

Tener presentes las singularidades lingüísticas y culturales en las evaluaciones internacionales de las competencias de los alumnos: ¿una nueva dimensión para PISA?

Gérard Bonnet

Delegación para las Relaciones Europeas e Internacionales. Dirección de la Evaluación y la Prospectiva. Ministerio de Educación Nacional de Francia

Resumen:

Este artículo intenta poner de relieve ciertas dificultades vinculadas a los objetivos y a las metodologías existentes en la evaluación comparativa de las competencias de los alumnos, y especialmente las puestas en marcha en el marco del proyecto PISA (*Programme for International Student Assessment*) de la OCDE. Examina la naturaleza de estas dificultades, que son a la vez lingüísticas y culturales en el sentido más amplio, los límites que imponen, y propone un acercamiento complementario. Este acercamiento se basa en un recurso teorizado y armonizado en los protocolos nacionales de evaluación concebidos en la o las lenguas de los alumnos. Este acercamiento podría ser utilizado como complemento de la metodología existente y permitiría a la vez relativizar los palmarés del país y proporcionar informaciones precisas para cada país, en términos pedagógicos, y no sólo en el campo de los sistemas.

Palabras clave: evaluación comparada, PISA, OCDE, sesgo cultural, Francia, contexto nacional.

Abstract: *Sustaining Linguistic and Cultural Diversity within the International Assessment on Students' Competencies: A New Dimension for PISA?*

This article sets out some specific difficulties related to present targets and methods concerning the comparative assessment of students' capacities, especially those carried out within the PISA study of the OECD. On the other hand, it inquires into the origin of these difficulties, which are both linguistic and cultural in a broad sense, and the limits resulting from them while suggesting a complementary approach. This approach lies on a specific resource which has been adapted to national procedures on assessment carried out in the students' mother tongues. Similarly, it could be used as a complement of present methodology, both as a way of diminishing the importance of countries' records and as a way of providing accurate information on each country from a pedagogical point of view, not just taking into account the different education systems.

Key words: comparative evaluation, PISA, OECD, cultural bias, methodology, France, national context.

INTRODUCCIÓN

Desde que existe hace más de cinco años el *Programme for International Student Assessment* (PISA), de la Organización para la Cooperación y Desarrollo Económico (OCDE), y tras la publicación de los resultados de los dos primeros ciclos, esta evaluación internacional de capacidades en el ámbito escolar ha alcanzado su carta de nobleza y se ha inscrito de manera perdurable en el marco comparativo. Su éxito, real en cuanto a sus resultados, pero también inteligentemente dirigido por los responsables y la población gracias a operaciones de comunicación muy logradas, ha impuesto la idea de que PISA ha conseguido una especie de perfección metodológica que sería difícil de superar.

No obstante, cualesquiera que sean los adelantos científicos que han permitido y marcado su desarrollo, PISA, como cualquier empresa humana, deberá evolucionar con el tiempo. Esta evolución metodológica necesaria pasa por el reconocimiento de los límites y la aceptación de nuevas perspectivas.

Una buena manera de trabajar en este sentido es preguntarse sobre lo que PISA podría aportar además por tratarse de la medida de competencias contextualizadas, con el fin de evaluar de la manera más profunda los resultados del sistema educativo de cada país. Para esto, es necesario un nuevo acercamiento. Éstos son los puntos que se van a discutir a continuación.

EL CONTEXTO Y LOS OBJETIVOS PISA

Para abordar la reflexión, es conveniente señalar que existen expertos que subrayan los límites y el sentido de las evaluaciones internacionales como PISA. Citemos, por ejemplo, a un autor americano que escribe en una revista mensual especializada en educación y muy influyente entre los miembros de la Administración y del Congreso de Washington:

Me he quedado estupefacto por la aceptación sin espíritu crítico de las evaluaciones internacionales en este país. Me parece que ciertos investigadores y especialistas en psicomotricidad, por otra parte competentes, abandonan toda facultad crítica cuando se trata de datos de la *International Association for the Evaluation of Educational Achievement* (IEA) –los trabajos TIMSS y PIRLS– o de la Organización para la Cooperación y el Desarrollo Económico (OCDE)-PISA. Para decirlo de manera técnica, se vuelven gagás. Está claro que existe una presión política para aceptar estos trabajos, pero veo también ahí una pequeña industria resplandeciente de la cual depende un cierto número de carreras (Bracey, 2005).

Una cuestión en particular ha sido ya objeto de discusiones que ilustran los interrogantes existentes sobre las enseñanzas de PISA: ¿en qué medida la evaluación de las competencias de los alumnos en este estudio comporta sesgos culturales diversos que harían difícil o errónea la comparación entre los países?

Como preámbulo a esta cuestión, conviene recordar el contexto y los objetivos particulares de PISA. Este proyecto tiene como fin producir periódicamente

indicadores comparativos de la eficacia de los sistemas educativos a través de la evaluación de las competencias en lectura, en matemáticas y en ciencias de pruebas representativas de alumnos de 15 años, organizada en ciclos de tres años. Esta organización en ciclos tiene especialmente por objetivo permitir un seguimiento de las evoluciones en el tiempo y apreciar el impacto de eventuales reformas de los sistemas educativos que tienden, a fin de cuentas, a promover.

PISA pone la mirada en el tipo de edad que llega al final de la escolaridad obligatoria en la mayoría de los países de la OCDE, cualquiera que sea su recorrido escolar, y cualesquiera que sean sus objetivos futuros (continuar los estudios o el comienzo de la vida laboral). Los alumnos son evaluados no por los conocimientos en el sentido estricto, sino por sus capacidades de movilizar y aplicar sus conocimientos en situaciones diversas, en ocasiones muy diferentes a las encontradas en el marco de la vida escolar. La utilización en inglés del término *literacy* (*reading literacy, science literacy*) pone bien de manifiesto que no son éstos los conocimientos disciplinarios a los que se apunta. Los ejercicios de evaluación propuestos son objeto de un compromiso en el ámbito internacional sobre lo que es considerado como necesario para el futuro ciudadano, principalmente para el acceso al mercado laboral. En esto, PISA se desmarca de las encuestas similares hechas desde hace numerosos años por la IEA, que buscan valorar las competencias de los alumnos en relación con los programas de enseñanza.

Esta orientación hacia las capacidades esencialmente utilitarias es por otra parte comprensible y legítima si se considera que la OCDE es ante todo una institución que se interesa por el desarrollo económico. Sin embargo, el objetivo económico no es más que uno entre otros objetivos en la mayoría de los sistemas educativos y lo que es evaluado no permite medir todo lo que los alumnos saben hacer. Esto conduce incluso con frecuencia a poner de relieve lo que ellos no saben hacer. Este último punto no es a pesar de todo negativo, ya que los resultados de PISA, que permiten especialmente revelar los puntos fuertes y los puntos débiles de los alumnos de cada país en el contexto internacional, proporcionan informaciones muy valiosas, complementarias de las evaluaciones nacionales efectuadas en los países donde existen. Con todo, no es justo, ni siquiera constructivo, ofrecer una imagen partidaria y parcial de lo que un país llega a enseñar a sus alumnos sobre la base de criterios restrictivos.

El dictamen nacional PISA 2003 establecido en la Suiza de habla francesa subraya la concepción particular de las singularidades culturales que prevalecen en esta encuesta.

PISA enfoca la medida de las mismas capacidades en el conjunto de países participantes. Realizar este objetivo implica de alguna manera definir las capacidades «universales» y neutralizar las singularidades culturales con el fin de asegurar la comparación entre los países. En otras palabras, las singularidades culturales se toman a la postre más como un obstáculo que hay que resolver que como un elemento que pudiera formar parte de las capacidades de los alumnos. Además, subrayaremos que PISA busca concretar las competencias necesarias para trabajar en la vida cotidiana, al margen de lo adquirido en la escuela, aunque la base de información sea el entorno escolar con instrumentos similares a los cuestionarios de tipo escolar (pruebas objetivas) (IRD, 2005, p. 24).

Notemos que ya se habían hecho observaciones parecidas a propósito de PISA 2000 (Bonnet, 2002).

Así pues, es en este marco en el que deben ser aclaradas algunas observaciones metodológicas que han sido expuestas a propósito de PISA, en particular por los expertos franceses.

¿POR QUÉ RESERVAS METODOLÓGICAS?

No es necesario recordar las razones que han llevado a Francia a hacer hincapié en estos problemas. La razón no está, como se oye decir con frecuencia, en que Francia busque mostrar su desacuerdo con la filosofía de PISA, o bien poner en duda los resultados obtenidos por sus alumnos en el 2000 y en el 2003.

Para rebatir este último punto, basta con observar que estos resultados se mantienen en la media europea y conformes a lo que muestran desde hace unos años las evaluaciones nacionales llevadas a cabo en el país. No justifican, en ningún caso, que nos aferremos a discutirlos para defender un orgullo nacional fuera de lugar, además de que no aportan nada fundamentalmente nuevo sobre el sistema educativo francés, en todo caso para los franceses.

Para convencerse de esto, no hay más que ver los resultados de PISA 2003 que se basaron en Francia en una muestra representativa de unos 5.000 alumnos, escolarizados básicamente en institutos, en segundo general y tecnológico, o en colegios, en clase de 3°. En el seno de 30 países de la OCDE, Francia alcanza un nivel global en «cultura matemática» significativamente (en el sentido estadístico del término) por encima de la media de estos países. Igualmente, en «cultura científica», su nivel está también significativamente por encima de la media. En «resolución de problemas», Francia exhibe sus mejores resultados, mientras que en «comprensión escrita» su nivel permanece en la media.

Esta evaluación del 2003 ha permitido, entre otras cosas, tomando los ejercicios de la evaluación del 2000, comparar por primera vez en el marco de PISA los conocimientos adquiridos por los alumnos en el transcurso del tiempo. En «comprensión escrita» y en «cultura matemática», el nivel de los alumnos franceses se mantiene igual a los resultados de hace tres años, mientras que en «cultura científica» progresa ligeramente. Señalemos de paso que esta comparación temporal debe ser tomada con mucha precaución, ya que es verdad que la muestra se basa en un número limitado de ítems.

Los niveles de los alumnos franceses son relativamente poco dispersos, pero se observan desviaciones muy importantes de la media entre alumnos de 3° y alumnos de 2° general y tecnológico en todos los dominios («cultura matemática», «comprensión escrita», «cultura científica»), así como entre chicas y chicos en «cultura matemática» y en «comprensión escrita». Los alumnos que están al día, es decir, aquéllos que no han repetido nunca, estén en 2° o en 3°, obtienen resultados muy parecidos. Estos resultados son igualmente idénticos, incluso mejores a los de otros países cuyos alumnos obtienen niveles más altos. Todo lo contrario, es cierto, para los que han repetido (Bourny, G. et al., 2004).

Esta breve exposición muestra con claridad que la situación en Francia, no siendo excepcionalmente positiva, tampoco tiene nada de deshonrosa. Por lo tanto, en ningún momento se ha planteado la cuestión de revisar los resultados obtenidos, incluso aunque sea indispensable recordar los límites de interpretación que se pueden hacer.

La sugestión según la cual Francia se opone sistemáticamente a PISA por razones ideológicas proviene sin duda de que por su parte la OCDE recalca con frecuencia que Francia hace una utilización bastante limitada de las indicaciones de PISA. Es verdad que, en comparación con lo que se hace en otros países (Alemania y Hungría, por ejemplo), ni la prensa francesa ni el Ministerio de Educación Nacional han difundido con exceso la publicación por la OCDE de los informes internacionales de PISA 2000 ó 2003. Se han incluido en la prensa presentaciones oficiales de los resultados, que tuvieron un eco positivo, se han hecho publicaciones y conferencias, sin que acapararan la atención de la Nación más que lo que era justificado.

La razón reside sin duda en que en Francia, al contrario que otros países, tanto la prensa como los políticos y los ciudadanos están acostumbrados a disponer de los resultados de la evaluación de los alumnos. La dirección de «La Evaluación y de la Prospectiva» del Ministerio de Educación Nacional organiza, en efecto, desde hace mucho tiempo, evaluaciones nacionales y participa en evaluaciones internacionales y europeas, que los resultados de PISA no hacen más que complementar y aclarar. El carácter «sensacional» que PISA presenta en algunos países no se justificaría en Francia donde su impacto se encuentra naturalmente condicionado y puesto en tela de juicio. Por otra parte, como se acaba de recordar, los resultados de los alumnos franceses, ligeramente superiores a la media de la OCDE tanto en el 2000 como en el 2003, no conducen a reacciones inmediatas de los responsables que desembocarían en reformas espectaculares. Por el contrario, Francia concentra los análisis de sus resultados no en una comparación poco fructífera entre países expresada en términos de palmarés, sino en una interpretación pedagógica delicada de las diferentes competencias medidas que permite enfocar las medidas de fondo, a largo plazo, en la enseñanza en clase.

Es pues en estos aspectos en los que hay que buscar las causas de la sensibilidad francesa a las cuestiones de sesgos culturales.

Por todo esto, conviene remontarse en el tiempo hasta mediados de los noventa, cuando varios países, coordinados por *Statistics Canada*, realizaron juntos y por primera vez una encuesta internacional para evaluar las competencias básicas de adultos en una situación de vida cotidiana, la *International Adult Literacy Survey* (IALS). Esta encuesta fue realizada cerca de los lugares de trabajo, donde vivían.

Después de haber participado en esta encuesta en 1995-1996, cuyos resultados y conclusiones la OCDE pidió publicar, Francia decidió *in extremis*, por razones de orden metodológico, retirar los datos franceses. En efecto, se publicaron los primeros resultados franceses en los que no se daba cuenta de la realidad en el país, ya que, según ellos, más del 40% de los franceses no dominaban las capacidades de lectura en su propia lengua.

La retirada de Francia de la encuesta fue decidida a la vista del informe remitido por tres expertos internacionales sobre la metodología y el desarrollo de la encuesta, informe que concluye afirmando la imposibilidad de comparar los resultados entre países. En él se observó que el cuestionario de evaluación estaba formado por una mitad de preguntas relacionadas con oficios y textos que provenían de Estados Unidos o de Canadá, lo que traduce a las claras el peso de los países anglo-sajones en este trabajo y explica los sesgos culturales, lingüísticos y de traducción advertidos.

Aunque un buen número de preguntas sesgadas hayan sido identificadas (Binckley y Pignal, 1998) con la ayuda de los procedimientos psicométricos habituales (Modelos de Respuestas al Ítem; Estadísticas de Mantel-Haenszel), otros sesgos fueron descubiertos por el análisis complementario de la versión francesa de los ítems (Guérin-Pace y Blum, 1999; Blum y Guérin, 2000). Como en el documento Siréci (1997), cuando las preguntas fueron traducidas, no fue posible determinar si una diferencia en los grados de éxito entre dos poblaciones era debido a que el nivel de dificultad de los ítems no eran equivalentes (sesgados) o al hecho de que estas poblaciones no tenían el mismo nivel de capacidad (impacto).

Otros trabajos realizados *a posteriori* han confirmado el papel determinante del contexto de la encuesta sobre los resultados. Por esta razón, el grado de éxito en los mismos cuestionarios propuestos a la misma clase de personas puede alcanzar 50 puntos más que en otro contexto, el de un reclutamiento o de una búsqueda de empleo, por ejemplo.

No cabe la menor duda de que la metodología utilizada en IALS está lejos de ser satisfactoria. La Comisión Europea confió, en 1996, al término de una oferta en el marco del programa LEONARDO, a la *Office for National Statistics* británico un estudio metodológico en relación con las condiciones de comparabilidad de la encuesta cuyas conclusiones, medidas, resaltan las dificultades (Carey, 2000).

Los sesgos encontrados en IALS eran a la vez de orden lingüístico y cultural, principalmente en este último punto, donde había una marcada influencia sobre las actitudes culturales en torno a la motivación en las personas consultadas. Hay motivos para pensar que tales factores están igualmente presentes en una encuesta como PISA en relación con las capacidades de los alumnos en el ambiente escolar, incluso si se muestran de forma diferente.

Por todo esto, los expertos franceses se mostraron particularmente cautelosos en estas cuestiones y estuvieron muy atentos a estos sesgos culturales y lingüísticos que pudieran existir en los formularios de evaluación de PISA, principalmente en los cuestionarios de lectura, en perjuicio de los esfuerzos realizados para corregirlos. Esta tendencia se encuentra reforzada por la designación de expertos que, en los diversos campos disciplinarios, son elegidos por la OCDE para elaborar, en relación con el consorcio encargado de conducir PISA, las orientaciones y el contenido de las pruebas de evaluación. Estos expertos son elegidos con mucha frecuencia entre el vivero informal de la OCDE y son, por una aplastante mayoría, oriundos de países de América del Norte y de países de la zona del Pacífico. La consecuencia está muy clara, reforzar la trayectoria de los sesgos

metodológicos y culturales en beneficio de un acercamiento anglo-sajón y de la primacía de la lengua inglesa.

Naturalmente, estas dudas por parte de Francia han sido fuertemente contestadas por los expertos y los responsables de PISA, que hacen constar que los países que han proporcionado algunos ítems de evaluación no los resuelven mejor, e incluso a veces lo hacen peor, que los demás países. Esto es sin duda exacto desde un punto de vista puramente estadístico. No obstante, la verdadera cuestión es saber cuáles hubieran sido los resultados con preguntas diferentes. En efecto, una de las principales críticas que se pueden hacer a esta metodología es que está enteramente concebida para responder a las exigencias de la estadística comparada. Esta preocupación es ciertamente legítima, pero conduce a guardar en la evaluación únicamente los ítems que «funcionan» de la misma manera en todos los países, por lo que en consecuencia lo que se evalúa está particularmente empobrecido y desprovisto del «sentido» cultural.

LA CUESTIÓN LINGÜÍSTICA Y CULTURAL

La cuestión de los sesgos o distorsiones lingüísticos está con frecuencia reducida a un problema de equivalencia de las traducciones del material de evaluación (textos e ítems, cuestionarios) entre su lengua de origen, con frecuencia el inglés, y las otras lenguas de transmisión de los tests.

A este respecto, conviene recordar que todo el material de evaluación utilizado por PISA es muy valioso por proceder de las dos lenguas originarias, a saber, de las dos lenguas oficiales de la OCDE, el inglés y el francés. Concretamente, esto significa que este material es apto para ser traducido a las otras lenguas a partir de estas dos fuentes de origen, consideradas como estrictamente equivalentes. Esto ya implica un cierto nivel de traducción fácil cuando la lengua original sea el inglés o el francés y exigirá doble esfuerzo cuando sea de otra lengua (por ejemplo, un texto original en coreano debe ser traducido al inglés y al francés –de hecho es principalmente traducido al inglés y después la traducción inglesa se traduce al francés). A partir de estas dos traducciones originales se debe traducir a continuación a otras lenguas utilizadas por los alumnos en las pruebas de PISA.

Además, una traducción, por más perfecta y fiel que sea, puede no presentar los mismos niveles de dificultad o tener las mismas connotaciones significativas que el original, cuando es utilizada en el contexto psicométrico de una prueba de evaluación.

Sin embargo, hay que resaltar con toda objetividad los esfuerzos alcanzados por los responsables de PISA a favor de reducir estos sesgos lingüísticos y culturales en relación con lo hecho por IALS. Estos esfuerzos son reales y han conseguido reducir sensiblemente algunos de estos sesgos, en particular mediante una metodología de traducción más esmerada. Ésta se basa esencialmente en varias traducciones del mismo documento realizadas de manera independiente por traductores profesionales que a continuación son «reconciliadas» y verificadas por expertos. El recurso a la traducción inversa (del texto traducido a su lengua original) puede ser igualmente útil bajo ciertas condiciones.

Por otra parte, se tiene un especial cuidado en la traducción de documentos a las dos lenguas originales (inglés y francés) gracias a una interacción casi simultánea entre los traductores y los que confeccionan las pruebas.

Además, a pesar del mucho interés puesto en los procesos de verificación realizados por el consorcio encargado de la puesta en marcha de PISA, la equivalencia lingüística de las pruebas expuestas a los alumnos depende en gran parte de las competencias que existen en los países en este campo, y podemos pensar que son bastante desiguales. Que los diferentes métodos psicométricos y estadísticos permitan al final resolver una parte del problema –principalmente, eliminando algunas preguntas- no es del todo satisfactorio desde el punto de vista intelectual.

El dispositivo lingüístico desarrollado en PISA es altamente complejo y costoso. Responde sin duda a las exigencias de una estadística comparativa, pero puede tener por efecto vaciar de su contenido cultural a una parte del material empleado.

Es ahí donde surge la pregunta principal. No se trata de saber si es posible o imposible que un texto traducido sea suficientemente equivalente al original como para permitir la comparación. Sabemos que por un lado existen diferencias irreductibles entre las lenguas y que por otra parte la traducción permite con bastante frecuencia establecer un puente entre estas lenguas. Sin embargo, en términos de comparación internacional, esta equivalencia tiene un precio (la reducción del contenido o la eliminación pura y simple de algunas preguntas) que nos conduce a no valorar más que algunas capacidades «universales» en detrimento de otras competencias o formas de medir que corresponden más a los objetivos de tal o cual sistema educativo.

Se puede razonablemente aceptar que el proceso de traducción, independientemente de su calidad, no es neutro, aunque sólo fuera como consecuencia del número de palabras necesarias en lenguas diferentes para decir «la misma cosa». Se ha señalado que, cuanto más se alarga un texto inglés en su traducción al francés, más difíciles son las preguntas para los alumnos francófonos en comparación con los alumnos anglófonos (Adams y Wu, 2002).

Por otro lado, los sesgos lingüísticos, ligados o no a la traducción, no son más que un aspecto del problema. Mientras que, en estas evaluaciones, la comprobación de los resultados y de las capacidades es difícil de separar, la influencia del ambiente cultural y escolar sí es fundamental. Entre lo que puede afectar a los resultados, señalemos el contenido de las enseñanzas, las modalidades de enseñanza y de aprendizaje y las modalidades de evaluación. Sobre este último punto, se han observado resultados diferentes en PISA según que los alumnos estén o no familiarizados con el cuestionario de elección múltiple o de respuestas abiertas. Tampoco hay que desestimar la importancia de la motivación de los alumnos y de su actitud general en el ambiente escolar, ya que, aunque se puedan explicar los factores sociales en el seno del mismo país, estos mismos factores pueden ser diferentes de un país a otro.

Un cierto número de expertos subrayan que PISA está fuertemente influida por «el modelo anglo-sajón» (Romainville, 2002). Otros, defensores de la metodolo-

gía en cuestión, ironizan con mucho gusto sobre esta crítica remarcando que esto mete en el mismo saco a la mayoría de los países anglófonos, escandinavos y del sur-este asiático (Lafontaine y Demeuse, 2002). Abstraerse del hecho de que no es la lengua el objetivo, sino una cierta manera de percibir el mundo y de enfocar la educación pueden muy bien en cierta medida ser comunes a estos países.

La sensibilidad de las evaluaciones en el contexto cultural está reforzada en PISA por el hecho de que lo esencial de los métodos estadísticos utilizados (los modelos de pregunta-respuesta, con frecuencia conocidos de manera impropia con el vocablo inglés de *Item response Theory*, ya que no son «teorías» en el sentido científico del término, sino simples modelos estadísticos) presuponen que la hipótesis de unidimensionalidad que los sustenta sea real. Así pues, aunque se puede aceptar que la unidimensionalidad se verifica por los ítems de la lengua dada, es, sin embargo, más discutible en el caso de ítems traducidos en lenguas diferentes (Goldstein, Bonnet y Rocher, pendiente de publicación).

De una manera más general, el recurso de esta técnica psicométrica no es unánime, como lo afirma el autor americano ya citado:

... PISA ha sido fuertemente criticada por su metodología que utiliza la teoría de pregunta-respuesta para construir escalas publicadas a este respecto. Los educadores americanos deberían mostrarse escépticos con relación a PISA, y me gustaría que el Ministerio Federal Americano de Educación y los investigadores en educación hicieran lo mismo (Bracey, 2005).

Una buena manera de ilustrar estadísticamente la influencia del contexto cultural sería elaborar una clasificación jerárquica de países según sus perfiles de éxito en los ítems de comprensión lectora, como lo hizo PISA 2000 (Rocher, 2004). Se constata, pues, que la hipótesis de unidimensionalidad no se verifica y que los análisis estadísticos fundados sobre los modelos de pregunta-respuesta no son suficientes para eliminar las diferencias y las similitudes culturales entre países. Se observa en esta clasificación que grupos de países fracasan o triunfan en los mismos ítems mientras que otros grupos de países fracasan o triunfan en ítems diferentes. Por esta razón, Francia, Bélgica e Italia y también los Países Bajos y Finlandia forman el mismo grupo, mientras que los países anglófonos están reunidos en otro grupo, próximo a Japón y a Corea. Dinamarca, Suecia, Islandia y Noruega reúnen por su parte las mismas características. Tal observación, a pesar de que no explica nada, tiende al menos a sugerir que los sesgos de tipo cultural existen perfectamente.

En su artículo, Rocher confirma por otra parte la influencia de ciertos factores mencionados más arriba, como los programas de enseñanza, demostrando de forma particular que los alumnos americanos resuelven mejor los ítems enfocados a textos de carácter profesional (demanda de empleo) que los alumnos franceses.

Por lo que respecta a la influencia de los modos de evaluación en la escuela, un análisis de los resultados franceses de PISA 2000 demostró que, de manera general, los alumnos franceses tienen tendencia a abstenerse de responder en las cuestiones abiertas, ya sea a causa de las dificultades de expresión escrita, ya sea

por miedo a dar una respuesta falsa, ya sea, en fin, porque no están normalmente acostumbrados a este tipo de cuestionamiento (Bourny et al., 2001)

Vemos así que los posibles sesgos lingüísticos debidos a la traducción no son más que uno de los aspectos del problema general de las limitaciones de esta metodología. En resumen, la elección para evaluar las capacidades «universales», válidas en el mundo globalizado que conocemos, y medir de manera estrictamente estadística, deja de lado el contexto lingüístico, cultural y escolar de los países afectados. Varios países han constatado este tipo de dificultad.

El informe suizo ya citado subraya, entre otros, un aspecto de la singularidad cultural:

En fin, como todo proyecto internacional, PISA se proyectó de manera que este conjunto de cuestiones sobre el entorno escolar pudiera aplicarse a todos los países. Las características específicas de cada sistema escolar nacional no se pueden tomar en cuenta de manera precisa. Se observa principalmente por el hecho de que una parte del cuestionario para las escuelas no estaba adaptado a la situación de los establecimientos helvéticos (IRDP, 2005, p. 24).

Otro ejemplo de la importancia determinante del contexto cultural se encuentra en Luxemburgo, país del plurilingüismo por excelencia. En este país, los alumnos no están escolarizados en luxemburgués, que es su lengua materna (excepto para los alumnos procedentes de la inmigración), sino que son escolarizados en alemán o en francés, que no son sus lenguas maternas. Estos alumnos tuvieron que pasar las pruebas de matemáticas de PISA 2000 en lengua alemana y en lengua francesa, en función de la clase de establecimiento frecuentado, utilizando una u otra lengua. Así pues, para ciertos alumnos la lengua de las pruebas no coincidía con la que, por razones personales o familiares, dominaban mejor. Sin embargo, se impuso esta regla, y los resultados luxemburgueses se resintieron y no se corresponden con la realidad del dominio de estas capacidades en el Gran Ducado.

Además de las informaciones sobre las competencias que PISA eligió para medir en razón de su importancia en todos los países concernientes, cada país tiene igualmente necesidad de evaluar de manera comparativa los conocimientos de sus alumnos en relación con un contexto cultural que le es propio. Si es útil saber que en tal país el mayor número de alumnos es capaz de adquirir, de manera equitativa, las capacidades necesarias para encontrar un empleo, es útil también saber si sus alumnos han asimilado las capacidades específicas que los distinguen culturalmente de otros países y que la escuela se empeña en enseñárselas, comparándolas con las de otros países similares o no.

Por esta clase de medida, la metodología de PISA está imperfectamente adaptada, ya que trata las singularidades culturales como un problema que pretende eliminar. Debería pues evolucionar a fin de proponer, no en lugar de los resultados que presenta, sino como complemento de éstos, un análisis más profundo de los resultados de los sistemas educativos que tome en cuenta el contexto nacional o regional.

Otra metodología, que tenga en cuenta la lengua de los alumnos y su cultura y no exclusivamente cimentada, como es el caso de PISA, sobre un cuestionario único traducido a varias decenas de lenguas y que refleja una cultura global, permitiría sin duda análisis más operativos en el plan pedagógico. Pues nunca hay que olvidar que, si un análisis económico y social de la educación es siempre útil, la finalidad primera de la escuela es la de formar a los jóvenes en su medio cultural históricamente constituido, lo que implica poder mejorarlos actualizando los contenidos y poniendo en marcha prácticas pedagógicas.

OTRO ACERCAMIENTO METODOLÓGICO

La dificultad para proponer tales acercamientos es que se tropieza con los hábitos y metodologías elaborados en el campo de las evaluaciones comparativas desde hace más de medio siglo, por lo que, como se ha visto, lo más importante es la estricta comparación estadística. El desarrollo de las técnicas psicométricas y muy especialmente de los modelos de pregunta-respuesta ha sido notable y ha marcado ampliamente la metodología comparativa de las capacidades imponiendo los límites mencionados.

Otro aspecto de esta dificultad es de orden financiero. Las instituciones, públicas o privadas, que desde hace años se comprometieron con la metodología actual, han dedicado a ello grandes sumas de dinero y esperan ahora, como es lógico, rentabilizar esta inversión. Estando en posición casi hegemónica en el mundo, se mantienen ahí mientras que las organizaciones internacionales (UNESCO, OCDE, Banco Mundial, etc.) llaman regularmente a sus expertos para pedir consejo. En un informe a la *Association for Educational Assessment in Europe* de noviembre de 2004, un famoso estadístico inglés presentó una excelente descripción del fenómeno (Goldstein, 2004).

Desde entonces, es muy difícil hacer valer que se puedan poner en práctica otros acercamientos metodológicos, en la medida en que estos expertos están bien situados, en nombre de su experiencia en este campo, para rechazar toda novedad en la que ellos no dominaran sus consecuencias. ¡En otros tiempos, otros expertos enviaban a la hoguera a los que pretendían afirmar que el sol no daba vueltas alrededor de la tierra!

Sin querer promover una revolución de tipo copernicano en el universo de las comparaciones internacionales, hay que señalar con insistencia que la metodología dominante, aunque debiera mantenerse exclusiva, tendría el inconveniente, no solamente de imponer una concepción de la educación muy específica, centrada en la adquisición de las competencias nacidas de la mundialización, sino también el de prohibir toda innovación. Estos creadores tendrían interés en defender lo que han elaborado. Para resolver este problema hay que aceptar que el futuro de las comparaciones internacionales se encuentra en los acercamientos plurales y complementarios más que en la supremacía total de una metodología.

La metodología modificada que podría ser la prevista se desmarca de la habitualmente utilizada en la búsqueda del respeto a los particularismos lingüísticos,

escolares y culturales de los países afectados. Se trata de tener en cuenta las singularidades culturales e intentar dar una imagen lo más fiel posible de las capacidades de cada uno. La diversidad y diferencia no son concebidas como obstáculos para la comparación, sino como parámetros que deben ser respetados si se quieren producir resultados contextualizados, es decir, útiles.

Ciertos sistemas educativos, particularmente en los países donde la tradición humanística tiene aún sentido, conceden una gran importancia a la adquisición de valores culturales, especialmente literarios o artísticos, que confirman la pertenencia de los alumnos a una sociedad determinada, a un sistema de valores específico y a una nación particular.

Además, los países tienen tradiciones de enseñanza, hábitos específicos de trabajo y de evaluación diferentes, que hay que comprender y respetar cuando se quieren medir las capacidades de los alumnos. Con seguridad, una de las singularidades más fuertes es la lingüística.

En cuanto se desea colocar la evaluación de las capacidades en el contexto de los programas de la enseñanza y de los modos de funcionamiento educativo nacionales, es conveniente respetar esta singularidad lingüística que es consustancial a ello. Para hacer esto, conviene avanzar la hipótesis de base de que es posible producir datos comparables a partir no solamente de un formulario de evaluación único común a todos los países y traducido a diferentes lenguas, sino construyendo, país por país, formularios nacionales fundados sobre la base de un marco común de evaluación que permite medir las mismas capacidades a iguales niveles de dificultad de un país y por consecuencia de una lengua a otra. Con claridad, esto quiere decir evitar traducir textos y cuestionarios, y también construir en cada lengua un instrumento de evaluación de igual dificultad y que mida las mismas dificultades.

Sin embargo, los especialistas en métodos psicométricos explican que esto no permite garantizar que sea comparable la medición de las capacidades. Es conveniente prever por añadidura un dispositivo de anclaje de instrumentos nacionales entre ellos, en donde uno de los componentes puede incluir algunos textos e ítems para traducir (y no traducidos –veremos más tarde la diferencia).

Tal metodología ha sido ya objeto de una reflexión y de una experimentación parcial en el escenario de dos proyectos europeos llevados a cabo en el marco de la *Red europea de los responsables de políticas de evaluación de los sistemas educativos*.¹

(1) La Red europea de los responsables de políticas de evaluación de los sistemas educativos fue creada a raíz de una reunión de altos funcionarios de educación durante la Presidencia francesa de la Unión Europea de 1995. Es una red intergubernamental cuyos miembros han sido designados por los gobiernos por mediación de sus representantes en el Comité de educación. Estos miembros tienen, en diferentes grados, responsabilidades en el campo de la evaluación y en la dirección de su sistema educativo en el ámbito escolar. Hoy día, esta Red reagrupa a los representantes de los 15 «antiguos» Estados miembros (también las dos comunidades belgas y Escocia) así como Islandia y Noruega. Está abierta a los nuevos Estados miembros, por lo que la mayoría ya ha designado a su representante. La Comisión Europea también está representada, mientras que Suiza y la IEA son observadores. La Dirección de la Evaluación y de la Prospectiva del Ministerio francés

Estos dos proyectos, titulados respectivamente *The Use of National Reading for International Comparisons: Ways of Overcoming Cultural Bias* (Bonnet et al., 2001) y *Culturally Balanced Assessment of Reading (C-BAR)* (Bonnet et al., 2003), han permitido poner al día el método descrito a continuación. Ocho países han participado en él. Los dos proyectos se unieron para proponer una metodología comparativa que se aplica a la evaluación de las capacidades de lectura de los alumnos de 15/16 años. Naturalmente, debería ser posible utilizar los principios generales de esta metodología para la evaluación de otros tipos de capacidades y para otros públicos.

Con el fin de garantizar la validez de la comparación, respetando el contexto cultural en el sentido más amplio, la metodología propuesta se basa en los cuatro principios siguientes:

UN ACERCAMIENTO PSICOLINGÜÍSTICO Y TRANSCULTURAL

Desde un punto de vista pedagógico, las actividades que los expertos en lectura y los docentes consideran que forman parte del acto de leer son más numerosas que las tareas requeridas para evaluar las situaciones de lectura de la vida cotidiana. Es muy especialmente en el nivel de capacidad de la enseñanza secundaria donde los alumnos pueden enfrentarse a pasajes literarios. La necesidad de proporcionar a los docentes informaciones sobre la eficacia de su enseñanza implica que el acto de leer sea analizado desde una perspectiva psicológica. De ahí, la elección de un punto de partida conceptual basado en un marco psicolingüístico y psicopedagógico.

La mayoría de los países han puesto en marcha modalidades de evaluación de los conocimientos adquiridos por los alumnos a través de tipos de instituciones variadas, donde han surgido diferentes formas de evaluación determinadas por las tradiciones y las estructuras nacionales. El objetivo es pues construir a partir de esta diversidad un marco común correspondiente a una comprensión compartida de estas culturas. Este acercamiento es el de la teoría de la psicología transcultural tal como se puso en práctica, principalmente, por el psicólogo canadiense Berry (1993).

encargada de la educación asegura su funcionamiento y su presidencia así como la jefatura de redacción, la publicación y la difusión de *Evaluación*, carta de información de la Red que se publica dos veces al año. Los objetivos de la Red son:

- Por una parte, permitir intercambios de información sobre las reformas y las innovaciones en curso o en proyecto en cada uno de los países europeos, en el marco de la evaluación y de la dirección de sistemas educativos;
- Por otra parte, definir, facilitar e iniciar acciones de cooperación europea en este mismo campo.

Para poner en marcha este segundo objetivo, la *Red* lanza y conduce proyectos que reagrupan a varios países, ya sea en un marco intergubernamental, ya sea en unión con convocatoria de oferta de la Comisión Europea.

<http://cisad.edc.educatio.fr/rev>

UNA ESTRATEGIA PARA TRATAR EL PROBLEMA DE LA TRADUCCIÓN

Por las razones expuestas más arriba, se debe evitar lo más posible la traducción en la construcción de los instrumentos principales, para suprimir a la vez los sesgos que conlleva y para respetar el contexto cultural. Se trabajará pues para construir un material de evaluación nacional elaborado según una metodología común que defina las capacidades que se van a evaluar y una tipología de los instrumentos que se van a utilizar (Nardi, 2003)

Se demostrará más adelante que el recurso de la traducción se mantiene como necesario para elaborar una partida de ítems de anclaje comunes. No obstante, esto se puede hacer en un marco teórico y práctico específico que permite limitar sensiblemente los posibles sesgos.

UN MARCO CONCEPTUAL COMÚN DE EVALUACIÓN DE LAS CAPACIDADES EN LECTURA

La competencia de los alumnos en lectura está afectada por una variedad de factores, tales como el interés y la dificultad de los textos, la influencia del contexto, la familiaridad con el tema abordado y su capacidad de utilizar los procesos para codificar y de dar un sentido al texto. Por este motivo, uno de los aspectos de la metodología de la que hablamos consiste en tratar de separar la naturaleza y el contexto de los textos de la capacidad de lectura del sujeto.

En concreto, se han extraído tres capacidades de lectura a los 15/16 años y tres niveles de dificultad para cada competencia. Es lo que describe el Cuadro I.

CUADRO I

	Nivel 1	Nivel 2	Nivel 3
Competencia A Encontrar informaciones e ideas explícitas.	Encontrar una o varias informaciones en función, eventualmente, de uno o varios criterios.	Encontrar una o varias informaciones o ideas, en función, si llega el caso, de criterios múltiples. Juzgar alguna información del texto si procede.	Encontrar y poner en orden o combinar conceptos e ideas y/o informaciones en un texto farragoso.
Competencia B Interpretar un sentido implícito.	Formular o aplicar categorías simples e interpretar el sentido de una parte limitada del texto. Identificar la idea o el sentido principal sobre la base de identificar el texto.	Identificar las ideas o temas subyacentes. Reconocer un desarrollo o un cambio en las ideas o en los temas y extraer conclusiones sobre la base de identificaciones en el texto.	Identificar y comprender los orígenes del sentido sobre la base de indicaciones en el texto. Interpretar el sentido de una parte del texto sobre la base del texto en su conjunto.
Competencia C Apreciación crítica de la forma y del fondo.	Identificar la manera en la que algunos aspectos de la forma corresponden al objetivo del texto.	Identificar la manera en la que algunos aspectos de la forma corresponden al objetivo del texto y se pronuncian sobre la eficacia de la relación.	Distinguir forma y fondo y apreciar la manera en que uno afecta al otro. Demostrar la eficacia sobre el punto de vista de la comunicación.

Los tres niveles de dificultad se definen con relación a los procesos necesarios para tratar las diferentes dimensiones de dificultad del texto, por ejemplo: si es simple, si es complejo, si es concreto, si es abstracto, si es específico, si es general, si es explícito, si es implícito...

Este cuadro conceptual de la evaluación de la lectura (cuadro de capacidades y de niveles de dificultades) debe ir acompañado de una clasificación de textos y de un cuadro de equivalencias que permita, sobre la base de un criterio común, determinar la naturaleza y comparar el grado de dificultad de los textos con el fin de seleccionar los textos equivalentes en varios países. Una taxonomía de tipos de cuestionarios que se van a utilizar en la evaluación completa el dispositivo. Cae por su propio peso que se deben elaborar consignas precisas y concretas para la corrección de las respuestas y la codificación de los ítems de los tests con el fin de asegurar igualmente la comparabilidad en este campo.

UN DISPOSITIVO DE ANCLAJE EN LOS CUESTIONARIOS NACIONALES

Una vez hecha la elección de recurrir para lo elemental a las consignas de evaluación elaboradas país por país, viene inevitablemente la pregunta sobre la comparabilidad psicométrica y estadística de los resultados de la evaluación. Este problema está en el corazón del método propuesto. Con el fin de dar una respuesta adaptada a esta dificultad real, se impone la necesidad de prever un dispositivo de anclaje. Este anclaje se puede hacer bien a través de temas, bien a través de cuestionarios. El anclaje a través de cuestionarios necesita la utilización de un cuestionario que tenga idéntica fama en cada país, mientras que el anclaje a través de temas necesita del recurso de temas cuyas competencias lingüísticas sean consideradas casi idénticas en varias lenguas.

El anclaje por cuestionarios es sin duda el más razonable. Se proponen dos variantes que permitan calibrar los formularios nacionales por lo que respecta a un cuestionario común externo, que no son mutuamente excluyentes.

La primera se apoya en la utilización de un cuestionario común traducido a todas la lenguas afectadas, un poco como en la metodología tradicional. A primera vista, es paradójico preconizar esta solución. En realidad, lo que se persigue aquí es muy diferente del método tradicional por al menos dos razones: 1: el número de cuestionarios y de formularios comunes traducidos es muy restringido, lo que permite controlar mejor los eventuales sesgos; 2: los textos utilizados como soportes en los cuestionarios son elegidos entre las publicaciones más conocidas que ya existen traducidas «naturalmente», es decir, traducidas de manera profesional para un público determinado y no de manera *ad hoc* para públicos con lingüística diferente por necesidades de una evaluación internacional. Tales textos no son concebidos como traducciones por quienes hablan estas lenguas en la medida en que, después de numerosos años, forman parte de su ambiente cultural. Por esta razón, más que de textos «traducidos» se trata de textos «en traducción». El matiz es importante.

Un cuestionario de anclaje de esta naturaleza es además particularmente homogéneo con los cuestionarios de lectura elaborados por los países. Los

formularios sobre textos, en sí mismos, aunque están traducidos de manera *ad hoc*, están elaborados en cada lengua según las reglas fijadas en común.

La segunda variante, que es útil para completar la anterior, consiste en utilizar un cuestionario ya traducido en diferentes lenguas nacionales y cuya variedad es conocida de manera indiscutible en el plano internacional. El que se propone es la versión 3 del subcuestionario de vocabulario de Wechsler adaptado para la lectura (*WISC III vocabulary sub-test*). Se presenta bajo la forma de una lista ordenada de palabras elaboradas según los criterios psicológicos. Ni el orden de las palabras ni las palabras mismas tienen por qué ser necesariamente las mismas en cada lengua (Wechsler, 1991 a y b). El *WISC III* proporciona así un cuestionario común que no presenta los mismos sesgos que un cuestionario traducido de manera *ad hoc*.

Es igualmente posible poder recurrir a un anclaje por temas, cuando tales temas existen, haciendo pasar los cuestionarios de los formularios nacionales elaborados según la presente metodología (por consiguiente, diferentes) por temas bilingües en las dos lenguas correspondientes. Puede tratarse, por ejemplo, de alumnos escolarizados en establecimientos donde la enseñanza se hace en las dos lenguas al mismo nivel (por ejemplo, en secciones bilingües de institutos nacionales o en institutos que algunos países tienen en otros como el Instituto Francés de Madrid). La dificultad estriba en estos casos en la identificación de temas bilingües en el sentido más estricto del término. Es evidente que no existe un número suficiente de temas para todas las combinaciones lingüísticas.

Los análisis estadísticos apropiados permiten a continuación, poniendo en relación los resultados de estos diferentes métodos de anclaje con sus formularios nacionales, verificar la comparación de los datos de la evaluación entre países.

Otros análisis, que utilizan especialmente métodos multidimensionales (análisis factoriales, *Structural Equation Modelling*) y, cuando sea preciso y convenga, con los datos de los métodos unidimensionales (como los modelos de pregunta-respuesta), permiten interpretar los resultados y estudiar su diferencia entre países de manera cualitativa y no solamente bajo la manera del palmarés internacional.

El detalle de esta metodología está disponible en Bonnet et al., 2003, accesible en:

<http://cisad.adc.education.fr/reva/pdf/cbarfinalreport.pdf>

CONCLUSIÓN

La metodología descrita aquí es sensiblemente diferente de la que prevalece de manera exclusiva en las evaluaciones internacionales contemporáneas. Puede dejar sin duda escépticos a ciertos especialistas poco inclinados a aceptar las innovaciones, tanto más cuando aún no se han podido hacer las pruebas de manera irrefutable. Sin embargo, se puede pensar que sería razonable aprobarla en su grandeza real. Los beneficios que se podrían sacar en materia de análisis pedagógico anclado a un contexto nacional vendrían a reforzar lo que PISA, TIMSS o PIRLS aportan ya a la comprensión de los resultados de los sistemas educativos, permitiendo identificar las medidas para remedios concretos al nivel de clase, país por país.

Hay que esperar que sea posible en el futuro enfocar su perfeccionamiento, su adaptación y su integración progresiva a las grandes evaluaciones internacionales con el fin de reforzar la pertenencia de estas últimas a los responsables de la educación. Por ello, convendrá en principio admitir que la pluralidad y la complementariedad de acercamientos metodológicos, como la diversidad cultural, no son pues un obstáculo sino una riqueza, la única capaz de hacer justicia a la diversidad de la experiencia humana.

REFERENCIAS BIBLIOGRÁFICAS

- ADAM, R.; WU, M. (eds.) (2002): *PISA 2000 Technical Report*. París, OECD.
- BERRY, J. W. (1993): «The way ahead: from indigenous psychologies to a universal psychology», en KIM, U.; BERRY, J. W. (eds.): *Indigenous psychology: Research and experience in cultural context* (pp. 260-276). Newbury Park Ca, Sage Publications.
- BINKLEY, M. R.; PIGNAL, J. R. (1998): «An analysis of items with different parameter across countries», en MURRAY, T. S.; KIRSCH, I. S.; JENKINS, L. B. (eds.): *Adult Literacy in OECD countries. Technical report on the first international adult literacy survey* (pp. 143-160). Washington DC, US department of Education National Center for Educational Statistics.
- BONNET, G.; BRAXMEYER, N.; HORNER, S.; LAPPAPAINEN, H. P.; LEVASSEUR, J.; NARDI, E.; RÉMOND, M.; VRIGNAUD, P.; WHITE, J. (2001): *The use of national reading tests for international comparisons: ways of overcoming cultural bias*. Paris, Ministère de l'éducation national, DPD Edition diffusion.
- BONNE, G. (2002): «Reflections in a critical eye; on the pitfalls of international assessment», en *Review Essay of «Knowledge and Skills for Life; First Results from PISA 2000» in Assessment in education: Principles, Policy and Practice*, vol. 9, 3, pp. 385-297. Abingdon, Carfax publishing.
- BONNET, G.; DAEMS, F.; GLOPPER, C. DE; HORNER, S.; LAPPALAINEN, H.-P.; NARDI, E.; RÉMOND, M.; RONBIN, I.; ROSEN, M.; SOLHEIM, R. G.; TONNESSEN, F.-E.; VERTECCHI, B.; VRIGNAUD, P.; WAGNER, A. K.; WHITE, J. (2003): *Culturally balanced Assessment of Reading. (C-BAR) (2003)*. Ministère de la Jeunesse, de l'Éducation et de la Recherche, p. 84. Paris, DEP Edition.
- BOURNY, G.; DUPÉ, C.; ROBIN, I.; ROCHER, T. (2002): *Les élèves de 15 ans. Premiers résultats d'une évaluation internationale des acquis des élèves (PISA)*. Note d'information, n° 01.52. Paris, DPD Edition diffusion.
- BOURNY, G. et al. (2004): *Les élèves de 15 ans. Premier résultats de l'évaluation internationale PISA 2003*. Note d'évaluation n° 04.12. Paris, MJENR/DEP, Edit.
- GUÉRIN-PACE, F.; BLUM, A. (1999): *L'illusion comparative : les logiques d'élaboration et d'utilisation d'une enquête internationale sur l'illettrisme*. Population, 54 (2). Paris, INED.
- BLUM, A.; GUÉRIN-PACE, F. (2000): *Des Lettres et des Chiffres*. Paris, Fayard.
- BRACY, G. W. (2005): «Put Out Over PISA», en *Phi Delta Kappan*. vol. 86, 10, pp. 797-798. Bloomington, In.
- CAREAY, S. (ed.) (2000): *Measuring Adult Literacy.- The International Adult Literacy Survey in the European Context*. London, Office for National Statistics.
- GOLSDTEIN, H. (2004): *The Education Word Cup: International comparisons of student achievement. Plenary talk to Association for Educational Assessment*. Europe, Budapest, Nov. 4-6.
- GOLSDTIEN, H.; BONNET, G.; ROCHER, T. (pendiente de publicación): *A study of procedures for the analysis of PISA 2000 reading data*.
- IRDP (2005): *PISA 2003: Compétences des jeunes Romands- Résultats de la seconde enquête PISA auprès des élèves de 9^{ième} année*. Neuchâtel.

- LAFONTAINE, D.; DEMEUSE, M. (2002): «Le bon (critique), la brute(médiatique) et les truants(anglo-saxons)», en *La Revue Nouvelle*, 3-4, tome 115. Bruxelles.
- NARDI, E. (2002): *Leggere a quindicenni anni. Riflessioni sulla ricerca PISA*. Milano, Franco Angelli.
- ROCHER, T. (2004): «La méthodologie des évaluations internationales de compétences», en *Psychologie et Psychométrie. Numéro Spécial: Mesure et Education*. vol. 24, 2/3, pp. 117-146. Paris, Editions et Applications Psychologiques.
- ROMAINVILLE, M. (2002): «Du bon usage de PISA», en *La Revue Nouvelle*, 3-4, tome 115. Bruxelles.
- SIRECI, S. G. (1997): *Problems and issues in linking assessments across languages*. Educational Measurement: Issues and Practice, 16.