

³ En este documento uso los términos occidental, eurocéntrico y cultura dominante como sinónimos.

Bibliografía

- AERA, APA y NCME (1999). *Standards for educational and psychological testing*, Washington, DC: American Educational Research Association.
- Apple, M. W. (2001). "Comparing neo-liberal projects and inequality in education", *Comparative Education*, vol. 37, núm. 4, pp. 409-423.
- Barrera, S. (1998). *Sistema de medición y evaluación de la calidad de la educación (SIMECAL) y la modalidad intercultural bilingüe en Bolivia (EIB)*, inédito, Bolivia.
- Benadusi, L. (2001). "Equity and education. A critical review of sociological research and thought", en Hutmacher *et al.* (eds.) *In pursuit of equity in education*, Dordrecht/ Boston/ Londres: Kluwer Academic Publishers.
- Bransford, J.; Brown, A. y Cocking, R. (eds.) (2000). *How people learn: brain, mind, experience, and school*. Expanded edition. Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC.
- Camilli, G. y Shepard, L. (1994). "Methods for identifying biased test items", *Measurement methods for the social sciences series*, vol. 4, Thousand Oaks, CA: Sage Publications.

- Cheng, K.M. (1996). "Quality assurance in education: The east asian perspective", en K. Watson, S. Modgil y C. Modgil (eds.) *Educational dilemmas: Debate and diversity*, vol 4: Quality in education, Londres: Cassell.
- Cizek, G. (1998). *Putting standardized tests to test*, Washington, DC: Thomas B. Fordham Foundation.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*, Cambridge, MA: Harvard University Press.
- De Haan, M. (1998-1999). *Learning as cultural practice: How children learning in mexican mazahua community*, Amsterdam:Thela Thesis Publishers.
- Diamond, J. y Spillane, J. (2004). "High-stakes accountability in urban elementary schools: challenging or reproducing inequality?", *Teachers College Record*, vol. 106, núm. 6, pp. 1145-1176.
- DOE (2003). *No child Left behind*, < <http://www.nclb.gov/>> Citado el 21 de abril de 2003.
- Falcon, R. (2002). *México descalzo*, México: Plaza Janés.
- Freedle, R. O. (2003). "Correcting the SAT's ethnic and social-class bias: A method for reestimating SAT scores", *Harvard Educational Review*, vol. 73, núm. 1, primavera.
- Fusarelli, L. (2004). "The potential impact of the No child left behind act on equity and diversity in American education", *Educational Policy*, vol. 18, núm 1, enero-marzo.
- Gopaul-McNicol, S. y Armour-Thomas, E. (2002). *Assessment and Culture: Psychological Tests with Minority Populations*, San Diego, CA: Academic Press.
- Grisay, A. (2003). "Translation procedures in OECD/PISA 2000 international assessment", *Language and Testing*, vol. 20, núm. 2, pp. 225-240.
- Gronmo, L.; Kjaernsli, M. y Lie, S. (2004). "Looking for cultural and geographical factors in patterns of responses to TIMSS items", en Papanastasiou (ed.) *Proceedings of the irc-2004 tims*, Chipre: Cyprus University Press, pp. 99-112.
- Hall, G. y Patrinos, H. (en prensa). *Indigenous peoples, poverty and human development in Latin America 1994-2004*, Washington, DC: World Bank.
- Hambleton, R. y Patsula, L. (2000). *Adapting tests for use in multiple languages and cultures*, Laboratory of Psychometric and Evaluation Research Report, Amherst, Mass.: Massachusetts University.
- Hambleton, R. (1993). *Translating achievement tests for use in cross-national studies*, Amherst, Massachusetts: International Association for the Evaluation of Educational Achievement and National Center for Educational Statistics.
- Hamilton, L. y Koretz, D. (2002). "Chapter 2: About tests and their use in test-based accountability systems", en Hamilton, *et al.* (eds.) *Making sense of test-based accountability in education*, Santa Monica, CA: Rand.
- Helms-Lorenz, M.; Van de Vijver, F. y Poortinga, Y. (2003). "Cross-cultural differences in cognitive performance and Spearman's hypothesis: g or c ?", *Intelligence*, vol. 31, pp. 9-29.
- IADB (2004). *Profile: Strategic Framework for Indigenous Development*. Indigenous Peoples and Community Development Unit. Inter-American Development Bank. <<http://www.iadb.org/sds/doc/IND%2DGN2295E2.pdf>> Citado en septiembre 15, 2004.

- ILO (2003). *Indigenous and Tribal Peoples Convention*. <<http://www.ilo.org/ilolex/english/convdisp1.htm>> Citado en mayo 6, 2003.
- INEGI (2000). *XII Censo General de Población y Vivienda*, México: INEGI.
- Joe, J. (1994). "Revaluating native-american concepts of development and education", en Greenfield y Cocking (eds.) *Cross-cultural Roots of Minority Child Development*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Kalantzis, M.; Cope, B. y Slade, D. (1989). *Minority languages and dominant culture: Issues of education, assessment and social equity*, Londres/Nueva York. The Falmer Press.
- Kirby, S. y Stecher, B (2004). "Chapter one: Introduction", en Stecher y Kirby (eds.) *Organizational improvement and accountability: lessons for education from other sectors*, Santa Monica, CA: Rand.
- Kudo, I. (2002). *Indigenous education in Peru: When opportunity speaks one language*. Santiago, Chile: UNESCO-OREALC.
- LLECE (1998). *Primer Informe de resultados*. Primer estudio internacional comparativo sobre lenguaje, matemática y factores asociados, para alumnos del tercer y cuarto grados de la educación básica, Santiago, Chile: UNESCO-OREALC.
- Maxwell, B. (1996). "Translation and cultural adaptation of the survey instruments", en Martin y Kelly (eds.) *Third international mathematics and science study (TIMSS) technical report*, volume I: *Design and development*, Chestnut Hill, MA: Boston College.
- McMeekin, R. (2000). "Implementing school-based merit awards: Chile's experience. Technical notes", *Education Reform and Management Publication Series*, vol. III, núm. 1, junio.
- McNeil, L. (2000). "Creating new inequalities: Contradictions of reform", *Phi Delta Kappan*, pp. 729-734, junio.
- McQueen, J. y Mendelovits, J. (2003). "PISA reading: cultural equivalence in a cross-cultural study", *Language and Testing*, vol. 20, núm. 2, pp. 208-224.
- Messick, S. (1993). "Validity", en Linn (ed.) *Educational measurement*, Nueva York: National Council on Measurement in Education and American Council on Education.
- Mier y Terán, M. y Romero, C. (2003). "Inequalities in mexican children's schooling", *Journal of Comparative Family Studies*, 34(3): 435-454.
- Muller, C. y Schiller, K. (2000). "Leveling the playing field? Students' educational attainment and states' performance testing", *Sociology of Education*, vol. 73, pp. 196-218.
- NCR-National Research Council (1997). *Educating language-minority children*, Washington, DC: National Academy Press.
- Padilla, A. (2001). "Issues in culturally appropriate assessment", en L. Suzuki, J. Ponterorro y P. Meller (eds.) *Handbook of Multicultural Assessment: Clinical, Psychological and Educational Applications*, San Francisco, CA: Jossey-Bass.
- Popham, W.J. (2000). *Modern educational measurement: Practical guidelines for educational leaders*, 3a ed., Boston: Allyn & Bacon.
- Price-Williams, D. (1975). *Explorations in cross-cultural psychology*, San Francisco, CA: Chandler & Sharp Publishers.
- Ravela, P. (2001). *¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina?* Santiago, Chile: PREAL.

- Roderick, M. *et al.* (1999). *Ending social promotion: Results from the first two years*, Chicago, Illinois: Consortium on Chicago School Research.
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*, Cambridge, Mass.: Perseus Books.
- Solano-Flores, G. y Nelson-Barber, S. (2001). "On the cultural validity of science assessments", *Journal of Research in Science Teaching*, vol. 38, núm. 5, pp. 553-573
- Suárez-Orozco, M. (1994). "Remaking of psychological anthropology", en Suárez-Orozco, *et al. The making of psychological anthropology II*. Forth Worth, Tx: Harcourt Brace College Publishers.
- Tharpe, R. (1994). "Intergroup differences among native americans in socialization and child cognition: An ethnogenetic analysis", en Greenfield y Cocking (eds.) *Cross-cultural roots of minority child development*, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Tirado, F., Backhoff, E., Larrazolo, N. y Rosas, M. (1997). "Validez predictiva del examen de habilidades y conocimientos básicos (EXHCOBA)", *Revista Mexicana de Investigación Educativa*, enero-junio 1997, vol 2, núm. 3, pp. 67-84.
- Treviño, E. y G. Treviño (2004). *Estudio sobre las desigualdades educativas en México: la incidencia de la escuela en el desempeño académico de los alumnos y el rol de los docentes*, colección Cuadernos de investigación, núm 5, México: Instituto Nacional para la Evaluación de la Educación.
- Treviño, E. (en prensa). *Buenas prácticas en el uso de evaluaciones educativas para decisiones de política en América Latina*, Santiago, Chile: UNESCO-OREALC.
- UNESCO (2000). *Informe subregional de América Latina: Evaluación de educación para todos en el año 2000*, Santiago, Chile: OREALC.
- Van de Vijver, F. (1997). "Meta-analysis of cross-cultural comparisons of cognitive test performance", *Journal of Cross-Cultural Psychology*, vol. 28, núm. 6, noviembre, pp. 678-709.
- Werner, E. (1979). *Cross-cultural child development: A view from the planet earth*, Monterey, CA: Brooks/Cole Publishing Co.
- Winkler, D. (2004). "Investigaciones sobre etnicidad, raza, género y educación en las Américas", en Winkler y Cueto (eds.) *Etnicidad, raza, género y educación en América Latina*, Santiago, Chile: PREAL.
- Wolff, L. (1998). *Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos*, Washington, DC/Santiago, Chile: PREAL.

Artículo recibido: 10 de noviembre de 2004

Dictamen: 1 de febrero de 2005

Segunda versión: 17 de junio de 2005

Revisión final: 20 de julio de 2005

Aceptado: 20 de septiembre de 2005