

RAE

1. **TIPO DE DOCUMENTO:** Trabajo de grado para optar por el título de Psicólogo(a)
2. **TÍTULO:** DISEÑO Y VALIDACIÓN DE UN MANUAL PARA LA ADAPTACIÓN DE PRUEBAS PSICOLOGICAS EN EL CONTEXTO COLOMBIANO.
3. **AUTORES:** Nicolás Yepes, Gabriela Flórez y Alejandro Barrero.
4. **LUGAR:** Bogotá D.C.
5. **FECHA:** Junio de 2015
6. **PALABRAS CLAVE:** Psicometría, instrumento, propiedades psicométricas, adaptación, traducción y manual.
7. **DESCRIPCIÓN DEL TRABAJO:** El objetivo de este proyecto es el diseño y validación de un manual para la adaptación de pruebas psicológicas en el contexto colombiano. Es una herramienta técnica que comprende los fundamentos psicométricos que contiene un instrumento. El resultado fue la estructura y la validez de contenido de las siguientes ocho fases; generalidades, autorización, contexto y lenguaje, traducción, validación por jueces, interpretación de los datos y análisis de los aspectos psicométricos. Obteniendo así una herramienta que responde a las necesidades actuales de psicólogos e investigadores que estén interesados en la adaptación de un instrumento.
8. **LÍNEAS DE INVESTIGACIÓN:** La investigación se enmarca desde dos clasificaciones, la primera de desarrollo tecnológico ya que toma los principios de la investigación empírica y sus hallazgos los traduce en herramientas prácticas para su utilización en campos aplicados. La segunda de tipo instrumental ya que se establece los contenidos y características de un manual de procedimientos de forma técnica, válida y confiable.
9. **METODOLOGÍA:** Comprende tres aspectos; a. unidad de análisis: son los componentes, pasos y conceptos que están implícitos en la estructura del manual. Los instrumentos son: cuadro matriz y el formato de validación de jueces. Por último, el procedimiento comprende a. Inicio: recopilación de la información, b. implementación: elaboración de las fases y juicio de un grupo de expertos y c. finalización: los resultados fueron sometidos al coeficiente de W Kendall y coeficiente de Laswhel además se realizó ajustes al manual según la evaluación cualitativa de los jueces.
10. **CONCLUSIONES:** La adaptación comprende un conjunto de pasos descritos en el manual contando con la evaluación de jueces expertos para así obtener la validez de contenido resultando una herramienta pertinente para los psicólogos. El proyecto reconoce la importancia de realizar adecuadas prácticas al adaptar una prueba, además fomenta la responsabilidad del manejo de instrumentos y la creación de pruebas psicológicas en el contexto colombiano.



Diseño y validación de un manual para la adaptación de pruebas psicológicas en el
contexto colombiano

Alejandro Barrero, Gabriela Flórez y Nicolás Yepes

Facultad de Psicología

Universidad de San Buenaventura Bogotá D.C.

Andersen Vera

2015

Tabla de contenido

Tabla de contenido	2
Resumen	3
Introducción	4
Marco Teórico	5
Justificación	25
Pregunta de investigación	26
Objetivo General	27
Objetivos específicos	27
Método	27
Instrumentos	28
Procedimiento	28
Unidad de Análisis	27
Aspectos éticos	28
Tipo de estudio	27
Resultados	29
Discusión	35
Referencias	39
Apéndice A	45
Apéndice B	46
Apéndices C (manual adaptación)	48

Resumen

El propósito del presente proyecto fue el diseño y validación de un manual para la adaptación de pruebas psicológicas en el contexto colombiano. Esta investigación está enmarcada dentro de los estudios de carácter instrumental y de desarrollo tecnológico, pues su objetivo fue la construcción de una herramienta técnica, comprendiendo los fundamentos psicométricos que contiene una prueba psicológica. Finalmente, el resultado fue la estructura, elaboración y la validez de contenido de las ocho fases descritas en el manual, obteniendo un producto para ser utilizado en estudios relacionados con la traducción, adaptación y estandarización de pruebas psicológicas.

Palabras clave: Psicometría, instrumento, propiedades psicométricas, adaptación, traducción y manual.

Summary

The intention of the present project is the design and validation of a manual for the adjustment of psychological tests in the Colombian context framed inside the studies of instrumental character, since his aim is the construction of a tool or technical instrument, by means of the analysis of the characteristics psychometric that contains an instrument. The results will be given in terms of the technical establishment of eight phases described in the manual, including his aim, the concept, the procedure and his applicability by means of the instructive one.

Key words: Psychometry, Psychological test, adjustment, psychometric properties, translation and manual.

MANUAL ADAPTACIÓN

La psicología es la disciplina que estudia el comportamiento humano y uno de sus objetivos es el diagnóstico y conocimiento de los rasgos o características de un individuo o grupo. Para lograr este propósito, la psicología hace uso de herramientas fundamentales como lo son la entrevista que brinda la información inicial en el abordaje de los fenómenos y problemáticas, así mismo permite hacer un filtraje de la información que se busca. En segunda instancia se encuentra la utilización de pruebas psicológicas, que también permiten conocer características relevantes de los sujetos.

La segunda herramienta mencionada se ha utilizado desde principios del siglo XX, siendo el enfoque cuantitativo quien más la utiliza, por su orientación a la predicción y control. En su conjunto, estas herramientas validan las hipótesis y miden las variables de interés por medio de puntuaciones, que pueden dar cuenta de las características de los evaluados (Coolican ,2005).

El uso de estas pruebas también permite obtener muestras de comportamiento de un dominio específico, construidas con un alto rigor metodológico es decir, si se rigen por una serie de reglas y particularidades fundamentadas principalmente por la psicometría, evidenciando una carga cultural propia de cada contexto (características sociales, económicas y culturales del país donde fueron creadas) (Cohen y Swerdlik, 2001).

Dentro de los países que se destacan en la construcción de pruebas de psicología, se encuentran Estados Unidos, España, México y Argentina, entre otros. Por su parte, para otros países como Colombia se ha identificado una baja producción y creación de pruebas, test, escalas o cuestionarios psicológicos, ya que los psicólogos e investigadores suelen adoptar pruebas ajenas al contexto, sin una adaptación previa ni una traducción, incumpliendo así con parámetros específicos que se ajusten al contexto e incurriendo en errores metodológicos (Hernández, Valencia y Rodríguez, 2003).

En internet se puede evidenciar la existencia de test de libre acceso, creados por personal no calificado, y que por lo tanto, no cumplen con estándares psicométricos mínimos para cualquier prueba, se identifica la validez y la confiabilidad, siendo utilizados indiscriminadamente. Así mismo, se utilizan instrumentos sin autorización de sus autores, haciendo modificaciones que afectan los procesos psicométricos de las

MANUAL ADAPTACIÓN

pruebas, trayendo consigo implicaciones legales para los psicólogos e investigadores (Vallo, 2010).

Otra situación que se identifica con lo descrito anteriormente, es la relacionada con los sujetos evaluados, a quienes constantemente se les debe dar claridad sobre extranjerismos que están presentes en las pruebas originales, generando confusiones al momento de responder. En este sentido, para afrontar los anteriores problemas expuestos que afectan e impiden realizar un proceso riguroso, resulta pertinente hacer una revisión de los contenidos que constituyen el quehacer de la psicometría.

Por tanto se puede decir que la psicometría es la disciplina encargada de entender el conjunto de métodos, técnicas y teorías implicadas en la medición de las variables psicológicas. Ésta, transforma los hechos en datos, con el fin de asignar valores numéricos a los sujetos, según sus respuestas. Desde el marco psicométrico se incluyen conceptos tales como validez, confiabilidad, estandarización, baremación, validación y el proceso de adaptación de pruebas psicológicas (Muñiz, 1998).

Así mismo, la psicometría está relacionada con la psicología experimental, al ser la base fundamental para iniciar el desarrollo de las pruebas psicológicas, consolidándose en países como Alemania y Gran Bretaña a finales del siglo XIX. Los primeros representantes fueron Wilhelm Wundt, Francis Galton y James Catell; este último inventó el término *prueba mental* y al igual que Galton, midió las funciones intelectuales mediante test de discriminación y tiempo de reacción (Ortiz, 2002).

Tales pioneros de la psicología, se interesaron por una formulación o medición aceptable de la inteligencia o el rendimiento intelectual, pero los primeros estudios realizados mostraban poca relación con las estimaciones y evaluaciones realizadas por profesores. Sin duda alguna, la consolidación de la medición se debe a Alfred Binet, quien en 1904 argumentó la necesidad de observar la ejecución de los individuos en los actos mentales complejos (Mora & Martin, 2007).

Desde esa época 1904, la psicología comienza la formulación, construcción y análisis de los test, mediante diversas teorías que guían el contenido del test de acuerdo con

MANUAL ADAPTACIÓN

avances estadísticos. Son estas teorías las que van a estimar las propiedades psicométricas y garantizar la fiabilidad y validez de las pruebas psicológicas.

Todas las teorías parten de la necesidad de encontrar un modelo estadístico que fundamentara las puntuaciones de los test y permitiera la estimación de los errores. Desde entonces, no sólo el juicio humano fue el determinante de los atributos psicológicos, sino que los instrumentos permitirían medir el comportamiento (Aragón, 2004).

Es por esto que para Avendaño y Medellín (2001) se necesita medir con precisión la conducta a través de modelos teóricos que permiten llevar a cabo todo proceso de medida, sentando así las bases para una realización adecuada. A continuación se resaltan dos modelos estadísticos que han tenido trascendencia en la construcción de las pruebas: la teoría clásica (TC) y la teoría de respuesta al ítem (TRI).

La teoría clásica (TC) se consolidó con los primeros trabajos de Charles Spearman en los años 1904 a 1913. Este psicólogo asumió que la puntuación de una persona obtenida en un test es su puntuación empírica, que se designa con la letra (x), y se conforma por dos componentes: la puntuación verdadera (v) y un error (e) que puede ser por causas externas (Muñiz, 2010).

De la misma forma, Spearman citado por Castelli (2007) añade dos supuestos al modelo. El primer supuesto explica que la puntuación verdadera de una persona es aquella que obtendría como media, si se le aplicara varias veces el test. El segundo supuesto entiende que el valor de la puntuación verdadera de una persona no tiene que ver con el error que afecta esa puntuación, ya que no hay una relación entre la puntuación y el alto o bajo tamaño de los errores.

La (TC) se identifica con tales supuestos y formula una definición de los test paralelos, siendo aquellos que miden lo mismo pero con distintos ítems. En suma, estos supuestos son la base de todas las formulas clásicas utilizadas habitualmente por los psicólogos, manejando los coeficientes de fiabilidad y validez para indicar que los test son precisos y con poco error de medida. Es por eso que antes de utilizar una prueba

MANUAL ADAPTACIÓN

psicológica bajo este modelo, se obtiene información sobre su confiabilidad siendo la consistencia de los resultados. Por otra parte la validez es el grado en el cual la prueba mide aquello para lo que fue diseñada (Aiken, 2003).

También es importante establecer los puntos centrales de la TC, los cuales explican el error de medición para así poder lograr la exactitud en la medida, ya que mientras este error sea mayor, será menor la confiabilidad del test. Para corroborar esta información, se utilizó el modelo de regresión lineal. Las características de esta teoría son: Los errores de medición son aleatorios, la media de error de medición es = 0, las puntuaciones verdaderas y los errores en diferentes pruebas no se correlacionan Gregory (2001).

La teoría clásica ha servido para el avance y desarrollo de pruebas psicométricas, pero a lo largo del tiempo ha sumado también detractores que han apoyado sus trabajos contradiciendo algunos de sus supuestos convencionalmente aceptados, además de hacer críticas a conceptos establecidos por ésta.

Para Hernández, Valencia y Rodríguez (2003), la definición de test paralelos es muy restrictiva, al evidenciar una dependencia muestral de las propiedades psicométricas de los tests, de las características de los ítems y de la puntuación observada de la longitud del test. Esta teoría ha acumulado en su recorrido seguidores y contradictores, pero siempre ha promovido el pensamiento del área psicométrica desde posiciones críticas que finalmente han permitido el desarrollo y sofisticación de sus propios supuestos y conceptos, como el caso de Gondra (1997) y su trabajo sobre el error de medida.

Según Gondra (1997) la TC especificó diversos aspectos a estudiar en el campo psicométrico, asumiendo el error de medida como parte de la construcción de las pruebas psicológicas, siendo importante controlar la mayoría de errores tanto estadísticos como ambientales. En este punto, se propone la teoría de los tests orientados al criterio, enfoque clásico propuesto por Lee J. Cronbach y sus colaboradores en el año 1963 como se citó en Gondra (1997) teniendo una gran importancia para el ámbito educativo. Además, este modelo fue relevante ya que utilizaron el análisis de varianza para la mayoría de sus cálculos y estimaciones.

MANUAL ADAPTACIÓN

En definitiva, la TC aportó diversas ideas como las anteriormente mencionadas, pero enfrentó un problema específico sobre la invarianza de las mediciones y las propiedades de los instrumentos de medida, apoyados en una suposición crucial sobre los errores no sistemáticos de medición. Todo esto, basado en que el error actúa como influencia aleatoria, o en otras palabras, el error de medición son sucesos que no se controlan del todo por el investigador sino que son situaciones ambientales o características de la persona a evaluar (Tornimbeni, Pérez y Olaz, 2008).

Por consiguiente, Muñiz, Elosua y Hambleton (2013) entienden que el test referido al criterio (TRC) no determina si la persona domina un campo de conocimiento, sino que evalúa el grado de conocimiento en un campo determinado. También exponen que los indicadores psicométricos clásicos no se ajustaban a la construcción de nuevos test, ni se podía calcular la fiabilidad y validez. Las causas por las que nació el test referido al criterio (TRC) fueron el interés por evaluar el nivel de habilidades de cada sujeto y la necesidad por conocer la eficacia de los programas educativos.

Prieto y Delgado (2003) señalan que la (TRC) ha propiciado el desarrollo de la psicometría en los siguientes ámbitos: Definir con claridad los objetivos o criterios de interés, potencia nuevas formas de evaluar la validez y la fiabilidad. Esta última es la medida en que los ítems son contestados correctamente.

Debido a las falencias que se presentaron dentro de estas dos teorías (TC) y (TRC), se comienza a analizar otro nuevo punto de vista, de ahí nace la teoría de la generalizabilidad (TG), solucionando algunos problemas que presenta la (TC), como el de la concepción unitaria e indiferenciada del error, cambiando el concepto tan limitado de confiabilidad por un concepto más amplio y reemplazando la postura de medidas paralelas por el de medidas aleatorias (Avendaño y Medellín, 2001).

Zúñiga y Montero (2007) afirman que este enfoque soluciona unos problemas del anterior modelo pero sigue exponiendo que las características de un instrumento dependen de la muestra, de la medida y del tipo de prueba utilizada. Por otro lado, su planteamiento se dirige a medir la confiabilidad de una prueba por medio de la cuantificación de cada una de las fuentes de variabilidad. Aquí el error se define como

MANUAL ADAPTACIÓN

condición utilizando el coeficiente de generalizabilidad, siendo la medida para estimar la confiabilidad.

En consecuencia, el psicólogo británico Charles Spearman citado por Cizek (2001) desarrolló un modelo de regresión lineal llamado coeficiente de correlación ρ , el cual expresa la relación lineal entre dos variables aleatorias: los valores próximos a 1 significan una correlación positiva y -1 indica una correlación fuerte-negativa. Spearman parte de la idea de que las puntuaciones de una persona varían de una prueba a otra, debido a las condiciones que se presentan en la situación de la prueba.

Ahora bien, la variabilidad de los ítems representa una fuente potencial de inconsistencia en la generalización, entre las que se define que:

- a) La fuente de variabilidad se encuentra en las diferencias sistemáticas entre las personas en el rasgo o constructo que se desea medir; es decir que la variabilidad entre los objetos de medida (normalmente las personas) se refleja en las diferencias de conocimiento, habilidades u otros atributos entre los examinados(as) (Brennan, 2001).
- b) La variabilidad es la diferencia en la dificultad de los ítems de la prueba. Algunos reactivos se consideran fáciles, intermedios o difíciles.
- c) La fuente de variabilidad se refleja en el nivel educativo y experiencias previas que las personas hayan tenido. Este emparejamiento entre las experiencias de una persona y un reactivo en particular, aumenta la variabilidad entre personas e incrementa la dificultad para generalizar, en términos del atributo específico que se desea medir.

Todo lo anterior, se suma a las diversas teorías que han surgido en el campo de la psicometría, como la teoría de respuesta al ítem (TRI), ya que fue un modelo estadístico que se centró en la información que otorga un determinado ítem. Sus principios fueron dados por Geog Rasch citado por Terrón (2010) centrándose en el análisis de respuesta a los test a nivel de cada ítem mediante rasgos latentes arrojados a través de modelos matemáticos, asumiendo que entre la respuesta de una persona a una prueba y el rasgo o habilidad existe una relación no lineal que se puede expresar en términos probabilísticos.

MANUAL ADAPTACIÓN

La teoría de respuesta al ítem (TRI), se refiere a las escalas unidimensionales, partiendo del supuesto de que cada ítem de la prueba debe medir una sola aptitud o rasgo. La TRI no contradice los supuestos de la teoría clásica, usando las herramientas de la estadística y las matemáticas para explicar el comportamiento de datos empíricos derivados de un instrumento psicológico. Cada ítem o pregunta puede ser modelado mediante una función matemática que se denomina curva característica del ítem (Rojas, Manrique y Gatica, 2004).

En efecto, la unidad básica es el ítem, no el test. En este sentido, las respuestas observadas dependen de la habilidad de cada persona o de las propias características de los ítems, exigiendo que todos y cada uno de los ítems midan una única dimensión, por lo que establece la curva característica del ítem (CCI), expresando la posibilidad de acertar un ítem en función de los valores en el rasgo. Por eso, la (CCI) no es la relación ítem-test sino ítem-rasgo, de manera que la probabilidad de acertar un ítem es independiente al acertar en otro requiriendo que los ítems sea dicotómicos o dicotomizados (Attoresi, Lozzia, Abal, Galibert y Aguerri, 2009).

Como se señaló anteriormente, la (TC) y la (TG) se basan en la distribución normal y la (TRI) utiliza diferentes funciones matemáticas. Algunos modelos usan la función de la ojiva normal y otros usan la función logística. Esta teoría a pesar de no ser reciente, ha tomado importancia en el campo de la psicometría, especialmente en la última década del siglo XX. En un caso concreto, para analizar el nuevo examen del estado realizado por el Instituto Colombiano para el Fomento de la Educación Superior (ICFES), se interpretó el uso de los puntajes según este modelo, cambiando los resultados y reflejando la competencia global del estudiante en el área medida (ICFES, 2000).

Teniendo en cuenta lo anterior, se reafirma la necesidad de reconocer los anteriores modelos teóricos que contribuyeron a la construcción rigurosa de las pruebas psicológicas así como diferentes aportes teóricos puestos en práctica en el quehacer profesional. Es necesario reconocer el aporte de autores como Galton, Pearson y Weldon, fundadores en 1901 de la revista *Biometrika* que desde entonces publica trabajos relacionados con la psicometría y la psicología. Igualmente, James Cattell,

MANUAL ADAPTACIÓN

fundó la Psychological Corporation para la producción industrial de tests en 1922, y Thorndike fundó en Estados Unidos en 1936 la revista *Psychometrika*, siendo de gran relevancia para la Psicometría (Arias, 1995).

En resumen, se entiende que los diferentes modelos teóricos son parte de la construcción de un instrumento de medición, ya que estas reglas sirven para realizar un proceso organizado, así como para clasificar y cuantificar los datos de la variable a medir. Son estándares internacionalmente aceptados, sirven de guía para desarrollar las directrices generales para la construcción y adaptación de una prueba garantizando que se cumplan los requisitos científicos en cualquier test de evaluación psicológica (Carretero y Pérez, 2007). Los contenidos que proporcionan un rigor científico a una prueba como la utilización de un procedimiento sistemático y estandarizado, se orientan hacia los siguientes conceptos: validez, confiabilidad, estandarización, baremación y validación, entre otros.

Históricamente, uno de los primeros usos de las pruebas fue la evaluación de individuos en determinadas áreas de contenido, y por ello se comparaba el contenido de esas categorías de pruebas con el del área que pretendían probar. Luego, el énfasis recayó en la predicción, por lo que ahora existen dos tendencias: una hacia el fortalecimiento de la orientación teórica y la otra hacia una estrecha vinculación entre la teoría y la verificación psicológicas mediante la comprobación empírica y experimental de las hipótesis (Argibay, 2006).

Uno de los conceptos más relevantes dentro de las pruebas psicológicas es la validez, definida por Aragón (2004) como el grado en que un instrumento mide lo que realmente se propone. Por lo que se refiere a la adecuación, significación y utilidad de las inferencias específicas que se derivan a partir de las puntuaciones de los test, por eso no se validan los instrumentos sino la interpretación de los datos. Se refiere a lo que mide una prueba y no puede expresarse en general sino que debe consignarse el uso particular para el que se plantea utilizar el instrumento. Todos los procedimientos utilizados para determinar la validez se interesan en las relaciones entre ejecución en las pruebas y otros

MANUAL ADAPTACIÓN

factores observados independientemente de las características de la conducta considerada.

Igualmente, se refiere a que es un juicio evaluativo global, constituido por la evidencia empírica y los supuestos teóricos que respaldan la interpretación y las acciones con base en los puntajes de las pruebas, abarcando no sólo los ítems sino la forma de responder de las personas dentro de un contexto determinado. No se evalúa la prueba sino las inferencias tales como el significado o utilidad, a partir de las inferencias específicas de las puntuaciones de los test de acuerdo con el responsable de este proceso (Prieto y Delgado, 2010).

La validez se distingue a partir de dos importantes consecuencias: a) el responsable de la validez de una prueba ya no es solo su constructor sino también el usuario, b) la validez de una prueba no se establece de una vez por todas sino que es resultado de los supuestos teóricos vistos desde un desarrollo evolutivo que comprende tanto los aspectos experimentales como estadísticos (Messick, 1995).

Por consiguiente, los test arrojan diversas consecuencias especialmente defendidas por Zwick (2000), quien integra un componente pragmático que enfatiza en: a) la importancia de la relación entre la connotación teórica y las connotaciones prácticas atribuidas a las puntuaciones; b) la necesidad de valorar la relevancia y la utilidad de las puntuaciones en cada uno de los usos propuestos; c) la necesidad de conocer y controlar consecuencias sociales del uso propuesto. Con ello, se busca un equilibrio entre la finalidad del test y los efectos derivados de su uso.

Específicamente el instrumento debe ser consistente y confiable ya que la validez es el aspecto que determina si las mediciones psicológicas realmente evalúan un fenómeno específico que se estudia. En psicología, este aspecto tiene gran importancia ya que a menudo resulta difícil coincidir en lo que significa realmente el constructo a medir porque no son manejables o accesibles como lo puede ser en la física o la química (Coolican, 2005).

MANUAL ADAPTACIÓN

Resumiendo, existen tres tipos de validez, de acuerdo a Elosua (2003): constructo, criterio y contenido. La primera evalúa las cualidades psicológicas, la segunda es la medida obtenida por un sujeto y la última se refiere al conjunto de reactivos que representan un contenido teórico, de conocimientos y habilidades. En el análisis de contenido evalúa las relaciones entre el constructo y el contenido del test; el objetivo es garantizar que la muestra sea relevante y representativa, teniendo en cuenta la definición del dominio, el estudio de su representación y relevancia.

Sireci y Zenisky (2006) definen el dominio como aquel que especifica las áreas que se pretenden evaluar. La representación y la relevancia en la evaluación de cada uno de los ítems en función de la definición; se debe agregar que el estudio de factores contextuales evita el uso sesgado del lenguaje o los significados que se exponen dentro del formato de los ítems incluyendo las instrucciones, la administración y la corrección de la prueba entre el examinador y el examinado, obligando a evitar fuentes de error.

La validez de contenido, se define como el grado en que los ítems representan el contenido que tratan de evaluar y se realiza a través de juicios de expertos; es principalmente utilizado con test de rendimiento, test educativos y test referidos al criterio. Se utilizan procedimientos tales como: cálculos de descriptivos (calcula la media y la desviación típica de todos los ítems) e índice de validez de contenido (según una fórmula matemática se valora las puntuaciones positivas y los ítems que serán eliminados) (Elosua, 2003).

Por otra parte, la validez de criterio es el grado en que el test correlaciona con variables ajenas al test (criterio), ya que es un indicador de aquello que el test pretende medir y se denomina coeficiente de validez a la correlación con un criterio externo. Dentro del concepto cabe distinguir los tipos de validez: validez externa e interna, validez concurrente y validez predictiva. La primera es la correlación del test con respecto a un criterio externo. Validez interna es cuando se correlaciona un test con otro que tenga validez reconocida y mida el mismo rasgo (Castejon, 1997).

En conjunto con la validez de contenido y criterio, se encuentra la validez de constructo y es definida por Pérez, Chacón y Moreno (2000) como algo más complejo,

MANUAL ADAPTACIÓN

al referirse al grado en que el instrumento de medida cumple con la hipótesis que se espera. Es la principal fuente de validez porque integra las consideraciones de validez de contenido y de criterio en un marco común para probar las hipótesis acerca de relaciones teóricamente relevantes. Es conceptualizada como un proceso científico de contrastación de hipótesis, donde se estudia tanto lo empírico como los juicios racionales y las hipótesis serían las inferencias realizadas a partir de las puntuaciones de los tests y la validación o acumulación de evidencias que apoyen dichas inferencias.

Lo cierto es que la validez es un proceso unitario, refiriéndose siempre al grado en que esa evidencia apoya las inferencias que se hacen a partir de las puntuaciones (APA, AERA, NCME, 1985, citado en Gómez, Hidalgo y Muñiz, 2002). Una variedad de evidencias pueden obtenerse de las puntuaciones producidas por un test, existiendo muchas formas de acumular evidencia para apoyar una inferencia específica tomando en cuenta la muestra y el contexto de aplicación.

Otra característica de una prueba es saber si fue diseñada realmente para cierta variable psicológica, lo cual conduce a contemplar el concepto de confiabilidad. Esta es una propiedad que se refiere al atributo de consistencia de medición, abarca desde la consistencia mínima de una medición hasta la casi perfecta respetabilidad de los resultados. Desde el planteamiento que propone Aiken (2003) la confiabilidad de un test es la precisión con que el test mide lo que mide, en una población determinada y en las condiciones normales de aplicación.

Los sinónimos de este término son la consistencia, replicabilidad y fiabilidad. Lo primero genera la misma puntuación o similar en el caso de un individuo, "en cuanto a la calificación tiende a replicarse o a reproducirse con cierto margen de error. Y la fiabilidad depende de que la prueba genere en buena medida la misma puntuación del individuo" (Hogan, 2004, p.87).

Por otra parte, y tomando como argumento lo expuesto Bonilla (2006) define la confiabilidad como la consistencia del proceso de medición o de los resultados, convirtiéndose en una condición necesaria pero no suficiente para que un instrumento tenga validez, lo cual expresa que un instrumento puede ser confiable sin tener validez.

MANUAL ADAPTACIÓN

También hay que reconocer los errores constantes y errores sistemáticos. La confiabilidad no explica errores constantes sino los errores no sistemáticos.

Es importante comprender la confiabilidad vista desde la teoría clásica, ya que se basa en la idea de que las puntuaciones de prueba son el resultado de la influencia de dos factores: el primero, contribuye a la consistencia como los atributos del individuo que el examinador intenta medir. El segundo, contribuye a la inconsistencia y abarca características o situaciones ambientales que afectan las puntuaciones de la prueba, generando imprecisiones en la medición.

En cuanto a las inconsistencias de una prueba psicológica, se entienden dos errores: los que no pueden ser controlados y los errores sistemáticos que son controlables y tienen alguna explicación por una fuente de variación sistemática (Polit y Hungler, 1997). Este último se refiere al error como todo dato diferente a la puntuación verdadera. Los errores que se han identificado son la selección de reactivos, la aplicación de la prueba, la calificación de la misma y los errores sistemáticos de medición.

El primero se refiere al conjunto particular de preguntas elegidas para una prueba, pudiendo ser o no equitativo para todos los individuos. Es prudente tener en cuenta que una prueba es una muestra y no una totalidad del conocimiento o la conducta de una persona (Gregory, 2001). Durante la aplicación de la prueba hay condiciones ambientales que ejercen influencia desfavorable en la precisión de la medición tales como la temperatura, iluminación, ruido, el propio evaluador, ansiedad, motivación, atención y nivel de fatiga del individuo. Por último, la calificación de la prueba debe contar con criterios definidos para las respuestas que se dan a cada una de las preguntas.

Para Aliaga (2006), el error es cualquier efecto irrelevante para los fines o resultados de la medición que influyen sobre la falta de confiabilidad de tal medición. Existen dos tipos de error: a) Error constante (sistemático), en el cual las mediciones que se obtienen con una escala son sistemáticamente mayores o menores que lo que realmente deben ser; b) Error causal, donde las mediciones alternativamente mayores o menores de lo que realmente deben ser.

MANUAL ADAPTACIÓN

Según el punto de vista de Terrón (2010), visto desde la teoría clásica, los errores causales o no sistemáticos de medición actúan como influencias aleatorias. Debido a esto, tienen la misma posibilidad de ser positivos o negativos y por tanto tendrán un promedio de cero, pero si estos errores se relacionan con otra puntuación serían sistemáticas y no aleatorias.

Es importante aclarar que los científicos han creado convenciones especiales para el coeficiente de confiabilidad y para el error estándar de medida. Estas convenciones han sido de gran utilidad para los investigadores y profesionales, pues facilitan la rápida ubicación de un problema presente en el instrumento de medición. El coeficiente de confiabilidad es identificado como (RXX) y el error estándar de medida como (EEM) (Aliaga, 2006).

Vale la pena precisar que la confiabilidad no es lo mismo que estabilidad porque se entiende que la falta de confiabilidad es resultado de una serie de errores de medición producidos por estados internos temporales, como la baja motivación o disposición de los sujetos evaluados y condiciones externas como un ambiente incómodo y con distracciones (Aiken, 2003).

Es necesario mencionar que cuando una prueba que tiene confiabilidad nula (error de medición de gran tamaño) producirá un coeficiente cercano a (0), mientras que una prueba completamente confiable (sin error de medición) producirá un coeficiente de confiabilidad de (1.0). Entre más cercano sea el valor del coeficiente a (1.0) será mejor ya que este es un índice de la consistencia real de las puntuaciones obtenidas (Gregory, 2001).

Dentro de la confiabilidad se debe revisar qué tan alto es el coeficiente de confiabilidad. Autores como Bonilla (2006) propone un coeficiente de .70 ideal para diferenciar grupos, y de .85 para diferenciar entre individuos. El coeficiente de confiabilidad, es un coeficiente de correlación entre dos grupos de puntajes e indica el grado en que los individuos mantiene sus posiciones dentro de un grupo.

MANUAL ADAPTACIÓN

Bonilla (2006) propone cuatro métodos para obtener el coeficiente de confiabilidad:

- a) Método de estabilidad o método test-retest: Se debe aplicar una prueba dos veces a un mismo grupo y se calculan las correlaciones entre los dos conjuntos de puntuaciones. Es necesario que pase un tiempo prudente entre la aplicación de la prueba y la post prueba, para evitar que haya recuerdos de esta. El resultado obtenido se le denomina coeficiente de estabilidad ya que sirve para conocer si las mediciones son estables durante la primera aplicación del test.
- b) Método de las formas equivalentes: Se aplican dos formas equivalentes o paralelas del test al mismo grupo de individuos, y las dos series de puntajes resultantes se correlacionan con el coeficiente de Pearson.
- c) Método de estabilidad y equivalencia: Se hacen dos grupos. Al primero se le entrega el formato A y seguidamente se entrega el formato B. Al segundo, se entrega primero el formato B y seguidamente el formato A.
- d) Método de la división por mitades emparejadas "Split-half". Se aplica el test una sola vez a una muestra. La prueba se puede dividir en dos mitades dependiendo de la cantidad de ítems o se puede dividir en números pares e impares. Luego, se obtiene las puntuaciones resultantes. Se aplica la fórmula de Sperman-Brownd (Bonilla, 2006).

Por otra parte, es necesario incluir el concepto de estandarización para la adaptación de un instrumento psicológico. Su propósito es determinar la distribución de puntuaciones crudas en la muestra (grupo norma) mediante instrucciones estándar de aplicación y calificación. "Existen instrucciones propias de cada prueba que indican la posición de una persona con respecto a la distribución de las calificaciones obtenidas por sujetos de la misma edad cronológica, grado, nivel socio-económico, sexo u otras características"(Aiken, 2003,p.268).

Otros autores conciben la estandarización como el proceso uniforme de aplicación, calificación e interpretación de un instrumento psicométrico y el cumplimiento de unas determinadas normativas para que este cumpla con las condiciones de validez cuando se utilice en una población amplia y representativa (Aliaga, 2006).

MANUAL ADAPTACIÓN

La forma más adecuada de estandarizar una prueba es empezar por categorizar o estratificar la población de una serie de variables demográficas, relacionadas con las calificaciones de la prueba. Cuando se emplea el muestro aleatorio estratificado se reduce la probabilidad de elegir una muestra atípica o sesgada, obteniendo una base mejor para interpretar las calificaciones de la prueba (Aiken, 1998).

Otra forma de selección de una muestra para la estandarización, es conocida como el muestreo por grupos, la cual consiste en dividir una región geográfica designada o alguna entidad relevante en bloques o grupos. Se elige al azar un porcentaje especificado de los grupos y dentro de cada uno se selecciona aleatoriamente una cantidad determinada de subunidades (escuelas, residencias entre otros). Ya en el último paso se administra la prueba a cada persona por subunidad o a una muestra aleatoria con características establecidas (Aiken, 2008).

Hogan (2004) propone una estrategia para elegir una muestra para la estandarización con individuos y con los reactivos, denominada muestra de reactivos, la cual consiste en aplicar distintas muestras de reactivos a diferentes personas seleccionadas al azar. En esta muestra se realiza un análisis de reactivos y se determinan normas basadas en calificaciones de muestras representativas.

De acuerdo con Aiken (1998) dentro de la estandarización se encuentran; las normas percentilares que corresponde a una tabla de percentiles, donde se encuentran las puntuaciones crudas, estas se transforman en percentiles y el porcentaje del grupo de norma inferior a una calificación en particular es el rango percentilar de dicha calificación. Además las tablas de normas de rangos percentilares suelen estar incluidas en los manuales adjuntos al instrumento psicométrico.

Después de haber desarrollado el concepto de estandarización y de resaltar la importancia dentro del accionar de la psicometría, se procede a describir dos procesos fundamentales: El desarrollo y validación de instrumentos. Para empezar, la validación hace referencia al proceso de recogida y evaluación de evidencia que sustenta el uso de un instrumento para determinado propósito.

MANUAL ADAPTACIÓN

De manera semejante, para Sireci y Padilla (2014 p.97) la definición del concepto de validación es la siguiente: “La validación es el proceso de aportar evidencias de que las evaluaciones mediante test y cuestionarios cumplen adecuada y apropiadamente los objetivos para los que se elaboran”. Sin embargo, este concepto suele confundirse con el concepto de validez. Se ha podido constatar que no existe un consenso entre los investigadores. La validación es importante en la medida que se hace necesario revisar las cualidades psicométricas del instrumento adaptado (Sireci, 1999, en Rodríguez, Jiménez & Caicedo 2007).

En consonancia con lo anterior, el presente trabajo toma la validez como "un juicio evaluativo global del grado en el que la evidencia empírica y la lógica teórica apoyan la concepción y conveniencia de las inferencias y acciones que se realizan basándose en las puntuaciones que proporcionan los test u otros instrumentos de medida" (Messick, 1989 en Barraza 2007).

Retomando el concepto de validación, la Asociación estadounidense de Psicología (1999) propone cinco fuentes de evidencia que pueden ser usados para la evaluación e interpretación de los resultados arrojados por el test con un propósito particular (Sireci y Padilla, 2014). La primera de estas consiste en las variadas conceptualizaciones de la evidencia de validación, basadas en el contenido de tests y describe diferentes tipos de estudios que pueden ser conducidos para proveer tal evidencia. La segunda fuente de evidencia se basa en la estructura interna, que explica diferentes abordajes estadísticos (Sireci y Faulkner-Bond, 2014).

La tercera fuente es aquella que se basa en las relaciones con otras variables. La cuarta fuente entrega evidencia a través de las consecuencias de la aplicación de pruebas, que se divide en consecuencias intencionales y no intencionales de un determinado programa de pruebas (Lane, 2014). La quinta y última fuente para encontrar evidencia de validez es una de las más difíciles para recopilar y analizar, y es aquella basada en procesos de respuesta (Sireci y Faulkner-Bond, 2014). El proceso de validación, tal como se ha mencionado anteriormente, cumple un rol clave para poder llevar a cabo debidamente un proceso de adaptación de instrumentos psicométricos.

MANUAL ADAPTACIÓN

Otro de los conceptos sustanciales en esta tarea es el referente a la baremación, este es un proceso de transformación de puntuaciones brutas o directas en puntuaciones derivadas (percentil y calificación estándar), resultado de una operación aritmética. La puntuación directa de un individuo en un test no es directamente interpretable, sino se relaciona con los contenidos incluidos en el test o al rendimiento de las demás personas que comparten el grupo normativo (Sireci y Padilla, 2014).

En consecuencia, los baremos consisten en asignar a cada probable puntuación directa un valor numérico (en determinada escala) que informa sobre la posición que ocupa la puntuación directa (y por tanto la persona que la obtiene) en relación con los que obtienen las personas que integran el grupo normativo donde se bareman las pruebas (Abad, Garrido, Olea y Ponsoda, 2006).

Otros autores también han contribuido al desarrollo del concepto, argumentando que los baremos son en esencia una tabla numérica que se utiliza para contrastar el resultado o puntaje directo (P.D.) que obtiene una persona en un cuestionario con el grupo al que pertenece (Espinosa, 2008). Por ejemplo, si el sujeto es un pandillero de los suburbios de Nueva Jersey, el baremo que se va a utilizar tiene que haber sido construido con una muestra de pandilleros de Nueva Jersey. Si, se compara con uno construido con pandilleros de las favelas de Río de Janeiro, la conclusión será errónea.

Un test psicológico puede presentar una o varias tablas de baremos. En ellas están presentados los puntajes directos y en los extremos derecho e izquierdo se observan una o dos escalas numéricas que sirven para transformar el P.D. en una misma escala para que sea posible comparar y analizar todos ellos, ya que cada prueba tiene un número distinto de ítems y por lo tanto, entre P.D. no se pueden así comparar. Para el caso de adaptación de una prueba a un nuevo contexto, es necesario construir de cero los baremos, de los cuales a continuación se nombran algunos (Saldarriaga, Jaimes, Aguilar Y Merino, 2002).

Baremos cronológicos: Edad Mental y Cociente Intelectual. Centiles o percentiles. Puntuaciones típicas: estándares, normalizadas, escalas T y D. Estaninos o eneatis. Los baremos se presentan generalmente en las siguientes escalas: Percentiles (P): va de 1

MANUAL ADAPTACIÓN

a 99. Centiles (P c): va de 1 a 99. Decatipos (DE): va de 1 a 10. Eneatipos (EN): va de 1 a 9. Desviación típica (S): va de 1 a 99. (Trabaja con una M= 50 y D.S.= 20). Lo más común en las pruebas que se encuentran disponibles a nivel comercial es encontrar baremos realizados en escala de centiles o estatinos (Abad et al, 2006).

Para las escalas que van de 1 a 99, los puntajes promedio o normales para la edad y nivel educativo del sujeto, están entre 40 y 60. Para los decatipos, los puntajes normales o promedio están entre 4 y 7. Para los eneatipos, los puntajes promedio o normales oscilan entre 4 y 6.

Los anteriores conceptos explican lo más relevante para iniciar un proceso de adaptación siguiendo los lineamientos de la Psicometría conceptos como la confiabilidad, validez, baremación estandarización entre otros, sirven para aplicar una prueba psicológica a un nuevo contexto y población.

Con todo lo expuesto hasta ahora, es importante hacer referencia a la conceptualización de la adaptación de pruebas, la cual es una práctica que abarca el estudio de diferentes tipos de pruebas, áreas de conocimiento y distintos idiomas. (Geisinger en 1994 citado por Sireci, 1999). Cabe resaltar que el término adaptación es preferible al de traducción, pues la adaptación va más allá de una simple traducción literal siendo un proceso más flexible.

La adaptación implica no sólo la traducción de un test, es un procedimiento riguroso que consiste en tomar una prueba disponible en un idioma y validarlo para usarlo en otro país con una cultura e idioma diferente al original. “Actualmente se identifica que los instrumentos de medición psicológica regularmente son extranjeros, generando la necesidad de realizar estandarizaciones que garanticen la fiabilidad de los resultados obtenidos con el mismo instrumento aplicado a otro contexto” (Zarate y Castro, 2009 p.182).

Así mismo, la adaptación es de gran interés debido a la aplicación masiva de pruebas, la comparación de resultados en estudios transculturales y en el estudio de temáticas como la calidad de vida y la investigación de mercados, entre otros (Hambleton y Li,

MANUAL ADAPTACIÓN

2004). Así mismo, se destaca el uso de pruebas adaptadas como un aspecto valioso, ya que permite generalizaciones mayores, como también comparaciones entre poblaciones, a pesar de las dificultades que puedan presentarse en el proceso (Chang, 2001).

Dada la extensión de este proceso de adaptación, se identifican algunas consideraciones, antes de adaptar una prueba. Hambleton, Merenda y Spielberger (2004) identifican como primera aseveración el alcance de preservar la equivalencia psicológica de una prueba en dos o más idiomas y grupos culturales. Posteriormente, se pueden requerir modificaciones a los formatos de presentación de los ítems y revisiones al tema que pretende medir la prueba, ya que son aspectos que sin duda afectan de una u otra manera la validez de los resultados obtenidos.

Un caso cotidiano que afecta la validez y confiabilidad de la prueba, es la aplicación de pruebas psicológicas extranjeras a la población colombiana. Debido a esta situación, se evidencian serias dificultades en el proceso de adaptación de pruebas, puesto que se encuentran errores lingüísticos e imprecisiones culturales y sociales, entre otros. Según Jaimes (2000), un adecuado proceso de adaptación se realiza a través de tres fases: a) definición de la población y selección de una muestra, b) traducción y adaptación de los ítems de la prueba c) elaboración de baremos.

Con relación a la primera fase, Sattler (1993) afirma que la muestra debe reflejar características demográficas como edad, escolaridad, sexo, región geográfica, nivel socioeconómico y origen étnico. Y con respecto a la adaptación de los ítems, el mecanismo que más se utiliza es la validación por expertos o profesionales con conocimiento en psicometría.

También se reconocen otros tipos de fases, que "establecieron tres niveles de adaptación de las pruebas psicológicas. El primero corresponde al de la aplicación, este es, la simple y llana traducción de un test de un idioma a otro. Este método asume la equivalencia de constructo. Desafortunadamente, es el método más utilizado en todo el mundo. Como se indicara anteriormente la sola traducción de una prueba no indica ningún nivel de equivalencia entre ambas versiones de la misma" (Van de Vijver y Leung 1997 en Mikulic 2007, p.42).

MANUAL ADAPTACIÓN

Un ejemplo de este proceso se evidencia en un estudio para evaluar la personalidad en una población con discapacidad visual de Bogotá, cuyos objetivos generales fueron adaptar el cuestionario de personalidad 16 PF- Forma C a un grupo de personas con discapacidad visual. Teniendo en cuenta las fases anteriores, se describe el contenido del tema, se reconoce la necesidad de construir baremos piloto, e igualmente se desarrolla la evaluación de la validez y confiabilidad implementando métodos de respuesta que combinen con técnicas como el uso de dispositivos de hardware (Jaimes, 2000).

Los resultados de la investigación anterior permiten reconocer que la adaptación de un test es un proceso mucho más complejo que la simple traducción de ese test en un idioma diferente. Se entiende que implica considerar no sólo las palabras utilizadas al traducir la prueba sino también las variables culturales involucradas. Es por eso que los investigadores se deben asegurar de que el proceso de adaptación reconozca las diferencias lingüísticas y culturales de dos o más poblaciones y basarse en las versiones adaptadas del test (Fernández y Scheffel, 2003).

En el proceso de traducción existen dos métodos comunes: la traducción directa y la inversa. La primera se refiere a la traducción del test desde el idioma original al nuevo idioma por parte de un grupo de personas, juzgando posteriormente la equivalencia entre las dos versiones. A la inversa, significa una traducción desde el idioma original al nuevo idioma, donde luego un segundo grupo realiza una traducción al idioma natal para poder determinar su equivalencia (Tornimbeni, et al, 2004). En consecuencia, una vez traducida la prueba, es necesario establecer si el test es equivalente al original. Hamblenton (2000) señala tres métodos:

a) Administración de la prueba original y de la prueba traducida a sujetos bilingües: Aplicar la prueba original y su traducción a sujetos que hablan ambos idiomas. Esto genera una desventaja ya que es difícil sostener si los sujetos son competentes en ambos idiomas.

b) Administración de la versión original y su traducción inversa a monolingües en el idioma original. Administrar la versión original del WAIS-III y la versión por traducción inversa. Existe una gran desventaja ante la posibilidad de una falta de independencia

MANUAL ADAPTACIÓN

entre los puntajes obtenidos, por un efecto de aprendizaje luego de administrar la primera versión de la prueba.

c) Administrar la versión de alguna prueba en inglés a sujetos cuyo idioma natal es el inglés y aplicar la versión en español a sujetos que hablen español. Una posible dificultad es asumir que los sujetos de ambas muestras poseen una habilidad comparable pero se resuelve con análisis desarrollados a partir de la teoría de respuesta al ítem.

El análisis puede desarrollarse dentro de la teoría de respuesta al ítem, y se centra en las probabilidades que tiene una persona en determinado nivel de habilidad de contestar un ítem de manera correcta. Sin embargo, si se utilizan o se analizan los comportamientos de los ítems desde la teoría clásica, se usa el método Delta de Angoff y Ford, sirviendo para detectar el sesgo basado en las diferencias o en valor de la dificultad del ítem, es decir este método elimina aquellos ítems que representan un funcionamiento diferencial dentro del instrumento psicológico (Bond, 2001).

Una vez obtenidos los datos a partir de alguno de los tres métodos, existen varias posibilidades estadísticas para ser analizado. La primera probabilidad es identificar el funcionamiento diferencial de ítems (FDI), debido a que se comportan de manera diferente por diversas muestras transculturales. En la investigación de Tanzer (1995), se compararon estudiantes australianos con singaporeanos, se administraron dos cuestionarios sobre auto-concepto académico en lectura y en matemáticas, dando como resultado que cuando las escalas de los ítems de competencia se trabajaban de forma individual, arrojaban grandes diferencias entre ambos grupos culturales.

La segunda alternativa es la traducción agregando la transformación, adición o substracción de algunos ítems de la escala original. Entre traductores se encuentran diferencias importantes en el significado de los ítems y en conjunto llegan a un acuerdo sobre los puntos discordantes. Posteriormente, el comité de revisión evaluará si el significado de los ítems re-traducidos coinciden con los de la escala original (Sánchez y Echeverri, 2004).

MANUAL ADAPTACIÓN

Un ejemplo dado por Bond (2001), al realizar la baremación del WISC-III, encontró que el nivel de dificultad original de los ítems pertenecientes a sub-escalas; comprensión, vocabulario e información no aplicaban para la población argentina, por lo cual se propuso un nuevo ordenamiento de los ítems, implicando una transformación a la escala original.

Finalmente, la opción ensamble puede emerger al momento de adaptar un instrumento de evaluación psicológica. En efecto el test patrón ha sido alterado de tal forma que se ha cambiado y se ha convertido en un nuevo instrumento. Esto suele ocurrir cuando numerosos ítems del test original son claramente inadecuados para representar el constructo a medir; un ejemplo es el de pruebas de denominación confrontacional utilizados en neuropsicología, donde se emplean láminas con dibujos de objetos que el evaluado debe nombrar. Estos objetos se presentan con distinta frecuencia de observación en la vida diaria de un individuo y por este motivo van a variar significativamente de una cultura a otra (Aiken, 1995).

Con este escenario y después de hacer la descripción de los distintos niveles en que se manifiestan los problemas de adaptación, es necesario buscar soluciones que permitan y garanticen dar respuesta a estas situaciones que distorsionan el proceso de adaptación. Por esto, se considera necesaria la elaboración y creación de pruebas en el contexto colombiano y la creación de un manual que establezca de manera clara las reglas psicométricas a seguir a la hora de adaptar una prueba al contexto colombiano.

Considerando las situaciones anteriormente mencionadas acerca de la utilización de las pruebas extranjeras aplicadas en otros países se comprende que el proceso de adaptación es muy relevante debido a las dificultades que se presentan cuando una prueba no es adaptada. Por consiguiente se concluye que la investigación a desarrollar es relevante por tres hechos fundamentales: Primero, por la necesidad de que en el país exista una mayor tendencia a la creación de pruebas psicométricas y la adaptación de instrumentos extranjeros; segundo, por la falta de manuales adecuados al idioma y a la cultura local; y tercero, por la propuesta de masificar la utilización de instrumentos

MANUAL ADAPTACIÓN

novedosos que al no estar adaptados a las exigencias del medio, no han podido ser utilizados.

Este proyecto resulta pertinente porque responde de manera directa a las necesidades que se tienen actualmente de tener una manual que, en primer lugar, se constituya en una guía para los investigadores que estén interesados en la adaptación de un instrumento o que por razones afines a la investigación así lo requiera. En segundo lugar, porque será una guía común a todos los investigadores que estén en proceso de adaptar un manual, permitiendo entender los procesos que se llevaron a cabo, y adicionalmente, posibilita el desarrollo de correcciones en falencias o adiciones que no se tuvieron presentes en la adaptación o traducción de un manual. De esta forma, se está dando respuesta de forma eficaz a la situación y contexto del país en esta materia.

Finalmente, tiene un impacto de escala progresiva, ya que se monitorearán a través de mediciones en el cercano como mediano y largo plazo. Se espera, en todo caso, que sea alto y genere eco dentro del círculo de la disciplina psicológica y sus diversos y plurales ámbitos. Las expectativas que se tienen para el campo psicométrico son prometedoras, pues se cree firmemente en la proliferación de este tipo de trabajos y tentativas con el fin de alcanzar en el transcurso del tiempo una gran explosión y repercusión para el mismo.

También engrosará la teoría psicométrica y sirve como herramienta de primera mano para los psicólogos que requieran seguir una guía para la adaptación de los instrumentos.

Pregunta de investigación

¿Es valida la estructura y contenidos del manual para la adaptación de pruebas psicológicas para el contexto colombiano?

MANUAL ADAPTACIÓN

Objetivos

General

Establecer el diseño y la validez de contenido de un manual de adaptación de pruebas psicológicas para el contexto Colombiano.

Específicos

Estructurar el diseño de un manual de adaptación de pruebas o test psicológicos para el contexto Colombiano.

Establecer la validez de contenido de un manual de adaptación de pruebas psicológica para el contexto Colombiano.

Método

Tipo de estudio

La presente investigación se enmarca desde dos clasificaciones, la primera hace referencia a la categoría propuesta por Colciencias, en cuanto que se puede considerar como de desarrollo tecnológico dado que su interés se centra en la creación de un manual de procedimientos (Colciencias, 2011), y la segunda clasificación, se puede establecer dentro de los estudios de tipo instrumental ya que lo que pretende es establecer los contenidos y características de un manual de procedimientos de forma técnica, válida y confiable. (León y Montero, 2007).

Unidad de Análisis

La muestra que se tendrá en cuenta en esta investigación se refiere a los componentes, pasos y conceptos que están implícitos en la estructura del manual, a saber: Introducción, justificación generalidades, autorización, contexto y lenguaje, traducción, validación de jueces, pilotaje, interpretación de los datos, análisis aspectos psicométricos de la prueba.

MANUAL ADAPTACIÓN

Instrumentos

Para la construcción del manual se tuvo en cuenta dos instrumentos: el cuadro matriz (Apéndice A), que refiere a los aspectos: componentes y conceptos que son pertinentes. El segundo es el formato de validación por jueces (Apéndice B) en el cual se registraron los aspectos necesarios siendo objeto de evaluación.

Procedimiento

Se describe los pasos desarrollados para la construcción del manual, definidos en los tres aspectos metodológicos a saber:

Inicio. Esta fase comprende toda la recolección de la información necesaria y pertinente, para la elaboración del manual, luego, a partir de la creación de la estructura del cuadro matriz, donde de forma técnica se sistematiza la información, se tabula y estandariza, de modo que dé cuenta de los componentes del manual.

Implementación. La segunda fase comprende la construcción del manual de procedimientos, el cual se realiza, con base en la información recolectada en el cuadro matriz, de acuerdo a los componentes definidos para el mismo. Una vez estructurado el manual, éste fue sometido a juicio de un grupo de expertos, tanto conceptuales, psicométricos y de investigación, y a partir de los resultados se realizaron las correcciones pertinentes.

Finalización. Una vez obtenido el proceso de evaluación por parte de los expertos, se obtuvieron los resultados, se hicieron los ajustes al manual, y se presentó la propuesta a manera de ejemplo guía para su aplicación y puesta a prueba por parte de los psicólogos.

Aspectos éticos

La presente investigación se realizara según la ley 1090 de 2006 que reglamenta el ejercicio del Psicólogo, utilizando como base el capítulo VI: sobre el uso del material psicotécnico, haciendo referencia al artículo 46. Indica que se debe utilizar test

MANUAL ADAPTACIÓN

psicológicos o instrumentos con procedimientos científicos debidamente comprobados. Dichos test deben cumplir con las normas propias para la construcción de instrumentos, estandarización, validez y confiabilidad.

Igualmente en el artículo 47. El psicólogo tendrá cuidado en la presentación de resultados, diagnósticos y demás inferencias hasta tanto estén debidamente validadas y estandarizadas. Y en el artículo 48, se comprende que los test psicológicos que se encuentren en fase de experimentación debe utilizarse con precaución. Es preciso hacer conocer a los usuarios sus alcances y limitaciones (MINISTERIO DE LA PROTECCIÓN SOCIAL, 2006).

Resultados

Para el desarrollo de esta investigación se realizó una revisión bibliográfica y teórica donde se ha podido aclarar conceptos que han servido de sustento para la construcción del manual para la adaptación de tests psicológicos. También se considera fundamental incluir los supuestos metodológicos que son las veintidós directrices estipuladas por la comisión internacional y los componentes del proceso de adaptación de un test propuesto por Muñiz, Elousa y Hamblenton (2013). El producto final consta de ocho fases, cada una con contenidos específicos relacionados con los diferentes temas como traducción y administración de test en contextos culturales y poblaciones distintas.

Ofreciendo al usuario la información técnica relevante acerca del concepto y las características. También existen recomendaciones para la interpretación de resultados. Además el lector encontrará una breve introducción y una justificación que informa acerca de la pertinencia e importancia del manual. El siguiente paso consistió en la evaluación por jueces expertos y se ajustó el manual según los requerimientos exigidos por los jueces. Para finalizar se encuentra como producto la propuesta técnica del manual. (Apéndice C).

Posteriormente se realizó una búsqueda exhaustiva de información, que resultó pertinente para la construcción del manual de adaptación de pruebas psicológicas, donde se tuvo en cuenta las directrices de la ITC (Muñiz, Elousa y Hamblenton, 2013) de estas

MANUAL ADAPTACIÓN

finalmente se desprendieron las ocho fases ajustadas con la introducción y la justificación del manual.

Para alcanzar el segundo objetivo se realizó el proceso de validez de contenido. En esta instancia se utilizaron las puntuaciones obtenidas de los jueces. De igual modo se realizó la tabulación de los datos recopilados según el formato de validación, además se procedió a obtener los diferentes descriptivos y coeficientes, para conocer cómo se distribuyeron los resultados que permiten el soporte y la viabilidad del manual.

Las cuatro categorías de evaluación del manual fueron: pertinencia, la cual evaluaba si cada aspecto del manual resulta oportuno y adecuado para su propósito, la redacción, que evaluaba si cada aspecto del manual está construido de forma lógica, coherente y clara; la suficiencia, que evaluaba si cada aspecto del manual cuenta con la información necesaria para su comprensión y la viabilidad que evaluaba si cada aspecto del manual puede ser factible para ser realizado. Se puntuaba en una escala Likert que contaba con cuatro opciones de respuesta, siendo uno la mínima puntuación, y cuatro la máxima. A partir de 3.3 hacia arriba, se consideró pertinente, suficiente, viable y con una redacción adecuada.

En primera instancia se realizó un análisis descriptivo del comportamiento de los diez componentes del manual (introducción, justificación, y las ocho fases), donde se hizo un análisis en conjunto, de acuerdo a las puntuaciones de los ocho jueces. En la Tabla 1. Se evidencia que el promedio de los jueces, para cada fase, supera los tres puntos. Lo cual sugiere una medición alta y positiva.

Además la Tabla 1 muestra que existe igual comportamiento en la mediana, evidenciando que más de la mitad de los jueces puntúa alto, lo que se corrobora con la moda. Se observa con la desviación, que no existe mucha dispersión en los datos, y aunque algunos jueces puntuaron (2) en algunas categorías, ninguno dio una puntuación de uno (1) a las diferentes fases. Es así como la media más elevada se encuentra en las fases 2 y 4, con (3.8), y la media más baja se ubica en la justificación con (3.3)

MANUAL ADAPTACIÓN

Tabla 1. Descriptivos proceso de validación por jueces expertos

Fases	Media	Mediana	Moda	S.	Min.	Max.
Introducción	3,6	3,7	4	.351	2	4
Justificación	3,3	3,4	3	.523	2	4
Fase 1.	3,5	3,6	4	.432	2	4
Fase 2.	3,8	3,8	4	.33	3	4
Fase 3.	3,5	3,5	4	.414	2	4
Fase 4.	3,8	3,8	4	.33	3	4
Fase 5.	3,7	3,6	4	.451	2	4
Fase 6.	3,6	3,5	4	.56	3	4
Fase 7.	3,6	3,7	4	.424	3	4
Fase 8.	3,7	3,6	4	.33	2	4

N=8

En segunda instancia se aplicó el coeficiente W de Kendall, para evidenciar el acuerdo entre los jueces en sus mediciones. Este indicador afirma que si se supera el puntaje de cero existe acuerdo y se espera que sea significativo, tal como se observa en la Tabla 2. Las puntuaciones de los 10 componentes estuvieron por encima de cero, indicando que existe acuerdo entre los jueces, siendo relevante la significancia, la cual se ubicó en los rangos de 0.01, 0.05 y 0.1.

Para esta medición, se cruzaron las cuatro categorías de evaluación (redacción, pertinencia, viabilidad y suficiencia), evaluadas por los ocho jueces, donde se midió el grado de acuerdo. Cabe mencionar que el índice de acuerdo alto se obtuvo en las fases

MANUAL ADAPTACIÓN

de justificación y la fase 8 (análisis de aspectos psicométricos), y la que tuvo menor valor fue la fase 4 (traducción), que de igual forma indica un acuerdo significativo.

Tabla 2. Índice de acuerdos entre jueces expertos para los componentes

Fases	W	Sig.
Introducción	.534	.037
Justificación	.683	.008
Fase 1.	.640	.012
Fase 2.	.315	.267
Fase 3.	.557	.029
Fase 4.	.250	.429
Fase 5.	.376	.160
Fase 6.	.501	.051
Fase 7.	.566	.027
Fase 8.	.621	.015

N=8

Para terminar de contrastar la información de la evaluación por jueces, se estableció el coeficiente de razón de validez de contenido (CVR) de Lawshell, tomando como una puntuación promedio 3.3 hacia arriba por cada fase, cuyo indicador establece que si el coeficiente arroja puntuaciones por encima de cero, significa que más de la mitad de los jueces consideran lo evaluado como esencial, es así como se observa en la Tabla 3. que los coeficientes de razón de validez de contenido para cada una de las 10 fases fueron altos.

MANUAL ADAPTACIÓN

Se puede observar que las fases donde concuerda el mayor número de calificaciones son la 2, 4, 5 y 7 seguidas de las fases 3, 6 y 8. Por último se encuentran las fases introducción, justificación y fase1.

Tabla 3. Coeficiente razón validez de contenido

Fases	N/e	CVR
Introducción	6	0.5
Justificación	6	0.5
Fase 1.	6	0.5
Fase 2.	8	1
Fase 3.	7	0,75
Fase 4.	8	1
Fase 5.	8	1
Fase 6	7	0,75
Fase 7.	8	1
Fase 8.	7	0,75

De acuerdo a los anteriores resultados, se puede mencionar que las puntuaciones fueron altas en el comportamiento de las mediciones de tendencia central, a su vez es claro que existió acuerdo entre los jueces expertos, cuyas mediciones de la W de Kendall fueron significativas, mostrando unas puntuaciones altas en el coeficiente de razón de validez de contenido. Lo que permitió hacer los ajustes al manual dejando un producto listo para ser utilizado en estudios relacionados con traducción, adaptación y estandarización de pruebas psicológicas.

MANUAL ADAPTACIÓN

Análisis cualitativo

Consecuente con la anterior información, se mencionan los aspectos relevantes de la evaluación cualitativa, realizada por los jueces, mediante un proceso de categorización. La categorización es la segmentación en elementos singulares, o unidades, que resultan trascendentes y significativas para lo que se busca en una investigación. La categorización se realiza por unidades de registro, estableciendo una unidad de sentido en un texto registrado por algún medio, es de carácter textual (Austin, 2008)

En cuanto al análisis de categorías apriorísticas se evidencian sugerencias importantes en el área de redacción, como modificación de ciertas palabras además de ampliación y profundización en determinados conceptos. Otra de las categorías destacadas ha sido la de comentarios y propuestas en aspectos psicométricos del manual. Ahondar, explicitar y delimitar conceptos y nociones ha sido la constante de la misma. Aspectos relacionados a la validez de criterio y de contenido son muestra de ello. Otra de las categorías que ha resaltado es la referente a los aspectos de traducción.

Como se ha comentado anteriormente, la redacción ha sido un punto de encuentro para las sugerencias de los jueces frente al manual, dentro de estas los conectores, la coherencia y afinidad entre párrafos, utilización de signos de puntuación como tildes, comas tanto por la necesidad de utilizarlos en ciertos momentos del texto, como por la sugerencia para suprimir algunos de estos si es pertinente para el mismo.

Otro aspecto sujeto de crítica y aportes importantes tras su evaluación ha sido la suficiencia, que registró distintos comentarios como la necesidad de ampliar en distintos elementos del manual, profundizar en aspectos tanto teóricos como procedimentales. También se enfatiza en las exigencias de la Comisión Internacional de Test en relación a cumplir con una explicitación detallada de cada fase del trabajo.

Todas las sugerencias anteriormente mencionadas fueron recibidas y acogidas como herramientas constructivas para el desarrollo del manual. Las sugerencias, recomendaciones y aportes han dado una nueva versión final del trabajo y está en sincronía con los intereses finales del trabajo. Ya que a través de la utilización de

MANUAL ADAPTACIÓN

categorías apriorísticas, sin dejar de lado los hallazgos emergentes de evidente relevancia, se llevaron a cabo las correcciones de mayor y menor pertinencia para lograr un resultado satisfactorio.

Tabla 4. Categorías de análisis cualitativo

Categoría	Definición	Objetivos
Redacción	Es una actividad comunicativa de primer orden, que implica la transmisión de una señal clara al receptor.	El objetivo es ordenar las ideas principales y secundarias mediante la coherencia y cohesión textual
Términos psicométricos	Divulgación de conocimientos que faciliten la comprensión, de ciertos conceptos.	El objetivo es proveer la información de algunos términos técnicos utilizados en psicometría.

Discusión

Las conclusiones de este trabajo se orientan según las directrices del Comité Internacional del Test (ITC) siendo el órgano coordinador de proteger el uso ético de los test. Además de revisar la adecuación y responsabilidad del manejo que hacen los profesionales de los instrumentos. Por esta razón, constituye un eje transversal para la adaptación de pruebas psicológicas de una cultura a otra, ya que es un proceso acorde con la función de la distancia cultural entre la cultura original y el contexto donde se pretende aplicar la prueba (Muñiz, Elousa & Hamblenton, 2013).

MANUAL ADAPTACIÓN

De acuerdo a Haig y Borsboom (2008) la adaptación se configura con base en la teoría de la medición psicológica, puesto que de ahí se deriva los estándares para diseñar, construir, validar, estandarizar y aplicar las pruebas psicológicas. Dichos estándares deben ser analizados por una comunidad científica que determine las propiedades psicométricas de la prueba, dada su relevancia para la inferencia y la teorización en Psicología.

A partir de lo anterior, se analizó el conjunto de pasos que debían pertenecer al manual para hacerlo pertinente y viable, siendo una herramienta útil que sirve para futuras adaptaciones de instrumentos en el contexto colombiano. Es por esto, que la investigación se consolidó a partir de la experticia de Muñiz, Elosua y Hambleton (2013) en el campo quien argumenta los principales pasos para la adaptación de una prueba psicológica.

Como: A) consideraciones legales o derechos de autor que incluye la propiedad intelectual B) valoración o significado del constructo C) características de la nueva población tanto lingüísticas, psicológicas y culturales del contexto D) la prueba piloto considera una muestra representativa de la población; esta fase recoge los datos sobre el test adaptado, posteriormente se delimita el grado de comparabilidad de puntuaciones, se efectúa el análisis del ítems, estudios de fiabilidad y validación para llevar a cabo las revisiones del test adaptado y unas correctas condiciones de la aplicación e interpretación como último procedimiento.

Además, es importante mencionar las cuatro categorías de evaluación del manual, que correspondían a la pertinencia siendo la categoría más relevante, evaluando si cada aspecto del manual resulta oportuno y adecuado para su propósito, la redacción determinaba si cada aspecto del manual está construido de forma lógica, coherente y clara; la suficiencia, evaluaba si cada aspecto del manual cuenta con la información necesaria para su comprensión y la viabilidad valoraba si cada aspecto del manual puede ser factible para ser realizado. Y se puntuaba en una escala Likert que contaba con cuatro opciones de respuesta, siendo uno la mínima puntuación, y cuatro la máxima.

MANUAL ADAPTACIÓN

Se considera que la elaboración del manual así como su validez de contenido, se basan en el análisis de cada uno de los diez componentes del manual, (introducción, justificación, y las ocho fases), cabe aclarar que el análisis de datos se ajustó a un análisis no paramétrico ajustada a una escala ordinal.

Por consiguiente, cada formato de validación mostro los resultados sobre la calidad de la fundamentación teórica, el análisis de ítems, la validez resultante, la fiabilidad de sus puntuaciones y la calidad de los baremos siendo evidencias del proceso de adaptación de test. E igualmente se presentan tanto resultados cuantitativos sobre los anteriores aspectos como información cualitativa que resalta los puntos fuertes del manual y los aspectos a mejorar. Finalmente, se incluye información sobre el proceso de revisión y algunas cuestiones a mejorar para futuras ediciones (Elosua, 2003).

Finalmente, el presente manual cumplió con los objetivos del proyecto puesto que según el análisis de los ocho jueces es pertinente en la totalidad de sus fases, sin embargo se deberán tener en cuenta recursos financieros, humanos y físicos suficientes para emprender este tipo de proyectos en el futuro.

Igualmente esta investigación reconoce la importancia de los tests psicométricos siendo una herramienta fundamental, dada su objetividad y economía de medios y tiempo, por eso los psicólogos deben empoderarse en mayor grado de la utilización e interpretación de los tests, a través de este manual se reconoce información para comprender mejor la adaptación de los tests.

Dando respuesta a los desafíos propuestos por Hamblenton et al. (2004), es necesario prestar atención al uso internacional de los tests, debido a la globalización y a los nuevos canales de comunicación, lo cual plantea todo un conjunto de problemas de adaptación de los test de unos países a otros. Esta internacionalización ha puesto de manifiesto, la necesidad de disponer de un marco general de evaluación que recoja las buenas prácticas de evaluación a través de nuevos modelos psicométricos y tecnologías para generar y analizar los tests.

MANUAL ADAPTACIÓN

Por otro lado es necesario mencionar que pocos, si alguno, de los instrumentos psicométricos de gran interés, le ha prestado atención a la posibilidad de hacer traducciones y adaptaciones en diferentes países. Es necesario mencionar que las ocho fases del manual han sido propuestas como un requerimiento mínimo para justificar el uso de instrumentos psicométricos traducidos y adaptados. Sólo cuando instrumentos psicológicos traducidos y adaptados han sido desarrollados de acuerdo a estas ocho fases puede ser justificable el considerar su uso en un contexto aplicado (Gudmundsson, 2009).

En vista del tiempo y esfuerzo involucrados en este proceso, y eventualmente estandarizando un instrumento, la pregunta que cabe hacerse es por qué no son construidos nuevos instrumentos en países que no son angloparlantes con más regularidad. La razón más comúnmente ofrecida es que al traducir y adaptar instrumentos se ahorra tiempo y dinero, y numerosas investigaciones sobre un instrumento en un país se vuelven disponibles en otro.

De cualquier forma es difícil que estos argumentos se sostengan. Puede consumir tanto tiempo y ser tan costoso realizar una adaptación, traducción y estandarización de un test como preparar uno original para medir constructos particulares. Además de esto, no puede asumirse que hallazgos en un instrumento original pueden ser generalizados al test traducido (Gudmundsson, 2009).

En vista del tiempo y dinero gastado en probar la validez de los estudios necesarios para confirmar la utilidad diagnóstica y varios tipos de validez de un instrumento traducido y validado, la asunción de que construir un nuevo instrumento podría ser más costoso puede no tener fundamentos. De todas maneras, esto podría cambiar, si publicadores de instrumentos psicológicos asumieran alguna responsabilidad por el uso adecuado de sus instrumentos en idiomas secundarios, y compartieran los costos de su traducción, adaptación y estandarización (Gudmundsson, 2009).

Referencias

- Abad, Garrido, Olea & Ponsoda (2006). Introducción a la psicometría. Teoría clásica de los tests y teoría de la respuesta al ítem. Universidad Autónoma de Madrid p. 1-150.
- Aiken, L. (1998). Test and examinations: Measuring abilities and performance. New York: Wiley.
- Aiken, L. (2003). Test psicológicos y evaluación. Ed. Pearson: Educación de México S.A de C.V.
- Aiken, L.R. (1995) Rating Scales and Checklists: Evaluating Behavior, Personality and Attitude. John Wiley & Sons. USA
- Aliaga, L. (2006). Psicometría: test psicométricos, confiabilidad y validez. En Quintana A. & Montgomery W. Psicología. Tópicos de actualidad. Lima UNMSM, p. 85-108.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education, (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.
- Aragón, L. (2004). Fundamentos psicométricos en la evaluación psicológica. Revista electrónica de Psicología Iztacala (4), p. 23-43.
- Argibay, J. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. Subjetividad y procesos cognitivos (8), p.15-33.
- Arias, R. (1995). Psicometría: teoría de los test psicológicos y educativos. Madrid: Capítulo1.
- Attoresi, H., Lozia, G., Abal, F., Galibert, M & Aguerri, M. (2009). Teoría de respuesta al ítem: Conceptos básicos y aplicación para la medición de constructos psicológicos. *Revista Argentina de clínica de Psicológica* (2), p. 179-188.
- Austin, T. (2008). Investigación Cualitativa. Metodología de investigación. Tema 9.

MANUAL ADAPTACIÓN

- Avendaño, B & Elvers, Lozano. (2001). Consideraciones sobre las teorías de medición: un ejemplo de aplicación bajo la teoría de respuesta al ítem (TRI). *Acta Colombiana de Psicología* (5), p. 87-98.
- Balluerka, N., Gorostiaga, A., Alonso, I & Haranburu, M. (2007). La adaptación de instrumentos de medida de unas culturas a otras: una perspectiva práctica. *Revista Psicotherma* (1), p. 124-133.
- Barraza, A. (2007). La consulta a expertos como estrategia para la recolección de evidencias de validez basadas en el contenido. Universidad Pedagógica de Durango 7 p. 5-14.
- Bond, L. (2001). Test Bias. *International encyclopedia of the social and behavioral sciences*. Elsevier Science Ltda.
- Bonilla, V. (2006). INEVA en acción. Vol. (2) 3 p. 1-3. Recuperado de www.ineva.uprrp.edu/boletin/v0002n0003.pdf
- Brennan, Robert L. (2001). Generalizability theory. *Educational and psychological measurement* (64), 391-418.
- Carretero, H & Pérez, C. (2007). Normas para el desarrollo y revisión de estudios instrumentales: consideraciones sobre la selección de test en la investigación psicológica. *International journal of clinical and health psychology* (7), 863-882.
- Castejón, J.L. (1997). Introducción a los métodos y técnicas de investigación y obtención de datos en Psicología. Sant Vicent del Raspeig, España: ECU.
- Castelli, Y. (2007). Bases teóricas: Teorías de la inteligencia. Recuperado de <http://www.ycastellif.cl/psicometria/art-inteligencia.htm>.
- Chang, C. (2001). Cross-cultural assessment: A call for test adaptation. Newsnotes of the association for assessment in counseling and education. Clemson University
- Cizek, G.J. (2001). Setting performance standards: concepts, methods and perspectives. Londres: LEA.

MANUAL ADAPTACIÓN

- Cohen, R.J & Swerdlik, M.E. (2001). Pruebas y evaluación psicológica: introducción a las pruebas y a la medición. Mc Geaw-Hill: México.
- Coolican, H. (2005). Métodos de investigación y estadística. (3a. Ed.). México; Manual Moderno.
- Departamento administrativo de ciencia, tecnología e innovación Colciencias (2011). Tipología de proyectos de carácter científico, tecnológico e innovación.
- Elosua, P. (2003). Sobre la validez de los test. *Revista Psicothema* (2), 315-321.
- Espinosa, J. (2008). Validación y estandarización de instrumentos. Universidad Nacional de Colombia 29 p. 1- 16.
- Fernández, A.L & Scheffel, D.L (2003). A study on the criterion validity of the Matiiis dementia rating scale. *International Journal of testing*, (3),49-58.
- Gómez, J., Hidalgo, M & Muñiz, J. (2002). Esquema conceptual y procedimientos para analizar la validez de las consecuencias del uso de los test. *Psicothema*, (1), p. 173-178.
- Gondra, J.M. (1997). Historia de la Psicología, Madrid, síntesis.
- Gregory, R. (2001). Evaluación psicológica; historia, principios y aplicaciones. Ed. Manual Moderno: México.
- Haig, B & Borsboom, D. (2008). On the conceptual foundations of Psychological measurement. *Psychology press*, 6, p. 1-6.
- Hambleton, R. & Li, S. (2004). Effective Implementation of the International Test Commission Guidelines for Adapting Test.
- Hambleton, R., Merenda, P & Spielberger. C. (2004). Adaption educational and psychological assesment . Hillsdale, NJ: Lawrence S. Erlbaum Publishers.
- Hambleton, R.K. (2000). Issues designs and technical guidelines for adapting tests in multiple languages and cultures.

MANUAL ADAPTACIÓN

- Hernández, E., Valencia, S. & Rodríguez, J. (2003). De la sección de psicotecnia al laboratorio de Psicometría: seis décadas de algo más que medición psicológica en Colombia. *Revista Avances en Medición*, 1 (1), p. 6-16.
- Hogan, T. (2004). Pruebas psicológicas: una introducción práctica. Ed Manual Moderno: México.
- ICFES. (2000). Admisión a la educación superior: algunos temas de discusión. Bogotá D.C.
- Jaimes, J. (2000). Adaptación de pruebas psicológicas y desarrollo de software para población limitada visual: cuestionario de personalidad 16 PF Forma C. *Acta Colombiana de Psicología* (5), p. 123-134.
- Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema*, p.127-135.
- León y Montero(2007) Guía para nombrar los estudios de investigación en Psicología Internacional Journal of Clinical and Health Psychology .Vol.7, (3) p.847-862.
- Messick, S. (1995). Validity of Psychological assesment: validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Pschychological association* (9), 741-749.
- Mikulic, I. (2007). Construcción y adaptación de pruebas psicológicas. Universidad de Buenos Aires.
- Ministerio de la Protección social (2006). Ley número 1090 de 2006: Código deontológico y bioético de la profesión de Psicología.
- Mora, J & Martin, M. (2007). La escala de inteligencia de Binet y Simon (1905) su recepción por la Psicología posterior. *Revista de la historia de la Psicología* (2), 307-313.
- Muñiz, J, Elosua, P &Hambleton, R. (2013). Directrices para la traducción y adaptación de los test. Segunda edición. *Revista Redalyc* (66), p. 1-8.
- Muñiz, J. (1998). La teoría clásica de los test. Ed. Pirámide. Madrid, España.

MANUAL ADAPTACIÓN

- Muñiz, J. (2010). La teoría de los test: teoría clásica y teoría de respuesta a los ítems. *Revista Papeles del Psicólogo* (1), p. 57-66.
- Ortiz, M. (2002). Introducción a la historia y conceptos básicos de la Psicometría.
- Pérez, J., Chacón, S. & Moreno, R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema* Vol. 12 (2), p. 442-446.
- Polit, D. & Hungler, B. (1997). Investigación científica en ciencias de la salud (5) México: Mc Graw-Hill.
- Prieto, G & Delgado, A. (2010). Fiabilidad y validez. *Papeles del Psicólogo*. (1), 67-74.
- Rodríguez, W., Jiménez, R & Caicedo, C. (2007). Protocolo de actitudes relacionadas con la ciencia: Adaptación para Colombia. Corporación Universitaria Iberoamericana.
- Rojas, M., Manríquez, G & Gatica, Y. (2004). Teoría de respuesta al ítem. *Revista ingeniería informática* (10), 1-12.
- Saldarriaga, J., Jaimes, C., Aguilar, E & Merino, M. (2002). Validez, confiabilidad y baremación del inventario de estrategias meta cognitivas en estudiantes universitarios. *Revista de Psicología* (1), p. 9-20.
- Sánchez, R & Echeverry, J. (2004). Validación de escalas de medición en salud. *Revista Salud Pública*, 6 (3), p.302-318.
- Sattler, J. (1993). Evaluación de la inteligencia infantil y habilidades especiales. Mexico: Manual Moderno.
- Sireci, S. & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema* (26) 1, p.100-107.
- Sireci, S. (1999). Guidelines for Adapting Certification Test for Use Across Multiple Languages. Recuperado de www.cesb.org/Guidelines/Adapting.

MANUAL ADAPTACIÓN

- Sireci, S., & Padilla, J. (2014). Validating assessments: Introduction to the Special Section. *Psicothema*, (26) 1, p. 97-99.
- 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.
- Tanzer, N.K. (1995). Cross- cultural bias in likert-type inventories: perfect matching structures and still biased. *European and Psychological Measurement*. (3), p. 194-201.
- Terrón, A. (2010). Introducción al análisis de la teoría de respuesta al ítem. San Francisco, California, Aidesoc.
- Tornimbeni, S., Perez, E & Olaz, F. (2008). Introducción a la Psicometría (1). Buenos Aires: Paidós
- Vallo, I. (2000). Directrices para la adaptación y uso de los test. Madrid; España.
- Van de Vijver, F. R. K. (1996). Translating Tests: Some Practical Guidelines. *European Psychologist*, 89-99.
- Verdugo, M., Arias, B., Gómez, L & Shalock, R. (2009). Manual de aplicación de la escala GENCAT de calidad de vida. Barcelona: Departamento de acción social y ciudadanía de la Generalitat de Catalunya.
- Zarate, L & Castro, C.G. (2009). Adaptación del cuestionario sobre el comportamiento en niños de 1 a 5 años de Achenbach. *Enseñanzas e investigación en Psicología*, 2, p. 179-191.
- Zúñiga, M & Montero, E. (2007). Teoría G: un futuro paradigma para el análisis de pruebas psicométricas. *Periodos electrónicos en Psicología* (21), p. 117-144.
- Zwick, R. (2000). The assessment of differential item functioning in computer adaptive test. En W. Van der Linden y C.A W. Glas (Eds). *Computerized adaptive testin. theory and practice*, 221-243.

APENDICE A

MATRIZ DEL MANUAL DE ADAPTACIÓN			
Es una guía práctica que explica paso a paso como se realiza el proceso de adaptación de un instrumento de evaluación. Comprende ocho fases, que van desde las generalidades que se deben tener en cuenta en el proceso, pasando por las especificaciones en cuanto a traducción, características lingüísticas y contextos culturales hasta llegar al análisis de los aspectos psicométricos de la prueba			
No.	FASES	DESCRIPCIÓN	PALABRAS CLAVE
1	Generalidades	Es una guía que define el proceso de adaptación y sus características	Teoría, operacionalización, ítem, características, confiabilidad, validez.
2	Autorización	Se reconoce los derechos de autor en Colombia y la protección de las pruebas.	Leyes, derechos de autor, copy-right, propiedad intelectual, permiso.
3	Contexto y lenguaje	Características de la población y el lenguaje de la prueba original traducida a otro país.	Lenguaje, cultura, contextos, social.
4	Traducción	Especifica las diferentes formas para una correcta traducción.	Traducción directa, inversa, idioma matriz, idioma objetivo y adaptación lingüística.
5	Validación de Jueces	Se relaciona con la selección y la evaluación.	Jueces, selección, evaluación e indicadores.
6	Pilotaje	Comprueba el diseño y los ítems pertinentes de la pruebas.	Comprobación, aplicación, y administración.
7	Interpretación de los datos	Reglas para identificar inconsistencias de la prueba.	Estadística inferencial, población, baremos.
8	Análisis aspectos psicométricos de la prueba	Procedimiento donde se codifica y se genera información de trabajo al público.	Sistematización, procedimiento y análisis de datos estadísticos

Apéndice B

**UNIVERSIDAD DE SAN BUENAVENTURA BOGOTÁ
PROGRAMA DE PSICOLOGÍA
SÍNTESIS FORMATO DE VALIDACIÓN POR JUECES**

Nombre:	
Formación:	
Fecha:	

INSTRUCCIONES

Señor Juez en este formato, encontrará, los aspectos a evaluar del Manual de Adaptación de Pruebas Psicológicas para el Contexto Colombiano, el cual contiene ocho fases, y se distribuyen por categorías.

Para la calificación de cada uno de los componentes del Manual, marque con una (X) en el espacio correspondiente según su criterio de evaluación. En la casilla de observaciones realice una evaluación cualitativa que dé cuenta del análisis de cada uno de los elementos que constituyen el Manual.

Estos son los criterios de Evaluación:

PERTINENCIA: Evalúa si cada aspecto del Manual resulta oportuno y adecuado para su propósito.

REDACCION: Evalúa si cada aspecto del Manual está construido de forma lógica, coherente y clara.

SUFICIENCIA: Evalúa si cada aspecto del Manual cuenta con la información suficiente para su comprensión.

VIABILIDAD: Evalúa si cada aspecto del Manual puede ser factible para ser realizado.

Pertinencia	1	No pertinente	2	Poco pertinente	3	Pertinente	4	Muy pertinente
Redacción	1	Muy confusa	2	Confusa	3	Clara	4	Muy clara
Suficiencia	1	Insuficiente	2	Poco suficiente	3	Suficiente	4	Muy suficiente
Viabilidad	1	Inviabile	2	Poco viable	3	Viable	4	Muy viable

Nota: ES IMPORTANTE, A PARTE DE LA CALIFICACIÓN CUANTITATIVA, REALIZAR LAS OBSERVACIONES CUALITATIVAS DE CADA UNO DE LOS ASPECTOS QUE INCLUYE EL MANUAL A EVALUAR.

MANUAL ADAPTACIÓN

COMPONENTE																
Aspectos a Evaluar	CRITERIOS															
	PERTINENCIA				REDACCIÓN				SUFICIENCIA				VIABILIDAD			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Introducción	Observaciones:															
Justificación																
Fase 1.																
Fase 2.																
Fase 3.																
Fase 4.																
Fase 5.																
Fase 6.																
Fase 7.																
Fase 8.																

Apéndice C

**MANUAL PARA LA ADAPTACIÓN DE PRUEBAS PSICOLÓGICAS EN EL
CONTEXTO COLOMBIANO**



**UNIVERSIDAD DE
SAN BUENAVENTURA
BOGOTÁ, D.C.**

NICOLÁS YEPES ALEJANDRO BARRERO Y GABRIELA FLÓREZ

UNIVERSIDAD DE SAN BUENAVENTURA BOGOTA

FACULTAD DE CIENCIAS HUMANAS Y SOCIALES

ANDERSSEN VERA

PROGRAMA DE PSICOLOGIA

2015

Tabla de contenido

Introducción	50
Justificación	50
Fases	51
Fase 1. Generalidades	51
Fase 2. Autorización (derechos de autor)	52
Fase 3. Contexto y lenguaje	56
Fase 4. Traducción	58
Fase 5. Validación por jueces	61
Fase 6. Pilotaje	63
Fase 7. Interpretación de los datos	69
Fase 8. Análisis aspectos psicométricos de la prueba	71
Bibliografía	74

INTRODUCCIÓN

En este manual se encontrará una guía práctica para desarrollar paso a paso el proceso de adaptación de una prueba o instrumento de medición utilizado en psicología. El manual comprende ocho fases, que van desde las generalidades que se deben tener en cuenta antes de adaptar una prueba o instrumento, pasando por las especificaciones iniciales como traducción, características lingüísticas y contextos culturales hasta llegar al análisis de los aspectos psicométricos. El manual toma como base las directrices del comité de la comisión internacional de test (ITC) y el marco integral propuesta por Muñiz, Elosua & Hamblenton (2013).

Por lo tanto, la presente propuesta se encuentra dirigida a estudiantes, profesores, investigadores y profesionales en el área de la Psicología, que estén interesados y requieran de una guía específica, para realizar un proceso técnico que conlleva la adaptación de pruebas o instrumentos de medición a un contexto específico.

JUSTIFICACION

La propuesta de la creación del manual está fundamentada en la importancia de contar con pruebas psicológicas que cumplan con las propiedades psicométricas exigidas para poder ser utilizadas de forma adecuada. Además busca facilitar procesos de adaptación y posteriormente estandarización de instrumentos extranjeros que en su mayoría solo llegan a ser traducidos al contexto colombiano, donde se utilizan los baremos que no corresponden a la población objetivo, conduciendo a información imprecisa que genera interpretaciones erróneas y eventualmente consecuencias no deseadas para la investigación.

Es así como la presente propuesta elucida de una forma técnica y con rigor metodológico los pasos que se requieren en el proceso de adaptación de pruebas o instrumentos de medición psicológica, convirtiéndose en una herramienta que sirva a estudiantes, profesores, investigadores y profesionales interesados en el desarrollo de este proceso.

FASES

FASE 1. GENERALIDADES

Aspectos que deberán tener en cuenta una vez haya escogido el test que se adaptará. El investigador analiza características relacionadas con la adaptación del test.

Palabras clave: teorías, operacionalización, características, recursos, procesos.

Contexto: Lo primero que debe realizar el investigador, es una revisión de las pruebas o instrumentos de medición psicológica que pueden ser útiles para los objetivos que se han planteado; una vez hecho esto, se debe escoger la prueba que más se adapte al criterio del investigador. En segundo lugar se debe estudiar cómo ha sido definido y operacionalizado el constructo que pretende medir la prueba. En tercer lugar se tienen que analizar las características de aplicación del test, donde se identifica el tiempo, las características ambientales, los rangos de edad, los materiales necesarios y las reglas que deben cumplir los profesionales y los sujetos evaluados. En cuarto lugar es necesario reconocer el método de medición, el tipo de escalas manejadas por el test, y las características de las respuestas que puede escoger el sujeto.

Además es necesario revisar cómo fueron construidas las tablas de puntajes y/o baremos, conocer cuáles son los coeficientes de validez y confiabilidad del test matriz. Una vez realizado los anteriores pasos es necesario reconocer que los ítems y estímulos son accesibles a la población objetivo.

Recomendación: Según las directrices previas de la ITC es necesario seguir las normas y prácticas profesionales relativas a la utilización de tests. Además se debe seleccionar el diseño de tests más pertinente para la adaptación. Evaluar la importancia del constructo o constructos medidos por la prueba en las poblaciones objetivo.

MANUAL ADAPTACIÓN

Referencias:

Elosua, P. Mujika, J. Almeida, L. & Hermosilla, D. (2014). Procedimientos analítico-rationales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento. *Revista Latinoamericana de psicología* (46) 2 p.117-126.

International Test Commission (2005). *International Guidelines on Test Adaptation*. [www.intestcom.org]

Muñiz, J, Elosua, P & Hambleton, R. (2013). Directrices para la traducción y adaptación de los test. Segunda edición. *Revista papeles del psicólogo* (66), p. 1-8.

FASE 2. AUTORIZACIÓN (DERECHOS DE AUTOR)

Reconocer los siguientes artículos que involucran lo propuesto por el congreso de la República de Colombia, acerca de los derechos de autor en el país y los derechos internacionales sobre copyright. Además se deben tener en cuenta los lineamientos del código deontológico del psicólogo

Palabras clave: derechos de autor, leyes.

Contexto: Según la Ley 44 de 1.993 que adiciona y modifica la Ley 23 de 1982 dictada por el Congreso colombiano, se decreta la protección de las siguientes creaciones expresadas: obras científicas, literarias y artísticas y cualquiera sea su destinación tales como; libros, folletos, escritos, conferencias entre otros. Dentro de esta ley se contempla el Artículo 5, donde se especifica que deben ser protegidas las obras independientes como las traducciones, adaptaciones y demás transformaciones realizadas sobre una obra de dominio privado, con autorización expresa del titular de la obra original.

Para llevar a cabo esta fase se deben tener en cuenta los siguientes lineamientos para adaptar o transformar una prueba psicológica distinta a los instrumentos. Los criterios de protección son: a) Protección a la forma y no a las ideas, b) La originalidad, c) el mérito y d) destinación de la obra, ausencia de formalidades para la protección, proteger la obra

MANUAL ADAPTACIÓN

independiente del objeto material en el que se encuentre fijada y dirección nacional de derechos de autor (Vega, 2010).

El traductor de la obra de dominio público, es autor de su propia versión pero no podrá oponerse a que se hagan traducciones distintas de la misma obra. Solo con el permiso expreso del autor adapta, modifica, extracta una obra del dominio privado, se convierte en autor sobre su propia adaptación pero salvo convención, es decir, si el autor original le niega el permiso, la persona no podrá darle publicidad a la modificación de la obra sin mencionar el título de la obra originaria y su autor (Congreso de la república, 1982).

De igual forma, en Colombia también existe un órgano institucional que es la Dirección Nacional de Derechos de Autor quienes protegen jurídicamente una obra, sus lineamientos son seguidos en la obra de Olarte y Rojas (2010) de acuerdo a este contexto normativo, se deben conocer los derechos morales, los cuales buscan proteger el vínculo entre el autor y su obra ya que constituye la expresión de su personalidad, estos derechos son intransferibles a otras personas. Por lo cual se considera que el autor puede: conservar la obra inédita, reivindicar la paternidad de la obra, oponerse a toda deformación, modificar la obra antes o después de su publicación, retirar la obra del mercado.

Así mismo los derechos patrimoniales, permiten al autor controlar la explotación de su obra, permitiendo autorizar o prohibir: la reproducción, comunicación y distribución pública, la traducción y adaptación de la obra y la importación de ejemplares sin su autorización.

Dentro de esta explicación acerca de los derechos de autor, cabe resaltar a la Asociación Americana de Investigación Educativa (1999) argumentando que el copyright es un requerimiento que tiene como meta demostrar el derecho con la ayuda de un registro que facilita al autor el diseño de un vínculo verificador entre su creación y su fecha, esto suele denominarse una prueba para hacer cumplir el derecho de autor.

MANUAL ADAPTACIÓN

El copyright es una obligación ética de los Psicólogos y demás profesionales para proteger sus investigaciones y principalmente la construcción de instrumentos. Los autores deben asumir que las pruebas son protegidas incluso sin una mención de propiedad intelectual. En consecuencia si una persona quiere reproducir o modificar una prueba se tiene que poner en contacto con el editor y solicitar el permiso de reproducir el material. De igual modo, los profesionales pueden pedir al consejo legal u otra entidad una orden protectora que prohíba hacer pública la obra o instrumento y a su vez la reproducción de copias.

Por otro lado, dentro un marco internacional, los derechos de autor han sido acogidos por los convenios internacionales como el convenio de Berna que proporciona una ley de propiedad intelectual, recurriendo a diferentes leyes nacionales y al derecho internacional privado. Sin embargo con la llegada de la tecnología digital en la época actual se ha facilitado la infracción y explotación de obras por internet, a esto se suma que las leyes nacionales del país propio de la prueba se convierten en insignificantes a medida que se traspasan fronteras y la ley de propiedad intelectual no se aplica de forma efectiva (Xalabarder, 2005).

Aparte del seguimiento cuidadoso de lo referente a los derechos de propiedad intelectual se debe estudiar de cerca la legislación Colombiana en lo relacionado al ejercicio profesional del psicólogo, donde se mencionan ciertos elementos que se vinculan estrechamente con la utilización de pruebas psicológicas. Es importante que usted haga una revisión cuidadosa y detallada de los aspectos legales relacionados a la protección de los derechos de autor que se encuentran reflejados en diversas instancias de la constitución política colombiana.

Según la ley 1090 de 2006 del código deontológico y Bioético del psicólogo de acuerdo el artículo 47. El psicólogo tendrá cuidado en la presentación de resultados, diagnósticos y demás inferencias hasta tanto estén debidamente validadas y estandarizadas. Y en el artículo 48, se comprende que los tests psicológicos que se encuentren en fase de experimentación deben utilizarse con precaución. Es preciso hacer conocer a los usuarios sus alcances y limitaciones (Ministerio De La Protección Social,

MANUAL ADAPTACIÓN

2006). Cuando se ha verificado este aspecto se debe proseguir con el reconocimiento de los derechos intelectuales del test.

Recomendaciones: En esta fase se sugiere identificar a quien le corresponde los derechos de autor, para así poder contactar con el autor original de la prueba o a la editorial y se les solicitaran los permisos pertinentes para poder hacer la adaptación del instrumento psicológico. Finalmente, se debe conseguir el permiso para poder modificar, añadir o traducir una prueba psicológica.

Referencias:

Asociación Americana De Investigación Educativa (2015). Permission for authors.

Recuperado el día 21 marzo de 2015 de <http://www.aera.net/Publications/PublicationsPermissions/tabid/13120/Default.aspx>

Dirección Nacional de Derechos de Autor (DNDA) ¿Qué es el derecho de autor?

Recuperado el día 21 marzo de 2015 de <http://www.derechodeautor.gov.co/web/guest/preguntas-frecuentes>

Ley N° 23 de 1982. Congreso De la Republica de Colombia: Régimen general de derechos de autor.

Ministerio de la Protección social (2006). Ley número 1090 de 2006: Código deontológico y bioético de la profesión de Psicología.

Olarte, J & Rojas, M. (2010). La protección del derecho de autor y los derechos conexos en el ámbito penal. Dirección Nacional De Derechos de Autor; Ministerio de Interior y de justicia.

Vega, A. (2010). Manual de Derechos de Autor. Dirección Nacional de Derechos de Autor: Bogotá.

Xalabarder, R. (2005). Copyright y derecho de autor ¿Convergencia internacional en un mundo digital? Revista de internet, derecho y política. Información recuperada de: <http://www.uoc.edu/idp/1/dt/esp/xalabarder02.pdf>

FASE 3. CONTEXTO Y LENGUAJE CARACTERÍSTICAS DE LA POBLACIÓN O MUESTRA

En esta fase encontrará las especificaciones correspondientes al contexto cultural y poblacional, como la atención especial que se debe prestar al momento de hacer un estudio transcultural de estas características

Palabras clave: Lenguaje, cultura, contexto.

Contexto: En esta fase, no se debe incursionar en el sesgo de constructo, que consiste en utilizar conceptualizaciones de una sola cultura sin tener en cuenta las diferencias con otras. Se debe estudiar cada ítem y que sus palabras y conceptos se adecuen a las distintas características etnográficas de las poblaciones a los que irá dirigido el instrumento, con el fin de evitar tendencias etnocéntricas.

En el caso de no encontrar una correcta homologación de un constructo con rasgos locales, elimine las palabras y conceptos que se limiten a características de una cultura o población en particular. Aunque la mejor alternativa para evitar propuestas como la anterior es comenzar el proceso de adaptación acompañado de un equipo de especialistas; lingüistas y expertos en diferentes culturas.

Otro detalle importante a señalar es que cuando los efectos de las diferencias culturales no son importantes para los propósitos del estudio estos deben ser en lo posible minimizados. Esto se refiere a que en muchas ocasiones las diferencias culturales son tan grandes y evidentes entre poblaciones que lo más conveniente resulta eliminar los ítems que comprendan las diferencias más específicas, esto debido a que entre mayores sean estas diferencias más improbable es que un instrumento pueda cubrirlas en toda su extensión, por lo tanto se debe concentrar la adaptación en los aspectos sobresalientes y que representen una mayor convergencia con sus contenidos principales (Hamblenton, 1994 citado en Van De Vijver, 1996).

Sumado a lo anterior, en la adaptación de un instrumento jamás se puede dar por sentado que las características comportamentales y culturales de una población son iguales. Por lo tanto se debe siempre hacer un estudio de cada constructo detallando los

MANUAL ADAPTACIÓN

cambios que presenta cuando se trata de adaptar a otro grupo social (Hamblenton, 2000).

Los agentes encargados de la adaptación de instrumentos nunca deben dejar de lado los dos factores de mayor importancia a saber: las diferencias lingüísticas y culturales de las poblaciones a las que va dirigido el test o prueba psicométrica. También se debe tener en cuenta que esto aplica para instrumentos en idioma español que a pesar de dirigirse a países hispanoparlantes manifiestan grandes diferencias culturales entre sus pueblos, así que la adaptación deberá realizarse con el mismo rigor si por ejemplo, el instrumento es de origen español y se adapta en países de América del Sur, de América Central u otros países de habla hispana.

El lenguaje usado por los agentes encargados de realizar la adaptación debe ser apropiado a todos los idiomas, culturas y poblaciones a los que va destinada su utilización. Esto se debe probar mediante todos los elementos del test, tanto los ítems como el manual de guía entre otros.

Si bien es necesario que no se olvide el hecho fundamental de adecuar el instrumento a todos los grupos culturales involucrados, tome en cuenta las siguientes guías prácticas que no se han mencionado todavía en el texto: No utilice palabras innecesarias. No utilice coloquialismos y metáforas. Evite las imprecisiones y utilice términos concretos, el evaluador debe asegurar que las personas entiendan el lenguaje planteado y comprendan claramente los significados (Van De Vijver, 1996).

Referencias:

- Balluerka, N., Gorostiaga, A., Alonso, I & Haranburu, M. (2007). La adaptación de instrumentos de medida de unas culturas a otras: una perspectiva práctica. *RevistaPsicotherma* (1), p. 124-133.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology* , 6, (2), p. 29-45.

MANUAL ADAPTACIÓN

Hambleton, R.K. (2000). Issues designs and technical guidelines for adapting tests in multiple languages and cultures.

Van de Vijver, F. R. K. (1996). Translating Tests: Some Practical Guidelines. *European Psychologist*, p. 89-99.

FASE 4. TRADUCCIÓN

En esta fase se especificarán las formas válidas para hacer una correcta traducción, teniendo en cuenta el idioma matriz y el idioma objetivo del test

Palabras clave: Traductor, traducción directa, traducción inversa, adaptación lingüística.

Contexto: Para obtener un resultado valido en esta fase y minimizar errores que afecten la equivalencia de constructo los autores Muñiz y Hambleton (1996) proponen dos tipos de traducción: traducción directa ("forward translation") y traducción inversa ("backward translation"), las cuales se describen a continuación.

Traducción directa "Forward translation". En este tipo de traducción el test psicológico Matriz (test original) es traducido a la forma Objetivo (test adaptado).

Para que este proceso se pueda desarrollar se debe contar con un grupo de traductores profesionales, estos deberán trabajar de forma independiente haciendo la traducción del test matriz. Posteriormente se deben suministrar las traducciones hechas a los jueces para confrontar y evaluarlas, adoptando la que mejor interprete los ítems del test matriz.

Otra propuesta para hacer este tipo de traducción inicia con la división de los traductores y lingüistas en grupos (grupo A, grupo B). El grupo A tiene la responsabilidad de hacer las traducciones de los ítems del test matriz. El grupo B será el encargado de revisar y evaluar las equivalencias entre los ítems traducidos y los ítems matrices. De esta forma se intenta minimizar los errores que se puedan cometer en las traducciones.

MANUAL ADAPTACIÓN

Al utilizar esta forma de traducción, el investigador deberá tener en cuenta los siguientes aspectos: Primero los traductores y lingüistas deben estar familiarizados con la cultura del test matriz y del test objetivo. Segundo se debe prestar atención a suposiciones y presunciones que pueden hacerse en la labor de traducción. Tercero se debe tener cuidado con las palabras, frases e ítems que presenten varios significados. Cuarto es necesario vigilar que aspectos innecesarios e inapropiados del test matriz no permanezcan en el test objetivo.

Traducción inversa "backward translation". Este tipo de traducción suele ser el más utilizado. Se divide el grupo de expertos en dos (grupo A y grupo B). El grupo A traduce los ítems del test matriz; obteniendo como resultado el test traducido. El grupo B toma el test traducido (test objetivo) y realiza una nueva traducción al idioma original. Por último se toman las dos versiones del test en el idioma original (test matriz y test retraducido) y se comparan para obtener equivalencias y determinar la calidad de las traducciones. Este proceso lo deben realizar los dos grupos en conjunto.

Además es importante contemplar los siguientes aspectos: Primero el resultado de la traducción del test matriz debe ser el que reciba especial atención pues será este el que se convertirá en el test objetivo; la retraducción al idioma original sirve para garantizar que no pierda la equivalencia de constructo. Segundo, se debe prestar atención a suposiciones y presunciones que pueden hacerse en la labor de traducción. Tercero, se debe tener cuidado con las palabras, frases e ítems que presenten varios significados. Cuarto, es necesario vigilar que aspectos innecesarios e inapropiados del test matriz no permanezcan en el test objetivo Mikulic (2007).

Se sugiere utilizar los dos métodos para realizar la traducción de los ítems. Pues permiten obtener una equivalencia de constructo más alta y confiable, disminuyendo la probabilidad de obtener traducciones inapropiadas, que puedan afectar posteriormente los resultados del test adaptado.

MANUAL ADAPTACIÓN

Recomendaciones para una correcta traducción:

Para el caso de los tests que fueron contruidos en países o culturas hispanohablantes es necesario realizar una adaptación lingüística. Puesto que existen diferencias fonéticas, léxicas y/o morfosintácticas, además de las socioculturales que debe tener en cuenta.

Se recomienda utilizar para el anterior caso el tipo de traducción directa (forward translation) pues permite contrastar las diferencias lingüísticas y los posibles significados que pueda llegar a tener una palabra.

Para el caso de los tests que contienen dentro de sus ítems-estímulos (figuras, dibujos, sonidos, texturas, formas armables, etc.) o están compuestos únicamente por estímulos que hacen parte del instrumento. Se debe realizar un estudio al interior de la población objetivo, donde se asegure que los estímulos utilizados en el test matriz sean familiares para todos ellos.

Recomendaciones: Recopilar evidencia sistemática de juicio tanto psicológica y lingüística, suficiente como para mejorar la precisión del proceso de traducción y adaptación. Se recomienda usar procedimientos de traducción estandarizada. Además se recomienda hacer las traducciones con un grupo de expertos ya que está comprobado que aumenta la calidad de las traducciones.

Asegurar que el diseño de la prueba permite utilizar técnicas estadísticas para establecer la equivalencia del ítem entre las distintas versiones idiomáticas del test. Es conveniente que se utilice un grupo de expertos bilingües, aunque constituye una tarea difícil.

Se deben utilizar técnicas estadísticas apropiadas para establecer la equivalencia entre las diferentes versiones del test e identificar aspectos o componentes problemáticos del instrumento que puedan resultar inadecuados a una o más poblaciones a las que va dirigido.

Se debe también asegurar que la validez del instrumento se mantiene de una población a otra ya que normalmente cuando se hacen estudios transculturales existen

MANUAL ADAPTACIÓN

grandes diferencias que obligan a modificar los ítems para que sean comprendidos y se adapten de manera precisa a cualquier población (Acosta, 2005).

También se debe garantizar evidencia estadística de la equivalencia entre ítems. Está claro que en estudios transculturales al momento de la adaptación se pueden presentar grandes diferencias en el contenido de los ítems. Así que debe proveerse este tipo de evidencia para todas las poblaciones a las que va dirigido el instrumento.

Referencias:

- Acosta, B. (2005). Tesis doctoral: La adaptación al español del inventario de autogobierno para su uso en México y España. Universidad de Granada.
- Mikulic, I. (2007). Construcción y adaptación de pruebas psicológicas. Universidad de Buenos Aires.
- Muñiz, J., & Hambleton, R.K. (1996). Directrices para la traducción y adaptación de los tests. Papeles del Psicólogo, 66, p. 63-70.

FASE 5.VALIDACIÓN POR JUECES

Hace referencia a la selección y evaluación de los jueces además de la validación de los indicadores de la prueba. La opinión de los expertos tiene por finalidad analizar la correspondencia entre el contenido de las pruebas y los constructos que intentan medir.

Palabras clave: Jueces, evaluación, ítems.

Contexto: Se selecciona una muestra intencional de diez profesionales afines al área de la Psicometría tanto metodológicos como conceptuales que cuentan con experiencia y conocimientos en la teoría y metodología de la prueba psicológica. Camargo y Pardo (2008) afirman que en esta fase se hará la revisión y validación de los indicadores por parte de diez jueces o más. Finalmente se deben hacer las correcciones del instrumento que sean pertinentes.

MANUAL ADAPTACIÓN

Es importante reconocer según lo afirma Rico (citado por González 2000, p. 21) “La evaluación es el enjuiciamiento sistemático de la valía o el mérito de un objeto. Esta definición se centra en que la evaluación siempre supone un juicio”. Dado que la misma revela o permite seleccionar, jerarquizar, diagnosticar y medir procesos tanto conceptuales para analizar los ítems y su consistencia y así poder determinar la pertinencia y claridad de estos.

Así mismo Lara y Sáyago (2012) incluyen la selección de tres jueces expertos en construcción de instrumentos, cuatro jueces en la variable a medir y tres jueces en el constructo que se pretende medir, quienes evalúan los ítems según criterio y forma. Después se entrega un formato de evaluación a los jueces que evalúa; comprensión (indicador claro y fácil de entender), correspondencia (cada ítem debe corresponder al aspecto a evaluar) y pertinencia (cada indicador debe ser evaluado por el experto).

Por último se realizan las correcciones pertinentes y reestructurar el formato del instrumento según las observaciones de los jueces. Después de esto se construye nuevamente el instrumento con los ítems pertinentes.

En el siguiente paso los jueces deben llevar a cabo un consenso donde se realiza la revisión y validación del instrumento para evaluar la calidad de la traducción, la confiabilidad y validez de los ítems en la prueba y observar que sean aplicables al contexto del test.

Recomendación: Para calcular la concordancia entre jueces se recomienda utilizar los estadísticos Kappa y Kendall, el primero es el índice o coeficiente utilizado en datos nominales. Y Kendall, es un coeficiente de acuerdo para datos ordinales, basándose en el grado de varianza de la suma de rangos obtenidos por el número de jueces (Escobar y Cuervo, 2008)

MANUAL ADAPTACIÓN

Referencias:

Camargo, I & Pardo, C. (2008). Competencias docentes de profesores de pregrado: diseño y validación de un instrumento de evaluación. *Universitas psychologica* (7), p. 441-457.

Escobar, J. & Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición* (6), p 27-36.

González S., (2000). Acerca de la evaluación en matemáticas. En revista *Educación y Cultura*. Instituto Colombiano para el Fomento de la Educación Superior (2003), Bogotá: ICFES.

Lara, M & Sáyago, G. (2012). Diseño, construcción y validación de un instrumento de medición sobre las características del liderazgo desde la teoría situacional integracionista. *Cuadernos Hispanoamericanos de Psicología* (1), p. 7-1

FASE 6. PILOTAJE

En esta fase se entrega la guía para realizar un proceso de pilotaje de manera correcta, el constructor del test conoce cuáles son sus elementos y características más relevantes para obtener datos de la prueba confiables

Palabras clave: Comprobación, aplicación previa.

Contexto: El pilotaje es un proceso que busca comprobar si son pertinentes los ítems y el diseño de la prueba, estima qué tan efectivos son los datos por medio de la aplicación previa del instrumento. Inostroza, Morales, Sanhueza y Vera (2008) argumentan que la prueba piloto es fundamental porque se conocen los resultados arrojados del instrumento y sugieren cambios a efectuar antes de realizar un muestreo a escala completa, al final del proceso se da paso a realizar la investigación formalmente ya que los datos son confiables, seguros y útiles. En este aspecto se tiene en cuenta la selección de la muestra, la aplicación de la prueba piloto a un número considerable de

MANUAL ADAPTACIÓN

sujetos y el análisis de los resultados mediante algún programa tecnológico y se analiza bajo la teoría de respuesta al ítem u otro modelo.

Dentro del pilotaje se deben tener en cuenta las siguientes etapas descritas por Andrade (2012):

Consensuar la estructura. Consiste en el revisión conjunta de los componentes y aspectos del test que se utilizarán en la versión adaptada.

Establecimiento de criterios de validez.

Para cumplir con los criterios de validez se deben establecer equivalencias entre los indicadores de validación del test matriz y los resultantes del test adaptado. Se deben tomar los siguientes conceptos:

La validez de constructo hace referencia a la coherencia del instrumento con la teoría de la cual parte y la capacidad de medir el constructo teórico para el cual fue diseñado, esto último se puede constatar mediante un análisis factorial o análisis correlacional (Martínez, Guerrero y Rey, 2012).

La validez ecológica hace referencia a la necesidad de emplear medidas que exploren el grado en que las funciones evaluadas se generalizan a la vida real. Dicho de otro modo, intenta evaluar las relaciones entre los rendimientos del sujeto en un test y la capacidad de funcionar en su vida cotidiana (Tirapu, 2010).

En ocasiones las situaciones experimentales de algunas pruebas neuropsicológicas son tan artificiales que los resultados obtenidos tienen muy poco valor a la hora de predecir el nivel de funcionamiento real. Es posible encontrar déficits sutiles que no provocan dificultades en las tareas de vida diaria y viceversa. Por ello, la tendencia a la evaluación neuropsicológica en la actualidad es desarrollar nuevos instrumentos que exploren conductas y actividades similares a las características del medio natural donde se desarrolla habitualmente la vida de los pacientes.

MANUAL ADAPTACIÓN

La validez de contenido se refiere al grado en que el test presenta una muestra adecuada de los contenidos a los que hace referencia, sin omisiones y sin desequilibrios de contenido. También, es el grado en que los ítems que componen el test representan el contenido que el mismo trata de evaluar. Por tanto, la validez de contenido se basa en (a) la definición precisa del dominio y (b) en el juicio sobre el grado de suficiencia con que ese dominio se evalúa (Pérez, 2008).

Selección de la muestra. En este apartado, el muestreo analiza qué parte de la población debe examinarse, con la finalidad de hacer inferencias sobre una muestra específica. El muestreo intencional o de conveniencia es uno de los más utilizados puesto que el investigador selecciona directamente los individuos o grupos típicos (Van der Vijer, 1996). Sin embargo es necesario recordar que el tipo de muestreo que se utilice depende de la preferencia del investigador que realiza el pilotaje.

Redacción de un instructivo para la realización del pilotaje. El instructivo cuenta con preguntas claras de acuerdo a una secuencia lógica para que los participantes lo puedan contestar fácilmente.

Aplicación de los formatos. Se debe tener un especial cuidado cuando se aplican los formatos de test, ítems, opciones de respuesta múltiple y distintos procedimientos de aplicación, supervisando que estos resulten familiares a todas las poblaciones a los que se dirige la prueba.

Es necesario prestar atención a los tipos de formatos de test ya que existen diversas características que pueden alterar su validez. Una de estas son los formatos de múltiple respuesta que en numerosas ocasiones presentan un significado distinto entre las alternativas.

Otro elemento a subrayar son pruebas que requieren un delicado equilibrio entre velocidad y precisión en las respuestas. Además de esta característica se encuentra que en diversas ocasiones existen poblaciones que tienen muy poca experiencia en la realización de tests y esto puede terminar generando diferencias relevantes entre grupos. Cuando existe un peligro real de que estos factores puedan afectar el desempeño en el

MANUAL ADAPTACIÓN

desarrollo de la prueba los creadores y administradores de la misma deben brindar información acerca de cómo ellos dieron solución al problema.

Seguido a esto, corresponde la revisión del contenido del ítem y/o los estímulos de los materiales buscando que sean familiares a las poblaciones a las cuales va dirigida la prueba. Se deben tratar de anticipar los posibles problemas que puedan presentarse en el proceso y tomarse las acciones respectivas como la preparación de materiales e instrucciones pertinentes.

Es recomendable omitir detalles que obstaculicen la comprensión por parte de algunas culturas, al igual que elementos como las instrucciones deben estar diseñadas de la forma más convencional posible. Conviene cuidar bien otro aspecto, relacionado a las condiciones físicas del ambiente donde se administra el test, ya que puede influenciar los resultados considerablemente cuando las condiciones de los ambientes varían significativamente creando por lo tanto situaciones difíciles de replicar a la hora de llevar a cabo la administración de una prueba.

Las instrucciones del instrumento deben ir en el idioma matriz y objetivo con el fin de evitar la influencia de fuentes no deseadas de variación a través de las poblaciones a las que se dirige el instrumento.

El manual del test debe especificar todos los aspectos del instrumento y su administración. En la misma no debe haber factores que obstaculicen la aplicación de la prueba. La relación administrador-administrado debe ser minimizada. Además, corresponde incluir reglas explícitas que son descritas en el manual original de la prueba.

Por consiguiente cuando un instrumento es traducido/adaptado para su uso en otra población debe proveerse documentación sobre los cambios acompañada con evidencia de la equivalencia. Las diferencias de puntajes entre muestras poblacionales no deben ser tomadas como verdaderas en primera instancia. El investigador tiene la responsabilidad de confrontar las diferencias con otra evidencia empírica.

MANUAL ADAPTACIÓN

Comparaciones a través de poblaciones solo pueden hacerse al nivel de invarianza que ha sido establecido por la escala sobre la cual los puntajes han sido reportados.

El desarrollador del instrumento debe proveer información específica sobre las maneras en las cuales los contextos sociocultural y ecológico de las poblaciones pueden afectar el desempeño en el instrumento y deben sugerir procedimientos para contabilizar esos efectos en la interpretación de resultados.

Recopilación y clasificación de la información obtenida. Para la organización del análisis de los datos se recurre a un software estadístico, que a la vez verifica la confiabilidad de la prueba y verifica su consistencia interna y se obtiene los índices psicométricos y descriptivos básicos, los cuáles son la consistencia interna de los elementos y las correlaciones de los ítems. Una vez obtenidos estos datos, se comparan con los de la prueba original, si existen valores diferentes o que no concuerdan, se prosigue a replantear el contenido de los ítems malogrados.

Interpretación de los resultados. Cuando se interpreten los datos que arroja la prueba piloto se presenta la siguiente estructura. Interpretación de los análisis de las distribuciones de puntuaciones totales en las escalas y sub-escalas, interpretación de ítems y consistencia interna, interpretación de la estructura factorial del inventorial y finalmente la interpretación de las relaciones entre las escalas y sub-escalas.

Las divergencias entre las puntuaciones resultantes por las muestras a las que se les aplicó la prueba, no deben utilizarse directamente. Los investigadores tienen la responsabilidad de comparar las diferencias con otra información científica. De igual forma, en esta fase se incluye la depuración de datos la cual busca inconsistencias como registros duplicados o incorrectos en la base de datos que se encontraron dentro del pilotaje.

Para Vallo (2000) el diseñador del test debería proporcionar información específica acerca de las distintas formas en la que los contextos socio-culturales y ecológicos de las poblaciones pueden afectar el rendimiento del test, y debería sugerir procedimientos para tener en cuenta estos efectos en la interpretación de los resultados.

MANUAL ADAPTACIÓN

Toma de decisiones para hacer algunos cambios de los instrumentos. Autores como Ding y Hershberger (2002), definen que la opinión de cada uno de los expertos evidencia juicios y valores de cualquier investigación. En este proceso cuenta la toma de decisiones basada en evidencia, estudios, publicaciones etc. Así mismo la imparcialidad de los expertos dependiendo del nivel de experticia y la diversidad del conocimiento. La decisión es adecuada si hay una cantidad considerable de expertos.

Referencias:

- Andrade, X. (2012). Informe de pilotaje de instrumentos de evaluación del desempeño docente. (Tesis de pregrado). Universidad Central del Ecuador.
- Bausela, E. (2005). SPSS: Un instrumento de análisis de datos cuantitativos. *Revista de informática educativas y medios visuales* (4), p.62-69.
- Callegaro, J., Figueiredo, B. & Ruschel, B. (2012). Adaptación y validación de instrumentos psicológicos entre culturas: algunas consideraciones. *Revista Scielo* (22), 53, 423-432.
- Ding, C & Herscheberger, S. (2002). Assesing content validity and content equivalence using structural equation modeling. *Structural equation modeling: A multidisciplinary journal*, 9 (2), p. 283-297.
- Escobar, J. & Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición*, 6, p. 27- 36.
- Inestroza, G., Morales, S., Sanhueza, C & V, M. (2008). Estudio para el diseño y pilotaje de una propuesta de formación de mentores de educadores de párvulos (Trabajo de grado). Universidad Católica de Temuco, Chile.
- Martínez, J, Guerrero, S & Rey, C. (2012). Evaluación de la validez de constructo y la confiabilidad del inventario de masculinidad y feminidad en adolescentes y adultos jóvenes colombianos. *Avances en la psicología Latinoamericana*, 30 (1), p. 170-188.

MANUAL ADAPTACIÓN

Pérez, J. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición* (6), p. 27-36.

Tirapu, J. (2010). Validez ecológica de la exploración de las funciones ejecutivas. *Informaciones psiquiátricas* (200), p. 1-12.

Vallo (2000). Propiedad psicométricas del test de optimismo life orientation test. *Psicothema* 3 (14), p. 673-680.

Van de Vijver, F. R. K. (1996). Translating Tests: Some Practical Guidelines. *European Psychologist*, p. 89-99.

FASE 7. INTERPRETACIÓN DE LOS DATOS

Esta fase consiste en interpretar los resultados, analizar su importancia y generalización, para finalmente valorar sus implicaciones. En su desarrollo se necesita de un psicómetra que realice un análisis estadístico y calcule los errores a nivel de significancia con respecto al tamaño de la muestra.

Palabras clave: Análisis, ajustes, generalización, resultados.

Contexto: En esta fase, es necesario ordenar los datos, describir la muestra y verificar la confiabilidad de la medición. Para este proceso se recurre a la estadística inferencial permitiendo hacer deducciones acerca de la probabilidad de las relaciones observadas en una muestra seleccionada de la población en general (Sim y Wright, 2005).

A la vez, se realiza un análisis de los datos recogidos durante la fase de pilotaje, centrándose en: los índices psicométricos, ordenamiento de los índices de dificultad y los índices globales.

Los índices psicométricos, comparan los índices de dificultad con los de discriminación obtenidos en la prueba original, a su vez el ordenamiento de los índices

MANUAL ADAPTACIÓN

de dificultad se basa en el manual del test original, comparando e interpretando cada ítem según el cuartil de dificultad que posee y por último los índices globales, se enfocan en el tamaño de la muestra, explorando que dimensión es relevante en la prueba, obteniendo un factor general del instrumento a través de un análisis de la varianza y el coeficiente alfa ordinal (Almeida y Lemos, 2006 citado en Elosua, Mujika, Almeida y Hermosilla, 2014).

Dentro del análisis de los ítems, se debe cuidar la validez de criterio y constructo; en cuanto a la primera, se debe hacer una descripción de la muestra y un análisis estadístico para determinar el grado de exactitud predictiva puesto que las medidas de tendencia central controlan los errores estándar y la validez de constructo proporciona el sustento que incluye los conocimientos o habilidades que se evalúan en la prueba y que resultan indispensables para el constructo que se quiere medir.

Al mismo tiempo en esta fase se tiene en cuenta la validez concurrente que consiste en el nivel hasta el cual la prueba mide lo mismo que miden otros tests del constructo a evaluar. Según lo mencionado por Argibay (2006) la validez concurrente busca la relación existente entre las puntuaciones de la prueba adaptada con alguna medición de criterio (otra prueba similar). Para este caso se suponen la prueba original y la prueba adaptada como mediciones adecuadas del mismo rasgo. (Casas, 2011). El inconveniente de esta clase de validez consiste en encontrar pruebas que funcionen como criterios efectivos y confiables (Aliaga, 2006).

Dentro de este proceso estadístico, se evalúa la validez, así como la confiabilidad y el error de medida de la prueba puesto que para cada puntaje global, parcial o combinado se deberá proporcionar información detallada sobre los niveles de confiabilidad, así como la forma de expresarla (componentes de varianza, coeficientes de correlación) y los errores estándar de medida, que faciliten al usuario la decisión de si el grado de precisión es suficiente para el uso que se pretenda dar a la prueba (Gempp, 2006).

MANUAL ADAPTACIÓN

Referencias:

- Aliaga, L. (2006). Psicometría: test psicométricos, confiabilidad y validez. En Quintana A. & Montgomery W. Psicología. Tópicos de actualidad. Lima UNMSM, p. 85-108.
- Argibay, J. (2006) Técnicas psicométricas. Cuestiones de validez y confiabilidad. Subjetividad y Procesos Cognitivos 8 p.15.
- Casas, L. (2011). Elaboración de instrumentos: validez de criterio. Recuperado el día 14 de Marzo de 2015 de <https://luiscasasvilchis.wordpress.com/2011/09/15/elaboracion-de-instrumentos-validez-de-criterio>
- Elosua, P., Mujika, J., Almeida, A & Hermsilla, D. (2014). Procedimientos analíticos- racionales en la adaptación de test. Adaptación al español de la batería de pruebas de razonamiento. *Revista latinoamericana de Psicología* 46 (2), 117-126.
- Gempp, R. (2006). El error estándar de medida y la puntuación verdadera de los test psicológicos; algunas recomendaciones prácticas. *Terapia psicológica* 24, (2), p. 117-130.
- Sim, J & Wright, C. (2005). The kappa statistic in reliability studies: Use, interpretation and sample size requirements. *Physical Therapy*, 85 (3), p. 257-268.

FASE 8. ANÁLISIS ASPECTOS PSICOMÉTRICOS DE LA PRUEBA

En esta fase se describen los aspectos psicométricos que debe tener un instrumento adaptado.

Palabras clave: sistematización, validez, confiabilidad, baremación.

El objetivo de un análisis psicométrico es la generalización de los datos obtenidos teniendo en cuenta los estimativos de parámetros poblacionales.

Contexto Se deben tratar de minimizar los errores en el proceso de adaptación, aplicando técnicas estadísticas apropiadas para establecer la equivalencia entre las diferentes versiones de un test. De igual forma se realiza una evaluación de la validez de

MANUAL ADAPTACIÓN

los resultados que arroja el test adaptado, se efectúa una comparación con la información del test original para conseguir minimizar las fuentes de variación no deseadas (ITC, 2005).

Para obtener la validez de contenido se escogen los ítems a través de una selecta evaluación. Cuando la variable posee distintas dimensiones, se realiza un muestreo probabilístico de ítems. Es necesario relacionar las puntuaciones de los ítems, debe existir correlaciones altas, específicamente entre ítems que miden una misma dimensión (Bohrnstedt, 1976 en Hernández, Fernández y Baptista, 2010), finalmente mediante mediciones estadísticas se comprueba la representatividad de la muestra.

Seguido a esto se debe obtener la validez de criterio. Este tipo de validez es más fácil de lograr, lo que se debe hacer es correlacionar la medición con el criterio, y este coeficiente resultante se asume como el de validez (Bohrnstedt, 1976 en Hernández, Fernández y Baptista, 2010).

Finalmente el último tipo de validez a tener en cuenta es el de constructo que consiste en encerrar toda evidencia que sustente la interpretación o el significado de los puntajes en una prueba. Para obtener esta validez el procedimiento más utilizado es el análisis factorial, el cual se lleva a cabo a través de la utilización de un software estadístico. También se puede obtener mediante métodos intrapruebas utilizando el coeficiente alfa de Cronbach o los coeficientes de Kuder-Richardson (Aliaga, 2006).

El siguiente paso es obtener la confiabilidad del instrumentado adaptado a través del método pre y posprueba, método de equivalencia, método Split half y el método de estabilidad y equivalencia. Más adelante se debe lograr la consistencia interna para lo cual se usan los coeficientes Kuder-Richardson KR20 y KR21. Tenga en cuenta que únicamente pueden aplicarse a tests que abarquen reactivos los cuales presenten respuestas dicotómicas. Otra forma de obtener la consistencia interna es a través del alfa de Cronbach. Sirve para ítems de alternativas múltiples (Bonilla, 2006).

Por último se debe realizar el proceso de baremación, el cual es la transformación de puntajes brutos en puntajes derivados (percentil y calificación estándar), producto de una

MANUAL ADAPTACIÓN

operación matemática. La puntuación directa de un sujeto en una prueba no es directamente interpretable, si no se vincula con la materia de estudio incluida en la prueba o con la performance de los demás individuos que hacen parte de la agrupación normativa (Sireci y Padilla, 2014).

Referencias

- Aliaga, L. (2006). Psicometría: test psicométricos, confiabilidad y validez. En Quintana A. & Montgomery W. Psicología. Tópicos de actualidad. Lima UNMSM, p. 85-108.
- Bonilla, V. (2006). INEVA en acción. Vol. (2) 3 p. 1-3. Recuperado de www.ineva.uprrp.edu/boletin/v0002n0003.pdf
- Hernández, R. Fernández, C. & Baptista, M. (2010). *Metodología de la investigación*, México D.F. México, McGrawHill.
- International Test Commission (2005). International Guidelines on Test Adaptation [www.intestcom.org]
- Tornimbeni, S., Pérez, E., Olaz, F & Fernández, A. (2004). Introducción a los test psicológicos (3) Editorial Brujas: Argentina.
- Sireci, S., & Padilla, J. (2014). Validating assessments: Introduction to the Special Section. *Psicothema*, Vol (26) 1, p. 97-99.

BIBLIOGRAFIA

- Acosta, B. (2005). Tesis doctoral: La adaptación al español del inventario de autogobierno para su uso en México y España. Universidad de Granada.
- Aliaga, L. (2006). Psicometría: test psicométricos, confiabilidad y validez. En Quintana A. & Montgomery W. Psicología. Tópicos de actualidad. Lima UNMSM, p. 85-108.
- Andrade. X. (2012). Informe de pilotaje de instrumentos de evaluación del desempeño docente. (Tesis de pregrado). Universidad Central del Ecuador.
- Argibay, J. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. Subjetividad y procesos cognitivos (8), p.15-33.
- Asociación Americana De Investigación Educativa (2015). Permission for authors. Recuperado el día 21 marzo de 2015 de <http://www.aera.net/Publications/PublicationsPermissions/tabid/13120/Default.aspx>
- Bausela, E. (2005). SPSS: Un instrumento de análisis de datos cuantitativos. *Revista de informática educativas y medios visuales* (4), p.62-69.
- Bonilla, V. (2006). INEVA en acción. Vol. (2) 3 p. 1-3. Recuperado de www.ineva.uprrp.edu/boletin/v0002n0003.pdf
- Callegaro, J., Figueiredo, B. & Ruschel, B. (2012). Adaptación y validación de instrumentos psicológicos entre culturas: algunas consideraciones. *Revista Scielo* (22), 53, 423-432.
- Camargo, I & Pardo, C. (2008). Competencias docentes de profesores de pregrado: diseño y validación de un instrumento de evaluación. *Universitas psychologica* (7), p. 441-457.
- Casas, L. (2011). Elaboración de instrumentos: validez de criterio. Recuperado el día 14 de Marzo de 2015 de <https://luiscasavilchis.wordpress.com/2011/09/15/elaboracion-de-instrumentos-validez-de-criterio>

MANUAL ADAPTACIÓN

- Ding, C & Herscheberger, S. (2002). Assessing content validity and content equivalence using structural equation modeling. *Structural equation modeling: A multidisciplinary journal*, 9 (2), p. 283-297.
- Dirección Nacional de Derechos de Autor (DNDA) ¿Qué es el derecho de autor? Recuperado el día 21 marzo de 2015 de <http://www.derechodeautor.gov.co/web/guest/preguntas-frecuentes>
- Elosua, P. Mujika, J. Almeida, L. & Hermosilla, D. (2014). Procedimientos analítico-rationales en la adaptación de tests. Adaptación al español de la batería de pruebas de razonamiento. *Revista Latinoamericana de psicología* (46) 2 p.117-126.
- Escobar, J. & Cuervo, A. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición* (6), p 27-36.
- Gempp, R. (2006). El error estándar de medida y la puntuación verdadera de los test psicológicos; algunas recomendaciones prácticas. *Terapia psicológica* 24, (2), p. 117-130.
- González S., (2000). Acerca de la evaluación en matemáticas. En revista Educación y Cultura. Instituto Colombiano para el Fomento de la Educación Superior (2003), Bogotá: ICFES.
- Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology* , 6, (2), p. 29-45.
- Hambleton, R.K. (2000). Issues designs and technical guidelines for adapting tests in multiple languages and cultures.
- Hernández, R. Fernández, C. & Baptista, M. (2010). *Metodología de la investigación*, México D.F. México, McGrawHill.
- Inestroza, G., Morales, S., Sanhueza, C & V, M. (2008). Estudio para el diseño y pilotaje de una propuesta de formación de mentores de educadores de párvulos (Trabajo de grado). Universidad Católica de Temuco, Chile.

MANUAL ADAPTACIÓN

International Test Commission (2005). International Guidelines on Test Adaptation. [www.intestcom.org]

Lara, M & Sáyago, G. (2012). Diseño, construcción y validación de un instrumento de medición sobre las características del liderazgo desde la teoría situacional integracionista. *Cuadernos Hispanoamericanos de Psicología* (1), p. 7-1

Ley N° 23 de 1982. Congreso De la Republica de Colombia: Régimen general de derechos de autor.

Martínez, J, Guerrero, S & Rey, C. (2012). Evaluación de la validez de constructo y la confiabilidad del inventario de masculinidad y feminidad en adolescentes y adultos jóvenes colombianos. *Avances en la psicología Latinoamericana*, 30 (1), p. 170-188.

Mikulic, I. (2007). Construcción y adaptación de pruebas psicológicas. Universidad de Buenos Aires.

Ministerio de la Protección social (2006). Ley número 1090 de 2006: Código deontológico y bioético de la profesión de Psicología.

Muñiz, J, Elosua, P & Hambleton, R. (2013). Directrices para la traducción y adaptación de los test. Segunda edición. *Revista papeles del psicólogo* (66), p. 1-8.

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

Olarte, J & Rojas, M. (2010). La protección del derecho de autor y los derechos conexos en el ámbito penal. Dirección Nacional De Derechos de Autor; Ministerio del Interior y de justicia.

Pérez, J. (2008). Validez de contenido y juicio de expertos: una aproximación a su utilización. *Avances en Medición* (6), p. 27-36.

MANUAL ADAPTACIÓN

- Sim, J & Wright, C. (2005). The kappa statistic in reliability studies: Use, interpretation and sample size requirements. *Physical Therapy*, 85 (3), p. 257-268.
- Sireci, S., & Padilla, J. (2014). Validating assessments: Introduction to the Special Section. *Psicothema*, Vol (26) 1, p. 97-99.
- Tirapu, J. (2010). Validez ecológica de la exploración de las funciones ejecutivas.
- Tornimbeni, S., Pérez, E., Olaz, F & Fernández, A. (2004). Introducción a los test psicológicos (3) Editorial Brujas: Argentina.
- Vallo, I. (2000). Directrices para la adaptación y uso de los test. Madrid; España.
- Van de Vijver, F. R. K. (1996). Translating Tests: Some Practical Guidelines. *European Psychologist*, p. 89-99.
- Vega, A. (2010). Manual de Derechos de Autor. Dirección Nacional de Derechos de Autor: Bogotá.
- Xalabarder, R. (2005). Copyright y derecho de autor ¿Convergencia internacional en un mundo digital? *Revista de internet, derecho y política*. Información recuperada de: <http://www.uoc.edu/idp/1/dt/esp/xalabarder02.pdf>