

Revista Latinoamericana de Psicología

www.elsevier.es/rlp



ORIGINAL

Procedimientos analítico-rationales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento

Paula Elosua^{a,*}, Josu Mujika^a, Leandro S. Almeida^b y Daniel Hermosilla^a

^a Universidad del País Vasco, San Sebastián, España

^b Universidad de Minho, Braga, Portugal

Recibido el 16 de octubre de 2013; aceptado el 29 de abril de 2014

PALABRAS CLAVE

Psicometría;
Adaptación de tests;
Fase analítico-razional;
Evaluación cross-cultural;
Batería de pruebas de razonamiento

Resumen

Adaptar un test entre culturas o idiomas requiere considerar aspectos legales, culturales, lingüísticos, métricos y de uso. La atención prestada por la literatura especializada a los aspectos metodológicos implicados en el estudio de la equivalencia métrica ha sido significativamente mayor que la dedicada a los procedimientos analítico-rationales previos a la fase de confirmación empírica. Sin embargo, la consideración de estos últimos es crucial en el proceso de adaptación. En esta línea, el objetivo de este trabajo es incidir en la relevancia de las etapas previas y describirlas, ofreciendo una sistematización que considera diez apartados. Todos ellos contribuyen a garantizar y favorecer la construcción de un test adaptado y equivalente, en la medida de lo posible, al test original. El proceso es ejemplificado con la adaptación al español de una prueba cognitiva construida en portugués para la población portuguesa, la *Bateria de Provas de Raciocínio*.

Copyright © 2013, Fundación Universitaria Konrad Lorenz. Publicado por Elsevier España, S.L.U. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

KEYWORDS

Psychometrics;
Test Adaptation;
Judgmental-analytical procedures
Cross-cultural Assessment;
Reasoning Test Battery

Judgmental-analytical procedures for adapting tests: Adaptation to Spanish of the Reasoning Tests Battery

Abstract

Adapting a test between cultures or languages requires taking into account legal, linguistic, metric, and use-related considerations. Significantly more attention has been paid to the methodological aspects involved in the study of metric equivalence than to judgmental-analytical procedures prior to the empirical confirmation stage. However, considering the latter is crucial in the adaptation process. Along these lines, this paper seeks to describe and focus on the relevance of the previous stages, thereby offering a systematization process that comprises ten sections. This approach contributes to

*Autor para correspondencia.

Correo electrónico: paula.elosua@ehu.es (P. Elosua).

ensuring the construction of a test adapted and equivalent in as much as possible to the original. This process is exemplified by means of a Spanish language adaptation of a cognitive test originally designed in Portuguese for the Portuguese population, the Reasoning Test Battery.

Copyright © 2013, Konrad Lorenz University Foundation. Published by Elsevier España, S.L.U. This is an open-access article distributed under the terms of the Creative Commons CC BY-NC ND Licence (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Desde las primeras adaptaciones de las escalas pioneras de Binet y Simon (Binet & Simon, 1905) y las actualizaciones llevadas a cabo por Terman-Merrill en los Estados Unidos para su uso en contextos lingüísticos y culturales diferentes a los originales, asistimos a un incremento continuo en el uso de tests adaptados (Elosua & Iliescu, 2012).

La necesidad de disponer de instrumentos de medida en distintos idiomas y culturas, junto con la exigencia de garantizar que el proceso que da lugar a un test adaptado cumple con criterios legales, éticos y psicométricos son el origen del surgimiento de un campo teórico relacionado con el estudio de la adaptación de tests (Elosua, 2005; Elosua & Muñiz, 2010; van de Vijver & Tanzer, 1997; Hambleton, 2001; Hambleton, Merenda, & Spielberger, 2005). Los avances sustantivos, metodológicos y técnicos en esta área han sido recogidos por organismos internacionales como la Comisión Internacional de Test (International Test Commission; <http://www.intestcom.org>), que los ha sistematizado y organizado en un conjunto de 20 directrices (Muñiz, Elosua, & Hambleton, 2013).

Su finalidad es doble: por un lado, ofrecer un marco integral que aborda los puntos clave en el proceso de adaptación—consideraciones previas, análisis de la propia adaptación, justificación técnica, evaluación e interpretación de las puntuaciones y elaboración del documento final (fig. 1)—y por otro, desterrar mitos y falsas creencias relacionadas con la adaptación de tests. Siguiendo a Hambleton (2009) podrían enumerarse cuatro nociones y prácticas comunes erróneas que influyen de forma negativa en la calidad del producto final:

1. *Hablar dos idiomas es suficiente condición para adaptar un test.* No es cierto, para adaptar un test, además de hablar dos idiomas (idioma origen e idioma destino), es necesario conocer las culturas, estar familiarizado con la variable medida, y estar mínimamente formado en las normas sobre redacción y corrección de ítems.
2. *Una buena traducción literal garantiza la equivalencia.* No es cierto. La adaptación de tests no se limita a presuponer que una traducción literal fiel cubre y garantiza que la versión adaptada sea equivalente a la versión original. En la adaptación de tests intervienen aspectos relacionados con el marco legal, diferencias culturales, divergencias lingüísticas, experiencia con el formato o familiaridad con las condiciones de aplicación que pueden ser causa de sesgo en la versión adaptada.
3. *La traducción inversa como procedimiento de verificación garantiza la equivalencia entre la versión original y la versión adaptada.* No es cierto. La retrotraducción o traducción inversa (*back-translation*, Brislin, 1986), utilizada de forma rutinaria como verificación de la calidad de la adap-

tación, considera que en una buena traducción la equivalencia entre la versión original y la retrotraducida generada por un traductor independiente es alta. Sin embargo, la situación inversa también es cierta, pues, habitualmente, las malas traducciones se apoyan en interpretaciones literales en lugar de en la adaptación de significados.

La verificación de la adaptación requiere la participación de un grupo de expertos en las áreas lingüística, cultural, psicométrica y sustantiva que a través de un proceso iterativo de depuración llega a una versión final consensuada. El proceso se iniciaría con traducciones independientes (mínimo dos) del idioma original al idioma destino (*forward translation*), que serán posteriormente revisadas y analizadas. El objetivo es lograr un test lingüísticamente correcto, culturalmente adaptado, que mida el constructo psicológico con precisión y validez, y utilice un lenguaje adecuado con respecto a las características de la población que se ha de evaluar.

4. *Los constructos evaluados son universales.* No es cierto. La hipótesis de generalización de un constructo a otras culturas ha de ser evaluada en cada situación. Los trabajos sobre medidas y modelos para la inteligencia (Stenberg, 2000), la personalidad (Cheung, 2004) o la calidad de vida (Skevington, Sartorius, Amir, & the WHOQOL-Group, 2004) muestran las peculiaridades asociadas a cada contexto sociocultural cuando se trata de definir y de operacionalizar la medición de la inteligencia, personalidad u otros constructos psicológicos.

Procedimientos analítico-rationales

Los errores citados afectan a las fases previas a la recogida de datos operacional y al consiguiente estudio de la equivalencia psicométrica (Elosua & Muñiz, 2010). Por lo tanto, su efecto habría de neutralizarse en estas mismas etapas. Existen pautas y procedimientos analítico-rationales diseñados con esta finalidad, que a través de un proceso depurativo e iterativo verifican la calidad lingüística, cultural, legal y sustantiva del test adaptado. Sus objetivos son básicamente: (a) explorar los aspectos legales implicados en la adaptación; (b) evaluar el grado de solapamiento del constructo en las poblaciones implicadas; (c) adaptar y verificar la adaptación con la ayuda de una comisión de expertos, y (d) obtener y analizar datos que permitan verificar la correcta comprensión de las tareas y de las instrucciones del test.

Marco legal y de uso

Es importante analizar los aspectos relacionados con la propiedad intelectual y con los usos y costumbres locales. En

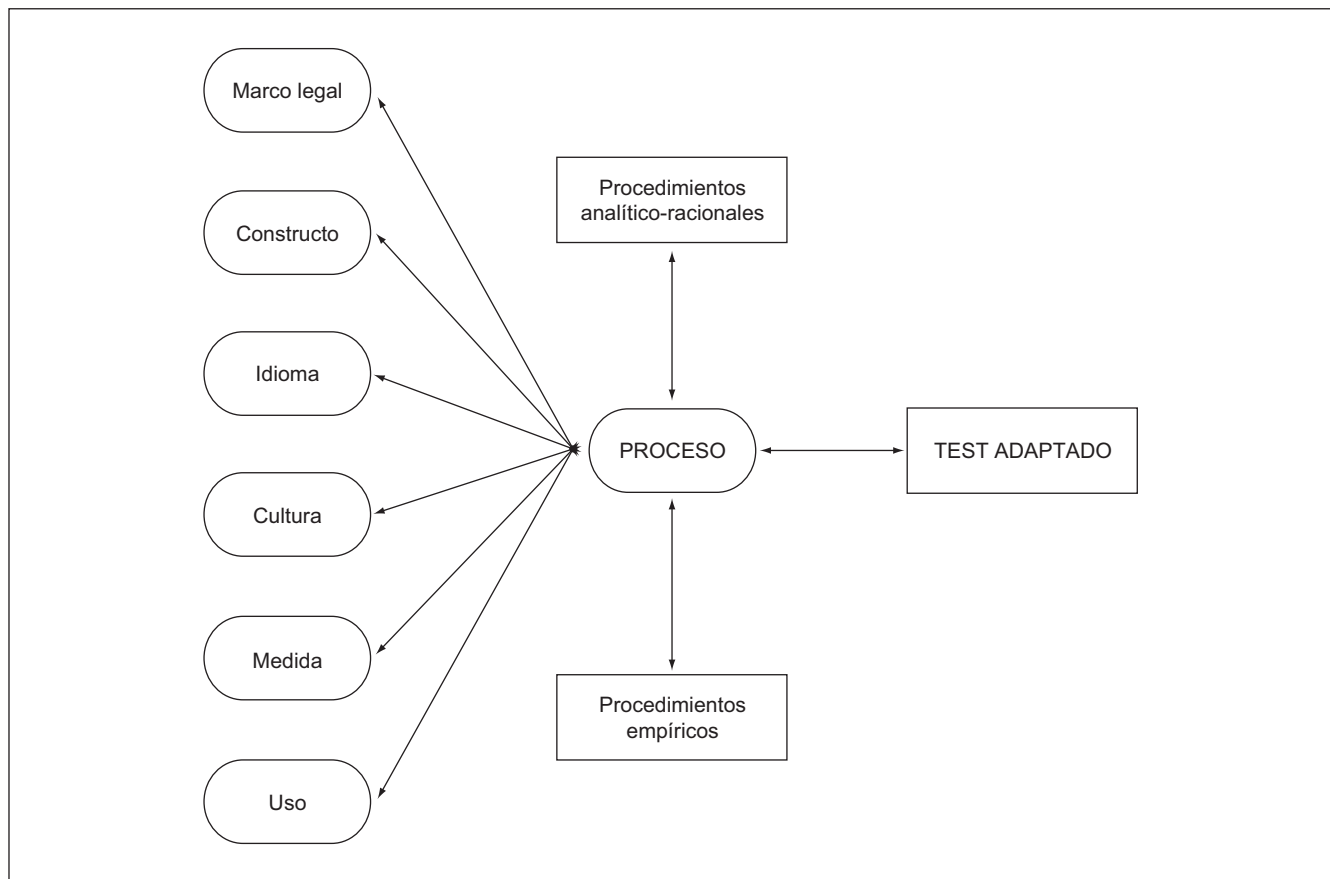


Figura 1. Proceso de adaptación (tomado de Muñiz, Elosua, & Hambleton, 2013).

un proceso de adaptación conviene comprobar sobre quién recae el derecho de la propiedad intelectual del cuestionario y en su caso, obtener los permisos legales necesarios que garanticen la autenticidad del producto y protejan el trabajo de adaptaciones no autorizadas.

Con respecto al marco ético en el que se desarrolla el proyecto, es necesario considerar los principios básicos para la investigación en seres humanos establecidos en el Código de Núremberg (1947), la Declaración de Helsinki (1964) o el Informe Belmont (1978). Con base en ellos conviene elaborar los siguientes documentos:

- **Presentación del proyecto.** Documento que incluye información sobre los objetivos de la investigación, entidades participantes, tratamiento de datos, garantía de confidencialidad de los datos y solvencia científica de los responsables de la investigación.
- **Convenio de colaboración** con los centros o instituciones en los que se fuera a administrar la prueba.
- **Modelo de consentimiento informado** que debería ser firmado por los participantes o, en su caso, por sus tutores.

Estudio teórico

Este punto reclama: (a) la necesidad de estudiar las características del constructo que se debe medir en la población destino, sin asumir la universalidad de los constructos entre

culturas, y (b) evaluar el grado o nivel de solapamiento entre el constructo en la población origen y en la población diana; sin este paso previo no es posible delimitar y definir el grado de equivalencia deseado (van de Vijver & Leung, 2011). Para alcanzar estos objetivos se recomienda: (a) la revisión bibliográfica sobre la variable medida en ambas poblaciones; (b) el estudio de las redes nomológicas del constructo en las poblaciones de interés (Cronbach & Meehl, 1955), y (c) comprobar el patrón de correlaciones convergentes con medias relacionadas y la presencia de bajos coeficientes de correlación entre medidas de diferentes constructos (validez discriminante).

Adaptación

Hace referencia al proceso iterativo de generación de la versión del test en la población destino. En esta fase interviene un grupo multidisciplinar de expertos formado por traductores, especialistas en la variable medida y en construcción de tests. La adaptación se lleva a cabo en varias etapas que se inician con al menos dos traducciones independientes del test al idioma destino (*forward translation*). Una vez obtenidas las dos versiones, se instruye a la comisión de expertos para que apliquen criterios de verificación de la adaptación (Elosua & López, 2007). Esta se lleva a cabo, en primer lugar, de forma individual, y posteriormente es discutida por la comisión. La retrotraducción —en caso

de que fuera utilizada— no es más que un método de verificación que habría de complementarse con el estudio de los siguientes puntos:

- **Equivalencia gramatical.** Cada idioma se caracteriza por poseer estructuras gramaticales propias (tipología lingüística) que no tienen necesariamente equivalentes en la lengua terminal. La adaptación entre idiomas pertenecientes a clases tipológicas diferentes origina, como mínimo, un cambio de formato (orden de los elementos de la oración, concordancia enunciado/alternativas, longitud de la oración...) que en ocasiones genera cambios en la dificultad del ítem en uno de los grupos.
- **Equivalencia semántica.** Analiza la equivalencia entre los significados connotativos de la voz en L1 y L2 prestando especial atención a la posible inadecuación y falta de correspondencia entre la palabra en la lengua original y la voz dada en la lengua terminal. Un término en un idioma posee normas de familiaridad, concreción o valencias afectivas que pueden no coincidir con su traducción literal a otro idioma. La presencia de diferentes valores en estas categorías puede producir falta de equivalencia psicométrica.
- **Relevancia cultural.** Hace referencia a la caracterización de un ítem como *etic* o *emic* (Berry, Poortinga, Segall, & Dasen, 1992; Pike, 1967). Existen voces que no pueden trasladarse directamente de un idioma a otro por falta de referentes equivalentes. En esta categoría se podrían incluir las costumbres locales, estructuras políticas, características geográficas o representaciones culturales (también los símbolos). Habría que analizar en cada caso, en función de los usos previstos para el test, la pertinencia de los ítems con elevada carga cultural y su adaptación a L2.
- **Adecuación lingüística.** No es suficiente que una voz esté correctamente adaptada, esta tiene que evaluarse con referencia a las características de la población destino, por ello, es importante considerar: (a) el registro lingüístico utilizado en la versión original y en la adaptada (científico/familiar), (b) la adecuación del lenguaje al perfil del evaluado, y (c) los factores que afectan a la legibilidad del texto, como longitud de los enunciados, número de palabras...
- **Formato y diseño.** La verificación ha de comprobar la correspondencia en la apariencia física de los ítems en las versiones original y adaptada.

Pretest cognitivo y prueba piloto

La fase de adaptación podría complementarse con técnicas de pretest cognitivo o entrevistas cognitivas (DeMaio & Rothgeb, 1996; Willis, 2005) para asegurarse de la correcta comprensión de las instrucciones y el contenido. Es importante comprobar que todas y cada una de las secciones del test (instrucciones, contenido, forma de respuesta) son bien entendidas por la población diana. Para ello, tras seleccionar un reducido número de participantes, se les interrogará incidiendo en: (a) si entienden cada una de las cuestiones, (b) qué entienden por cada una de ellas, (c) se les pedirá que las repitan utilizando palabras propias, (d) se les pedirá que expliquen cómo y por qué eligen la respuesta correcta o en su caso, elaboran la repuesta, y (e) se les preguntará por cualquier aspecto de las instrucciones o

palabra que no entiendan bien o que pueda resultarles ofensiva o gratuita.

Prueba piloto

Antes de la fase operacional de validación y baremación es importante disponer de datos obtenidos en una pequeña muestra representativa de la población a la que va dirigida la prueba. Los datos piloto permitirán: (a) recoger *in situ* las reacciones de las personas que realizan el cuestionario, (b) asegurarse de que los ítems se comprenden correctamente, (c) cerciorarse de que las instrucciones están bien redactadas y se entienden bien, (d) registrar el tiempo necesario para la ejecución de la prueba, (e) recoger información sobre posibles errores de contenido o formato, y (f) obtener datos para un primer análisis de ítems que indiquen la dirección y sentido de los índices psicométricos más relevantes:

- **Índices psicométricos clásicos.** Los índices de dificultad y los índices de discriminación permiten obtener información inicial sobre el funcionamiento general del test. Conviene revisar aquellos valores extremos como indicadores de potenciales problemas.
- **Ordenamiento de los índices de dificultad.** Es aconsejable construir tablas comparativas entre el orden de dificultad de los ítems en la prueba original y en la versión adaptada. Si el ordenamiento es altamente discordante, sería necesario revisar y tal vez modificar el contenido de los ítems.
- **Índices globales.** El estudio de la consistencia interna de los elementos, su dimensionalidad y el análisis de las correlaciones entre escalas parciales ofrece información que podrá compararse con la aportada en la documentación de la prueba original. En caso de encontrar valores discordantes, se procedería a una reconsideración del contenido de los ítems afectados.

Una vez cumplimentadas estas etapas, y habiendo modificado en cada caso el contenido de acuerdo con cada criterio, se obtiene la versión del test que se administrará a una muestra amplia de la población para la fase de confirmación empírica (Muñiz et al., 2013).

En este contexto teórico, el objetivo del trabajo es ilustrar los procedimientos analítico-rationales aplicados a la adaptación al español de la batería de pruebas de razonamiento (BPR; Almeida & Lemos, 2006).

Método

Instrumento

La BPR es un desarrollo del *Test de Raisonment Differentiel* construido por Georges Meuris (1969) en Bélgica. La finalidad de la batería es evaluar la capacidad de razonamiento de los niños y adolescentes a partir de tareas con contenidos diferenciados. Si bien la primera versión de la batería incluía en su título la palabra “diferencial”, los resultados de los análisis factoriales llevados a cabo sobre versiones previas aconsejaron eliminar la connotación diferencial, pues de forma sistemática se concluía la presencia de un

factor general relacionado con el razonamiento general (Almeida & Lemos, 2006).

La BPR se construye sobre las teorías jerárquicas de organización de las habilidades cognitivas (Cattell, 1963, 1971; Vernon, 1961) y en consecuencia, combina factores cognitivos generales y factores específicos (Almeida, 2002; Horn & Noll, 1997). Tomando como referencia el modelo sobre los tres estratos o teoría de Cattell-Horn-Caroll, el razonamiento general se asocia con el factor g o gf (inteligencia fluida) y los contenidos concretos con factores específicos relacionados con funciones cognitivas o tipos de información que se han de procesar (por ejemplo, verbal, figurativa, numérica) que definen factores de segundo orden (Carroll, 2003). La BPR combina la evaluación de aspectos generales de la inteligencia, en particular, los procesos de razonamiento que se pueden asumir como esenciales en la definición del factor g (Almeida, 2002) y otros componentes de la inteligencia más asociados a las aptitudes específicas evaluadas en las baterías multifactoriales de inteligencia.

En términos de razonamiento, las pruebas presentan tareas de analogías, series que se han de completar y resolución de problemas; con relación al contenido, las pruebas están constituidas por figuras geométricas sin significado (figurativo-abstracto), significado de palabras (verbal), secuencia de números (numérico), cubos en movimiento (espacial) y situaciones prácticas (concreto-mecánico). La versión actual portuguesa es una batería de pruebas que cuenta con tres versiones diferentes en función del nivel de escolarización del alumno: BPR5/6 constituida por cuatro pruebas, BPR7/9 y BPR10/12 formadas por cinco pruebas. En la tabla 1 se presenta la estructura de las tres versiones de la BPR (Almeida & Lemos, 2006) que se adaptaron a la población española.

Aspectos legales y formales

Marco legal

Los autores de la adaptación se pusieron en contacto con los autores de la prueba para obtener los permisos necesarios que garantizaran la legalidad de la versión adaptada.

Aspectos éticos

Se elaboraron los siguientes documentos:

- **Presentación del proyecto** para su evaluación por los centros y tutores de los estudiantes participantes en la investigación.
- **Convenio de colaboración** con los centros educativos en los que se fuera a administrar la prueba, que fueron firmados por ambas partes.

- **Modelo de consentimiento informado** que debería ser firmado por los tutores de los estudiantes a los que se fuera a administrar la prueba.

Estudio teórico

Evaluación del constructo

Se llevó a cabo una revisión bibliográfica exhaustiva sobre los tests de razonamiento publicados en España y en Portugal, y sobre la batería BPR (Almeida & Lemos, 2006; Primi & Almeida, 2000). Se revisaron las redes nomológicas relacionadas con el factor g tanto en España como en Portugal, y los estudios que analizaban la relación entre el factor g y el rendimiento académico (Almeida, Guisande, Primi, & Lemos, 2008; Lemos, Abad, Almeida, & Colom, 2013).

Tal y como se esperaba, no se encontraron indicios que indicaran la diferente estructura del modelo jerárquico de la inteligencia entre Portugal y España.

Análisis de los materiales y del formato de los ítems

Se analizaron el formato de los ítems, la tarea demandada y las instrucciones para poder establecer el grado de comparabilidad entre la población original (portuguesa) y la española. Se evaluó la correspondencia entre los ciclos educativos y las distintas formas de la BPR. Las diferencias estructurales entre los sistemas educativos se refirieron a la duración de los ciclos. En el proceso de adaptación se modificaron los títulos de las formas de la BPR, y se añadió un curso educativo más a la forma 1. La tabla 2 muestra la concordancia entre las versiones originales y las adaptadas, así como los cursos y ciclos destinatarios de cada una de ellas.

Se concluyó la pertinencia de generalización de las variables cognitivas medidas por la BPR y se adaptó el título de cada una de las versiones de la BPR del portugués al español.

Adaptación lingüística y cultural

Comité de expertos

Se formó un equipo multidisciplinar constituido por dos traductores profesionales, dos orientadores pedagógicos pertenecientes a los centros educativos públicos que participaron en el proyecto de adaptación, y dos investigadores y profesores de psicometría con experiencia en la adaptación de tests y modelos psicométricos.

Adaptación del test

Los traductores profesionales adaptaron de forma independiente el test al español. Los miembros del comité de expertos verificaron las adaptaciones de forma independiente siguiendo para ello las pautas establecidas en la tabla 3 que

Tabla 1 Estructura de la batería de pruebas de razonamiento

		Razonamiento abstracto	Razonamiento verbal	Razonamiento espacial	Razonamiento numérico	Razonamiento práctico	Razonamiento mecánico
BPR5/6	N.º ítems	20	20	-	15	15	-
BPR7/9BPR 10/12	N.º ítems	25	25	20	20	-	25

BPR: batería de pruebas de razonamiento.

se presentaron en forma de cuadrícula para cada uno de los ítems analizados. Posteriormente, a lo largo de varias sesiones de trabajo, se llegó a una versión consensuada en la que se primó la adecuación de los contenidos a las características de la población destino.

- **Equivalencia gramatical.** El portugués y el español son lenguas lingüísticamente cercanas; ambas son lenguas romances pertenecientes a la familia lingüística indoeuropea, lo cual minimiza el impacto de la falta de equivalencia gramatical.
- **Equivalencia semántica.** Con el fin de salvaguardar la equivalencia semántica de los ítems se propusieron modificaciones en la prueba de razonamiento verbal. Se trata

de una tarea de analogías en las que hay que inferir la relación que existe entre varias palabras, utilizando, en cada caso, reglas de sinonimia, antonimia o de relación lógica por funcionalidad, proximidad. Cuando se optó por modificar el contenido, se mantuvo la estructura y tipo de la analogía original y se buscó que la familiaridad con el contenido no causará una mayor/menor dificultad en la tarea. Se construyó una tabla que documentó los cambios realizados (tabla 4).

- **Relevancia cultural.** La prueba no incluye referencias locales.
- **Adecuación lingüística.** La versión adaptada se envió a un centro educativo para que, antes de la administración

Tabla 2 Correspondencia entre las formas, ciclos educativos y cursos de las formas de la batería de pruebas de razonamiento

	Forma 1		Forma 2		Forma 3	
	Ciclos	Cursos	Ciclos	Cursos	Ciclos	Cursos
Portugal	2 Enseñanza básica	5, 6	3 Enseñanza básica	7, 8, 9	Secundaria	10, 11, 12
España	2 y 3 Enseñanza primaria	4, 5, 6	1 y 2 Educación secundaria	1, 2, 3	2 Educación secundaria, bachiller, grados formativos	4 B1-2

Tabla 3 Puntos en el proceso de verificación

Verificación	Sí	No	No procede
¿Pertencen los idiomas a familias lingüísticas diferentes?			
¿Se ha añadido información en el ítem que no aparece en la versión original?			
¿Se ha omitido información en el ítem que aparece en la versión original?			
¿Es correcta la traducción?			
¿Se ha cambiado el orden sintáctico del enunciado o de las opciones que originen un nivel de dificultad mayor en una de las versiones?			
¿Es la longitud del texto similar en ambas versiones?			
¿El uso de la sufijación ha originado cambios en la dificultad del ítem?			
¿Existen polisemias en la versión adaptada que añadan/reduzcan la dificultad del ítem?			
¿Se han adaptado los contenidos específicos de una cultura/idioma a sus equivalentes en el idioma destino?			
¿Es el aspecto visual del ítem igual al original? (formato, negrillas, subrayados, gráficos, tablas, apariencia, otros aspectos visuales)			

Tabla 4 Ejemplo de cambios sobre la forma 2 de la batería de pruebas de razonamiento

Versión en portugués		Versión en español	
Orden ítem	Contenido	Orden ítem	Contenido
18	ensaio-exibição-treino	7	Suiza-país-África
	repetição-vitória-desporto-preparação-competição		unión-comunidad-península-isla-continente
2	quartel-soldado-fábrica	11	estrella-constelación-cerdo
	trabalho-indústria-esforço-máquina-operário		carne-manada-barro-caserío-piara
23	juntar-fundir-reunir	15	juntar-fundir-aproximar
	relacionar-incorporar-associar-aliar-misturar		asociar-incorporar-acercar-aliar-mezclar

piloto, fuera evaluada por los profesores de los estudiantes a los que iba dirigida la prueba. La selección del centro fue incidental.

- **Formato y diseño.** Las pruebas se maquetaron en su totalidad con la ayuda de un gabinete de arquitectos. Los ítems gráficos fueron meticulosamente analizados para garantizar su calidad gráfica y hermandad con los ítems originales.

Fase piloto

Muestra piloto

Tras un muestreo incidental, se contactó con los centros colaboradores. La muestra piloto estuvo formada por 64 estudiantes para la forma 1, 76 estudiantes respondieron a la forma 2, y 46 lo hicieron a la forma 3. La distribución por curso educativo y sexo puede consultarse en la tabla 5.

Contacto y administración

- **Contacto con los centros participantes.** Se contactó con los centros seleccionados, para fijar reuniones informativas con la dirección del centro, y un calendario de administración de las pruebas piloto.
- **Formación de personal y administración.** Se formó a un grupo de colaboradores para la administración de los tests que fueron aplicados durante el horario lectivo en presencia de uno de los miembros de la comisión de expertos y del tutor de cada curso.

Durante la administración, se controló: (a) tiempo de ejecución; (b) se comprobó que los ítems se entendieran bien, y (c) se cotejó que las instrucciones fueran comprendidas correctamente. Los maestros de algunos de los cursos sugirieron “recortar” el tiempo dedicado a la explicación del test y a los ejemplos de entrenamiento previos que acompañaban cada tarea. Sugerencia que se incorporó al test final.

Análisis preliminares

Los datos empíricos recogidos durante la fase piloto permitieron llevar a cabo estudios preliminares sobre: (a) la distribución de las respuestas a los ítems; (b) su dificultad; (c) estructura dimensional, y (d) consistencia interna.

- **Índices psicométricos clásicos.** Los índices de dificultad y los índices de discriminación se compararon con los obtenidos en la prueba original (tabla 6; aportamos los datos de la prueba de razonamiento abstracto perteneciente a la forma 2 de la BPR).

- **Ordenamiento de los índices de dificultad.** Se construyeron tablas comparativas entre el orden de dificultad que ocupaban los ítems en la prueba original y el orden que ocupaban sus versiones adaptadas. Esta información se obtuvo a partir del manual del test original. Para facilitar la comparación e interpretación, se asignó cada ítem a un cuartil de dificultad (tabla 6). Se detectaron algunas diferencias en la prueba de comprensión verbal, debido a las cuales se modificó el orden de presentación de algunos de los ítems (tabla 4).

- **Índices globales.** Si bien el tamaño de la muestra no es elevado, se ejecutó un estudio exploratorio de la dimensionalidad por medio de un análisis de componentes principales. El primer componente extraído tuvo un valor propio de 3.18 y explicó un porcentaje de varianza igual al 63.75%. Los resultados son concordantes con la teoría original subyacente a la construcción de la BPR (Almeida & Lemos, 2006), que justifica la existencia de un factor general de razonamiento.

La consistencia interna de las escalas se analizó con el coeficiente alfa ordinal (tabla 7) (Elosua & Zumbo, 2008). Los resultados fueron óptimos, con valores situados entre 0.82 en la prueba de razonamiento mecánico, y 0.93 en la prueba de razonamiento numérico.

Comentarios

Es conocido y asumido por la comunidad científica que la adaptación de tests no es meramente una cuestión lingüística ni tampoco una cuestión estrictamente cuantitativa; exige la conjunción de aspectos legales, culturales, conceptuales, lingüísticos y métricos que han de acometerse desde perspectivas de estudio analítico-rationales y empíricas.

La necesidad de adaptar tests entre culturas e idiomas, junto a la exigencia de que la medición cumpla con estrictos criterios éticos y científicos han impulsado la colaboración internacional en el desarrollo de un marco teórico y aplicado relacionado con la adaptación de tests. El máximo exponente de esta cooperación lo definen las directrices para la adaptación de tests desarrolladas desde la Comisión Internacional de Test.

Las directrices de la Comisión Internacional de Test aún nan las pautas que se han de seguir para asegurar el máximo nivel de equivalencia entre las versiones original y adapta-

Tabla 5 Distribución de la muestra piloto por curso educativo y sexo

Edades	Curso educativo	Mujeres	Varones	Totales
8-9	Educación primaria 4	11	11	22
10-11	Educación primaria 5	9	11	20
11-12	Educación primaria 6	7	14	21
12-13	Educación secundaria 1	27	17	44
13-14	Educación secundaria 2	6	13	19
14-15	Educación secundaria 3	5	8	13
16-18	Bachiller	23	24	46
	Totales	88	98	186

da 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 la prueba 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 sobre los cambios llevados a cabo en el test adaptado.

En este marco, el presente trabajo se ha centrado en las fases previas a la confirmación empírica de equivalencia psicométrica. Tras revisar algunos de los errores y concepciones equívocas sobre la adaptación de tests, se concluye que estas afectan a las fases iniciales del proceso de adaptación y que por ello, es necesario evaluarlas y corregirlas antes de acometer la fase operacional. Los pasos previos a la etapa de confirmación empírica centran su interés en los contextos éticos y legales de la adaptación, la representación del constructo en la población destino, la importancia de una adaptación que tenga en cuenta la cultura, la gramática, la semántica, la adecuación lingüística y la psicométrica, y la conveniencia de una prueba piloto. Todos ellos

Tabla 6 Estadísticos descriptivos de los ítems de la prueba de razonamiento abstracto

Ítem	Media	Varianza	Índice discriminación	Cuartil de dificultad adaptada	Cuartil de dificultad original
Ítem 1	.86	.35	.53	1	1
Ítem 2	.93	.25	.45	1	1
Ítem 3	.76	.43	.61	1	1
Ítem 4	.82	.39	.67	1	1
Ítem 5	.75	.44	.36	1	1
Ítem 6	.86	.35	.44	1	1
Ítem 7	.71	.46	.24	2	2
Ítem 8	.74	.44	.39	2	2
Ítem 9	.66	.48	.44	2	2
Ítem 10	.68	.47	.58	2	2
Ítem 11	.55	.50	.34	3	2
Ítem 12	.33	.47	.34	3	2
Ítem 13	.63	.49	.41	2	3
Ítem 14	.72	.45	.56	2	3
Ítem 15	.51	.50	.43	3	3
Ítem 16	.41	.49	.43	3	3
Ítem 17	.36	.48	.44	3	3
Ítem 18	.26	.44	.54	4	3
Ítem 19	.18	.39	.48	3	4
Ítem 20	.12	.33	.31	4	4
Ítem 21	.12	.33	.28	4	4
Ítem 22	.13	.34	.28	4	4
Ítem 23	.08	.27	.28	4	4
Ítem 24	.13	.34	.35	4	4
Ítem 25	.04	.20	.11	4	4

Tabla 7 Matriz de correlaciones y estadísticos de las escalas parciales de la forma 2 de la batería de pruebas de razonamiento

	RA	RV	RE	RN	n	Media	Varianza	Consistencia interna
RA	-	-	-	-	25	12.34	24.20	.92
RV	.71**	-	-	-	25	12.93	20.73	.88
RE	.61**	.50**	-	-	20	9.03	13.68	.91
RN	.59**	.57**	.65**	-	19	6.72	18.02	.93
RM	.48**	.50**	.40**	.42**	25	8.72	9.16	.82

**Las correlaciones son significativas ($p < .01$).

RA: razonamiento abstracto; RV: razonamiento verbal; RE: razonamiento espacial; RN: razonamiento numérico; RM: razonamiento mecánico.

Tabla 8 Tabla resumen de procedimientos analítico-rationales

1. Contactar con el editor/constructor del test
2. Respetar las normas de uso y de investigación con seres humanos
 - a. Presentación del proyecto
 - b. Convenio de colaboración
 - c. Consentimiento informado
3. Recoger información sobre la variable/constructo que se ha de medir
 - a. Estudio de las redes nomológicas
 - b. Revisión bibliográfica
 - c. ¿Es el constructo equivalente en las poblaciones origen y destino?
4. Análisis de los materiales y formato de los ítems
 - a. La tarea y los materiales, ¿poseen el mismo nivel de familiaridad en ambas culturas?
5. Formar un grupo de expertos multidisciplinar que incluya especialistas en el área que se ha de medir, en la construcción de cuestionarios, en los idiomas y culturas implicados y que sea próxima a la población a la que va destinada la prueba
6. Adaptar el test
 - a. Llevar a cabo al menos dos traducciones hacia adelante
 - b. Verificar la calidad de la adaptación
 - i. Comprobar la equivalencia gramatical
 - ii. Comprobar la equivalencia semántica
 - iii. Comprobar la equivalencia cultural
 - iv. Comprobar la adecuación lingüística
 1. Registro lingüístico
 2. Comprobar características como legibilidad, longitud, dificultad
 3. Vigilar el formato y la apariencia
 - v. Comprobar que las instrucciones y el contenido sean similares a la prueba original y suficientes para una correcta comprensión
7. Pretest cognitivo
8. Muestreo. Contacto con los centros colaboradores y la administración de cuestionarios
9. Recoger datos de una muestra piloto
 - a. Controlar la administración
 - i. Verificar el tiempo de administración
 - ii. Verificar la correcta comprensión de la tarea
 - iii. Verificar la correcta comprensión de las instrucciones
 - b. Obtener índices psicométricos y descriptivos básicos
 - i. Calcular los estadísticos descriptivos y coeficientes de discriminación
 - ii. Comprobar que la dificultad de los ítems es similar en orden en la prueba original y en la prueba adaptada
 - iii. Comprobar con una pequeña muestra que el contenido de los ítems es comprendido de forma correcta y no da lugar a ambigüedades
 - iv. Llevar a cabo estudios exploratorios sobre la calidad de los datos: dimensionalidad, consistencia
10. Documentar las modificaciones sobre el test original

son aspectos que en ocasiones quedan relegados y cuya falta de consideración puede acarrear consecuencias de difícil solución.

Las pautas iniciales respaldan el establecimiento de estudios analítico-rationales y de evaluación de contenidos. Su finalidad es ofrecer evidencias que repercutan positivamente en la validación de los tests (Elosua, 2003) y a su vez, que ahorren dinero y tiempo al intentar controlar los errores más comunes que inciden sobre el proceso de adaptación. Es cierto, como señalan Malda, van de Vijver, Srinivasan, Transler, Sukumar y Rao (2008), que no existen requerimientos mínimos en la aplicación de las directrices previas a los estudios operacionales confirmatorios; sin embargo, cada vez es mayor el clamor que parte de la comunidad psicométrica en su defensa y aplicación (Muñiz et al., 2013).

El resultado de esta investigación ha permitido definir y sistematizar en diez puntos las fases analítico-rationales que es conveniente seguir en la adaptación de un test y en su verificación. Los puntos, resumidos en la tabla 8, ofrecen al investigador pautas concretas que pueden utilizarse como guía durante el complejo trabajo de adaptar un test.

El presente artículo ha ejemplificado las pautas iniciales en la adaptación de tests con la traducción al español de la BPR. El estudio de las variables cognitivas medidas por la BPR, la consideración de los aspectos legales y éticos en torno a las investigaciones con humanos, el trabajo de la comisión de expertos que tuvo en cuenta factores gramaticales, semánticos, culturales y de adecuación a la muestra, junto con los datos obtenidos en la prueba piloto avalan las conclusiones preliminares de que el nivel de equivalencia entre la prueba original y la prueba adaptada es elevado. Los resultados permiten continuar con el proyecto de adaptación con amplias garantías científicas y éticas.

Agradecimientos

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

Referencias

- Almeida, L.S. (2002). As aptidões na definição e avaliação da inteligência: O concurso da análise fatorial. *Paidéia, (Ribeirão Preto)*, 12(23), 5-17.
- Almeida, L.S., Guisande, M.A., Primi, R., & Lemos, G. (2008). Inteligencia y rendimiento escolar. *European Journal of Education and Psychology*, 1, 5-16.
- Almeida, L.S., & Lemos, G. (2006). *Bateria de Provas de Raciocínio: Manual Técnico*. Braga: Universidade do Minho, Centro de Investigação em Psicologia.
- Berry, J.W., Poortinga, Y.H., Segall, M.H., & Dasen, P.R. (1992). *Cross-cultural psychology: Research and applications*. Cambridge: Cambridge University Press.
- Binet, A., & Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année psychologique*, 11, 191-336.
- 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.

- Carroll, J.B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about 10 broad factors. En H. Nyborg (Ed.). *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. (pp. 5-21). Amsterdam, Holanda: Pergamon.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*, 1-22.
- Cattell, R.B. (1971). *Intelligence: Its structure, growth and action*. Boston, MA: Houghton Mifflin.
- Cheung, F. (2004). Use of Western and Indigenously Developed Personality Tests in Asia. *Applied Psychology An International Review*, *53*, 173-191.
- Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.
- DeMaio, T.J., & Rothgeb, J.M. (1996). Cognitive Interviewing Techniques: In the Lab and in the Field. En N. Schwarz & S. Sudman (Eds.). *Answering questions. Methodology for determining cognitive and communicative processes in survey research*. (pp. 177-196). San Francisco, CA: Jossey-Bass, Inc.
- Elosua, P. (2003). Sobre la validez de los test. *Psicothema*, *15*, 315-321.
- Elosua, P. (2005). Evaluación progresiva de la invarianza factorial entre las versiones original y adaptada de una escala de auto-concepto. *Psicothema*, *17*, 356-362.
- Elosua, P., & Iliescu, D. (2012). Tests in Europe. Where we are and where we should to go. *International Journal of Testing*, *12*, 157-175.
- Elosua, P., & López, A. (2007). Potential DIF sources in the adaptation of tests. *International Journal of Testing*, *7*(1), 39-52.
- Elosua, P., & Zumbo, B. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema*, *20*, 896-901.
- Elosua, P., & Muñoz, 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.
- Hambleton, R.K. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment*, *17*, 164-172.
- Hambleton, R.K. (2009). *International Test Commission Guidelines for Test Adaptation, Second Edition*. Comunicación presentada en el 11th European Congress of Psychology, Oslo, Noruega.
- Hambleton, R.K., Merenda, P., & Spielberger, C. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence Erlbaum Publishers.
- Horn, J., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. En D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues*. Nueva York, NY: The Guilford Press.
- Lemos, G., Abad, F.J., Almeida, L.S., & Colom, R. (2013). Sex differences on and non-gintellectual performance reveal potential sources of STEM discrepancies. *Intelligence*, *14*, 11-18.
- Malda, M., van de Vijver, F.J.R., Srinivasan, K., Transler, C., Sukumar, P., & Rao, K. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures. *Psychology Science Quarterly*, *50*, 451-468.
- Meuris, G. (1969). *Tests de raisonnement différentiel*. Bruselas, Bélgica: Editest.
- Muñoz, J., Elosua, P., & Hambleton, R.K. (2013). Directrices para la traducción y adaptación de los test. (2.^a ed). *Psicothema*, *25*, 149-155.
- Pike, K.L. (1967). *Language in relation to a unified theory of structure of human behavior*. La Haya, Holanda: Mouton.
- Primi, R., & Almeida, L.S. (2000). Estudo de Validação da Bateria de Provas de Raciocínio (BPR-5). *Psicologia: Teoria e Pesquisa*, *16*(2), 165-173.
- Skevington, S.M., Sartorius, N., Amir, M., & the WHOQOL-Group (2004). Developing methods for assessing quality of life in different cultural settings. *Social Psychiatry and Psychiatric Epidemiology*, *39*, 1-8.
- Sternberg, R.J. (2000). *The handbook of intelligence*. Nueva York, NY: Cambridge University Press.
- Van de Vijver, F.J.R., & Leung, K. (2011). Equivalence and bias: A review of concepts, models and data analytic procedures. En D. Matsumoto & F.J. R. van de Vijver (Eds.). *Cross-cultural research methods in Psychology* (pp. 17-45). Cambridge: Cambridge University Press.
- Van de Vijver, F.J.R., & Tanzer, N.K. (1997). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *47*, 263-279.
- Vernon, E. (1961). *The structure of human abilities*. Londres, Inglaterra: Methuen.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.