Psicothema

Psicothema 2013, Vol. 25, No. 2, 151-157
doi: 10.7334/psicothema2013.24

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

# Directrices para la traducción y adaptación de los tests: segunda edición

José Muñiz[1], Paula Elosua[2] y Ronald K. Hambleton[3]
[1] Universidad de Oviedo, [2] Universidad del País Vasco y [3] University of Massachusetts

## Resumen

**Antecedentes:** en los últimos años la adaptación de los tests de unas culturas a otras se ha incrementado en todos los ámbitos evaluativos. Vivimos en un entorno cada vez más multicultural y multilingüe en el que los tests se utilizan como apoyo en la toma de decisiones. El objetivo de este trabajo es presentar la segunda edición de las directrices de la Comisión Internacional de Tests (ITC) para la adaptación de los tests de unas culturas a otras. **Método:** un grupo de seis expertos internacionales revisaron las directrices originales propuestas por la Comisión Internacional de Tests, teniendo en cuenta los avances habidos en el campo desde su formulación inicial. **Resultados:** la nueva edición está compuesta por veinte directrices agrupadas en seis apartados: directrices previas, desarrollo del test, confirmación, aplicación, puntuación e interpretación y documentación. Se analizan los diferentes apartados, y se estudian las posibles fuentes de error que pueden influir en el proceso de traducción y adaptación de los tests. **Conclusiones:** se proponen veinte directrices para guiar la traducción y adaptación de los tests de unas culturas a otras. Se discuten las perspectivas futuras de las directrices en relación con los nuevos desarrollos en el ámbito de la evaluación psicológica y educativa.

***Palabras clave:*** traducción, adaptación, test, directrices, Comisión Internacional de tests.

## Abstract

***International Test Commission Guidelines for test translation and adaptation: Second edition.*** **Background:** Adapting tests across cultures is a common practice that has increased in all evaluation areas in recent years. We live in an increasingly multicultural and multilingual world in which the tests are used to support decision-making in the educational, clinical, organizational and other areas, so the adaptation of tests becomes a necessity. The main goal of this paper is to present the second edition of the guidelines of the International Test Commission (ITC) for adapting tests across cultures. **Method:** A task force of six international experts reviewed the original guidelines proposed by the International Test Commission, taking into account the advances and developments of the field. **Results:** As a result of the revision this new edition consists of twenty guidelines grouped into six sections: Precondition, test development, confirmation, administration, score scales and interpretation, and document. The different sections are reviewed, and the possible sources of error influencing the tests translation and adaptation analyzed. **Conclusions:** Twenty guidelines are proposed for translating and adapting tests across cultures. Finally we discuss the future perspectives of the guidelines in relation to the new developments in the field of psychological and educational assessment.

***Keywords:*** Translation, adaptation, test, guidelines, International Test Commission.

La adaptación de tests y cuestionarios para su uso en contextos lingüísticos y culturales diferentes a aquellos en que fueron construidos es una práctica tan antigua como los propios tests que se remonta a la aparición de las primeras escalas de Binet y Simon (1905) en los albores del siglo XX. Su incremento en las últimas décadas es el reflejo de un medio social marcado por los contactos entre culturas e idiomas y en el que los tests y cuestionarios asisten diariamente en los ámbitos educativo, social, jurídico o clínico, entre otros, en la toma de decisiones individuales o grupales (Muñiz y Hambleton, 1996). El interés por la adaptación de instrumentos de medida no se circunscribe únicamente a un ámbito evaluativo, una revisión de los veinticinco tests más utilizados en la práctica profesional española (clínica, educativa u organizacional) deja patente que de ellos diecisiete son adaptaciones de versiones construidas en otro idioma, en su mayoría del inglés (Elosua, 2012), y lo mismo puede decirse en relación con otros países (Elosua y Iliescu, 2012; Evers et al., 2012). Una simple ojeada a las publicaciones recientes en español pone de manifiesto esta misma tendencia (Calvo et al., 2012; Iturbide, Elosua y Yenes, 2010; Nogueira, Godoy, Romero, Gavino y Cobos, 2012; Ortiz, Navarro, García, Ramis y Manassero, 2012; Rodríguez, Martínez, Tinajero, Guisande y Paramo, 2012). Dentro del área e intereses de la psicología de las organizaciones y de los recursos humanos es cada vez mayor la necesidad mostrada por las corporaciones multinacionales y organismos internacionales de disponer de pruebas de acreditación o de selección que puedan utilizarse en distintos países o en distintos idiomas. Del mismo modo, el impacto social de las evaluaciones educativas internacionales como PISA y TIMMS, que utilizan pruebas adaptadas a más de cuarenta idiomas, deja clara la importancia de un correcto proceso de adaptación de los instrumentos de medida. Consciente de esta necesidad, la Comisión Internacional de Tests (*International Test Commission*, ITC) inició el año 1994 un proyecto de elaboración de directrices relacionadas con la adaptación de tests y cuestionarios. Este proyecto (Hambleton, 1994, 1996; Muñiz y Hambleton, 1996) dio origen a un conjunto

Received: December 14, 2012 • Accepted: February 15, 2013
Corresponding author: José Muñiz
Facultad de Psicología
Universidad de Oviedo
33003 Oviedo (Spain)
e-mail: jmuniz@uniovi.es

de veintidós directrices que, agrupadas en cuatro apartados (Contexto, Construcción y Adaptación, Aplicación e Interpretación), intentaban prevenir sobre las distintas fuentes de error intervinientes en el proceso de adaptación de tests y ofrecían vías para controlarlas. El documento ha sido citado en más de quinientas ocasiones en publicaciones científicas y profesionales (Hambleton, 2009), lo cual es un claro indicador del impacto de las directrices de la ITC para la adaptación de los tests.

En los últimos años se han producido avances importantes en el campo de la adaptación de los tests, tanto desde un punto de vista metodológico y psicométrico como sustantivo (Hambleton, Merenda y Spielberger, 2005; Matsumoto y van de Vijver, 2011; van de Vijver y Tanzer, 1997). Cabe reseñar los desarrollos metodológicos y técnicos en el establecimiento de la equivalencia intercultural (Byrne, 2008; Elosua, 2005), así como una mayor toma de conciencia sobre la importancia de los estudios analítico-racionales durante el proceso de adaptación, solo por citar un par de ejemplos. Estos y otros avances han dado lugar a la necesidad de revisar las directrices originales a la luz de los nuevos desarrollos. Para llevar a cabo la revisión se formó un grupo de trabajo multidisciplinar en el seno de la ITC coordinado por el profesor Ronald K. Hambleton y compuesto por representantes de varias asociaciones de psicólogos: Dave Bartram (Reino Unido), Giray Berberoglu (Turquía), Jacques Gregoire (Bélgica), José Muñiz (España) y Fons van de Vijver (Holanda).

Las directrices propuestas ofrecen un marco integral (figura 1) en el que se aborda el estudio de las fases previas a la adaptación, el análisis de la propia adaptación, de su justificación técnica, de la evaluación e interpretación de las puntuaciones y de la elaboración del documento final. Se trata de veinte reglas agrupadas en seis categorías que quedan resumidas en la tabla 1.

El objetivo de las directrices es que el producto final del proceso de adaptación consiga con respecto a la prueba original el máximo nivel de equivalencia lingüística, cultural, conceptual y métrica posible, y para ello son concebidas como un patrón que guía a los investigadores y profesionales en las pautas a seguir. El proceso es global en naturaleza y abarca la totalidad de fases y cuestiones a considerar durante el proceso de traducción, desde cuestiones legales relacionadas con los derechos de la propiedad intelectual del test a adaptar, hasta aspectos formales que atañen a la redacción del manual que documenta los cambios introducidos. Todos ellos son importantes, y a todos ellos se habrá de prestar atención.

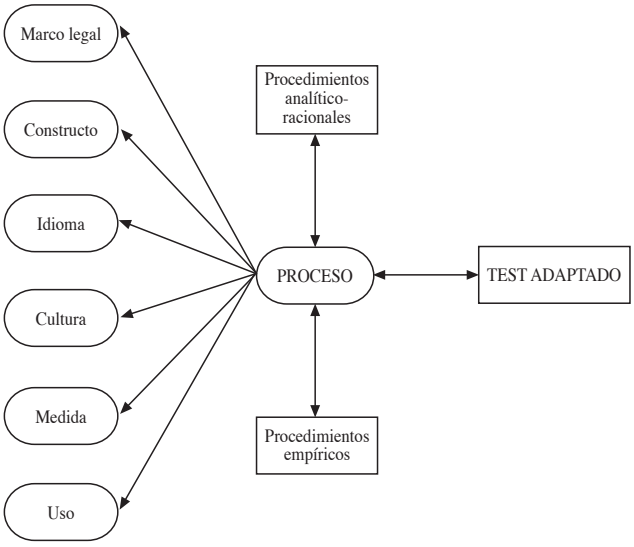


***Figura 1.*** *Componentes en el proceso de adaptación de un test*

*Directrices previas*

Fijan su atención sobre dos cuestiones previas a la ejecución de cualquier adaptación y que atañen a su correcta planificación: la comprobación del registro de la propiedad intelectual y el estudio de la relevancia del constructo. Comprobar sobre quién recae el derecho de la propiedad intelectual del instrumento y en su caso obtener los permisos legales permitirán garantizar la autenticidad del producto final y proteger el trabajo de adaptaciones no autorizadas. El segundo aspecto, ya tratado en la primera edición de las directrices, se refiere al estudio de las características del constructo a medir en la población diana. El interés en este punto alerta sobre las consecuencias de asumir sin más la universalidad de los constructos entre culturas, y aconseja evaluar el grado o nivel de solapamiento entre el constructo en la población origen y en la población diana como único medio para delimitar y definir el nivel de equivalencia deseado.

*Directrices sobre el desarrollo del test*

Guían durante el proceso de adaptación y desarrollo del test, y ofrecen pautas para superar algunos de los malentendidos más comunes relacionados con el uso de la traducción literal como garantía de equivalencia, o el excesivo peso otorgado a la traducción inversa (*back-translation*) (Brislin, 1986) como procedimiento de verificación de la calidad de la adaptación. Es habitual considerar que en una buena traducción la equivalencia entre la versión original y la versión retro-traducida generada por un traductor independiente es muy alta. Esta consideración, sin embargo, no es garantía de validez de la versión diana, es más, en una mala traducción el grado de equivalencia entre la versión original y la versión retro-traducida

*Tabla 1*
Categorías y aspectos analizados por las Nuevas Directrices de la Comisión Internacional de Tests (ITC) para la traducción y adaptación de los tests

| **Categorías** | **Número de directrices** | **Aspectos analizados** |
|---|---|---|
| Previas | 5 | Marco legal<br>Diseño<br>Evaluación del constructo |
| Desarrollo | 5 | Adaptación lingüística<br>Adaptación cultural<br>Estudios piloto |
| Confirmación | 4 | Recogida de datos<br>Equivalencia<br>Fiabilidad<br>Validación |
| Aplicación | 2 | Administración |
| Puntuación e interpretación | 2 | Interpretación de las puntuaciones<br>Comparabilidad |
| Documentación | 2 | Cambios entre versiones<br>Uso correcto |

puede ser muy alta. La razón de esta singularidad se debe a que habitualmente las malas traducciones se apoyan en traducciones literales en lugar de en una esmerada adaptación de significados. Por ejemplo, los resultados de las traducciones automatizadas originan buenas retro-traducciones, a pesar de ello, nunca las utilizaríamos sin una evaluación y ajuste de la versión en el idioma destino.

Para considerar los factores lingüísticos y culturales a tener en cuenta durante la adaptación se aconseja la implementación de un procedimiento iterativo de depuración que se inicia con varias traducciones independientes hacia adelante, que luego serán revisadas por un comité mixto en el que conviene incluir, además de traductores con conocimientos de los idiomas y culturas, a especialistas en el campo de evaluación que analicen la adecuación de la versión adaptada. Corrección lingüística y adecuación práctica son conceptos complementarios que es necesario compatibilizar.

Las directrices de desarrollo incluyen también un apartado que resalta la importancia de las pruebas piloto (Downing, 2006; Downing y Haladyna, 2006; Schmeiser y Welch, 2006; Wilson, 2005). La disponibilidad de datos obtenidos en una pequeña muestra permite analizar, estudiar y corregir aspectos relacionados con la adaptación en desarrollo. Las pruebas piloto permiten, entre otras cosas: a) recoger "in situ" las reacciones de las personas que realizan la prueba; b) asegurarse de que los ítems e instrucciones son correctamente comprendidos; c) registrar el tiempo necesario para la ejecución del cuestionario; d) recoger información sobre posibles errores de contenido o formato que se pueden corregir antes de pasar a la fase operacional; y e) obtener datos que permitirán llevar a cabo un primer análisis de ítems que indiquen la dirección y sentido de los índices psicométricos más relevantes. Por ejemplo, es interesante analizar las medias aritméticas de los ítems, o índices de dificultad en su caso, y compararlas con los valores de la prueba original; para ello podrían compararse los cuartiles de dificultad de cada uno de los ítems de la prueba, como primera aproximación a posteriores estudios de equivalencia (Elosua, Bully, Mujika y Almeida, 2012).

*Directrices de confirmación*

Este grupo de directrices hacen referencia a aspectos técnicos relacionados con las propiedades psicométricas del test adaptado y a su equivalencia con respecto al test original. Proponen llevar a cabo estudios de equivalencia métrica entre las versiones original y adaptada que determinarán el grado de relación entre cada uno de los ítems que componen la prueba y la dimensión que representan. Si la relación funcional no es equivalente entre las versiones original/adaptada la comparabilidad entre escalas se verá amenazada. Son varios los modelos teóricos que permiten acometer los estudios de equivalencia, tales como los modelos de ecuaciones estructurales, los modelos de teoría de respuesta a los ítems, o los procedimientos para la detección del funcionamiento diferencial de los ítems. Solo la evaluación de la equivalencia permitirá conocer el nivel de comparabilidad entre puntuaciones (Elosua y Muñiz, 2010). Las directrices de confirmación proponen llevar a cabo estudios sobre fiabilidad y estudios de validación (Elosua, 2003; Lissitz, 2009; Muñiz, 2003).

*Directrices sobre aplicación*

La forma en la que se aplica un test influye en las propiedades psicométricas de las puntuaciones obtenidas, tales como su fiabilidad y validez. Las relaciones de los aplicadores con las personas a las que se pasa el test (*rapport*), la forma de dar las instrucciones de la prueba, y en general las interacciones aplicador-examinado deben de cuidarse al máximo. Como señala Hambleton (1996), los aplicadores: a) deben ser elegidos entre personas de la población a la que se aplica el test; b) estar familiarizados con los distintos matices de la cultura de que se trate; c) tener experiencia y aptitudes para la aplicación de tests; y d) conocer la importancia de seguir al pie de la letra los procedimientos reglados para la aplicación de los tests. Deben de programarse sesiones de entrenamiento riguroso para los aplicadores.

*Directrices sobre puntuación e interpretación*

Las directrices sobre puntuación e interpretación alertan sobre los riesgos derivados de la tentación de comparar directamente puntuaciones obtenidas en contextos culturales o lingüísticos diferentes por medio de escalas adaptadas. La comparación de puntuaciones ha de circunscribirse al nivel de equivalencia psicométrico empíricamente demostrado con la aplicación de las directrices de confirmación. Si no puede demostrarse la existencia de equivalencia métrica entre todos los ítems que componen las escalas original y adaptada, las puntuaciones obtenidas no podrán compararse directamente. El problema de la comparación entre puntuaciones se agrava con su interpretación. Como señalan algunos autores (Hambleton y Bollwark, 1991; Westbury, 1992), los estudios comparativos deberían de usarse para comprender las semejanzas y diferencias entre los grupos analizados, pero nunca para establecer comparaciones sin más. Y no es adecuado establecerlas porque raramente encontraremos dos comunidades que sean equiparables completamente en aspectos tan influyentes como motivación a la hora de hacer las pruebas, curriculas escolares, valores culturales, nivel de vida, políticas educativas, oportunidades de acceso a la educación, etc.

*Directrices sobre documentación*

Finalmente, para interpretar las puntuaciones el psicólogo debe de disponer de una documentación exhaustiva acerca de cómo se llevó a cabo el proceso de adaptación. El manual del test deberá incluir todo tipo de detalles del proceso adaptativo y de los cambios y modificaciones llevados a cabo sobre el test original, que en determinadas circunstancias pueden dar las claves interpretativas de un resultado (Prieto y Muñiz, 2000).

A continuación se ofrecen las veinte directrices agrupadas en las seis secciones citadas.

*Directrices para la traducción/adaptación de tests*

1. Directrices previas
   - DP1. Antes de comenzar con la adaptación hay que obtener los permisos pertinentes de quien ostente los derechos de propiedad intelectual del test.
   - DP2. Cumplir con las leyes y prácticas profesionales relativas al uso de tests que estén vigentes en el país o países implicados.
   - DP3. Seleccionar el diseño de adaptación de tests más adecuado.
   - DP4. Evaluar la relevancia del constructo o constructos medidos por el test en las poblaciones de interés.

DP5. Evaluar la influencia de cualquier diferencia cultural o lingüística en las poblaciones de interés que sea relevante para el test a adaptar.

2. Directrices de desarrollo
   DD1. Asegurarse, mediante la selección de expertos cualificados, de que el proceso de adaptación tiene en cuenta las diferencias lingüísticas, psicológicas y culturales entre las poblaciones de interés.
   DD2. Utilizar diseños y procedimientos racionales apropiados para asegurar la adecuación de la adaptación del test a la población a la que va dirigido.
   DD3. Ofrecer información y evidencias que garanticen que las instrucciones del test y el contenido de los ítems tienen un significado similar en todas las poblaciones a las que va dirigido el test.
   DD4. Ofrecer información y evidencias que garanticen que el formato de los ítems, las escalas de respuesta, las reglas de corrección, las convenciones utilizadas, las formas de aplicación y demás aspectos son adecuados para todas las poblaciones de interés.
   DD5. Recoger datos mediante estudios piloto sobre el test adaptado, y efectuar análisis de ítems y estudios de fiabilidad y validación que sirvan de base para llevar a cabo las revisiones necesarias y adoptar decisiones sobre la validez del test adaptado.
3. Directrices de confirmación
   DC1. Definir las características de la muestra que sean pertinentes para el uso del test, y seleccionar un tamaño de muestra suficiente que sea adecuado para las exigencias de los análisis empíricos.
   DC2. Ofrecer información empírica pertinente sobre la equivalencia del constructo, equivalencia del método y equivalencia entre los ítems en todas las poblaciones implicadas.
   DC3. Recoger información y evidencias sobre la fiabilidad y la validez de la versión adaptada del test en las poblaciones implicadas.
   DC4. Establecer el nivel de comparabilidad entre las puntuaciones de distintas poblaciones por medio de análisis de datos o diseños de equiparación adecuados.
4. Directrices sobre la aplicación
   DA1. Preparar los materiales y las instrucciones para la aplicación de modo que minimicen cualquier diferencia cultural y lingüística que pueda ser debida a los procedimientos de aplicación y a los formatos de respuesta, y que puedan afectar a la validez de las inferencias derivadas de las puntuaciones.
   DA2. Especificar las condiciones de aplicación del test que deben seguirse en todas las poblaciones a las que va dirigido.
5. Directrices sobre puntuación e interpretación
   DPI1. Interpretar las diferencias de las puntuaciones entre los grupos teniendo en cuenta la información demográfica pertinente.
   DPI2. Comparar las puntuaciones entre poblaciones únicamente en el nivel de invarianza establecida para la

*Tabla 2*
Listado para el control de calidad de la traducción-adaptación de los ítems (tomado de Hambleton y Zenisky, 2011)

**Generales**
1. ¿El ítem tiene el mismo significado o muy parecido en los dos idiomas?
2. ¿El tipo de lenguaje del ítem traducido tiene una dificultad y familiaridad comparables al del idioma original?
3. ¿Introduce la traducción cambios en el texto (omisiones, sustituciones o adiciones) que puedan influir en la dificultad del ítem?
4. ¿Hay diferencias entre la versión original del ítem y la traducida en relación con el uso de metáforas, giros o expresiones coloquiales?

**Formato del ítem**
5. ¿El formato del ítem, incluyendo los aspectos físicos, es el mismo en los dos idiomas?
6. ¿La longitud del enunciado y de las alternativas de respuesta, cuando las haya, tienen una longitud similar en ambas versiones?
7. ¿El formato del ítem y la tarea a realizar por la persona evaluada son de una familiaridad similar en las dos versiones?
8. ¿Si se destacó una palabra o frase (negrita, cursiva, subrayado, etc.) en la versión original, se hizo también en el ítem traducido?
9. En el caso de tests educativos, ¿hay una respuesta correcta en ambas versiones del ítem?

**Gramática y redacción**
10. ¿Hay alguna modificación de la estructura gramatical del ítem, tal como la ubicación de las oraciones o el orden de las palabras, que pueda hacer el ítem más o menos complejo en una versión que en otra?
11. ¿Existen algunas pistas gramaticales que puedan hacer el ítem más fácil o más difícil en la versión traducida?
12. ¿Existen algunas estructuras gramaticales en la versión original del ítem que no tienen equivalente en la versión traducida?
13. ¿Existen algunas referencias al género u otros aspectos que puedan dar pistas sobre el ítem en la versión traducida?
14. ¿Hay palabras en el ítem que tengan un significado unívoco, pero que en la versión traducida puedan tener más de un significado?
15. ¿Hay cambios en la puntuación entre las dos versiones que puedan hacer que el ítem sea más fácil o difícil en la versión traducida?

**Pasajes** (cuando haya)
16. Cuando se traduce un pasaje, ¿las palabras y frases de la versión traducida transmiten el mismo contenido e ideas que la versión original?
17. ¿Describe el pasaje individuos o grupos de forma estereotipada en relación con su ocupación, emociones, situación u otro aspecto?
18. ¿La forma en la que está escrito el pasaje es controvertida o polémica, o puede ser percibido de forma denigrante u ofensiva?
19. ¿El pasaje incluye contenidos o requiere habilidades que pueden ser poco habituales en cualquiera de los dos idiomas o grupos culturales?
20. Aparte de los cambios exigidos por la traducción, ¿los gráficos, tablas u otros elementos son iguales en las dos versiones del ítem?

**Cultura**
21. ¿Los términos utilizados en el ítem en el idioma original han sido adaptados de forma adecuada al contexto cultural de la versión traducida?
22. ¿Existen diferencias culturales que tengan un efecto diferencial sobre la probabilidad de que una respuesta sea elegida en la versión original y la traducida?
23. Las unidades de medida y las monedas (distancia, etc.) de la versión original del ítem ¿están convenientemente adaptadas en la versión traducida?
24. Los conceptos implicados en el ítem ¿están al mismo nivel de abstracción en las dos versiones?
25. El concepto o constructo del ítem ¿es igual de familiar y tiene el mismo significado en las dos versiones?

escala de puntuación utilizada en las comparaciones.

6. Directrices sobre la documentación
   - DC1. Proporcionar documentación técnica que recoja cualquier cambio en el test adaptado, incluyendo la información y las evidencias sobre la equivalencia entre las versiones adaptadas.
   - DC2. Proporcionar documentación a los usuarios con el fin de garantizar un uso correcto del test adaptado en la población a la que va dirigido.

Para ayudar en la aplicación empírica de las directrices Hambleton y Zenisky (2011) proponen veinticinco preguntas para responder sobre cada uno de los ítems de la prueba adaptada (tabla 2). No cubren de forma exhaustiva todos los aspectos incluidos en las directrices, pero constituyen una buena primera aproximación para detectar posibles lagunas en la calidad de la traducción-adaptación realizada.

## Discusión y conclusiones

Se han presentado las nuevas directrices para la traducción y adaptación de los tests elaboradas por la Comisión Internacional de Tests (ITC). Es ya conocido y asumido por la comunidad científica que la adaptación de tests no es meramente una cuestión lingüística, y que exige la conjunción de aspectos culturales, conceptuales, lingüísticos y métricos que han de acometerse desde perspectivas de análisis tanto analítico-racionales como empíricas. Las directrices de la ITC aúnan en un documento sencillo las pautas a seguir para asegurar el máximo nivel de equivalencia entre las versiones original y adaptada de un test, que podrían resumirse en: a) consideraciones legales previas que afectan a la propiedad intelectual; b) valoración del constructo en la población diana; c) diseños de adaptación que tengan en cuenta las características lingüísticas, psicológicas y culturales del texto adaptado, así como su adecuación práctica; d) la importancia de las pruebas piloto; e) la selección cualitativa y cuantitativa adecuada de la muestra de adaptación; f) la importancia de los estudios de equivalencia; g) la delimitación del grado de comparabilidad entre puntuaciones; h) la importancia de unas correctas condiciones de aplicación e interpretación; e i) la información exhaustiva sobre los cambios llevados a cabo en el test adaptado. Estas directrices constituyen una actualización y reorganización de las publicadas originalmente (Hambleton, 1996; Hambleton et al., 2005; Muñiz y Hambleton, 1996), tratando de aprovechar la experiencia recogida desde la publicación de la primera edición. Pero la evolución de la evaluación en el ámbito de las Ciencias de la Salud, en general, y de Psicología y Educación, en particular, se ha acelerado notablemente en los últimos años, por lo que las nuevas directrices necesariamente van a requerir futuras revisiones a la luz de los nuevos avances.

Se comentan a continuación algunas de las vías de desarrollo futuro de la evaluación psicológica y educativa, siguiendo las líneas de lo expuesto en Muñiz y Fernández-Hermida (2010) y Muñiz (2012), y centrándose fundamentalmente en los tests, instrumentos claves en el proceso de evaluación. Bien se puede decir que la gran fuerza que está moldeando la evaluación psicológica en la actualidad son las nuevas tecnologías de la información, y en especial los avances informáticos, multimedia e Internet. Expertos como Bennet (1999, 2006), Breithaupt, Mills y Medican (2006), o Drasgow, Luecht y Bennet (2006) consideran que las nuevas tecnologías están influyendo sobre todos los aspectos de la evaluación psicológica, tales como el diseño de los tests, la construcción de los ítems, la presentación de los ítems, la puntuación de los tests y la evaluación a distancia. Todo ello está haciendo cambiar el formato y contenido de las evaluaciones, surgiendo la duda razonable de si los tests de papel y lápiz tal como los conocemos ahora serán capaces de resistir este nuevo cambio tecnológico. Nuevas formas de evaluación emergen, pero los tests psicométricos seguirán siendo herramientas fundamentales, dada su objetividad y economía de medios y tiempo (Phelps, 2005, 2008).

Según Hambleton (2004, 2006), seis grandes áreas atraerán la atención de investigadores y profesionales en los próximos años. La *primera* es el uso internacional de los tests, debido a la globalización creciente y a las facilidades de comunicación, lo cual plantea todo un conjunto de problemas de adaptación de los tests de unos países a otros (Byrne et al., 2009; Hambleton et al., 2005). Esta línea de desarrollo es la que justifica y motiva precisamente la nueva edición de las directrices que aquí se presenta. Esta internacionalización ha puesto de manifiesto la necesidad de disponer de un marco general de evaluación que recoja las buenas prácticas de evaluación. Por ello, el Instituto Internacional de Estandarización (ISO) ha desarrollado una nueva norma (ISO-10667) que recoge la normativa a seguir para una evaluación adecuada de las personas en entornos laborales y organizacionales (ISO, 2011). En España se ha formado un Comité en el seno de AENOR, liderado por el Consejo General de Colegios Oficiales de Psicólogos, que está traduciendo la norma, por lo que no tardará en estar operativa en nuestro país. La *segunda* es el uso de nuevos modelos psicométricos y tecnologías para generar y analizar los tests. Cabe mencionar aquí toda la nueva psicometría derivada de los modelos de Teoría de Respuesta a los Items (TRI), los cuales vienen a solucionar algunos problemas que no encontraban buena solución dentro del marco clásico, pero como siempre ocurre, a la vez que se solucionan unos problemas surgen otros nuevos que no estaban previstos (Abad, Olea, Ponsoda y García, 2011; De Ayala, 2009; Elosua, Hambleton y Muñiz, en prensa; Hambleton, Swaminathan y Rogers, 1991; Muñiz, 1997; Van der Linden y Hambleton, 1997). La *tercera* es la aparición de nuevos formatos de ítems derivados de los grandes avances informáticos y multimedia (Irvine y Kyllonen, 2002; Shermis y Burstein, 2003; Sireci y Zenisky, 2006; Zenisky y Sireci, 2002). Ahora bien, no se trata de innovar por innovar, antes de sustituir los viejos por los nuevos formatos hay que demostrar empíricamente que mejoran lo anterior, las propiedades psicométricas como la fiabilidad y la validez no son negociables. La *cuarta* área que reclamará gran atención es todo lo relacionado con los tests informatizados y sus relaciones con Internet. Mención especial merecen en este campo los Tests Adaptativos Informatizados, que permiten ajustar la prueba a las características de la persona evaluada, sin por ello perder objetividad o comparabilidad entre las personas, lo cual abre perspectivas muy prometedoras en la evaluación (Olea, Abad y Barrada, 2010). La evaluación a distancia o tele-evaluación es otra línea que se abre camino con rapidez, lo cual plantea serios problemas de seguridad de los datos y de las personas, pues hay que comprobar que la persona que se está evaluando es la que realmente dice ser, sobre todo en contextos de selección de personal o de pruebas con importantes repercusiones para la vida futura de la persona evaluada (Bartram y Hambleton, 2006; Leeson, 2006; Mills, Potenza, Fremer y Ward, 2002; Parshall, Spray, Kalohn y Davey, 2002; Williamson, Mislevy y Bejar, 2006; Wilson, 2005). Dentro de esta línea tecnológica también

merecen especial mención los avances relativos a la corrección automatizada de ensayos, que plantea interesantes retos (Shermis y Burstein, 2003; Williamson, Xiaoming y Breyer, 2012). En *quinto* lugar cabe señalar un campo que puede parecer periférico pero que está cobrando gran importancia, se trata de los sistemas a utilizar para dar retroalimentación (feedback) de los resultados a los usuarios y partes legítimamente implicadas. Es fundamental que éstos comprendan sin equívocos los resultados de las evaluaciones, y no es obvio cuál es la mejor manera de hacerlo, sobre todo si se tienen que enviar para su interpretación y explicación del profesional, como ocurre en numerosas situaciones de selección de personal, o en la evaluación educativa (Goodman y Hambleton, 2004). Finalmente, es muy probable que en un futuro haya una gran demanda de *formación* por parte de distintos profesionales relacionados con la evaluación, no necesariamente psicólogos, aunque también, tales como profesores, médicos, enfermeros, etc. No se trata de que estos profesionales puedan utilizar e interpretar los tests propiamente psicológicos, sino que demanden información para poder comprender y participar en los procesos evaluativos y de certificación que se desarrollan en su ámbito laboral.

Éstas son algunas líneas de futuro sobre las que muy probablemente girarán las actividades evaluadoras en un futuro no muy lejano, no se trata de hacer una relación exhaustiva ni mucho menos, sino indicar algunas pistas para orientarse en el mundo rápidamente cambiante de la evaluación psicológica. Si Heráclito tenía razón al decir que *todo fluye*, en el caso de la evaluación psicoeducativa ese fluir es ciertamente vertiginoso. Las directrices aquí presentadas para la traducción-adaptación de los tests constituyen un eje transversal que atraviesa todas estas líneas de futuro esbozadas, pues en cualquiera de las circunstancias citadas siempre cabe la posibilidad de que las pruebas utilizadas hayan de ser adaptadas de unas culturas a otras, variando el calado de las adaptaciones en función de la distancia cultural entre la cultura original y aquella a la que se pretende adaptar la prueba.

Agradecimientos

Este trabajo ha sido financiado por el Ministerio español de Economía y Competitividad (PSI2011-28638, PSI2011-30256) y por la Universidad del País Vasco (GIU12-32).

Referencias

Abad, F.J., Olea, J., Ponsoda, V., y García, C. (2011). *Medición en Ciencias Sociales y de la Salud*. Madrid: Síntesis.

Bartram, D., y Hambleton, R.K. (Eds.) (2006). *Computer-based testing and the Internet: Issues and advances*. Chichester: Wiley.

Bennett, R.E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice, 18*(3), 5-12.

Bennett, R.E. (2006). Inexorable and inevitable: The continuing story of technology and assessment. En D. Bartram y R.K. Hambleton (Eds.), *Computer-based testing and the Internet: Issues and advances*. Chichester: Wiley.

Binet, A., y Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectual des anormaux. *L'Année Psychologique*, *11*, 191-336.

Breithaupt, K.J., Mills, C.N., y Melican, G.J. (2006). Facing the opportunities of the future. En D. Bartram y R.K. Hambleton (Eds.), *Computer-based testing and the Internet* (pp. 219-251). Chichester: John Wiley and Sons.

Brislin, R.W. (1986). The wording and translation of research instruments. En W.J. Lonner y J.W. Berry (Eds.), *Field methods in cross-cultural psychology* (pp. 137-164). Newbury Park, CA: Sage Publications.

Byrne, B. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*, 872-882.

Byrne, B.M., Leong, F.T., Hambleton, R.K., Oakland, T., van de Vijver, F.J., y Cheung, F.M. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology, 3*(2), 94-105.

Calvo, N., Gutiérrez, F., Andión, O., Caseras, X., Torrubia, R., y Casas, M. (2012). Psychometric properties of the Spanish version of the self-report personality diagnostic questionnaire-4+ (PDQ-4+) in psychiatric outpatients. *Psicothema, 24*(1), 156-160.

De Ayala, R.J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.

Downing, S.M. (2006). Twelve steps for effective test development. En S.M. Downing y T.M. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Lawrence Erlbaum Associates.

Downing, S.M., y Haladyna, T.M. (Eds.) (2006). *Handbook of test development*. Mahwah, NJ: LEA.

Drasgow, F., Luecht, R.M., y Bennett, R.E. (2006). Technology and testing. En R.L. Brennan (Ed.), *Educational measurement*. Westport, CT: ACE/Praeger.

Elosua, P. (2003). Sobre la validez de los tests. *Psicothema, 15*(2), 315-321.

Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de autoconcepto. *Psicothema, 17*(2), 356-362.

Elosua, P. (2012). Tests publicados en España: usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo, 33*(1), 12-21.

Elosua, P., Bully, P., Mujika, J., y Almeida, L. (2012, julio). Practical ways to apply the ITC precondition, and development guidelines in adapting tests. Spanish adaptation of "Bateria de Provas de Raciocinio". *Paper presented at the* V *European Congress of Methodology, Santiago de Compostela*.

Elosua, P., Hambleton, R., y Muñiz, J. (in press). *Teoría de la Respuesta al ítem aplicada con R*. Madrid: La Muralla.

Elosua, P., e Iliescu, D. (2012). Tests in Europe. Where we are and where we should to go? *International Journal of Testing, 12*, 157-175.

Elosua, P., y Muñiz, J. (2010). Exploring the factorial structure of the self-concept: A sequential approach using CFA, MIMIC and MACS models, across gender and two languages. *European Psychologist*, *15*, 58-67.

Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J.R., et al. (2012). Testing practices in the 21st century: Developments and European psychologists's opinions. *European Psychologist*, *17*(4), 300-319.

Goodman, D.P., y Hambleton, R.K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education, 17*, 145-220.

Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*(3), 229-244.

Hambleton, R.K. (1996). Adaptación de tests para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas. En J. Muñiz (Ed.), *Psicometría* (pp. 207-238). Madrid: Universitas.

Hambleton, R.K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17*(3), 164-172.

Hambleton, R.K. (2004). Theory, methods, and practices in testing for the 21st century. *Psicothema, 16*(4), 696-701.

Hambleton, R.K. (2006). *Testing practices in the 21st century*. Key Note Address, University of Oviedo, Spain, March 8th.

Hambleton, R.K. (2009, julio). International Test Commission Guidelines for Test Adaptation, second edition. *Paper presented at the 11th European Congress of Psychology, Oslo*.

Hambleton, R.K., y Bollwark, J. (1991). Adapting tests for use in different cultures: Technical issues and methods. *Bulletin of the International Test Commission*, *18*, 3-32.

Hambleton, R.K., Merenda, P., y Spielberger, C. (Eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence Erlbaum Publishers.

Hambleton, R.K., Swaminathan, H., y Rogers, J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: Sage.

Hambleton, R.K., y Zenisky, A.L. (2011). Translating and adapting tests for cross-cultural assessments. En D. Matsumoto y F.J.R. van de Vijver (Eds.), *Cross-cultural research methods in psychology*. Nueva York: Cambridge University Press (pp. 46-70).

ISO (2011). *Procedures and methods to assess people in work and organizational settings (part 1 and 2)*. Ginebra: ISO.

Irvine, S., y Kyllonen, P. (Eds.) (2002). *Item generation for test development*. Mahwah, NJ: Lawrence Erlbaum.

Iturbide, L.M., Elosua, P., y Yenes, F. (2010). Medida de la cohesión en equipos deportivos. Adaptación al español del Group Environment Questionnaire (GEC). *Psicothema, 22*, 482-488.

Leeson, H.V. (2006). The mode effect: A literature review of human and technological issues in computerized testing. *International Journal of Testing, 6*, 1-24.

Lissitz, R.W. (Ed.) (2009). *The concept of validity: Revisions, new directions, and applications*. Charlotte, NC: Information Age Publishing.

Matsumoto, D., y van de Vijver, F.J.R. (Eds.) (2011). *Cross-cultural research methods in psychology*. Nueva York: Cambridge University Press.

Mills, C.N., Potenza, M.T., Fremer, J.J., y Ward, W.C. (Eds.) (2002). *Computer-based testing: Building the foundation for future assessments*. Hillsdale, NJ: LEA.

Muñiz, J. (2003). *Teoría clásica de los tests*. Madrid: Pirámide.

Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.

Muñiz, J. (2012). Perspectivas actuales y retos futuros de la evaluación psicológica. En C. Zúñiga (Ed.), *Psicología, sociedad y equidad*. Santiago de Chile: Universidad de Chile.

Muñiz, J., y Fernández-Hermida, J.R. (2010). La opinión de los psicólogos españoles sobre el uso de los tests. *Papeles del Psicólogo, 31*, 108-121.

Muñiz, J., y Hambleton, R.K. (1996). Directrices para la traducción y adaptación de tests. *Papeles del Psicólogo, 66*, 63-70.

Nogueira, R., Godoy, A., Romero, P., Gavino, A., y Cobos, M.P. (2012). Propiedades psicométricas de la versión española del Obssesive Belief Questionnaire-Children VersiOn (OBQ-CV) en una muestra no clínica. *Psicothema, 24*(4), 674-679.

Olea, J., Abad, F., y Barrada, J.R. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del Psicólogo*, *31*(1), 94-107.

Ortiz, S., Navarro, C., García, E., Ramis, C., y Manassero, M.A. (2012). Validación de la versión española de la escala de trabajo emocional de Frankfurt. *Psicothema, 24*(2), 337-342.

Parshall, C.G., Spray, J.A., Kalohn, J.C., y Davey, T. (2002). *Practical considerations in computer-based testing*. New York: Springer.

Phelps, R. (Ed.) (2005). *Defending standardized testing*. Londres: LEA.

Phelps, R. (Ed.) (2008). *Correcting fallacies about educational and psychological testing*. Washington: APA.

Prieto, G., y Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo, 77*, 65-71.

Rodríguez, M.S., Martínez, Z., Tinajero, C., Guisande, M.A., y Páramo, M.F. (2012). Adaptación española de la Escala de Aceptación Percibida (PAS) en estudiantes universitarios. *Psicothema , 24*(3), 483-488.

Shermis, M.D., y Burstein, J.C. (Eds.) (2003). *Automated essay scoring: A cross-disciplinary perspective*. Hillsdale, NJ: LEA.

Schmeiser, C.B., y Welch, C. (2006). Test development. En R.L. Brennan (Ed.), *Educational measurement* (4th edition). Westport, CT: American Council on Education/Praeger.

Sireci, S., y Zenisky, A.L. (2006). Innovative items format in computer-based testing: In pursuit of construct representation. En S.M. Downing y T.M. Haladyna (Eds.), *Handbook of test development*. Hillsdale, NJ: LEA.

Van de Vijver, F.J.R., y Tanzer, N.K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 47*(4), 263-279.

Westbury, I. (1992). Comparing American and Japanese achievement: Is the United States really a low achiever? *Educational Researcher*, *21*, 10-24.

Williamson, D.M., Mislevy, R.J., y Bejar, I. (2006). *Automated scoring of complex tasks in computer based testing*. Mahwah, NJ: LEA.

Williamson, D.M., Xi, X., y Breyer, J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice, 31*(1), 2-13.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: LEA.

Zenisky, A.L., y Sireci, S.G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education, 15*, 337-362.