

EVALUACIÓN PSICOLÓGICA
Títulos publicados

1. A. Howell y C. Amatruda, *Diagnóstico del desarrollo normal y anormal del niño*
7. D. W. Harris, *El test de Goodenough*
8. H. A. Murray, *Test de Apercepción Temática (TAT)*
9. W. Klopfer y D. Kelly, *Técnica del Psicondiagnóstico de Rorschach*
10. L. Bender, *Test Gestáltico Visomotor*
13. L. Bellak y S. S. Bellak, *Test de Apercepción Infantil con figuras animales (CAT-A)*
14. L. Bellak y S. S. Bellak, *Test de Apercepción Infantil con figuras humanas (CAT-H)*
18. E. Anstey, *Test de dominós*
26. E. F. Hammer, *Test proyectivos gráficos*
33. R. Stora y otros, *El test del árbol*
35. G. F. Kuder, *Cuestionario general de intereses*
56. G. H. Bennett y otros, *Tests de Aptitudes Diferenciales (DAT)*
57. J. C. Raven, *Test de Matrices Progresivas*
61. M. M. Casullo y otros, *Proyecto de vida y decisión vocacional*
62. F. J. Sainz y L. Gorospe, *El test de Rorschach y su aplicación en la psicología de las organizaciones*
63. M. M. Casullo (comp.), *Evaluación psicológica en el campo de la salud*
64. G. J. Neimeyer (comp.), *Evaluación constructivista*
65. M. M. Casullo (comp.), *Evaluación psicológica en el campo socioeducativo*
67. T. Millon, *Inventario Millon de Estilos de Personalidad (MIPS)*
68. H. Phillipson, *Test de Relaciones Objetales (TRO)*
69. T. B. Brazelton y J. K. Nugent, *Escala para la evaluación del comportamiento neonatal*
70. T. Kirchner, M. Torres y M. Forns, *Evaluación psicológica: modelos y técnicas*
71. J. P. Choca y E. Denburg, *Guía para la interpretación del MCMI (Millon Clinical Multiaxial Inventory)*
72. R. A. Neimeyer, *Métodos de evaluación de la ansiedad ante la muerte*
73. National Center for Clinical Infant Programs, *Clasificación diagnóstica: 0-3*
74. M. M. Casullo (comp.), *Aplicaciones del MMPI-2 en los ámbitos clínico, forense y laboral*
75. A. N. Cayssials, *La escala de inteligencia WISC-III en la evaluación psicológica infanto-juvenil*
76. T. Millon, *Inventario Millon de Estilos de Personalidad (MIPS)*
78. A. Castro Solano, M. M. Casullo y M. Pérez, *Aplicaciones del MIPS en ámbitos laborales, educativos y médicos*
79. A. Castro Solano, *Técnicas de evaluación psicológica en los ámbitos militares*
80. E. Pérez, J. Pássera, F. Olaz y M. Osuna, *Orientación, información y educación para la elección de la carrera*
81. V. Feld, I. Taussik y C. Azaretto, *Pro-cálculo. Test para la evaluación del procesamiento del número y el cálculo en niños*
82. A. T. Beck, R. Steer y G. Brown, *BDI-II Inventario de depresión de Beck*
83. S. Defior Citoler, L. Fonseca y B. Gottheil, *Test de lectura y escritura en español (LEE)*
84. D. Burin, M. Drake y P. Harris (comps.), *Evaluación neuropsicológica en adultos*
85. S. Tornimbeni, E. Pérez y F. Olaz, *Introducción a la psicometría*

SILVIA TORNIMBENI
EDGARDO PÉREZ
FABIÁN OLAZ

Con la colaboración de
NURIA CORTADA DE KOHAN
ALBERTO FERNÁNDEZ
MARCOS CUPANI

INTRODUCCIÓN A LA PSICOMETRÍA



PAIDÓS

Buenos Aires
Barcelona
México

Alberto Fernández

7.1. Por qué adaptar tests

La utilización de tests psicológicos creados en otros contextos culturales es una práctica habitual en todo el mundo. Este fenómeno es particularmente frecuente en los países de las regiones con menor desarrollo científico, tales como Latinoamérica y África. Sin embargo, también ocurre en las naciones más desarrolladas, donde usualmente se utilizan los tests construidos en los Estados Unidos, sin duda el país en el que se producen la mayoría de los instrumentos de medición psicológica.

El uso de un test en un contexto cultural diferente al original genera diversas dificultades. El idioma, la familiaridad con los estímulos del test (ítems) y las diferentes características de las muestras de estandarización son ejemplos de fuentes de posibles sesgos en la medición transcultural de constructos psicológicos.

Existe sesgo en la medición cuando las diferencias individuales en las puntuaciones de un test no reflejan las diferencias reales en un rasgo o habilidad. Tomemos como ejemplo una hipotética investigación en la cual se midiese la capacidad de denominar objetos o animales. Si entre los ítems estuvieran incluidas las figuras de un canguro y un oso koala es más probable que una muestra de estudiantes australianos (lugar en el que viven estos animales) obtenga puntuaciones superiores a las de una muestra de estudiantes nigerianos e incluso a la de muchos de los lectores de este texto. Estos datos no estarían demostrando una mayor capacidad de denominación de los estudiantes australianos, sino que sugerirían que el indicador empleado en la medición está sesgado. Es decir, existe un elemento diferente (en este ejemplo, la familiaridad con el estímulo presentado) de la capacidad (de denominación en este caso) que influye en el desempeño en la prueba. Por ende, no se estaría midiendo la habilidad en forma equivalente en ambos casos, a pesar de estar utilizando la misma prueba. La equivalencia puede definirse por lo tanto de manera opuesta al sesgo, es decir, se manifiesta cuando las puntuaciones de un test reflejan diferencias entre las personas evaluadas que existen verdaderamente en el rasgo en cuestión.

La existencia de sesgo en un test puede conducir a obtener resultados gravemente erróneos. En una prueba utilizada en psicología clínica, por ejemplo, se podría inferir la presencia de un trastorno de personalidad cuando el rasgo o comportamiento así diagnosticado es normal en la cultura del individuo examinado.

Además de la obvia necesidad de contar con instrumentos adecuados para la práctica psicológica, la adaptación de tests a diferentes culturas obedece a otras razones, de índole científicas y prácticas.

Entre las primeras, es importante tener en cuenta que la mayoría de las teorías psicológicas contemporáneas han sido desarrolladas en el marco de la cultura occidental, más precisamente en universidades norteamericanas. Asimismo, el

proceso de validación de dichas teorías se realiza preferentemente en investigaciones que utilizan muestras de jóvenes universitarios de raza blanca.

En la actualidad se reconoce la necesidad de demostrar la "universalidad" de las teorías, si es que ello fuese posible. De este modo, para poder determinar si determinado constructo psicológico existe en otras culturas es necesario contar con tests equivalentes, es decir, que midan el mismo constructo en las diferentes culturas donde va a ser utilizado.

Las razones prácticas se relacionan con la dinámica de la globalización y los fenómenos migratorios, principalmente dentro de los países más desarrollados. Las personas que son evaluadas en un proceso de selección de personal, por ejemplo, provienen de diferentes partes del mundo y, para que esa evaluación sea justa, es necesario contar con instrumentos adecuados (equivalentes o sin sesgo). Lo mismo ocurre en los casos en los que se administran tests a individuos pertenecientes a minorías étnicas de un país.

7.2. Fuentes de sesgo

Van de Vijver y Tanzer (1997) identificaron diferentes fuentes de sesgo que se describen a continuación.

Sesgo de constructo

Este sesgo se presenta "cuando el constructo medido no es idéntico en diferentes grupos culturales" (Van de Vijver y Tanzer, 1997, p. 264.). Comportamientos morales que en algunas sociedades pueden ser normales en otras pueden constituir un verdadero rasgo de rigidez y asemejarse a una conducta obsesivo-compulsiva. McCrae, Yik, Trapnell, Bond y Paulhus (1998) encontraron, por ejemplo, importantes diferencias entre los perfiles de personalidad de estudiantes canadienses y estudiantes chinos utilizando versiones equivalentes del NEO PI-R. Los estudiantes chinos obtuvieron puntajes significativamente menores en algunas facetas de la escala Extraversión de este último test.

Sesgo metodológico

Este tipo de sesgo reconoce tres formas:

- a) El *sesgo de muestra*, que ocurre cuando las muestras son incomparables entre sí. La cantidad de años de escolaridad que poseen los individuos de una muestra constituye una variable determinante en el desempeño de los mismos en un test determinado, especialmente si se trata de un test de ejecución máxima (Heaton, Grant y Matthews, 1991). Los tests de razonamiento lógico o matemático, por ejemplo, presentan una dificultad considerablemente mayor para las personas con baja escolaridad. El nivel socio-cultural, la motivación, el sexo y la edad de los sujetos son otras de las variables que pueden hacer incomparables a dos muestras.

Fernández y Marcopulos (2004) compararon los estudios normativos de un test de atención en diez países y advirtieron que una de las principales dificultades para comparar los puntajes residía en las diferencias entre las muestras. Así, observaron que la edad media de los ancianos de la muestra neozelandesa estaba

1,6 desviaciones estándar por encima de la muestra danesa. Sin embargo, en la muestra danesa el 80% de los individuos tenía siete o menos años de educación mientras que la muestra neozelandesa tenía un promedio de diez años de educación. De este modo, no era posible comparar los puntajes de ambas muestras puesto que el nivel educativo (variable que tiene una gran influencia en el desempeño en este test) era muy diferente. Por consiguiente, las diferencias observadas en los puntajes del test posiblemente no reflejaban diferencias reales en la habilidad atencional sino el nivel educativo de los grupos. Es probable que la inclusión de una muestra de individuos daneses con el mismo nivel educativo que los neozelandeses hubiera reducido notablemente esas diferencias.

- b) El sesgo en el instrumento puede provenir de las características del test. La familiaridad que los sujetos tengan con los ítems presentados es un aspecto de gran importancia. Algunos estímulos tales como objetos, dibujos, figuras u otros elementos utilizados en algunas culturas no existen en otras o son irrelevantes. El ítem de ejemplo en el subtest de Ordenamiento del WISC-III (Wechsler, 1994) que muestra a una mujer frente a una máquina expendedora de latas de gaseosa tiene muy poco valor en culturas árabes, por ejemplo, o aun en zonas rurales de nuestro país. Las máquinas expendedoras de gaseosas no son comunes en todo el mundo y mucho menos su utilización, lo que puede hacer incomprensible dicha lámina para muchos individuos.

El idioma es otra fuente de sesgo del instrumento. La traducción es un problema considerable y requiere una metodología específica que se explicará más adelante. Los problemas pueden ser aún mayores cuando, por ejemplo, los idiomas son tan distintos como el inglés y el árabe, en los cuales la lectura se realiza de izquierda a derecha y de derecha a izquierda, respectivamente. En algunos tests que incluyen completamiento de frases, comprensión lectora o cancelación de letras, la disposición del texto tiene mucha relevancia. El problema es aún mayor cuando, por ejemplo, el estímulo del test es un carácter del alfabeto, como en el caso del Test del Trazo (TT, Trail Making Test), donde se deben conectar en forma alternada letras y números siguiendo el orden alfabético y un orden numérico ascendente. Incluso entre las lenguas occidentales existen diferencias en los caracteres del alfabeto: el inglés no contiene la "ñ", el alfabeto sueco contiene las siguientes vocales, a, e, i, o, u, y, å, ä, ö; y en portugués existen la á y la á, entre otros ejemplos. En el caso del TT, Axelrod, Aharon-Peretz, Tomer y Fisher (2000) han adaptado una versión hebrea de este test utilizando los caracteres propios de ese idioma.

Los materiales o estímulos de respuesta constituyen otra fuente posible de sesgo del instrumento. Las láminas de respuestas del Test de Matrices Progresivas, de Raven (1993), que implican completar una secuencia lógica seleccionando una figura entre un grupo de alternativas, incluyen la figura faltante al final de la segunda fila con lo que se asume una lectura de izquierda a derecha. Este hecho fue observado por Carpenter, Just y Shell (1990), y constituye una severa desventaja para los individuos árabes, quienes involuntariamente van a intentar resolver la prueba de derecha a izquierda, el modo en que se lee su idioma.

- c) Finalmente, la última variedad de sesgo metodológico es el de administración. Esta categoría incluye los problemas de comunicación, es decir, dificultades para que el entrevistado entienda las instrucciones del entrevistador ya sea por el tipo de palabras utilizadas, la forma de suministrar las instrucciones o un inadecuado manejo del idioma por parte

del examinador o del examinado. También comprende los cambios introducidos en el modo de administración de la prueba puesto que, frecuentemente, los manuales de tests incluyen instrucciones que no son adecuadas para la población evaluada. Los administradores del test, entonces, optan por "adaptar" esas instrucciones según su criterio personal, lo cual puede conducir a severas distorsiones en la interpretación de los resultados obtenidos. En la investigación mencionada de Fernández y Marcopulos (2004) se observó que algunas modificaciones en la metodología para cronometrar un test de velocidad producían puntuaciones medias muy diferentes en una muestra de individuos daneses y otra de individuos de la Argentina.

Sesgo de ítem

El sesgo del ítem se genera cuando éste último posee diferentes significados en las culturas consideradas. Ciertos grupos culturales pueden obtener puntajes significativamente distintos en un ítem determinado a pesar de obtener un puntaje total similar en el test. La deseabilidad social o la relevancia cultural, entre otros factores, pueden producir el sesgo de ítem. Tanzer (1995), por ejemplo, demostró que, aunque la estructura factorial de un test de autoconcepto académico era semejante en muestras de estudiantes australianos y singapurenses, existían diferencias sustanciales entre ambos colectivos cuando se comparaban los puntajes obtenidos en algunos ítems específicos.

7.3. La influencia del lenguaje

El lenguaje es otro factor que afecta el desempeño en la evaluación transcultural que utiliza tests psicológicos. Lau y Hoosain (1999) demostraron que los individuos que hablan chino rinden más que los sujetos que hablan japonés en una prueba de cálculo mental. Estos últimos, a su vez, superaron en su desempeño a las personas angloparlantes. Los autores pudieron demostrar que estas diferencias estaban relacionadas con la duración de los dígitos cuando son pronunciados, lo que a su vez está vinculado con la memoria de trabajo. Cuando les administraron una prueba de supresión articulatoria,¹ las diferencias entre estos tres grupos desaparecieron, lo que apoya la hipótesis de que la menor duración de los dígitos en el japonés que en el inglés les otorgaba ventaja a los japoneses y, a su vez, la menor duración de la pronunciación de los dígitos en chino comparado con el japonés e inglés, les daba ventaja a los chinos sobre los otros dos grupos.

Esta investigación demostró claramente cómo el idioma puede producir diferencias en el rendimiento en un test determinado y, por consiguiente, advierte sobre la inconveniencia de utilizar baremos extranjeros. En este caso específico, si los investigadores japoneses o chinos hubieran utilizado baremos elaborados en Inglaterra para evaluar el rendimiento de un individuo en un test de cálculo mental (algo que fácilmente podría ocurrir) es probable que los examinados hubiesen calificado dentro de rangos normales cuando en realidad algunos podrían tener un déficit.

1. La capacidad de la memoria de trabajo puede estimarse a través de la amplitud de dígitos, es decir, la cantidad máxima de dígitos que puede recordar una persona en un ensayo. La supresión articulatoria se ejecuta para suprimir el componente fonológico de esta memoria a través de la repetición constante de una misma sílaba, por ejemplo "bah".

7.4. Métodos de adaptación

Actualmente se reconoce que la adaptación de un test es un proceso mucho más complejo que la mera traducción a un idioma diferente. La traducción del inglés al español del siguiente ítem de un inventario de personalidad: "I wouldn't enjoy vacationing in Las Vegas" por "No disfrutaría tomando mis vacaciones en Las Vegas" es correcta. Sin embargo este ítem probablemente tendrá un significado distinto para personas que no viven en los Estados Unidos. Es evidente que una traducción correcta no asegura un significado unívoco de un ítem en diferentes culturas.

Van de Vijver y Leung (1997) establecieron tres niveles de adaptación de las pruebas psicológicas. El primero corresponde a la aplicación, esto es, la simple traducción de un test asumiendo a priori la equivalencia de constructo. Desafortunadamente, éste es el método más utilizado y, como se afirmó anteriormente, el problema es que la mera traducción de un test no garantiza la equivalencia entre ambas versiones (la original y la traducida).

La segunda opción es la adaptación. En este caso, a la traducción se agrega la transformación, adición o sustracción de algunos ítems de la escala original. En efecto, algunos ítems pueden cambiar su significado a través de las culturas y por lo tanto necesitan ser modificados o eliminados. Asimismo, los ítems nuevos que no existían en la versión original del test pueden representar mejor el constructo en la población en la cual se administrará la versión adaptada. En una investigación se descubrió que el nivel de dificultad original de algunos ítems pertenecientes a los subtests de Comprensión, Vocabulario e Información del WISC-III (Wechsler, 1994) no era apropiado para una muestra de estudiantes argentinos, por lo que se propuso un nuevo ordenamiento de los ítems (Baldo, 2001).

Finalmente, en la opción de ensamble (*assembly*) el instrumento original se modifica tan profundamente que prácticamente se transforma en un nuevo test. De allí el término ensamble (unión, integración), puesto que este proceso implica una integración del instrumento original con los nuevos elementos incorporados. Esto acontece cuando muchos de los ítems del test original son claramente inadecuados para representar el constructo a medir. Esta situación es común en los tests de denominación confrontacional, donde se presenta una lámina con dibujos de objetos que el examinado debe nombrar. Las láminas de estas pruebas son ordenadas desde objetos fáciles a difíciles de denominar. Como puede advertirse, este ordenamiento puede variar considerablemente de una cultura a otra. Este fue el caso de la adaptación argentina del Test de Denominación de Boston (Allegrí y colaboradores, 1997). En la versión original, por ejemplo, la figura de una bellota está ubicada en el lugar número 32, mientras que en la versión argentina esa lámina se encuentra en el número 59 (el test consta de 60 láminas) y esta situación se presentaba en muchos ítems.

El ensamble también es recomendable cuando un constructo no está representado de forma adecuada por la versión original de un test que se desea usar en otra cultura. Los enfoques indigenistas² de la medición de la personalidad, por ejemplo, han promovido el desarrollo de tests distintos para medir rasgos de la personalidad no contemplados en las teorías desarrolladas en occidente, tales como la de los cinco grandes factores.

2. "Psicología indigenista es el estudio de la conducta humana y los procesos mentales dentro de un contexto cultural que descansa sobre los valores, conceptos, sistemas de creencia, y otros recursos del grupo étnico o cultural que se investiga" (Ho, 1998: 94).

Tal es el caso del Inventario Chino de Evaluación de la Personalidad (Ho, 1998), que mide dimensiones indigenistas de la personalidad tales como "armonía", junto a los otros factores reconocidos en occidente.

En algunos casos, los límites entre "ensamble" y construcción de un nuevo test son poco nítidos, tal como puede apreciarse en este último ejemplo.

En el proceso de adaptación de un test hay una secuencia de actividades: en primer lugar es necesario traducir la versión original al idioma de la cultura meta del test; en segundo término se requiere introducir las modificaciones que sean necesarias a la nueva versión; y, finalmente, debemos verificar mediante un diseño de investigación apropiado que las dos versiones (original y adaptada) son equivalentes. A continuación revisaremos brevemente cada uno de estos pasos y concluiremos con un ejemplo concreto de adaptación de una prueba.

Técnicas de traducción de un test

El proceso de traducción es complejo e implica más que la traducción literal de las palabras escritas a un nuevo lenguaje. Existen dos métodos fundamentales: la traducción directa (*for-ward translation*) e inversa (*backward translation*).

En el método de traducción directa un traductor o, preferentemente, un grupo de traductores, traducen el test desde el idioma original al nuevo. Luego otro grupo de traductores juzga la equivalencia entre las dos versiones. De este modo pueden realizarse las correcciones pertinentes a las dificultades o errores identificados.

En el caso de la traducción inversa, el método más utilizado, un grupo de traductores realiza una traducción desde el idioma original al nuevo idioma; luego un segundo grupo de traductores toma el test traducido (en el nuevo idioma) y vuelve a traducirlo al idioma original. Seguidamente, se realizan las comparaciones entre la versión original y la versión retraducida al idioma original para determinar su equivalencia.

Ambos métodos poseen ventajas y desventajas que no serán analizadas en este texto introductorio. El lector interesado en profundizar esta problemática específica puede consultar el texto clásico de Hambleton (1994).

Diseños experimentales para verificar la equivalencia de tests e ítems

Una vez que se ha traducido adecuadamente el test, es necesario establecer si esta versión traducida es equivalente a la original. Para ello se requiere un diseño experimental y el análisis de los datos obtenidos a través de ese diseño. Hambleton (1994) indicó que existen básicamente tres métodos:

- a) *Administración del test original y traducido a individuos bilingües.* En este caso se les administra ambas versiones del test (la original y la traducida al nuevo idioma) a personas que hablen ambos idiomas. Si, por ejemplo, quisiéramos adaptar un test de inteligencia desde el inglés al español, administraríamos la versión en inglés y la versión en español a individuos que hablen ambos idiomas, y luego verificaríamos la equivalencia de las puntuaciones de ambas formas. Este método posee ventajas y limitaciones. Entre las primeras se puede mencionar que permite controlar las diferencias de los participantes en el test (en su inteligencia, por ejemplo), puesto que ambas versiones del test son administradas a las mismas personas. Entre las desventajas, Hambleton señaló que este diseño está basado en la premisa de

que los individuos son igualmente competentes en ambos idiomas, lo cual es difícil de sostener. Es probable, entonces, que puedan observarse diferencias entre los resultados de ambas versiones debido a una menor capacidad de algunas personas para entender los ítems en alguna de las dos lenguas. La segunda desventaja de este diseño es que no puede asegurarse que los bilingües posean el mismo nivel de competencia que la población general. Por el hecho de conocer otro idioma, es probable que se trate de personas con una mayor capacidad intelectual o mejor educación. Hambleton (1994.) también propuso una variación de este método, que conserva las mismas ventajas y desventajas pero que es más fácil de implementar, y consiste en administrar al azar una de las versiones del test (en inglés o en español, siguiendo nuestro ejemplo anterior) a cada participante bilingüe.

- b) *Administración de la versión original del test y su traducción inversa a monolingües en el idioma original.* Siguiendo el ejemplo anterior, se les administraría la versión original del test de inteligencia en inglés y la versión obtenida por traducción inversa a personas cuyo idioma natal es el inglés. La equivalencia entre los ítems se determinaría comparando el desempeño de cada individuo en ambas versiones. Nuevamente, la ventaja de este diseño está en el control de las diferencias en las características de los participantes. Una limitación es que no permite obtener datos de la versión en el idioma meta del test (español, en este ejemplo). Otra desventaja reside en la posible falta de independencia entre los puntajes obtenidos, puesto que es probable que exista un efecto de aprendizaje luego de la administración de la primera versión de la prueba, especialmente si es la original. La administración al azar de una de las versiones en primer lugar puede atenuar la influencia de la variable "aprendizaje".
- c) *Administración de la versión original a monolingües que hablan el idioma original y de la versión traducida a monolingües que hablan el idioma al que ha sido traducido el test.* Continuando con el ejemplo anterior, en este caso se administraría la versión en inglés del test a sujetos cuya lengua materna es el inglés y se administraría la versión en español a personas cuyo idioma materno es el español. Una posible dificultad de este diseño reside en asumir que los sujetos de ambas muestras poseen una inteligencia (u otro rasgo medido) comparable. Sin embargo, Hambleton (1994) sugirió que este obstáculo puede subsanarse si los análisis son desarrollados en el contexto de la teoría de respuesta al ítem, que permite utilizar distintos conjuntos de ítems para obtener las mismas estimaciones del rasgo o aptitud.

Una vez obtenidos los datos por medio de estos diseños de investigación existen varias posibilidades de análisis estadístico. Básicamente, el análisis estará destinado a identificar que las propiedades psicométricas esenciales (confiabilidad, validez) del test original se mantengan en la versión adaptada. Para ello es necesario replicar esos estudios psicométricos del test original en la población meta de adaptación (consistencia interna, estructura interna mediante análisis

factorial, estabilidad, evidencias de validez convergente-discriminante y relacionada con criterio, por ejemplo).

También es importante verificar la posible existencia de funcionamiento diferencial del ítem (FDI), es decir, corroborar que las propiedades psicométricas de los ítems (y no solamente de las puntuaciones totales del test) no varíen en las diferentes muestras. El FDI es una aplicación importante de la teoría de respuesta al ítem que revisaremos en el último capítulo.

En una investigación donde se compararon estudiantes australianos con estudiantes singapurenses a propósito del auto-concepto académico en lectura y matemática (Tanzer, 1995), pudo observarse que, a pesar de que la prueba mostraba la misma estructura factorial en ambos grupos, algunos ítems de la escala competencia/facilidad evidenciaban FDI. El autor conjeturó que esas diferencias entre ambas muestras podrían atribuirse a un factor de "modestia", una virtud deseable dentro de la cultura oriental que provocaba en los estudiantes singapurenses mayor resistencia a expresar una actitud de auto-confianza. En otro estudio, Galibert, Aguerri, Lozzia y Abal (2006) analizaron el funcionamiento diferencial de los ítems de una escala construida para medir un constructo semejante a Responsabilidad correspondiente al modelo de los cinco factores de personalidad. En esa investigación, en una muestra de estudiantes universitarios de Psicología, encontraron que el ítem "Aunque me sienta cansado finalizo la tarea que me impuse" evidencia posible interacción entre el constructo medido y el sexo.

El EDI se detecta con varios estadísticos que, esencialmente, consisten en refutar o confirmar la hipótesis nula que indica su ausencia. El método más utilizado en la actualidad es el de Mantel y Haenszel (1959) que fuera adaptado para el FDI por Holland y Thayer (1988).

Ejemplo de adaptación de un test

La Comisión Internacional de Tests (International Test Commission) ha elaborado 22 pautas para adaptar tests de una cultura a otra (Hambleton, 1994). Esta serie de recomendaciones comprende cuatro secciones: contexto, desarrollo y adaptación del instrumento, administración e interpretación.

- a) Contexto: Por ejemplo, "los efectos de las diferencias culturales que no sean relevantes para los objetivos centrales del estudio deberían minimizarse en la medida de lo posible".
- b) Desarrollo y adaptación del instrumento: Por ejemplo, "los constructores/editores de tests deberían asegurar que el proceso de adaptación tiene en cuenta las diferencias lingüísticas y culturales entre las poblaciones a las que se dirigen las versiones adaptadas del test".
- c) Administración: Por ejemplo, "los constructores y administradores de tests deberían tratar de prever los problemas de comprensión que pueden presentarse en la administración del test y tomar las medidas oportunas para evitarlos mediante la preparación de materiales e instrucciones adecuados".

Interpretación: Por ejemplo, "cuando se adapta un test para utilizarlo en otra población, debe facilitarse la documentación sobre los cambios introducidos,

así como los datos que permitan verificar la equivalencia entre las versiones".

El establecimiento de estos estándares constituye un importante marco conceptual de referencia para el arduo trabajo de adaptar tests. Aunque estas pautas no se abordarán en detalle aquí, se analiza a continuación un ejemplo de aplicación de las mismas al proceso de adaptación de un test.

En un estudio reciente, El Hassan y Jammal (2005) adaptaron a la cultura del Líbano el Test de Comprensión Auditiva del Lenguaje - Revisado (Carrow-Woolfok, 1985). Este test, construido en los Estados Unidos, consta de tres partes que miden la comprensión del significado común y literal de las palabras, así como la comprensión de morfemas gramaticales y el significado de oraciones complejas. Es apropiado para niños de 3 a 10 años y sus ítems consisten en seleccionar entre varias figuras aquella que mejor representa el estímulo verbal provisto por el examinador. Estos investigadores realizaron una traducción directa desde el inglés al árabe y propusieron las siguientes modificaciones en los ítems para adaptar los materiales:

- a) tradujeron al árabe las palabras y frases que aparecían en algunas láminas;
- b) cambiaron la figura del billete de un dólar por un billete de mil libras libanesas en el ítem 34;
- c) la tercera figura del ítem 3 en la sección II se modificó para mostrar la imagen de un campesino libanes en lugar de un campesino occidental;
- d) se modificó la figura que representaba una pasta dental occidental, introduciendo una figura similar a la pasta dental que utilizan los niños en el Líbano;
- e) en un ítem que incluía figuras de guantes de béisbol, las mismas se reemplazaron por imágenes de guantes comunes.

Se administró entonces la versión adaptada a una muestra piloto de 50 niños. De acuerdo a los resultados obtenidos, se realizó un reordenamiento de los ítems según la frecuencia de dificultad en la versión árabe. De esta manera, algunos ítems cambiaron radicalmente su ubicación respecto de la versión original en inglés. Por ejemplo, el ítem 22 ("cuatro") pasó a ser el número 15 en la versión en árabe y el ítem 8 ("el gato está entre las sillas") pasó a la posición 14. Los investigadores observaron que los ítems que incluían preposiciones y verbos parecían ser más difíciles que aquellos que incluían sustantivos y descripciones de cantidad.

Finalmente, se realizaron estudios psicométricos esenciales (consistencia interna, validez) y los investigadores estandarizaron el test para utilizarlo con niños libaneses. Éste es un claro ejemplo de un test desarrollado en inglés y adaptado rigurosamente a una lengua (y una cultura) tan disímil como la árabe.

REFERENCIAS BIBLIOGRÁFICAS

- Abad, F.; Garrido, J.; Olea, J. y Ponsoda, V. (2006). Introducción a la psicometría. Teoría Clásica de los Tests y Teoría de Respuesta al Ítem. Madrid: Universidad Autónoma de Madrid. Inédito.
- Abelson, R. (1998). *La estadística razonada*. Buenos Aires: Paidós.
- Adorno, T.; Frenkel-Brunswik, E.; Levinson, D. y Sanford, R. (1950). *The authoritarian personality*. Nueva York: Harper & Row.
- Aftanas, M. S. (1988). Theories, models, and standard systems of measurement. *Applied Psychological Measurement*, 12, 325-338.
- Aiken, A. (2003). *Tests psicológicos y evaluación*. México: Prentice Hall.
- Allegri, R. F.; Mangone, C. A.; Fernández, A.; Rymberg, S.; Taragano, F. E. y Baumann, D. (1997). Spanish Boston Naming Test norms. *The Clinical Neuropsychologist*, 11, 416-420.
- American Psychological Association (APA) (1999). *Standards for psychological and educational tests*. Author: Washington, D.C.
- Anastasi, A. y Urbina, S. (1998). *Tests psicológicos*. México: Prentice Hall Latinoamericana.
- Angoff, W. (1988). Validity: An envolving concept. *Applied Measurement in Education*, 1, 215-222.
- Aron, A. y Aron, E. (2001). *Estadística para psicología*. Buenos Aires: Prentice Hall, 2ª ed.
- Axelrod, B.N.; Aharon-Peretz, J.; Tomer, R. y Fisher, T. (2000). Creating interpretation guidelines for the Hebrew Trail Making Test. *Applied Neuropsychology*, 7, 186-188.
- Baker, F. (2001). *The basics of item response theory*. ERIC Clearinghouse on Assessment and Education: College Park, MD.
- Baldo, M. (2001). Normas regionales y análisis de ítems del test de inteligencia para niños de Wechsler (WISC-III). *Evaluar*, 1, 29-52.

- Bandura, A. (1987). *Pensamiento y acción*. Madrid: Martínez Roca.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. Nueva York: Freeman.
- Bandura, A. (2001). Guía para la construcción de escalas de autoeficacia. *Evaluar*, 2, 7-38.
- Bayley, N. (1993). *Manual of the Bayley Scales of Infant Development* (2nd edition). San Antonio, TX: The Psychological Corporation.
- Beck, S.; Steer, R. y Brown, G. (1996). *Beck Depression Inventory-II. Manual*. San Antonio, TX: The Psychological Corporation.
- Bem, S. (1974). The measurement of psychological androgyny. *Journal of Consulting & Clinical Psychology*, 42, 165-172.
- Bennett, G.; Seashore, H. y Wesman, A. (2000). *Tests de Aptitudes Diferenciales (DAT-5). Manual*. Madrid: TEA Ediciones.
- Binet, A. y Simon, T. (1916). *The development of intelligence in children*. Baltimore: Williams & Wilkins.
- Bloom, B. (1966). *Taxonomy of educational objectives. The classification of educational goals*. Nueva York: McKay.
- Bogardus, E. (1925). Measuring social distances. *Journal of Applied Sociology*, 9, 299-308.
- Bond, L. (1996). Norm and criterion-referenced testing. *Practical Assessment, Research & Evaluation*, 2. <www.ericae.net/pare/getvn.asp>. Documento recuperado el 13/11/99.
- Briggs, S. R. y Cheek, J. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106-148.
- Bunge, M. (1983). *La investigación científica*. Barcelona: Ariel.
- Bunge, M. y Ardila, R. (2002). *Filosofía de la psicología*. Barcelona: Ariel.
- Butcher, J.; Dahlstrom, W.; Graham, J.; Telegen, A. y Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Caballo, V. (1987). Evaluación y entrenamiento de las habilidades sociales: una estrategia multimodal. Tesis doctoral (inédita). Universidad Autónoma de Madrid.
- Campbell, N. R. (1938). Measurement and its importance for philosophy. *Symposium of the Aristotelian Society*. Londres: Harri-son.
- Campbell, D. T. y Fiske, A. W. (1954). Convergent and discriminant validation by the multirait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

- Campbell, D. P. y Hansen, J. C. (1981). *Manual for the Strong-Campbell Interest Inventory*. Stanford, CA: Stanford University Press.
- Carpenter, P.; Just, M. y Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Annual Psychological Review*, 97, 404-431.
- Carretero-Dios, H. y Pérez, C. (2005). Normas para el desarrollo y revisión de estudios instrumentales. *International Journal of Clinical and Health Psychology*, 5, 521-551.
- Carroll, J. (1993). *Human cognitive abilities*. Londres: Cambridge University Press.
- Carrow-Woolfok, E. (1985). *Test for Auditory Comprehension of Language-Revised*. Allen (TX): DLM Teaching Resources.
- Carver, C. y Scheier, M. (1996). *Perspectives on personality*. Needham Heights, MA: Allyn & Bacon.
- Casullo, M.; Cayssials, A.; Liporace, M.; De Diuk, L.; Arce Michel, J. y Álvarez, L. (1994). *Proyecto de vida y decisión vocacional*. Buenos Aires: Paidós.
- Cattell, R. (1966). The Scree Test for the number of factors. *Multivariate Behavioral Research*, 1, 141-161.
- Cattell, R. (1967). The theory of fluid and crystallized intelligence. *British Journal of Educational Psychology*, 37, 209-224.
- Charter, R. (1999). Sample size requirements for precise estimates of reliability, generalizability and validity coefficients. *Journal of Clinical and Experimental Neuropsychology*, 21, 559-566.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, Nueva York: Lawrence Earlbaum Associates.
- Cohen, R. y Swerdlik, M. (2000). *Pruebas y evaluación psicológicas. Introducción a las pruebas y a la medición*. México: McGraw Hill.
- Cooley, W. y Lohnes, P. R. (1976). *Evaluation research in education*. Nueva York: Wiley.
- Cortada de Kohan, N. (1998). La teoría de la respuesta al ítem y su aplicación al test verbal Buenos Aires. *Interdisciplinaria*, 15, 101-129.
- Cortada de Kohan, N. (1999). *Teorías psicométricas y construcción de tests*. Buenos Aires: Lugar Editorial.
- Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 96-104.

- Costa, P. y Mc Crae, R. (1999). *NEO-PI-R. Manual*. Madrid: TEA ediciones.
- Costello, A. y Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, 7, 1-9.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. (1998). *Fundamentos de la evaluación psicológica*. Madrid: Biblioteca Nueva.
- Cronbach, L. y Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Cronbach, L. y Gleser, G. (1972). *The dependability of behavioral measurements. Theory of generalizability for scores and profiles*. Nueva York: Wiley.
- El Hassan y Jammal (2005). Validation for the test for auditory comprehension of language-revised (TACL-R) in Lebanon. *Assessment in Education*, 2, 183-202.
- Embretson, S. y Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Eysenck, H. J. (ed.). (1981). *A model for personality*. Nueva York: Springer.
- Eysenck, H. J. y Eysenck, S. B. (1997). *Cuestionario revisado de Personalidad de Eysenck. EPQ-R. Manual*. Madrid: TEA.
- Exner, J. (1993). *The Rorschach: A comprehensive system. Volume 1: Basic foundations*. Nueva York: Wiley.
- Fernández, A. L.; Marino, J. C.; Villacorta, L. y Pérez, E. (2000). Uso y desarrollo de tests en Argentina. Facultad de Psicología. Universidad Nacional de Córdoba, Argentina. Inédito.
- Fernández, A. L. y Scheffel, D. L. (2003). A study on the criterion validity of the Mattis Dementia Rating Scale. *International Journal of Testing*, 3, 49-58.
- Fernández, A. L. y Marcopulos, B. (2004). About the influence of culture in neuropsychological testing: the case of the Trail Making Test. En revisión.
- Ferreyra, R. (1982). Nuevas modalidades para la evaluación educativa. Pruebas referidas a criterio. Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba. Inédito.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 30, 469-479.
- Fogliatto, H. (1991). *Cuestionario de Intereses Profesionales (CIP). Manual*. Buenos Aires: Guadalupe.

- Fogliatto, H. y Pérez, E. (2003). Sistema de Orientación Vocacional Informatizado. *SOVI 3. Manual*. Buenos Aires: Paidós.
- Folstein, M.; Folstein, S. y McHugh, P. (1975). "Mini-Mental State": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research*, 19, 189-198.
- Galibert, M. S.; Aguerri, M. E.; Lozzia, G. y Abal, F. J. (2006). Análisis del DIF en una escala de voluntad de trabajo mediante los procedimientos de Mantel-Haenszel y Breslow-Day, regresión logística y el criterio ETS. *Actas de las Jornadas de Investigación de la Facultad de Psicología*. Buenos Aires: Universidad Nacional de Buenos Aires.
- Gardner, H. (1994). *Estructuras de la mente. La Teoría de las Inteligencias Múltiples*. México: Fondo de Cultura Económica.
- Gardner, H. (1999). *Intelligence Reframed*. Nueva York: Basic Books.
- Gardner, R. C. (2003). *Estadística para psicólogos usando SPSS para Windows*. México: Prentice Hall.
- Gesell, A. y Amatruda, C. (1971). *Diagnóstico del desarrollo normal y anormal del niño*. Buenos Aires: Paidós.
- Glaser, R. (1963) Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*, 18, 519-521.
- Glass, G.; McGaw, B. y Smith, M. (1981). *Meta-analysis in Social Research*. California: Sage Publications.
- Glutting, J. (2002). Some psychometric properties of a system to measure ADHD among college students. *Measurement and Evaluation in Counseling and Development*, 34, 194-209.
- Goldberg, L. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. En M. Mervielde, I.; Deary, F.; De Fruyt, E. y Ostendorf, F. (eds.), *Personality Psychology in Europe, vol. 7* (pp. 7-28). Tilburg, Holanda: Tilburg University Press.
- Gómez Benito, J. (1987). *Meta-análisis*. Barcelona: P.P.V.
- Grasso, L. (1969). Posibilidad de prever el rendimiento académico a través de una evaluación de conocimientos previos en la Universidad Nacional de Córdoba. Tesis de Licenciatura en Psicología. Facultad de Filosofía y Humanidades. Universidad Nacional de Córdoba. Inédita.
- Grasso, L. (1984). Una escala para la medición de la autoestima en ancianos. *Nueva Revista de Psicología*, 1 (1), 49-62.
- Grasso, L. (1999). Introducción a la estadística en Ciencias Sociales y del Comportamiento. Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba. Inédito.

- Guilford, J. (1967). *The nature of human intelligence*. Nueva York: McGraw-Hill.
- Gulliksen, H. (1950). *Theories of mental tests*. Nueva York: Wiley.
- Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 229-244.
- Hambleton, R. K. y Rogers, H. J. (1991). *Fundamentals of item response theory*. Beverly Hills, CA: Sage.
- Heaton, R. K.; Chelune, G. J.; Talley, J. L.; Kay, G. G. y Curtiss, G. (1991). *Wisconsin Card Sorting Test Manual*. Odessa, FL: Psychological Assessment Resources.
- Heaton, R.; Grant, I. y Matthews, C. (1991). *Comprehensive norms for an expanded Halstead-Reitan battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources.
- Herrera Rojas, A. (1998). Notas sobre Psicometría. Bogotá: Universidad de Colombia. Inédito.
- Himmel, E. (1979). Tendencias actuales en la evaluación del rendimiento escolar. *Revista de Tecnología Educativa*, 2, 3-7.
- Ho, D. (1998). Indegenous psychologies. Asian perspectives. *Journal of Cross Cultural Psychology*, 29, 88-103.
- Hogan, T. (2004). *Pruebas psicológicas: una introducción práctica*. México: El Manual Moderno.
- Holland, J. (1994). *Self-Directed Search. Manual*. Odessa, FL: Psychological Assessment Resources.
- Holland, J. (1997). *Making vocational choices*. Englewood Cliffs, NJ: Prentice-Hall.
- Holland, P. W. y Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. En H. Wainer y H. I. Braun (eds.), *Test Validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hood, A. y Johnson, R. (2002). *Assessment in Counseling*. Alexandria, VA: American Counseling Association.
- Horn, J. (1965). A rationale and test for the number the factors in factor analysis. *Psychometrika*, 30, 179-185.
- Jensen, A. (1998). *The g factor. The science of mental ability*. Westport: Praeger.
- Johnson, W. y Bouchard, T. (en prensa). The structure of human intelligence: It is verbal, perceptual and image rotation (VPR), not fluid and crystallized. *Intelligence*.
- Johnson, S. (1994). Scholastic Assessment Tests (SAT). En R. Stern-

- berg (ed.), *Encyclopedia of human intelligence* (pp. 956-960). Nueva York: Macmillan.
- Juan-Espinosa, M. (1997). *Geografía de la inteligencia humana*. Madrid: Pirámide.
- Kahn, J. H. (2006). Factor analysis in Counseling Psychology research, training and practice: Principles, advances and applications. *The Counseling Psychologist*, 34, 1-36.
- Kaplan, R. y Saccuzzo, D. (2006). *Pruebas psicológicas. Principios, aplicaciones y temas*. México: Thompson Learning.
- Kehoe, (1995). Writing multiple-choice test items. *Practical Assessment, Research & Evaluation*, 7. <www.ericae.net/pare> Documento recuperado el 11/13/99.
- Kelly, J. (1987). *Entrenamiento de las habilidades sociales*. Madrid: Desclee de Brower.
- Kerlinger, F. y Lee, H. (2002). *Investigación del comportamiento. Métodos de investigación en las ciencias sociales*. México: McGraw-Hill.
- Kline, P. (2000). *Handbook of Psychological Testing*. Londres: Routledge.
- Kolb, B. y Wishaw, I. (1986). *Fundamentos de neuropsicología humana*. Barcelona: Labor.
- Kuder, E. y Zitowski, G. (1991). *Kuder Occupational Interest Survey. General Manual*. Adel, IA: National Career Assessment Services.
- Lau, C.W. y Hoosain, R. (1999). Working memory and language difference in sound duration: a comparison of mental arithmetic in Chinese, Japanese, and English. *Psychologia. An International Journal of Psychology in the Orient*, 42, 139-144.
- Ledesma, R. y Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: an easy-to-use computer program for carrying out Parallel Analysis. *Practical Assessment, Research & Evaluation*, 12 (2), 1-11.
- Lent, R. (2004). Toward a unifying theoretical and practical perspective of well-being and psychosocial adjustment. *Journal of Counseling Psychology*, 51, 482-509.
- Lent, R.; Brown, S. y Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest: Choice and performance. *Journal of Vocational Behavior*, 45, 79-122.
- Lezak, M. D. (1995). *Neuropsychological assessment*. Nueva York: Oxford University Press.
- Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140.

- Loelhin, J. C. (1992). *Genes and environment in personality development*. Nuevabury Park, CA: Sage.
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Lawrence Erlbaum Associates.
- Lord, F. y Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Mac Combs, B. L. y Branan, L. (1990). *Social skills for job success: Teacher's guide*. Baltimore, MD: Educational Press.
- Mantel, N. y Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of The National Cancer Institute*, 22, 719-748.
- Martínez Arias, R. (1995). *Psicometría*. Madrid: Síntesis Psicológica.
- Mc Crae, R. R.; Yik, M. S. M.; Trapnell, P. D.; Bond, M. H. y Paulhus, D. L. (1998). Interpreting personality profiles across cultures: bilingual, acculturation, and peer rating studies of chinese students. *Journal of Personality and Social Psychology*, 74, 1041-1055.
- McGrew, K.; Flanagan, D.; Keith, T. y Vanderwood, M. (1997). Beyond g: the impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the school. *School Psychology Review*, 26, 189-210.
- Merenda, P. (1997). A guide to the proper use of Factor Analysis in the conduct and reporting of research: pitfalls to avoid. *Measurement and Evaluation in Counseling and Evaluation*, 30, 156-163.
- Mitchell, J. (1990). *An introduction to the logic of psychological measurement*. Hillsdale: Erlbaum.
- Moreno, R.; Martínez, R. y Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, 6, 3, 490-497.
- Multon, R.; Brown, S. y Lent, R. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, 38, 30-38.
- Muñiz, J. (1998). La medición de lo psicológico. *Psicothema*, 10, 1, 1-21.
- Muñiz, J. (2001). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Muraki, E. (1993). *POLYFACT* (computer program). Princeton, NY: Educational Testing Service.
- Murat, F. (1984). Actitud hacia la matemática. *Nueva Revista de Psicología*, 1, 13-25.
- Murat, F. (1985). *Evaluación del comportamiento humano*. Universidad Nacional de Córdoba.

- Norman, W. (1963). Toward an adequate taxonomy of personality attributes. *Journal of Abnormal and Social Psychology*, 65, 574-583.
- Nunnally, J. (1991). *Teoría psicométrica* (2ª ed.). Buenos Aires: Paidós.
- Nunnally, J. y Bernstein, I. (1995). *Teoría psicométrica*. México: McGraw Hill.
- Oesterlind, S. (1990). Establishing criteria for meritorius test items. *Educational Research Quality*, 3, 26-30.
- Olea, J.; Ponsoda, V. y Prieto, G. (eds.) (1999). *Tests adaptativos informatizados*. Madrid: Pirámide.
- Oliden, P. E. (2003). Sobre la validez de los tests. *Psicothema*, 15, 315-321.
- Orwin, R. G. (1983). A fail safe N for effect size. *Journal of Educational Statistics*, 8, 157-159.
- Padua, J. (1979). *Técnicas de investigación aplicadas a las Ciencias Sociales*. México: Fondo de Cultura Económica.
- Pajares, F.; Hartley, J. y Valiente, G. (2001). Response format in Writing Self-Efficacy Scales. Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development*, 33, 214-221.
- Pérez, E. (2001). Construcción de un Inventario de Autoeficacia para Inteligencias Múltiples. Tesis Doctoral. Córdoba, Argentina: Universidad Nacional de Córdoba. Inédita.
- Pérez, E. y Gay, D. (1991). La utilización de pruebas psicológicas en la ciudad de Córdoba. Facultad de Psicología. Universidad Nacional de Córdoba, Argentina. Inédito.
- Pérez, E.; Cupani, M. y Ayllón, S. (2005). Predictores de rendimiento académico en la escuela media: habilidades, autoeficacia y rasgos de personalidad. *Avaliação Psicológica*, 4, 1, 1-12.
- Plomin, R.; DeFries, J.; McClearn, G. y McGuffin, P. (2002). *Genética de la conducta*. Barcelona: Ariel Ciencia.
- Popham, W. J. (1975). *Educational evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- Prieto, G. y Delgado, A. (1999). Medición cognitiva de las aptitudes. En Olea, V.; Ponsoda, J. y Prieto, G. (eds.). *Tests informatizados. Fundamentos y aplicaciones* (pp. 28-57). Madrid: Pirámide.
- RASCAL (1989). *Rasch Item Calibration Program*. St. Paul: Assessment System Corporation.
- Rasch, G. (1963). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark's Paedagogiske Institut.
- Raven, J. (1993). *Test de Matrices Progresivas. Manual*. Buenos Aires: Paidós.

- Reise, S.; Waller, N. y Comrey, A. (2000). Factor analysis and scale revision. *Practical Assessment*, 12, 287-297.
- Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie*, 28, 286-340.
- Richaud de Minzi, C. (2005). Estilos parentales y afrontamiento en niños. *Revista Latinoamericana de Psicología*, 37, 1, 47-58.
- Rorschach, H. (1921). *Psychodiagnostik*. Berne, Switzerland: Bircher.
- Rosenthal, R. y Di Matteo, M. R. (2001). Meta-Analysis: Recent Developments in Quantitative Methods for Literature Reviews. *Annual Reviews Psychology*, 52, 59-82.
- Rosenthal, R. y Di Matteo, M. R. (2002). Meta-Analysis. En H. Pashler et al. (eds.). *Handbook of Experimental Psychology* (3ª ed.) (pp.391- 428). Nueva York: John Wiley & Sons.
- Roth, G. y Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, 9, 250-257.
- Rowell, R. K. (1996). Partitioning predicted variance into constituent parts: How to conduct regression commonality analysis. En Thompson, B. (ed.). *Advances in social science methodology*, vol. 4 (pp. 33-43). Greenwich, CT: JAI Press.
- Russell, M. T. y Karol, D. L. (2000). *16PF-5. Manual*. (5ª ed.) Madrid: TEA.
- Samejina, F. (1973). Homogeneous case of the continuous response model. *Psychometrika*, 38, 203-219.
- Santisteban, C. (1990). *Psicometría*. Madrid: Norma.
- Shearer, B. (1999). *Multiple Intelligence Developmental Assessment Scale (MIDAS). Professional Manual*. Kent: University of Kent Press.
- Spearman, C. (1927). *The nature of "intelligence" and the principles of cognition*. Londres: MacMillan.
- Spielberger, C. (1983). *State-Trait Anxiety Inventory for Adults. Manual*. Redwood City, CA: Mind Garden.
- Statistical Package for the Social Sciences (SPSS) (1995). *Statistical Package for the Social Sciences Reference Guide*. Chicago: Autor.
- Sternberg, R. (1987). *Inteligencia humana*. Vol. I. Barcelona: Paidós.
- Stevens, S. (1949). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stroop, J. R. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, 18, 643-662.
- Tabachnick, B. y Fidell, L. (2001). *Using multivariate statistics*. Nueva York: Harper & Row.

- Tanzer, N. K. (1995). Cross-cultural bias in likert-type inventories: perfect matching structures and still biased? *European Journal of Psychological Assessment*, 11, 194-201.
- Taub, G. (2001). A confirmatory analysis of the Wechsler adult intelligence scale-third edition: is the verbal-performance discrepancy justified? *Practical Assessment, Research and Evaluation*, 6, 1-11.
- Tellegen, A. (1988). Personality similarity in twins reared apart and together. *Journal of Personality and Social Psychology*, 54, 1031-1039.
- Tetreau, B. y Trahan, M. (1986). *Test Visuel d'Interets Tetreau-Trahan. Manuel d'usage*. Montreal: SECOROP.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis*. Washington, DC: American Psychological Association.
- Thompson, B. y Borrello, G. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement*, 28, 203-209.
- Thompson, B. y Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: An historical overview and some guidelines. *Educational and Psychological Measurement*, 56, 197-208.
- Thorndike, R. L. (1989). *Psicometría aplicada*. México: Limusa.
- Thurstone, L. L. (1935). *The vectors of the mind*. Chicago: University of Chicago Press.
- Tokar, D. M.; Fischer, A. R. y Subich, L. M. (1998). Personality and vocational behavior: A selective review of the literature, 1993-1997. *Journal of Vocational Behavior*, 53, 115-153.
- Tornimbeni, S. y González, C. (1997). *Construcción de una escala de actitudes ante la investigación en Psicología*. Facultad de Psicología. Universidad Nacional de Córdoba. Inédito.
- Tracey, T. (2002). Personal Globe Inventory: measurement of the spherical model of interests and competence beliefs. *Journal of Vocational Behavior*, 60, 113-172.
- Tyler, R. (1978). *The Florida Accountability Program*. Washington DC: National Education Association.
- Van de Vijver, F. J. F. y Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Nuevabury Park, CA: Sage.
- Van de Vijver, F. J. R. y Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*, 47, 263-279.
- Van der Linden, W. y Hambleton, R. (1997). *Handbook of modern Item Response Theory*. Nueva York: Springer.

- Velandrino, A. (1998). *Análisis de datos en ciencias sociales*. Murcia: DM Editora.
- Vernon, P. (1964). *The structure of human abilities*. Londres: Metluen.
- Walsh, W. y Betz, N. (1990). *Tests and assessment*. Englewood Cliffs, NJ: Prentice Hall.
- Wechsler, D. (1994). *Test de Inteligencia para niños (WISC-III)*. Manual Técnico. Buenos Aires: Paidós.
- Wechsler, D. (1999). *WAIS-III. Test de Inteligencia para adultos*. Manual Técnico. Buenos Aires: Paidós.
- Wechsler, D. (2005). *WISC-IV. Escala de Inteligencia de Wechsler para niños*. Manual. Madrid: TEA.
- Wissler, C. (1901). *The correlation of mental and physical tests*. Nueva York: Columbia University.
- Wolf, F. (1986). *Meta-analysis: Quantitative methods of research synthesis*. Beverly-Hills, CA: Sage.
- Woodcock, R.; McGrew, K. y Mather, N. (2001). *Woodcock-Johnson III. Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.
- Woolfolk, A. (2006). *Psicología educativa*. México: Pearson Educación.
- Yerkes, R. (1921). Psychological examining in the United States Army. *Memoirs of the National Academy of Science*, vol. 15.